

NON-CONVEX OPTIMIZATION FOR LINEAR SYSTEM WITH PREGAUSSIAN MATRICES
AND RECOVERY FROM MULTIPLE MEASUREMENTS

by

YANG LIU

(Under the direction of Ming-Jun Lai)

ABSTRACT

The extremal singular values of random matrices in ℓ_2 -norm, including Gaussian random matrices, Bernoulli random matrices, subgaussian random matrices, etc, have attracted major research interest in recent years. In this thesis, we study the q -singular values, defined in terms of the ℓ_q -quasinorm, of pregaussian random matrices. We give the upper tail probability estimate on the largest q -singular value of pregaussian random matrices for $0 < q \leq 1$, and also the lower tail probability estimate. Particularly, these estimates show that the largest q -singular value is of order $m^{1/q}$ with high probability for pregaussian random matrices of size m by m . Moreover, we also give probabilistic estimates for the smallest q -singular value of pregaussian random matrices. In addition, we also present some results on the largest p -singular value for $p > 1$, and some numerical-experimental results as well.

Compressed sensing, a technique for recovering sparse signals, has also been an active research topic recently. The extremal singular values of random matrices have applications in compressed sensing, mainly because the restricted isometry constant of sensing matrices depends on them. We prove that the pregaussian random matrices with m much less than N but much larger than $N^{q/2}$ have the q -modified restricted isometry property for $0 < q \leq 1$ with overwhelming probability. As a result, we show that every sparse vector can be recovered

as a solution to the ℓ_q -minimization problem with overwhelming probability if m is much less than N but much larger than $N^{q/2}$.

In compressed sensing, we also show that the real and complex null space properties (NSP) are equivalent for the sparse recovery by ℓ_q -minimization and more generally for the NSP for the joint-sparse recovery from multiple measurements via ℓ_q -minimization. These results answer the open questions raised by Foucart and Gribonval. We also extend Berg and Friedlander's theorem on NSP for recovery from multiple measurements. As a consequence of the equivalence on NSP and the extension, we give a necessary and sufficient condition for the uniqueness of the solution to the multiple-measurement-vector non-convex optimization problem.

INDEX WORDS: Optimization, Random Matrices, Sparse Recovery, Null Space Property

NON-CONVEX OPTIMIZATION FOR LINEAR SYSTEM WITH PREGAUSSIAN MATRICES
AND RECOVERY FROM MULTIPLE MEASUREMENTS

by

YANG LIU

B.S., Guangzhou University, China, 2003

M.S., Sun Yat-sen University, China, 2006

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

© 2010

Yang Liu

All Rights Reserved

NON-CONVEX OPTIMIZATION FOR LINEAR SYSTEM WITH PREGAUSSIAN MATRICES
AND RECOVERY FROM MULTIPLE MEASUREMENTS

by

YANG LIU

Approved:

Major Professor: Ming-Jun Lai

Committee: Edward Azoff
Alexander Petukhov
Andrew Sornborger
Jingzhi Tie

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2010

DEDICATION

To my parents, Ye-Ren Liu and Guo-E Long, to my sister Fen Liu, and to my brother Zhen Liu.

ACKNOWLEDGMENTS

First of all, I want to give the greatest thanks to my advisor, Prof. Ming-Jun Lai, for his guidance, encouragement, and support. He is such a great advisor who has inspired me so much in my mathematical research and my career. Without him, this thesis would never have been possible. His teaching has made the research subjects interesting to me, and he has motivated me to do cutting edge research, particularly to work on open problems in the area. I would like to express my utmost gratitude to him for the wonderful advisory work on my completion of the thesis and on helping me to achieve career goals.

I am very grateful to Prof. Edward Azoff and Prof. Jingzhi Tie for their very helpful advices in the course of my doctoral study at the University of Georgia, especially the valuable suggestions and helpful discussions concerning some part of the work in the thesis, for their generosity on their time, and furthermore for helping me in shaping my career. Moreover, I am also very thankful to Prof. Alexander Petukhov and Prof. Andrew Sornborger for serving on my advisory committee and for stimulating my interest in the applied side of mathematics. In addition, thanks to many other faculty members in the Department of Mathematics for the support through my years here.

My sincere thanks also go to my friends here, who were always there to help whenever I needed them, including Jennifer Belton, Qianying Hong, Tin Kong, Leopold Matamba, Jie Yu, and many others. It is a big blessing in my life having these precious friends. There are also many friends out there who have made my life in Athens more enjoyable. I wish to thank them for their help in my life as well, which I think is also very important to my academic life here.

Last but not the least, I am deeply indebted to my family. They have supported me through my studies and life, and I could never have come this far without their constant

support, encouragement, and patience. Their endless love has been the source of my courage and strength to overcome difficulties and to pursue my dreams.

PREFACE

The work in this thesis is mainly on optimization and random matrices. This thesis is organized in the following way: Chapter 1 gives an introduction of the thesis; the following three chapters are devoted to the study of generalized singular values of pregaussian matrices, including Chapter 2 on the largest q -singular value of pregaussian random matrices for $0 < q \leq 1$, Chapter 3 on the smallest q -singular value of pregaussian random matrices for $0 < q \leq 1$, and additionally Chapter 4 on the largest p -singular value of pregaussian random matrices for $p > 1$; in Chapter 5, we study the q -modified restricted isometry property for $0 < q \leq 1$; and lastly, Chapter 6 is on the null space property for recovery from multiple measurements via q -minimization.

CONTENTS

| | Page |
|--|------|
| ACKNOWLEDGMENTS | v |
| PREFACE | vii |
| LIST OF FIGURES | x |
| LIST OF TABLES | xi |
| CHAPTER | |
| 1 INTRODUCTION | 1 |
| 2 PROBABILISTIC ESTIMATE ON THE LARGEST q -SINGULAR VALUES OF PRE- GAUSSIAN RANDOM MATRICES | 14 |
| 2.1 PRELIMINARIES | 14 |
| 2.2 LINEAR COMBINATION OF PREGAUSSIAN RANDOM VARIABLES | 16 |
| 2.3 THE LARGEST q -SINGULAR VALUE OF PREGAUSSIAN RANDOM MA- TRICES | 18 |
| 3 PROBABILISTIC ESTIMATE ON THE SMALLEST q -SINGULAR VALUES OF PRE- GAUSSIAN RANDOM MATRICES | 32 |
| 3.1 SINGULAR VALUE OF PREGAUSSIAN RANDOM MATRICES IN ℓ_2 | 32 |
| 3.2 THE q -SINGULAR VALUES OF AN $m \times n$ MATRIX | 34 |
| 3.3 THE SMALLEST q -SINGULAR VALUE OF RANDOM MATRICES | 43 |
| 3.4 UPPER TAIL PROBABILITY OF THE SMALLEST q -SINGULAR VALUE | 49 |
| 4 THE PROBABILISTIC ESTIMATES ON THE LARGEST p -SINGULAR VALUE OF PREGAUSSIAN RANDOM MATRICES FOR $p > 1$ | 53 |

| | | |
|-----|--|-----|
| 4.1 | INTRODUCTION | 53 |
| 4.2 | LOWER TAIL PROBABILITY OF LARGEST p -SINGULAR VALUE FOR $p > 1$ | 53 |
| 4.3 | UPPER TAIL PROBABILITY OF THE LARGEST p -SINGULAR VALUE FOR $p > 1$ | 55 |
| 5 | MODIFIED RESTRICTED ISOMETRY PROPERTY AND SPARSE RECOVERY . . | 58 |
| 5.1 | 1-MODIFIED RESTRICTED ISOMETRY PROPERTY | 58 |
| 5.2 | q -MODIFIED ISOMETRY PROPERTY | 64 |
| 5.3 | ON THE SPARSE RECOVERY | 67 |
| 6 | NULL SPACE PROPERTY FOR RECOVERY FROM MULTIPLE MEASUREMENTS VIA ℓ_q -MINIMIZATION | 69 |
| 6.1 | INTRODUCTION | 69 |
| 6.2 | NULL SPACE PROPERTY FOR RECOVERY VIA ℓ_q MINIMIZATION . . . | 70 |
| 6.3 | REAL VERSUS COMPLEX NULL SPACE PROPERTIES FOR ℓ_q | 74 |
| 6.4 | ON RECOVERY FROM MULTIPLE MEASUREMENTS | 77 |
| 6.5 | THE NULL SPACE PROPERTY FOR RECOVERY FROM MULTIPLE MEASUREMENTS | 79 |
| | BIBLIOGRAPHY | 82 |
| | APPENDIX | |
| A | EXAMPLES ON THE DUALITY | 89 |
| B | NON-CONVEX GEOMETRY UNDER LINEAR MAP | 101 |
| C | SOME NUMERICAL EXPERIMENTS ON p -SINGULAR VALUE FOR $p > 1$ AND q -SINGULAR VALUE FOR $0 < q \leq 1$ OF RANDOM MATRICES | 106 |

LIST OF FIGURES

| | | |
|-----|---|-----|
| A.1 | Minimization of the maximum of $ y_1 $ and $ y_1 + y_2 $ | 94 |
| B.1 | Parallelograms | 104 |
| B.2 | The graph of difference function $f(\theta)$ | 105 |
| C.1 | Largest 2-singular value of Gaussian random matrices | 106 |
| C.2 | Largest 1-singular value of Gaussian random matrices: Experiment 1 | 107 |
| C.3 | Largest 1-singular value of Gaussian random matrices: Experiment 2 | 107 |
| C.4 | Largest 1-singular value of Gaussian random matrices: Experiment 3 | 108 |
| C.5 | Largest ∞ -singular value of Gaussian random matrices | 108 |
| C.6 | Largest $\frac{1}{3}$ -singular value of Gaussian random matrices | 109 |
| C.7 | Largest $\frac{1}{4}$ -singular value of Gaussian random matrices | 109 |
| C.8 | Largest $\frac{1}{4}$ -singular value of rectangular Gaussian random matrices | 110 |

LIST OF TABLES

- 3.1 Results obtained on the probabilistic estimates on the largest and smallest q -singular values of pregaussian random square matrices for $0 < q \leq 1$ 52

CHAPTER 1

INTRODUCTION

The largest singular value and the smallest singular value of random matrices in ℓ_2 -norm, including Gaussian random matrices, Bernoulli random matrices, subgaussian random matrices, etc, have attracted major research interest in recent years, and have applications in compressed sensing, a technique for recovering sparse or compressible signals. For instance, [Soshnikov and Fyodorov'2002, [50]] and [Soshnikov'2005, [51]], studied the largest singular value of random matrices, and [Rudelson and Vershynin'2008, [43]], [Rudelson and Vershynin'2008, [44]], [Tao and Vu'2009, [59]], and some others, studied the smallest singular values.

In the study of the asymptotic behavior of eigenvalues of symmetric random matrices, Wigner symmetric matrix is a typical example, whose upper (or lower) diagonal entries are independent random variables with uniform bounded moments. Wigner proved in [Wigner'1958, [62]] that the normalized eigenvalues are asymptotically distributed in the semicircular distribution. Precisely, let A be a symmetric gaussian random matrix of size $n \times n$ whose upper diagonal entries are independent and identically-distributed copies of the standard gaussian random variable, then the probability distribution function of the eigenvalues of $\frac{1}{\sqrt{n}}A$ is asymptotically

$$p(x) := \begin{cases} \frac{1}{2\pi} \sqrt{4-x^2} dx & \text{for } |x| \leq 2 \\ 0 & \text{for } |x| > 2 \end{cases} \quad (1.1)$$

as the size n goes to infinity. This is the well-known Wigner's Semicircle law. More generally, for a random matrix whose entries are independent and identically-distributed (i.i.d.) copies

of a complex random variable with mean 0 and variance 1, Tao and Vu showed in [Tao and Vu'2008, [56]] and [Tao and Vu'2010, [60]] that the eigenvalues of $\frac{1}{\sqrt{n}}A$ converges to the uniform distribution on the unit circle as n goes to ∞ , and that holds not only for the random matrices with real entries but also for complex entries. Their result has also generalized Girko's circular law in [Girko'1984, [27]] and solved the circular law conjecture open since the 1950's, that the smallest eigenvalue converges to the uniform distribution over the unit disk as n tends to infinity (see also [4]).

For random matrices whose entries are i.i.d. random variable satisfying certain moment conditions, the largest singular value was studied in [Geman'1980, [26]] and [Yin, Bai, and Krishnaiah'1988, [65]]. Furthermore, the distribution of the eigenvalue of Wishart matrices, $W_{N,n} = AA^*$, where $A = A_{N,n}$ is an $N \times n$ random Gaussian random matrix, was studied in [Soshnikov'2002, [50]]. More generally, Seginer in [Seginer'2000, [47]] compared the Euclidean operator norm of a random matrix with i.i.d. mean zero entries to the Euclidean norm of its rows and columns. Later, Latala in [Latala'2005, [34]] gave the upper bound on the expectation (or average value) of largest singular value namely the norm of any random matrix whose entries are independent mean zero random variables with uniformly bounded fourth moment.

The condition number, which is the ratio of the largest singular value over the smallest singular value of a matrix, is critical to the stability of linear systems. In [Edelman'1988, [17]], the distribution of the condition number of Gaussian random matrices, was particularly investigated in numerical experiments. As a typical example of subgaussian random matrices, the invertibility of Bernoulli random matrices was also studied. In [Tao and Vu'2007, [55]], the probability of Bernoulli random matrices to be singular is shown to be at most $(\frac{3}{4} + o(1))^n$, where n is the size of the matrices. Their result shows that the probability of the smallest singular value of Bernoulli random matrices to be zero is exponentially small as n tends to infinity.

The recent studies of the smallest singular value have also been motivated, in a large sense, by some open questions or conjectures. In [Spielman and Teng'2002, [52]], the following conjecture was proposed in the International Congress of Mathematicians in 2002.

Conjecture 1.0.1 (Spielman-Teng). *Let ξ be Bernoulli random variable, in other words, $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = \frac{1}{2}$. Then*

$$\mathbb{P}(s_n(M(\xi)) \leq \frac{t}{\sqrt{n}}) \leq t + c^n \quad (1.2)$$

for all $t > 0$ and some $0 < c < 1$.

In the breakthrough work on the estimate on the smallest singular value, [Rudelson and Vershynin'2008, [44]], Rudelson and Vershynin obtained the upper tail probabilistic estimate on the smallest value in ℓ_2 -norm for square matrices of centered random variables, with unit variance and appropriate moment assumptions. In particular, they proved the Spielman-Teng conjecture up to a constant. The lower tail probabilistic estimate on the smallest value in ℓ_2 -norm for square matrices was estimated in [Rudelson and Vershynin'2008, [43]]. These results have shown that the smallest singular value of the $n \times n$ subgaussian random matrices is of order $n^{-\frac{1}{2}}$ in high probability for large n . In a more explicit way, the distribution of the smallest singular value of random was given in [Tao and Vu'2009, [59]] by using property testing from combinatorics and theoretical computer science. The pregaussian matrices were used to recover sparse image in [Foucart and Lai'2010, [21]]. Very recently, Rudelson and Vershynin gave a comprehensive survey on the extreme singular values of random matrices in [Rudelson and Vershynin'2010, [46]].

In this thesis, we will study the probabilistic estimates on the largest q -singular value defined in terms of the in ℓ_q -quasinorm as

$$s_1^{(q)}(A) := \sup_{x \in \mathbb{R}^N, x \neq 0} \frac{\|Ax\|_q}{\|x\|_q} \quad (1.3)$$

and the smallest q -singular value defined accordingly (see [49] for instance) of pregaussian matrices, as they have become important to the ℓ_q -approach in compressed sensing, because of the restricted isometry property (see [6] and [21]), that will be studied in later chapters.

Before proceeding, we would like to mention some notations we will use in this thesis. $X = O(Y)$ denotes $X < CY$ for some constant $C > 0$. As conventional in probability theory, $\mathbb{P}(\mathcal{E})$ denotes the probability that the event \mathcal{E} occurs, and $\mathbb{E}(X)$ denotes the expectation of the random variable X .

It is well-known that the classic singular value is defined in terms of ℓ_2 -norm, then a natural question would be what if one defines the singular value by the ℓ_q -quasinorm for $0 < q \leq 1$ and ℓ_p -norm for $p > 1$. The first three chapters of the thesis will be devoted to the study of generalized singular values of pregaussian matrices, including Chapter 2 on the largest q -singular value of pregaussian random matrices for $0 < q \leq 1$, Chapter 3 on the smallest q -singular value of pregaussian random matrices for $0 < q \leq 1$, and additionally Chapter 4 on the largest p -singular value of pregaussian random matrices for $p > 1$. Our main results include the tail probability estimates on these generalized singular values of pregaussian matrices.

There were some remarkable results by other researchers on the largest singular values of random matrices in the ℓ_2 -norm. For example, Geman in [26] and Yin, Bai, and Krishnaiah in [65] showed that the largest singular value of random matrices of size $m \times N$ with independent entries of mean 0 and variance 1 tends to $\sqrt{m} + \sqrt{N}$ almost surely. However, we want to study the singular values of random matrices in terms of the ℓ_q -quasinorm for $0 < q \leq 1$, because there are some advantages of using general ℓ_q -norm to study the singular value of random matrices, as suggested in [Foucart and Lai'2009, [20]] and [Foucart and Lai'2010, [21]]. In this thesis, we study the q -singular values of random matrices whose entries are independent and identically-distributed copies of a pregaussian random variable, defined in terms of the ℓ_q -quasinorm. We obtain the decay on the upper tail probability of the largest q -singular value $s_1^{(q)}$ for all $0 < q \leq 1$, defined in terms of the non-convex ℓ_q -quasinorm, as the number of rows of the matrices becomes very large. This result is stated as

Theorem 1.0.2 (Upper tail probability of the largest q -singular value, $0 < q \leq 1$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix*

with i.i.d. copies of ξ in its entries, then for every $0 < q < 1$,

$$\mathbb{P}\left(s_1^{(q)}(A) \geq Cm^{\frac{1}{q}}\right) \leq \exp(-C'm) \quad (1.4)$$

for some $C, C' > 0$ only dependent on the pregaussian random variable ξ .

We have also obtained the lower tail probability of the largest q -singular value $s_1^{(q)}$ for all $0 < q \leq 1$, based on a linear bound for partial binomial expansion.

Theorem 1.0.3 (Lower tail probability of the largest q -singular value, $0 < q \leq 1$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $0 < q < 1$ and any $\varepsilon > 0$, there exists $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right) \leq \varepsilon \quad (1.5)$$

in which K only depends on q, ε and the pregaussian random variable ξ .

In particular, these estimates show $s_1^{(q)}(A) \sim m^{\frac{1}{q}}$ with high probability for $m \times N$ pregaussian random matrix A .

On the smallest singular value of random matrices, Rudelson and Vershynin first showed the following results in [43],

Theorem 1.0.4 (Rudelson-Vershynin). *If A is a matrix of size $n \times n$ whose entries are independent random variables with variance 1 and bounded fourth moment. Then for any $\delta > 0$, there exists $\epsilon > 0$ and integer $n_0 > 0$ such that*

$$\mathbb{P}\left(s_n(A) \leq \frac{\epsilon}{\sqrt{n}}\right) \leq \delta,$$

for all $n \geq n_0$.

Later, they proved the following theorem in [44].

Theorem 1.0.5 (Rudelson-Vershynin). *Let A be an $n \times n$ matrix whose entries are i.i.d. centered random variables with unit variance and fourth moment bounded by B . Then, for*

every $\delta > 0$ there exist $K > 0$ and n_0 which depend (polynomially) only on δ and B , and such that

$$\mathbb{P}\left(s_n(A) > \frac{K}{\sqrt{n}}\right) \leq \delta,$$

for all $n \geq n_0$.

For the smallest q -singular value of an $n \times n$ pregaussian random matrix, we have obtained the estimate on the lower tail probability in this thesis.

Theorem 1.0.6 (Lower tail probability of the smallest q -singular value, $0 < q \leq 1$). *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $n \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $\varepsilon > 0$, there exists some $\gamma > 0$ such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \gamma n^{-\frac{1}{q}}\right) < \varepsilon, \quad (1.6)$$

where γ only depends on q , ε and the pregaussian random variable ξ .

On the upper tail probability of the smallest q -singular value, we have

Theorem 1.0.7 (Upper tail probabilistic estimate on the smallest q -singular value). *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $n \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $K > e$, there exist some $C > 0$, $0 < c < 1$, and $\alpha > 0$ only dependent on pregaussian random variable ξ , q , such that*

$$\mathbb{P}\left(s_n^{(q)}(A) > K n^{-\frac{1}{2}}\right) \leq \frac{C (\ln K)^\alpha}{K^\alpha} + c^n. \quad (1.7)$$

In particular, for any $\varepsilon > 0$, there exist some $K > 0$ and n_0 , such that

$$\mathbb{P}\left(s_n^{(q)}(A) > K n^{-\frac{1}{2}}\right) < \varepsilon \quad (1.8)$$

for all $n \geq n_0$.

However, we strongly believe the probabilistic estimate on the smallest singular value in (1.8) may be improved and conjecture that

$$\mathbb{P}\left(s_n^{(q)}(A) > K n^{-\frac{1}{q}}\right) < \varepsilon \quad (1.9)$$

under the assumptions in Theorem 1.0.7.

While extending the q -singular value for $0 < q \leq 1$ to p -singular value for $p > 1$, we would like to investigate the probability analysis on the p -singular value for $p > 1$ as well. In [36], the lower tail probability of the ℓ_p -norm of a sequence of independent, centered, Gaussian random variables was estimated. Very recently, on random matrices, the extremal singular values defined in terms of ℓ_2 were studied in [43], [44], [45], [59], etc, as we mentioned earlier. However, the p -singular value, defined in terms of ℓ_p -norm for $p > 1$ will be the main topic in Chapter 4 of this thesis. In this respect, we have the following two theorems.

Theorem 1.0.8 (Lower tail probability of the largest p -singular value, $p > 1$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $p > 1$ and any $\varepsilon > 0$, there exists $\gamma > 0$ such that*

$$\mathbb{P} \left(s_1^{(p)}(A) \leq \gamma m^{\frac{1}{p}} \right) \leq \varepsilon \quad (1.10)$$

in which γ only depends on p , ε and the pregaussian random variable ξ .

The duality theorem on the largest p -singular value allows us to extend these results from $1 < p < 2$ to $p > 2$ and vice versa, and thus we are able to obtain the probabilistic estimates on the growth rate of the largest p -singular value for $p \geq 1$, including $p = \infty$, as the size of the random matrices grows.

Random matrices are often used as the sensing matrices in the optimization problem for compressed sensing. Foucart and Lai defined the new quasinorm

$$\|x\|_{f,q} := \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \quad (1.11)$$

for any pregaussian distribution probability density function f and $q > 0$ and introduced the q -modified restricted isometry property, $0 < q \leq 1$ based on the quasinorm. The extreme singular values of random matrices, including the largest singular value and smallest singular

value, become important for the restricted isometry property introduced in [6] and the generalized restricted property defined in [21], called q -modified restricted isometry property. In Chapter 5 of this thesis, we study the modified the q -modified restricted isometry property, $0 < q \leq 1$. We prove that the pregaussian random matrices, whose entries are independent and identically-distributed pregaussian random variables (see e.g. [10]), with $N^{\frac{q}{2}} \ll m \ll N$, have the q -modified restricted isometry property with overwhelming probability. As a result, we show that every sparse vector can be recovered as a solution to the ℓ_q -minimization problem with overwhelming probability if $N^{\frac{q}{2}} \ll m \ll N$ for $0 < q \leq 1$.

Theorem 1.0.9. *Suppose that A is an $m \times N$ matrix whose entries are independent and identically-distributed copies of a symmetric pregaussian random variable with probability density function f . Then*

$$\mathbb{P} \left((1 - \varepsilon) m \|x\|_{f,q}^q \leq \|Ax\|_q^q \leq (1 + \varepsilon) m \|x\|_{f,q}^q \right) \geq 1 - 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2} \quad (1.12)$$

for any $0 < \varepsilon < 1$ and $0 < q \leq 1$, and some $\kappa > 0$ dependent of f and q .

Chapter 6 of this thesis is devoted to the null space property in compressed sensing, especially for multiple measurement vector problem (MMV). In compressed sensing, we want to recover a sparse or compressible signal via solving the minimization problem,

$$\text{minimize}_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to } Ax = b, \quad (1.13)$$

in which $\|x\|_0$ is the the number of non-zero entries of the vector x , namely the sparsity of x . Since $\|x\|_0$ can be approximated by $\|x\|_q^q$, one can use the ℓ_q -approach (see [24], [33] and [20]), which is considering the following ℓ_q -minimization problem with $0 < q \leq 1$,

$$\text{minimize}_{x \in \mathbb{R}^N} \|x\|_q \quad \text{subject to } Ax = b. \quad (1.14)$$

Foucart and Lai presented some numerical experimental results which indicate that the q -method performs better than other available methods and a sufficient condition on the matrix of an underdetermined linear system which guarantees the solution of the system

with minimal q -quasinorm (cf. [20]). The ℓ_q -method in unconstrained minimization can be also used to generate the sparse solution to underdetermined linear system. See [33]. These have motivated us to consider the null space property for ℓ_q -minimization, as it characterizes the uniqueness of the solution to the ℓ_q -minimization problem.

Instead of single measurement vector, multiple measurement vectors have been used in many fields of technology, for instance, neuromagnetic imaging [15] and communication channels (see e.g. [14]), as pointed out in [8]. In general situations, precisely, the multiple-measurement-vector (MMV) problem, which is

$$\text{minimize } \|X\|_{q,p} \text{ subject to } AX = B, \quad (1.15)$$

for an $m \times N$ matrix A and an $m \times r$ matrix B , in which

$$\|X\|_{q,p} := \left(\sum_{j=1}^N \|X^{j \rightarrow}\|_p^q \right)^{\frac{1}{q}}, \quad (1.16)$$

where $X^{j \rightarrow}$ is the j -th row of X , for $p > 0$ and $q > 0$.

In recent literature, there were many significant studies on the MMV problem. Berg and Friedlander in [8] studied the joint-sparse recovery from multiple measurements through the MMV problem, as an extension of the single-measurement-vector (SMV) problem (1.13). They implemented the ReMBo (Reduce MMV and Boost) algorithm combined with ℓ_1 -minimization, introduced and also performed in [39] and [40], based on the reduction to SMV problem. Some other researchers also studied the algorithmic approaches to solve the MMV convex optimization problem that is the MMV problem (1.15) with $q \geq 1$ for the recovery from multiple measurements. For instance, Malioutov, Cetin and Willsky in [38] used the interior point implementation to solve the MMV problem with $p = 2$ and $q = 1$, in which (1.16) is neither linear nor quadratic, efficiently in a second-order cone (SOC) programming. Tropp in [61] developed another algorithmic approach called convex relaxation, by replacing ℓ_0 -quasinorm by ℓ_1 -norm, and it under certain conditions provides good solutions to MMV problem for $p = \infty$ and $q = 1$. For $q = 1$ and general $p \geq 1$, Chen and Huo in [12] showed that

the orthogonal matching pursuit for MMV (OMPMMV) can find the sparsest solution to the MMV problem with computational efficiency. In addition to the above studies, Cotter, Rao and others in [15] extended the matching pursuit (MP) and Focal Underdetermined System Solver (FOCUSS), introduced earlier in [22] for tomographic source reconstruction in neural electromagnetic inverse problems and designed to obtain sparse solutions by successively solving quadratic optimization problems (cf. [64]), to solve the MMV problem for $p \leq 1$ and $q = 2$.

For the SMV non-convex ℓ_q -minimization problem (5.49), Foucart and Lai in [20] compared the numerical experiments results by their ℓ_q -algorithm with other algorithm, such as the regularized orthogonal matching pursuit, the ℓ_1 -minimization, the reweighted ℓ_1 -minimization, and the comparisons showed that ℓ_q -method performs better.

However, we will study the multiple-measurement-vector (MMV) non-convex optimization problem, that is

$$\text{minimize } \|X\|_{q,p} \quad \text{subject to } AX = B, \quad (1.17)$$

in which $0 < q \leq 1$.

The null space property has been used to quantify the error of approximations (Cohen, Dahmen, and DeVore' 2009, [13]), and it also guarantees the exact recovery. The ℓ_1 null space property is the sufficient condition for unique recovery through ℓ_1 -minimization. Foucart and Gribonval in [Foucart and Gribonval'2009, [19]] proved that the real null space and complex null space property are equivalent for the sparse recovery achieved by ℓ_1 -minimization. They considered the convex optimization problem

$$\text{minimize}_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{subject to } Az = y. \quad (1.18)$$

in which A is a real-valued measurement matrix. But a real-valued measurement matrix is also a complex-valued one, they proved the following result

Theorem 1.0.10 (Foucart-Gribonval). *For a measurement matrix $A \in \mathbb{R}^{m \times N}$ and a set $S \subseteq \{1, 2, \dots, N\}$, the real null space property that*

$$\sum_{j \in S} |u_j| < \sum_{j \in S^c} |u_j| \quad (1.19)$$

for all $(u_1, u_2, \dots, u_N)^T \in (\ker_{\mathbb{R}}(A)) \setminus \{0\}$ is equivalent to the complex null space property, in view of $\ker_{\mathbb{C}} A = \ker_{\mathbb{R}} A + i \ker_{\mathbb{R}} A$, that

$$\sum_{j \in S} \sqrt{u_j^2 + v_j^2} < \sum_{j \in S^c} \sqrt{u_j^2 + v_j^2} \quad (1.20)$$

for all $(u_1, u_2, \dots, u_N)^T \in \ker_{\mathbb{R}}(A)$ and $(v_1, v_2, \dots, v_N)^T \in \ker_{\mathbb{R}}(A)$ with either of the vectors non-zero.

Replacing ℓ_1 -norm by ℓ_q -quasinorm in the ℓ_1 null space property, the condition is then called ℓ_q null space property, that becomes the sufficient condition for unique recovery through solving the ℓ_q -minimization problem. On the other hand, the well-known null space property (cf. [16] and [24]) for the standard ℓ_1 minimization in the setting of single measurement vector (SMV) has been extended to this setting of multiple measurement vectors (MMV). In the study on these aspects, we show that the real null space and complex null space property are equivalent for the sparse recovery achieved by ℓ_q -minimization and more generally that for the joint-sparse recovery from multiple measurements, which answer the open questions raised in [19].

Berg and Friedlander in [8] showed the null space property for recovery from multiple measurements via ℓ_1 -minimization. In [8], the following result is proved.

Theorem 1.0.11 (Berg-Friedlander). *Let A be a real matrix of $m \times N$ and $S \subset \{1, 2, \dots, N\}$ be a fixed index set. Denote by S^c the complement set of S in $\{1, 2, \dots, N\}$. Let $\|\cdot\|$ be any norm. Then all $\mathbf{x}^{(k)}$ with support $\mathbf{x}^{(k)}$ in S for $k = 1, \dots, r$ can be uniquely recovered using the following*

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^{(k)} \in \mathbb{R}^N \left\{ \sum_{j=1}^N \|(x_{1,j}, \dots, x_{r,j})\| : \text{subject to } A\mathbf{x}^{(k)} = \mathbf{b}^{(k)}, k = 1, \dots, r \right\} \\ & k = 1, \dots, r \end{aligned} \quad (1.21)$$

if and only if all vectors $(\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(r)}) \in (N(A))^r \setminus \{(0, 0, \dots, 0)\}$ satisfy the following

$$\sum_{j \in S} \|(u_{1,j}, \dots, u_{r,j})\| < \sum_{j \in S^c} \|(u_{1,j}, \dots, u_{r,j})\|, \quad (1.22)$$

where $N(A)$ stands for the null space of A .

We extend their theorem on the null space property from ℓ_1 to ℓ_q , $0 < q \leq 1$, for recovery from multiple measurements. As a consequence of the extension, we give a necessary and sufficient condition for the uniqueness of the solution to the multiple-measurement-vector (MMV) non-convex optimization problem. Precisely, we prove the following

Theorem 1.0.12 (Null space property for MMV recovery via ℓ_q -minimization). *Let A be an $m \times N$ matrix and $S \subset \{1, 2, \dots, N\}$ be an index set. Then for any $0 < q \leq 1$, all $X_0 \in \mathbb{R}^{N \times r}$ with the support of the rows of X_0 contained in S can be uniquely recovered via solving the MMV non-convex optimization problem (1.17) for $p = 2$ if and only if $\|u_S\|_q < \|u_{S^c}\|_q$ for all $u \in \ker(A) \setminus \{0\}$.*

The above theorem characterizes the uniqueness of the solution to the MMV non-convex optimization problem (1.17), and it would simplify the algorithm checking the uniqueness of the solution to (or the exact recovery through) the MMV non-convex optimization problem.

In fact, we consider a joint recovery from multiple measurement vectors via

$$\text{minimize } \sum_{j=1}^N \left(\sqrt{x_{1,j}^2 + \dots + x_{r,j}^2} \right)^q : \text{ subject to } A\mathbf{x}^{(1)} = \mathbf{b}^{(1)}, \dots, A\mathbf{x}^{(r)} = \mathbf{b}^{(r)} \quad (1.23)$$

for a given $0 < q \leq 1$, where $\mathbf{x}^{(k)} = (x_{k,1}, \dots, x_{k,N})^T \in \mathbb{R}^N$ for all $k = 1, \dots, r$, and this is actually ((1.17)) for when $p = 2$.

Written as equivalent conditions, the theorem on the exact recovery we mainly prove is the following

Theorem 1.0.13. *Let A be a real matrix of size $m \times N$ and $S \subset \{1, 2, \dots, N\}$ be a fixed index set. Fix $p \in (0, 1]$ and $r \geq 1$. Then the following conditions are equivalent:*

(a) *All $\mathbf{x}^{(k)}$ with support in S for $k = 1, \dots, r$ can be uniquely recovered using (1.23);*

(b) For all vectors $(\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(r)}) \in (N(A))^r \setminus \{(0, 0, \dots, 0)\}$

$$\sum_{j \in S} \left(\sqrt{u_{1,j}^2 + \dots + u_{r,j}^2} \right)^q < \sum_{j \in S^c} \left(\sqrt{u_{1,j}^2 + \dots + u_{r,j}^2} \right)^q; \quad (1.24)$$

(c) For all vector $\mathbf{z} \in N(A)$ with $\mathbf{z} \neq 0$,

$$\sum_{j \in S} |z_j|^q < \sum_{j \in S^c} |z_j|^q, \quad (1.25)$$

where $\mathbf{z} = (z_1, \dots, z_N)^T \in \mathbb{R}^N$.

Our theorem can be applied to develop algorithms, as condition (c) significantly reduces the complexity of verification of condition (b) for unique recovery from multiple measurement vectors by using (1.23).

CHAPTER 2

PROBABILISTIC ESTIMATE ON THE LARGEST q -SINGULAR VALUES OF PREGAUSSIAN RANDOM MATRICES

2.1 PRELIMINARIES

In this section, we are going to give an introduction on some basics on the Gaussian random variable, subgaussian random variable, and pregaussian random variable. There are many references on these standard terminologies, such as [9], [10], [21], and [23].

The Gaussian random variable is well-known as a basic random variable widely used in the probability and statistics studies, and is also an important random variable because of the central limit theorem in probability theory. A symmetric Gaussian random variable X obeys the standard normal distribution,

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (2.1)$$

where σ^2 is the variance, and its moment generating function is

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} p_X(x) dx = e^{\frac{\sigma^2 t^2}{2}}. \quad (2.2)$$

Subgaussian random variable is more general than Gaussian random variable.

Definition 2.1.1 (Subgaussian random variable). A random variable X is called subgaussian, if its moment generating function satisfies

$$M_X(t) \leq e^{\frac{\sigma^2 t^2}{2}} \quad (2.3)$$

for all $t \in \mathbb{R}$ and some $\sigma \geq 0$.

A typical example of subgaussian random variable is the centered Bernoulli random variable.

Example 2.1.2 (Bernoulli random variable). A Bernoulli random variable X with mean 0 and variance 1 is a discrete random variable with probability mass function

$$p_X(x) = \begin{cases} \frac{1}{2}, & x = 1 \\ \frac{1}{2}, & x = -1. \end{cases} \quad (2.4)$$

The moment generating function of X

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{2}(e^t + e^{-t}) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{t^{2k}}{(k)!} = e^{t^2}, \quad (2.5)$$

so Bernoulli random variable X is subgaussian.

More generally, pregaussian random variable is defined as

Definition 2.1.3 (Pregaussian random variable). A random variable X is called Pregaussian if there exists some $T > 0$ such that the moment generating function of X satisfies

$$M_X(t) \leq e^{\frac{\sigma^2 t^2}{2}} \quad (2.6)$$

for all $t \in [-T, T]$ and some $\sigma \geq 0$.

Pregaussian random variable can be also characterized by the growth rate of its moments.

Proposition 2.1.4 (Pregaussian random variable and growth rate of moments, [10]). *A random variable X is pregaussian if and only if*

$$\mathbb{E}(|X|^{2k}) \leq (2k)! \lambda^{2k} \quad (2.7)$$

for some constant $\lambda > 0$.

Let us also see an example for pregaussian random variable.

Example 2.1.5 (Laplacian random variable). A Laplacian random variable X with mean 0 has its probability distribution function

$$p_X(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right) \quad (2.8)$$

for some $\lambda > 0$, and $\mathbb{E}\left(|X|^{2k}\right) = \Gamma(2k+1)\lambda^{2k} = (2k)!\lambda^{2k}$.

The Bernstein inequality is useful in obtaining estimates on the tail probability of the sum of independent variables, provided the growth rate of the moments of the variables are bounded appropriately.

Theorem 2.1.6 (Bernstein's inequality, [10]). *Let X_1, \dots, X_n be independent random variables. Suppose that $\mathbb{E}X_i = 0$ and*

$$\mathbb{E}|X_i^k| \leq \frac{\mathbb{E}X_i^2}{2} H^{k-2} k!, \quad (2.9)$$

for every positive integer k and some $H > 0$, $i = 1, \dots, n$. Then

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i\right| \geq t\right\} < 2 \exp\left(-\frac{t^2}{2(tH + \sum_{i=1}^n \mathbb{E}X_i^2)}\right), \quad (2.10)$$

for all $t > 0$.

Remark 2.1.7. In particular, let X_1, \dots, X_n be independent pregaussian random variables, then they satisfy the assumptions of Bernstein inequality. Indeed, by $\mathbb{E}|X_i^k| \leq (\mathbb{E}|X_i^{2k}|)^{\frac{1}{2}}$, Proposition 2.1.4 implies (2.9), and using the Taylor expansion of the exponential function in $\mathbb{E}(e^{tX_i})$, one can deduce $\mathbb{E}X_i = 0$, $i = 1, \dots, n$.

2.2 LINEAR COMBINATION OF PREGAUSSIAN RANDOM VARIABLES

Let's consider the linear combination of pregaussian random variables in this section.

Lemma 2.2.1 (Linear combination of pregaussian random variables). *Let a_{ij} , $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, N$, be pregaussian random variables and $\eta_i = \sum_{j=1}^N a_{ij}x_j$, then η_i are pregaussian random variables for $i = 1, 2, \dots, m$.*

Proof. Since a_{ij} are pregaussian random variables, then $\mathbb{E}a_{ij} = 0$ and there are constants $\lambda_{ij} > 0$ such that $\mathbb{E}|a_{ij}|^k \leq k!\lambda_{ij}^k$ for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, N$, then

$$\mathbb{E}\eta_i = \sum_{j=1}^N x_j \mathbb{E}a_{ij} = 0 \quad (2.11)$$

and using the convexity of the function $x \mapsto x^k$ for positive integer k , we have

$$\begin{aligned} \mathbb{E}\left(|\eta_i|^k\right) &= \mathbb{E}\left(\left|\sum_{j=1}^N a_{ij}x_j\right|^k\right) \\ &\leq \|x\|_1^k \mathbb{E}\left(\sum_{j=1}^N \frac{|x_j|}{\|x\|_1} |a_{ij}|\right)^k \\ &\leq \|x\|_1^k \sum_{j=1}^N \left(\frac{|x_j|}{\|x\|_1} \mathbb{E}|a_{ij}|^k\right). \end{aligned} \quad (2.12)$$

Thus the expectations of the pregaussian random variables a_{ij} give that

$$\mathbb{E}\left(|\eta_i|^k\right) \leq \|x\|_1^k \sum_{j=1}^N \frac{|x_j|}{\|x\|_1} k!\lambda_{ij}^k = \|x\|_1^k k! \left(\max_j (\lambda_{ij})\right)^k = k! \left(\|x\|_1 \max_j (\lambda_{ij})\right)^k \quad (2.13)$$

for all integers $k \geq 1$.

□

Note that the product of pregaussian random variables may not be pregaussian. For instance, let's consider the square of a pregaussian random variable. Let $\xi_i := \eta_i^2$. However, ξ_i is not necessary to be pregaussian. For example, let η_i be Laplace random variables with probability density functions

$$p_{\eta_i}(\eta_i) = \frac{1}{2\lambda_i} \exp\left(-\frac{|\eta_i|}{\lambda_i}\right). \quad (2.14)$$

η_i is pregaussian, as we know from the previous example. However, the k -th moment of ξ_i

$$\mathbb{E}\left(|\xi_i|^k\right) = \mathbb{E}\left(|\eta_i|^{2k}\right) = \Gamma(2k+1) \lambda_i^{2k} = (2k)! (\lambda_i^2)^k. \quad (2.15)$$

It has a growth pattern different from $k!\lambda^k$ for any $\lambda > 0$, as $k!\lambda^k = o\left((2k)! (\lambda_i^2)^k\right)$ when k goes to ∞ , that one can see by Stirling's formula in one way. In another way, the moment-generating function of ξ_i is

$$\mathbb{E}(\exp(t\xi_i)) = \mathbb{E}(\exp(t\eta_i^2)) = \int_{-\infty}^{\infty} \frac{1}{2\lambda_i} \exp\left(t\eta_i^2 - \frac{|\eta_i|}{\lambda_i}\right) d\eta_i = \infty \quad (2.16)$$

for any $t > 0$. By the definition or the property of pregaussian random variable, we know that ξ_i is not pregaussian, as its moments grow faster than the moments of any pregaussian random variable.

2.3 THE LARGEST q -SINGULAR VALUE OF PREGAUSSIAN RANDOM MATRICES

2.3.1 LARGEST q -SINGULAR VALUES AND THEIR PROPERTIES

One can define the largest q -singular value by

$$s_1^{(q)}(A) := \sup \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^N \text{ with } x \neq 0 \right\}. \quad (2.17)$$

for an $m \times N$ matrix A . Firstly, one can derive the following

Lemma 2.3.1. *For $q \geq 1$, (2.17) defines a norm on the space of $m \times N$ matrices and*

$$\max_{1 \leq j \leq N} \|a_j\|_q \leq s_1^{(q)}(A) \leq N^{\frac{q-1}{q}} \max_{1 \leq j \leq N} \|a_j\|_q, \quad (2.18)$$

in which a_j , $j = 1, 2, \dots, N$, are the column vectors of A . For $0 < q < 1$, (2.17) defines a quasinorm on the space of $m \times N$ matrices and $\left(s_1^{(q)}(A+B)\right)^q \leq \left(s_1^{(q)}(A)\right)^q + \left(s_1^{(q)}(B)\right)^q$ for any $m \times N$ matrices A and B , moreover, $s_1^{(q)}(A) = \max_{1 \leq j \leq N} \|a_j\|_q$.

Proof. It follows that $s_1^{(q)}(A)$, $q \geq 1$, defines a norm from the known norm ℓ_q for $q \geq 1$.

Moreover, by the Minkowski inequality and discrete Hölder inequality, when $q > 2$,

$$\begin{aligned} \|Ax\|_q &= \left\| \sum_{j=1}^N x_j a_j \right\|_q \leq \sum_{j=1}^N |x_j| \cdot \|a_j\|_q \\ &\leq \|x\|_p \left(\sum_{j=1}^N \|a_j\|_q^q \right)^{\frac{1}{q}} \\ &\leq \|x\|_p N^{\frac{1}{q}} \max_{1 \leq j \leq N} \|a_j\|_q \\ &\leq \|x\|_q N^{\frac{1}{p}} \max_{1 \leq j \leq N} \|a_j\|_q \end{aligned} \quad (2.19)$$

for all $x \in \mathbb{R}^N$, in which $\frac{1}{p} + \frac{1}{q} = 1$, i.e. $p = \frac{q-1}{q}$, and likewise, when $1 \leq q < 2$,

$$\|Ax\|_q \leq \sum_{j=1}^N |x_j| \cdot \|a_j\|_q \leq \|x\|_1 \max_{1 \leq j \leq N} \|a_j\|_q \leq \|x\|_q N^{\frac{1}{p}} \max_{1 \leq j \leq N} \|a_j\|_q \quad (2.20)$$

for all $x \in \mathbb{R}^N$. Thus

$$\|Ax\|_q \leq \|x\|_q N^{\frac{q-1}{q}} \max_{1 \leq j \leq N} \|a_j\|_q \quad (2.21)$$

for $q \geq 1$, which yields $s_1^{(q)}(A) \leq N^{\frac{q-1}{q}} \max_{1 \leq j \leq N} \|a_j\|_q$. Choosing x to be the standard basis vectors of \mathbb{R}^N gives us $\max_{1 \leq j \leq N} \|a_j\|_q \leq s_1^{(q)}(A)$.

The latter part of the claim follows from the fact that ℓ_q for $0 < q < 1$ is a quasinorm and $\|x + y\|_q^q \leq \|x\|_q^q + \|y\|_q^q$ for all $x, y \in \mathbb{R}^N$ if $0 < q < 1$. Similarly to the case when $q \geq 1$, we have for $0 < q < 1$,

$$\|Ax\|_q^q \leq \sum_{j=1}^N |x_j|^q \cdot \|a_j\|_q^q \leq \|x\|_q^q \max_{1 \leq j \leq N} \|a_j\|_q^q, \quad (2.22)$$

which implies $s_1^{(q)}(A) \leq \max_{1 \leq j \leq N} \|a_j\|_q$. On the other hand, we also have $\max_{1 \leq j \leq N} \|a_j\|_q \leq s_1^{(q)}(A)$. Thus $s_1^{(q)}(A) = \max_{1 \leq j \leq N} \|a_j\|_q$ in the case of $0 < q < 1$. \square

Remark 2.3.2. In particular, if $q = 1$, (2.18) implies that $s_1^{(1)}(A) = \max_j \|a_j\|_1$, but if $q = \infty$, (2.18) implies $\max_{1 \leq j \leq N} \|a_j\|_\infty \leq s_1^{(\infty)}(A) \leq N \max_{1 \leq j \leq N} \|a_j\|_\infty$. To have a better estimate on $s_1^{(\infty)}$, we need to consider each component $\sum_{j=1}^N x_j a_{i,j}$, where $(a_{i,j})_{m \times N} := A$, in the column vector $\sum_{j=1}^N x_j a_j$. In this way,

$$\|Ax\|_\infty = \max_i \left| \sum_{j=1}^N x_j a_{i,j} \right| \leq \left(\max_{1 \leq j \leq N} |x_j| \right) \left(\max_{1 \leq i \leq m} \sum_{j=1}^N |a_{i,j}| \right) = \|x\|_\infty \max_{1 \leq i \leq m} \sum_{j=1}^N |a_{i,j}|, \quad (2.23)$$

and one can choose a vector x_0 consisting of ± 1 such that $\|Ax_0\|_\infty = \|x_0\|_\infty$, so $s_1^{(\infty)}(A) = \max_i \sum_{j=1}^N |a_{i,j}|$. The example of $s_1^{(\infty)}$ suggests that it would be more advantageous to use row vectors instead of column vectors in estimating $s_{1,q}$ when q is very large. On the other hand, we can see that $s_1^{(1)}(A) = s_1^{(\infty)}(A^T)$.

In general, for the relation between $s_1^{(q)}$ and $s_1^{(p)}$, $\frac{1}{p} + \frac{1}{q} = 1$, $q \geq 1$, one can deduce the following lemma on the general rectangular matrix.

Lemma 2.3.3. *For any $q \geq 1$ and $m \times N$ matrix A ,*

$$s_1^{(q)}(A) = s_1^{(p)}(A^T) \quad (2.24)$$

in which $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. By the discrete Hölder inequality, we know that if $\frac{1}{p} + \frac{1}{q} = 1$ then

$$|\langle Ax, y \rangle| \leq \|Ax\|_q \quad (2.25)$$

for all $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^m$ with $\|y\|_p = 1$, and furthermore the equality holds for some y_0 with $\|y_0\|_p = 1$. Thus

$$\|Ax\|_q = \sup_{y \in \mathbb{R}^m, \|y\|_p=1} |\langle Ax, y \rangle|. \quad (2.26)$$

By the definition of the largest q -singular value,

$$\begin{aligned} s_1^{(q)}(A) &= \sup_{x \in \mathbb{R}^N, \|x\|_q=1} \|Ax\|_q \\ &= \sup_{x \in \mathbb{R}^N, \|x\|_q=1} \sup_{y \in \mathbb{R}^m, \|y\|_p=1} |\langle Ax, y \rangle|. \end{aligned} \quad (2.27)$$

In the same way, we have

$$s_1^{(p)}(A^T) = \sup_{y \in \mathbb{R}^m, \|y\|_p=1} \sup_{x \in \mathbb{R}^N, \|x\|_q=1} |\langle Ax, y \rangle|. \quad (2.28)$$

Finally, using $\langle Ax, y \rangle = \langle A^T y, x \rangle$ and exchanging taking the supremums, one can get $s_1^{(q)}(A) = s_1^{(p)}(A^T)$. \square

Remark 2.3.4. In the study of the distribution of the largest q -singular value of square random matrices when $q \geq 2$, the above lemma allows us to use the the distribution of the largest singular value $s_1^{(p)}$ for $1 \leq p \leq 2$.

2.3.2 SOME ESTIMATES ON THE UPPER TAIL PROBABILITY OF THE LARGEST q -SINGULAR VALUE

The distribution of the singular values of random matrices has been an interesting topic in recent years, especially for the two extremal singular values, the largest one and the smallest one, for Gaussian random , subgaussian random matrices, etc, see for instance, [48], [44], [59]. Here we are going to study the largest q -singular value for $m \times N$ matrix A are i.i.d.

copies of a symmetric pregaussian random variable as m and N are large. Theorem 5.2.1 tells us that

$$\mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_{f,q}} \geq (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \leq 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2}, \quad (2.29)$$

where

$$\|x\|_{f,q} := \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}}, \quad (2.30)$$

for any $0 < \varepsilon < 1$ and $0 < q \leq 1$, and some $\kappa > 0$ dependent of f and q . To get an estimate on the largest q -singular value, one needs to use the finite cover of the unit sphere in ℓ_q and the union bound for probability.

First we need to convert the probabilities involving $\|x\|_{f,q}$ into those containing $\|x\|_q$ instead.

Lemma 2.3.5. *Given $0 < q \leq 1$, for every $0 < \varepsilon < 1$,*

$$\mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_q} \geq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \leq 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2} \quad (2.31)$$

for some $\kappa > 0$ dependent of f and q .

Proof. By Lemma 5.1.5 and (5.40), we know

$$\|x\|_{f,q} \leq C_q^{\frac{1}{q}} \|x\|_2 \leq C_q^{\frac{1}{q}} \|x\|_q, \quad (2.32)$$

which gives

$$\frac{\|Ax\|_q}{\|x\|_{f,q}} \geq C_q^{-\frac{1}{q}} \frac{\|Ax\|_q}{\|x\|_q}. \quad (2.33)$$

It follows that

$$\begin{aligned} \mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_{f,q}} \geq (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) &\geq \mathbb{P} \left(C_q^{-\frac{1}{q}} \frac{\|Ax\|_q}{\|x\|_q} \geq (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \\ &= \mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_q} \geq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right), \end{aligned} \quad (2.34)$$

and then by (2.29) we obtain

$$\mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_q} \geq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \leq 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2}, \quad (2.35)$$

where $C_q = \frac{(\sqrt{2}\sigma)^q}{\sqrt{\pi}} \Gamma(\frac{q+1}{2})$ (see (5.40)). □

The cardinality of a finite cover of the unit sphere in ℓ_q for $q \geq 1$ and more generally any norm on \mathbb{R}^N are estimates in [41], and for ℓ_q and more generally $\|\cdot\|_{f,q}$ are given in [21]. It allows us to derive the following

Theorem 2.3.6. *Suppose that A is an $m \times N$ matrix whose entries are i.i.d. copies of a pregaussian random variable with an even probability density function f . Then for any given $0 < q \leq 1$, the tail probability for the largest q -singular value*

$$\mathbb{P}\left(s_1^{(q)}(A) \leq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}}\right) \quad (2.36)$$

exponentially decays for any $\varepsilon > 0$, provided that $m \gg N^{1+\frac{q}{2}}$.

Proof. By Lemma 2.2 in [21], there exists a finite set $\mathcal{U}_q \subseteq \mathcal{S}_q$, in which \mathcal{S}_q is the unit ball in \mathbb{R}^N in ℓ_q -norm such that

$$\min_{u \in \mathcal{U}_q} \|x - u\|_q^q \leq \varepsilon \quad (2.37)$$

for all $x \in \mathcal{S}_q$ and

$$\text{card}(\mathcal{U}_q) \leq \left(1 + \frac{2}{\varepsilon}\right)^{\frac{N}{q}}. \quad (2.38)$$

Then for any $x \in \mathcal{S}_q$ there is some $u_0 \in \mathcal{U}_q$, such that $|x - u_0| \leq \varepsilon$. Therefore

$$(s_1^{(q)}(A))^q = \sup_{x \in \mathcal{S}_q} \|Ax\|_q^q \leq \sup_{u \in \mathcal{U}_q} \|Au\|_q^q + \varepsilon^q (s_{1,q}(A))^q, \quad (2.39)$$

which yields

$$s_1^q(A) \leq (1 - \varepsilon^q)^{-\frac{1}{q}} \sup_{u \in \mathcal{U}_q} \|Au\|_q. \quad (2.40)$$

Therefore by (2.38), Lemma 2.3.5, and the union probability,

$$\begin{aligned} \mathbb{P}\left(s_1^{(q)}(A) \geq C_q^{\frac{1}{q}} \left(\frac{1+\varepsilon}{1-\varepsilon^q}\right)^{\frac{1}{q}} m^{\frac{1}{q}}\right) &\leq \mathbb{P}(\sup_{u \in \mathcal{U}_q} \|Au\|_q \geq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}}) \\ &\leq \sum_{u \in \mathcal{U}_q} \mathbb{P}(\|Au\|_q \geq C_q^{\frac{1}{q}} (1 + \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}}) \\ &\leq \left(1 + \frac{2}{\varepsilon}\right)^{\frac{N}{q}} \cdot 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2}. \end{aligned} \quad (2.41)$$

But $(1 + \frac{2}{\varepsilon})^{\frac{N}{q}} \leq e^{\frac{2N}{q\varepsilon}}$, thus

$$\mathbb{P}\left(s_1^{(q)}(A) \geq C_q^{\frac{1}{q}} \left(\frac{1 + \varepsilon}{1 - \varepsilon^q}\right)^{\frac{1}{q}} m^{\frac{1}{q}}\right) \leq 2 \exp\left(-\kappa m N^{-\frac{q}{2}} \varepsilon^2 + \frac{2N}{q\varepsilon}\right). \quad (2.42)$$

Thus if $m \gg N^{1+\frac{q}{2}}$ then

$$\mathbb{P} \left(s_1^{(q)}(A) \geq C_q^{\frac{1}{q}} \left(\frac{1+\varepsilon}{1-\varepsilon^q} \right)^{\frac{1}{q}} m^{\frac{1}{q}} \right) = \mathbb{P} \left(s_{1,q}(A) \geq C_q^{\frac{1}{q}} (1+\varepsilon_0) m^{\frac{1}{q}} \right) \quad (2.43)$$

for $\varepsilon_0 > 0$ exponentially decays. \square

Remark 2.3.7. In [21] it is shown that for $0 < q \leq \frac{1}{3}$,

$$\mathbb{P} \left(\frac{\|Ax\|_q}{\|x\|_{f,q}} \geq (1+\varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \leq 2e^{-\kappa m \varepsilon^2}, \quad (2.44)$$

by the same method of using the finite covering of the unit ball of ℓ_q , we have

$$\mathbb{P} \left(s_1^{(q)}(A) \leq C_q^{\frac{1}{q}} (1+\varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}} \right) \leq 2e^{-cm\varepsilon^2} \quad (2.45)$$

for some $c > 0$ dependent of f and q for $0 < q \leq \frac{1}{3}$.

2.3.3 ESTIMATE ON THE UPPER TAIL PROBABILITY OF THE LARGEST q -SINGULAR VALUE IN GENERAL

To obtain estimates on the upper tail probability of the largest q -singular value for general $0 < q \leq 1$, let us start with the random matrices which are Bernoulli ensemble.

Theorem 2.3.8 (q -singular value of Bernoulli ensemble). *Let ξ be the Bernoulli random variable normalized to have mean 0 and variance 1, and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then $s_1^{(q)}(A) = m^{\frac{1}{q}}$ for all $0 < q \leq 1$.*

Proof. That is because we know from Lemma 2.3.1 that $s_1^{(q)}(A) = \max_j \|a_j\|_q$ for $0 < q \leq 1$ and we have $\|a_j\|_q = m^{\frac{1}{q}}$ for all j since every entry of A is Bernoulli with mean 0 and variance 1. \square

More generally, for pregaussian random variables, we need to consider the distribution or the probability properties of $(\sum_{i=1}^m |a_{ij}|^q)^{\frac{1}{q}}$ when a_{ij} are i.i.d. copies of a pregaussian random variable. The following theorem on $q = 1$ follows from the Bernstein inequality and the estimate of the probability of pregaussian random variables on the domain outside of the m -dimensional cross polytopes, which are also the balls in the ℓ_1 -norm on \mathbb{R}^m .

Theorem 2.3.9 (Upper tail probability of the largest 1-singular value). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times m$ matrix with i.i.d. copies of ξ in its entries, then*

$$\mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right) \leq \exp(-C'm) \quad (2.46)$$

for some $C, C' > 0$ only dependent on the pregaussian random variable ξ .

Proof. Since a_{ij} are i.i.d. copies of the pregaussian random variable ξ , then $\mathbb{E}a_{ij} = 0$ and there exists some $\lambda > 0$, such that $\mathbb{E}|a_{ij}|^k \leq k!\lambda^k$ for all k . Since a_{ij} has variance 1, then $\mathbb{E}a_{ij}^2 = 1$. Therefore

$$\mathbb{E}|a_{ij}^k| \leq \frac{\mathbb{E}a_{ij}^2}{2} H^{k-2} k! \quad (2.47)$$

for $H := 2\lambda^3$ and all $k \geq 2$. By the Bernstein inequality, we know

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_{ij}\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2(m+tH)}\right) = 2 \exp\left(-\frac{t^2}{2(m+2t\lambda^3)}\right) \quad (2.48)$$

for all $t > 0$. In particular, when $t = Cm$,

$$\mathbb{P}\left(\sum_{j=1}^n a_{ij} \geq Cm\right) \leq 2 \exp\left(-\frac{C^2 m}{4C\lambda^3 + 2}\right), \quad (2.49)$$

in which a condition on C will be determined later.

On the other hand, by Lemma 2.3.1,

$$s_1^{(1)}(A) = \max_j \|a_j\|_1 = \sum_{i=1}^m |a_{ij_0}| \quad (2.50)$$

for some j_0 . Furthermore, for any $t > 0$, by the probability of the union,

$$\mathbb{P}\left(\sum_{i=1}^m |a_{ij_0}| \geq t\right) \leq \sum_{(\epsilon_1, \dots, \epsilon_m) \in \{-1, 1\}^m} \mathbb{P}\left(\sum_{i=1}^m \epsilon_i a_{ij_0} \geq t\right). \quad (2.51)$$

But $-a_{ij_0}$ has the same pregaussian properties as a_{ij_0} , precisely, $\mathbb{E}(-a_{ij_0}) = 0$, $\mathbb{E}|-a_{ij_0}|^k \leq k!\lambda^k$. Thus we have

$$\begin{aligned} \mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right) &\leq m \mathbb{P}\left(\sum_{i=1}^m |a_{ij_0}| \geq Cm\right) \\ &\leq 2^{m-1} m \mathbb{P}\left(\left|\sum_{j=1}^n a_{ij_0}\right| \geq Cm\right) \\ &\leq 2^m \exp\left(-\frac{C^2 m}{4C\lambda^3 + 2} + \ln m\right) \\ &= \exp\left(-\left(\frac{C^2}{4C\lambda^3 + 2} - \ln 2e\right) m\right). \end{aligned} \quad (2.52)$$

To have an exponential decay for the probability $\mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right)$, we require that $\frac{C^2}{4C\lambda^3+2} - \ln 2e > 0$, for which

$$C > 2\lambda^3 \ln 2e + \sqrt{4(\ln 2e)^2 \lambda^6 + 2 \ln 2e}. \quad (2.53)$$

Finally, choosing $C' = \frac{C^2}{4C\lambda^3+2} - \ln 2e$, we get (2.46). \square

The previous theorem allows us to estimate the largest q -singular value for $0 < q < 1$.

Theorem 2.3.10 (Upper tail probability of the largest q -singular value, $0 < q < 1$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times m$ matrix with i.i.d. copies of ξ in its entries, then for any $0 < q < 1$,*

$$\mathbb{P}\left(s_1^{(q)}(A) \geq Cm^{\frac{1}{q}}\right) \leq \exp(-C'm) \quad (2.54)$$

for some $C, C' > 0$ only dependent on the pregaussian random variable ξ .

Proof. By the discrete Hölder inequality,

$$\|a_j\|_q^q = \sum_{i=1}^m |a_{ij}|^q \leq m^{1-q} \left(\sum_{i=1}^m |a_{ij}| \right)^q = m^{1-q} \|a_j\|_1^q \quad (2.55)$$

for any $0 < q < 1$, which implies $\|a_j\|_q \leq m^{\frac{1}{q}-1} \|a_j\|_1$. It follows that

$$s_1^{(q)}(A) = \max_j \|a_j\|_q \leq m^{\frac{1}{q}-1} \max_j \|a_j\|_1 = m^{\frac{1}{q}-1} s_1^{(1)}(A). \quad (2.56)$$

From (2.46), we have

$$\begin{aligned} \mathbb{P}\left(s_1^{(q)}(A) \geq Cm^{\frac{1}{q}}\right) &\leq \mathbb{P}\left(m^{\frac{1}{q}-1} s_1^{(1)}(A) \geq Cm^{\frac{1}{q}}\right) \\ &= \mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right) \\ &\leq \exp(-C'm) \end{aligned} \quad (2.57)$$

for some $C, C' > 0$. \square

2.3.4 THE LOWER TAIL PROBABILITY OF THE q -SINGULAR VALUE

For the lower tail probability of $s_1^{(q)}(A)$, we know about it for the Bernoulli ensemble from Theorem 2.3.8, and can derive the exponential decay for

$$\mathbb{P}\left(s_1^{(q)}(A) \leq C_q^{\frac{1}{q}} (1 - \varepsilon)^{\frac{1}{q}} m^{\frac{1}{q}}\right) \quad (2.58)$$

for random matrices whose entries are independent and identically-distributed copies of a pregaussian random variable for any $\varepsilon > 0$ and $0 < q < \frac{1}{3}$ by using finite covering of the unit sphere in ℓ_q and moreover the decay for $0 < q < 1$ under some conditions analogue to the ones Theorem 2.3.6 that we imposed. But to attain a lower tail probability similar to Theorem 2.3.9 for all $0 < q < 1$, the method we used in the proof of Theorem 2.3.9 becomes somewhat infeasible. However, we can still give similar estimates on the lower tail probability with some different techniques.

Lemma 2.3.11 (Linear bound for partial binomial expansion with even integer power). *For every positive integer n ,*

$$\sum_{k=n+1}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} \leq 8x \quad (2.59)$$

for all $x \in [0, 1]$.

Proof. For every $x \in [\frac{1}{8}, 1]$, we have

$$\sum_{k=n+1}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} \leq \sum_{k=0}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} = 1 \leq 8x. \quad (2.60)$$

But for $x \in [0, \frac{1}{8}]$, let

$$f(x) := \sum_{k=n+1}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} \quad (2.61)$$

then

$$f(x) + f(1-x) + \binom{2n}{n} x^n (1-x)^n = 1, \quad (2.62)$$

and in particular,

$$f\left(\frac{1}{2}\right) = \frac{1}{2^{2n}} \sum_{k=n+1}^{2n} \binom{2n}{k}. \quad (2.63)$$

By De Moivre-Stirling's formula (see e.g. [18]) and furthermore the estimate in [42],

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\lambda_n}, \quad (2.64)$$

in which $\frac{1}{12n+1} < \lambda_n < \frac{1}{12n}$. Therefore,

$$\begin{aligned} \binom{2n}{n} &= \frac{\sqrt{2\pi \cdot 2n} \left(\frac{2n}{e}\right)^{2n} e^{\lambda_{2n}}}{\left(\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\lambda_n}\right)^2} \\ &= \frac{4^n}{\sqrt{\pi n}} e^{\lambda_{2n} - 2\lambda_n} \\ &\leq \frac{4^n}{\sqrt{\pi n}} e^{\frac{1-36n}{24n(12n+1)}} \\ &\leq \frac{4^n}{\sqrt{\pi n}}. \end{aligned} \tag{2.65}$$

Since $\binom{2n}{k} \leq \binom{2n}{n}$ for $n+1 \leq k \leq 2n$, then

$$f(x) \leq \sum_{k=n+1}^{2n} \binom{2n}{n} x^k (1-x)^{2n-k} \leq \sum_{k=n+1}^{2n} \binom{2n}{n} x^k \leq n \binom{2n}{n} x^{n+1} \tag{2.66}$$

for all $x \in [0, 1]$. Using (2.65), we then have

$$f(x) \leq 4^n \sqrt{\frac{n}{\pi}} x^{n+1}. \tag{2.67}$$

Now let $g(x) := 4^n \sqrt{\frac{n}{\pi}} x^n$, then

$$\ln(g(x)) = n \ln(4x) + \frac{1}{2} \ln\left(\frac{n}{\pi}\right). \tag{2.68}$$

Thus if $x \in [0, \frac{1}{8}]$,

$$\begin{aligned} \ln(g(x)) &\leq \frac{1}{2} \ln n - n \ln 2 - \frac{1}{2} \ln \pi \\ &\leq \frac{1}{2} n - n \ln 2 \\ &\leq 0. \end{aligned} \tag{2.69}$$

Hence, we obtain

$$f(x) \leq xg(x) \leq x \leq 8x \tag{2.70}$$

for all $x \in [0, \frac{1}{8}]$. \square

Remark 2.3.12. The coefficient in the linear bound can be definitely improved, because one can give sharper estimates for (2.69) on an interval with the right endpoint larger than $\frac{1}{8}$ and thus the coefficient in the linear bound will be less than 8. But for our purpose of using it to estimate the probabilities later, the linear bound obtained will be sufficient.

Considering odd integers together, we have in general

Lemma 2.3.13 (Linear bound for partial binomial expansion). *For every positive integer n ,*

$$\sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} x^k (1-x)^{n-k} \leq 8x \quad (2.71)$$

for all $x \in [0, 1]$.

To prove this lemma, we can modify the proof of Lemma 2.3.11.

Proof. It suffices to show that for every positive integer n ,

$$\sum_{k=n+1}^{2n+1} \binom{2n+1}{k} x^k (1-x)^{2n-k+1} \leq 8x \quad (2.72)$$

for all $x \in [0, 1]$.

Same as (2.60),

$$\sum_{k=n+1}^{2n+1} \binom{2n+1}{k} x^k (1-x)^{2n-k+1} \leq 1 \leq 8x. \quad (2.73)$$

for all $x \in [\frac{1}{8}, 1]$.

Let $f_1(x) := \sum_{k=n+1}^{2n+1} \binom{2n+1}{k} x^k (1-x)^{2n-k+1}$, then

$$f_1(x) \leq n \binom{2n+1}{n+1} x^{n+1} \leq n \binom{2n+2}{n+1} x^{n+1} \quad (2.74)$$

for $x \in [0, 1]$, similar to (2.66). Taking advantage of (2.65), we have

$$\binom{2n+2}{n+1} \leq \frac{4^{n+1}}{\sqrt{\pi(n+1)}}. \quad (2.75)$$

It follows that

$$f_1(x) \leq \frac{4^{n+1}n}{\sqrt{\pi(n+1)}} x^{n+1}. \quad (2.76)$$

Let $g_1(x) := \frac{4^{n+1}n}{\sqrt{\pi(n+1)}}x^n$, then

$$\begin{aligned} \ln(g_1(x)) &= (n+1)\ln 4 + n\ln x + \ln n - \frac{1}{2}\ln(n+1) - \frac{1}{2}\ln \pi \\ &\leq n\ln(4x) + \frac{1}{2}\ln(n+1) + \ln 4 - \frac{1}{2}\ln \pi \\ &\leq n\ln(4x) + \frac{1}{2}n + \ln 4 - \frac{1}{2}\ln \pi. \end{aligned} \quad (2.77)$$

Thus if $x \in [0, \frac{1}{8}]$,

$$\begin{aligned} \ln(g_1(x)) &\leq \frac{1}{2}n - n\ln 2 + \ln 4 - \frac{1}{2}\ln \pi \\ &\leq \ln 4 - \frac{1}{2}\ln \pi \\ &\leq \ln 8. \end{aligned} \quad (2.78)$$

So

$$f_1(x) = xg_1(x) \leq 8x \quad (2.79)$$

for all $x \in [0, \frac{1}{8}]$, and that completes the proof. \square

The above lemma can be applied to estimate probabilities.

Lemma 2.3.14. *Suppose $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d copies of a random variable ξ , then for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^n |\xi_i| \leq \frac{n\varepsilon}{2}\right) \leq 8\mathbb{P}(|\xi| \leq \varepsilon). \quad (2.80)$$

Proof. First, we have the relation on the probability events that

$$\left\{(\xi_1, \dots, \xi_n) : \sum_{i=1}^n |\xi_i| \leq \frac{n\varepsilon}{2}\right\} \quad (2.81)$$

is contained in

$$\bigcup_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \{(\xi_1, \dots, \xi_n) : |\xi_{i_1}| \leq \varepsilon, \dots, |\xi_{i_k}| \leq \varepsilon, |\xi_{i_{k+1}}| > \varepsilon, \dots, |\xi_{i_n}| > \varepsilon\} := \mathcal{E} \quad (2.82)$$

where $\{i_1, i_2, \dots, i_k\}$ is a subset of $\{1, 2, \dots, n\}$ and $\{i_{k+1}, \dots, i_n\}$ is its complement.

Let $x = \mathbb{P}(|\xi| \leq \varepsilon)$, then by the union probability,

$$\mathbb{P}(\mathcal{E}) = \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} x^k (1-x)^{n-k}, \quad (2.83)$$

and applying Lemma 2.3.13, we have

$$\mathbb{P}(\mathcal{E}) \leq 8x = 8\mathbb{P}(|\xi_1| \leq \varepsilon). \quad (2.84)$$

Since the event (2.81) is contained in the event (2.82),

$$\mathbb{P}\left(\sum_{i=1}^n |\xi_i| \leq \frac{n\varepsilon}{2}\right) \leq \mathbb{P}(\mathcal{E}) \leq 8\mathbb{P}(|\xi_1| \leq \varepsilon). \quad (2.85)$$

□

Now we are ready to give a lower tail probability for the 1-singular values.

Theorem 2.3.15 (Lower tail probability of the largest 1-singular value). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for any $\varepsilon > 0$, there exists $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(1)}(A) \leq Km\right) \leq \varepsilon \quad (2.86)$$

in which K only depends on ε and the pregaussian random variable ξ .

Proof. Since a_{ij} is pregaussian with variance 1, then for any $\varepsilon > 0$, there is some $\delta > 0$, such that

$$\mathbb{P}(|a_{ij}| \leq \delta) \leq \frac{\varepsilon}{8}. \quad (2.87)$$

But we know

$$s_1^{(1)}(A) = \sum_{i=1}^m |a_{ij_0}| \quad (2.88)$$

for some j_0 . Therefore, by Lemma 2.3.14,

$$\mathbb{P}\left(s_1^{(1)}(A) \leq \frac{\delta}{2} \cdot m\right) = \mathbb{P}\left(\sum_{i=1}^m |a_{ij_0}| \leq \frac{m\delta}{2}\right) \leq 8\mathbb{P}(|a_{ij}| \leq \delta) \leq \varepsilon. \quad (2.89)$$

Thus letting $K = \frac{\delta}{2}$, we obtain (2.46).

For general $0 < q < 1$, we have

□

Theorem 2.3.16 (Lower tail probability of the largest q -singular value). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for any $\varepsilon > 0$, there exists $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right) \leq \varepsilon \quad (2.90)$$

in which K only depends on q , ε and the pregaussian random variable ξ .

Proof. We can use the same method used in the proof of Theorem 2.3.15. First, for any $\varepsilon > 0$, there is some $\delta > 0$, such that

$$\mathbb{P}(|a_{ij}|^q \leq \delta) \leq \frac{\varepsilon}{8}. \quad (2.91)$$

Therefore, By Lemma 2.3.1,

$$s_1^{(q)}(A) = \left(\sum_{i=1}^m |a_{ij_0}|^q\right)^{\frac{1}{q}} \quad (2.92)$$

for some j_0 . Then by Lemma 2.3.14,

$$\mathbb{P}\left(s_1^{(q)}(A) \leq \left(\frac{\delta}{2}\right)^{\frac{1}{q}} m^{\frac{1}{q}}\right) = \mathbb{P}\left(\sum_{i=1}^m |a_{ij_0}|^q \leq \frac{m\delta}{2}\right) \leq 8\mathbb{P}(|a_{ij}|^q \leq \delta) \leq \varepsilon. \quad (2.93)$$

Thus let $K = \left(\frac{\delta}{2}\right)^{\frac{1}{q}}$, then (2.90) follows. \square

Remark 2.3.17. From Theorem 2.3.10 and Theorem 2.3.16, we know that $s_1^{(q)}(A) \sim m^{\frac{1}{q}}$ in probability for $m \times m$ pregaussian random matrix A .

CHAPTER 3

PROBABILISTIC ESTIMATE ON THE SMALLEST q -SINGULAR VALUES OF PREGAUSSIAN RANDOM MATRICES

In this section, we will derive some probability tail bounds for the smallest singular value for pregaussian random matrices in ℓ_2 -norm first, and then we'll make some generalizations to the smallest singular value for pregaussian random matrices ℓ_q .

3.1 SINGULAR VALUE OF PREGAUSSIAN RANDOM MATRICES IN ℓ_2

As we know, pregaussian random variable is a more general type of random variables than subgaussian random variable. A subgaussian random variable ξ has its moment generating function

$$\mathbb{E}e^{t\xi} = e^{O(t^2)} \quad (3.1)$$

for all $t \in (-\infty, \infty)$. By relaxing the condition on the domain of t , one can define a pregaussian random variable. Specifically, (3.1) holds on a bounded and centered interval for a pregaussian random variable.

As pointed out by Buldygin and Kozachenko in [10], pregaussian random variable has the following property

Lemma 3.1.1. *If ξ is a pregaussian random variable, then for any $p > 0$, $\mathbb{E}|\xi|^p < \infty$, and $\mathbb{E}\xi = 0$.*

Proof. By the definition of pregaussian random variable, $\mathbb{E}e^{t\xi} = e^{O(t^2)}$ for all $t \in [-T, T]$ for some $T > 0$, then we have

$$\mathbb{E}e^{|T\xi|} \leq \mathbb{E}e^{T\xi} + \mathbb{E}e^{-T\xi} = 2e^{O(T^2)} < \infty, \quad (3.2)$$

Let $n := \lfloor p \rfloor + 1$. Then applying Hölder's inequality we have

$$\mathbb{E}|\xi|^p \leq (\mathbb{E}(|\xi|^p)^{\frac{n}{p}})^{\frac{p}{n}} (\mathbb{E}1^{\frac{n}{n-p}})^{\frac{n-p}{n}} = (\mathbb{E}|\xi|^n)^{\frac{p}{n}} \quad (3.3)$$

But it follows from (3.2) that $\mathbb{E}|\xi|^n \leq \frac{n!}{T^n} \mathbb{E}e^{T|\xi|} < \infty$. So $\mathbb{E}|\xi|^p < \infty$ for any $p > 0$.

Let $L = \inf \left\{ c > 0 : \mathbb{E}e^{t\xi} \leq e^{c\lambda^2}, t \in [-T, T] \right\}$, then

$$\mathbb{E}e^{t\xi} \leq e^{Lt^2} \quad (3.4)$$

for all $t \in [-T, T]$. However, $\frac{d}{dt}|_{t=0} \mathbb{E}e^{t\xi} = \mathbb{E}\xi$ gives that

$$\mathbb{E}e^{t\xi} = 1 + t\mathbb{E}\xi + o(\lambda) \quad (3.5)$$

for small λ , and by the Taylor expansion,

$$e^{Lt^2} = 1 + Lt^2 + o(t^2). \quad (3.6)$$

Thus (3.4) implies

$$t\mathbb{E}\xi + o(t) \leq Lt^2 + o(t^2). \quad (3.7)$$

Dividing (3.7) by $t > 0$ and taking $t \rightarrow 0^+$, we get $\mathbb{E}\xi = 0$. \square

Thus ξ has mean 0, and the moment bound condition is satisfied sufficiently for pregaussian random variable ξ .

In [59], Tao and Vu proved the following theorem on the universality for the least singular value,

Theorem 3.1.2 (Tao-Vu). *Let ξ be \mathbb{R} -normalized, and suppose $\mathbb{E}|\xi|^{C_0} < \infty$ for some sufficiently large absolute constant C_0 . Then for all $t > 0$, we have*

$$\mathbb{P}(n\sigma_n(M_n(\xi))^2 \leq t) = \int_0^t \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-(x/2 + \sqrt{x})} dx + O(n^{-c}) \quad (3.8)$$

where $c > 0$ is an absolute constant. The implied constants in the $O(\cdot)$ notation depend on $\mathbb{E}|\xi|^{C_0}$ but are uniform in t .

By Lemma 3.1.1, we know $\mathbb{E}|\xi|^{C_0} < \infty$ if ξ is pregaussian. Additionally normalizing ξ to have variance 1, by the theorem of the distribution of smallest singular values, we have the following

Corollary 3.1.3. *Let ξ be pregaussian random variable with variance 1 and $M_n(\xi)$ be $n \times n$ matrix whose entries are i.i.d. copies of ξ . Then*

$$\mathbb{P}(s_n(M_n(\xi)) > \frac{t}{\sqrt{n}}) = e^{-t^2/2-t} + O(n^{-c}) \quad (3.9)$$

for all $t > 0$ and some absolute constant $c > 0$.

3.2 THE q -SINGULAR VALUES OF AN $m \times n$ MATRIX

Now let's consider in general the k -th q -singular value first. The k -th q -singular value is defined in [49] as

Definition 3.2.1. The k -th q -singular value of an $m \times n$ matrix A ,

$$s_k^{(q)}(A) := \inf \left\{ \sup \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in V \setminus \{0\} \right\} : V \subseteq \mathbb{R}^n, \dim(V) \geq n - k + 1 \right\}. \quad (3.10)$$

Remark 3.2.2. Since the norm of A restricted on a subspace is less or equal to the norm of A restricted on another subspace which contains the first subspace,

$$s_k^{(q)}(A) = \inf_{V \subseteq \mathbb{R}^n, \dim(V)=n-k+1} \sup_{x \in V, \|x\|_q=1} \|Ax\|_q. \quad (3.11)$$

Also, it is easy to see that

$$s_1^{(q)}(A) \geq s_2^{(q)}(A) \geq \dots \geq s_{\min(m,n)}^{(q)}(A) \geq 0. \quad (3.12)$$

The two extremal q -singular values, $s_{1,q}$ and $s_{\min(m,n),q}$ are of special interest in various studies, for which the largest q -singular value has been discussed earlier and the smallest q -singular value will be studied in the following sections.

If $m \geq n$, then the n -th q -singular value is the smallest q -singular value, which can also be expressed in another way.

Lemma 3.2.3. *Let A be a $m \times n$ matrix with $m \geq n$, then the smallest q -singular value*

$$s_n^{(q)}(A) = \inf \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\}. \quad (3.13)$$

Proof. By the definition

$$\begin{aligned} s_n^{(q)}(A) &= \inf \left\{ \sup \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in V \setminus \{0\} \right\} : V \subseteq \mathbb{R}^n, \dim(V) \geq 1 \right\} \\ &\leq \inf \left\{ \sup \left\{ \frac{\|Av\|_q}{\|v\|_q} : v \in V \setminus \{0\} \right\} : V = \text{span}(x) : x \in \mathbb{R}^n \setminus \{0\} \right\} \\ &= \inf \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\}. \end{aligned} \quad (3.14)$$

We also know the infimum can be achieved by considering the unit ℓ_q -sphere in the finite dimensional space, and so the claim follows. \square

Remark 3.2.4. In particular, if A is an $n \times n$ invertible matrix, then

$$\begin{aligned} s_n^{(q)}(A) &= \inf \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\} \\ &= \frac{1}{\sup \left\{ \frac{\|A^{-1}x\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\}} \\ &= \frac{1}{s_1^{(q)}(A^{-1})}. \end{aligned} \quad (3.15)$$

In compressed sensing, we usually have a matrix with dimension $m \times N$, $m \ll N$, in the underdetermined system. So let us make a remark on the singular value of a $m \times n$ matrix when $m < n$.

Remark 3.2.5. If $m < n$, in some literature, for instance [49], it is put that k runs from 1 through n . In fact, in this case, by the definition of k -th q -singular value we know that $s_{k,2} = 0$ for $k = m+1, \dots, n$ because there are $n-m$ linearly independent vectors in the null space of the $m \times n$ matrix with $m < n$, which can span an $n-k+1$ -dimensional subspace of \mathbb{R}^n for $k = m+1, \dots, n$. On the other hand, the diagonal matrix in the singular value decomposition of an $m < n$ matrix with $m < n$ in ℓ_2 does not have the diagonal entries $s_{k,2}$, $k = m+1, \dots, n$, so here we assume that the smallest q -singular value is the m -th q -th singular value but not the n -th q -th singular value for the case $m < n$.

There are some properties on the smallest singular value $s_{\min(m,n),q}$, which are given in the following lemmas.

Lemma 3.2.6 (Full rank). *An $m \times n$ matrix A has full rank, if and only if $s_{\min(m,n)}^{(q)}(A) > 0$.*

Proof. If $m \geq n$, by Lemma 3.2.3, we have

$$\begin{aligned} s_{\min(m,n)}^{(q)}(A) &= s_n^{(q)}(A) \\ &= \inf \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\} \\ &= \inf \left\{ \|Ax\|_q : \|x\|_q = 1 \right\}. \end{aligned} \quad (3.16)$$

But $\mathcal{S}_q := \{x \in \mathbb{R}^n : \|x\|_q = 1\}$ is compact and the map $x \rightarrow \|Ax\|_q$ is continuous, then $\|Ax\|_q$ achieves its minimum at some $x_0 \in \mathcal{S}_q$. Thus

$$s_{\min(m,n)}^{(q)}(A) = \|Ax_0\|_q > 0 \quad (3.17)$$

because A has rank n in this case.

If $m < n$, then the null space of A , denoted as \mathcal{N} , has dimension $n - m$ since A has rank m in this case. Therefore for any $V \subseteq \mathbb{R}^n$ with $\dim(V) \geq n - m + 1$, we have

$$\dim(\mathcal{N}^\perp) + \dim(V) \geq (n - (n - m)) + (n - m + 1) = n + 1 > \dim(\mathbb{R}^n). \quad (3.18)$$

By the inclusion-exclusion principle, we know that there a vector $x_V \in V$ with $\|x_V\|_q = 1$ such that $x_V \in \mathcal{N}^\perp \cap V \cap \mathcal{S}_q$ and so $\|Ax_V\|_q > 0$. Hence by the definition of the m -th singular value,

$$\begin{aligned} s_{\min(m,n)}^{(q)}(A) &= s_m^{(q)}(A) \\ &\geq \inf \left\{ \|Ax_V\|_q : V \subseteq \mathbb{R}^n, \dim(V) \geq n - m + 1 \right\} \\ &\geq \inf \left\{ \|Ax\|_q : x \in \mathcal{N}^\perp \cap \mathcal{S}_q \right\} \\ &> 0 \end{aligned} \quad (3.19)$$

since $\mathcal{N}^\perp \cap \mathcal{S}_q$ is compact.

For the converse, assume that the $m \times n$ matrix A has rank not larger than $\min(m, n) - 1$, then the null space \mathcal{N} has dimension not less than $n - \min(m, n) + 1$. Hence

$$\sup \left\{ \frac{\|Ax\|_q}{\|x\|_q} : x \in V \setminus \{0\} \right\} = 0 \quad (3.20)$$

and $\dim(V) \geq n - \min(m, n) + 1$ when V is the null space. Thus by the definition $s_{\min(m,n)}^{(q)}(A) = 0$. So if $s_{\min(m,n)}^{(q)}(A) > 0$ then A has full rank. \square

Furthermore, in matrix approximation theory, the singular values of a matrix are related to the matrices of a certain rank closest to it.

Theorem 3.2.7 (Schmidt-Mirsky, [2]). *Let $s_k^{(2)}(A)$ be the k -th singular value of an $m \times N$ matrix A in ℓ_2 . Then*

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = s_{k+1}^{(2)}(A). \quad (3.21)$$

We have shown that $\max_j \|a_j\|_q \leq s_1^{(q)}(A) \leq n^{\frac{q-1}{q}} \max_j \|a_j\|_q$ in Lemma 2.3.1 for the largest q -singular value for an $m \times n$ matrix A . Nevertheless, the smallest q -singular value does not rely much on a single column or row of the matrix, though we have the very rudimentary estimate $s_{\min(m,n)}^{(q)}(A) \leq \min_j \|a_j\|_q$ since one can just choose x to be the basis vectors of \mathbb{R}^N such that $\|Ax\|_q = \min_j \|a_j\|_q$ and then choose the subspace in \mathbb{R}^n spanned by x and other $m - n$ linearly independent null vectors if $m \leq n$.

Analogue to Lemma 2.3.3, for the smallest q -singular value, we have

Lemma 3.2.8. *For any $q \geq 1$ and $n \times n$ matrix A ,*

$$s_n^{(p)}(A) = s_n^{(q)}(A^T) \quad (3.22)$$

in which $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We consider the following two cases.

One case is $s_n^{(p)}(A) = 0$. By Lemma 3.2.6, we know that A is not full rank, and then so does A^T . Applying Lemma 3.2.6 again, $s_n^{(q)}(A^T) = 0$.

The other case is $s_n^{(p)}(A) > 0$, then A is invertible, by Remark 3.2.4 and Lemma 2.3.3, we know the claim holds. \square

An immediate corollary of this lemma for rectangular matrices is

Corollary 3.2.9. *For any $q \geq 1$ and $m \times n$ matrix A which has $n - m$ columns if $m \leq n$ or $m - n$ rows if $m \geq n$ are zeroes,*

$$s_{\min(m,n)}^{(p)}(A) = s_{\min(m,n)}^{(q)}(A^T) \quad (3.23)$$

in which $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. The claim will follow from the lemma, by using the natural embedding of $\mathbb{R}^{\min(m,n)}$ into $\mathbb{R}^{\max(m,n)}$ and the natural projection from $\mathbb{R}^{\max(m,n)}$ onto $\mathbb{R}^{\min(m,n)}$.

Indeed, without loss of generality, we assume that $m \leq n$ and the last $\min(m, n) = m$ columns are zeroes. Let A_1 be the submatrix of A formed by the first m columns. Because of the compactness of the sphere of unit ℓ_p -norm and Lemma 3.2.8, there are some $x \in \mathbb{R}^m$ with $\|x_0\|_p = 1$ and $y \in \mathbb{R}^m$ with $\|y_0\|_q = 1$ such that

$$\|A_1 x_0\|_p = \inf_{x \in \mathbb{R}^m, \|x\|_p=1} \|A_1 x\|_p = \inf_{y \in \mathbb{R}^m, \|y\|_q=1} \|A_1^T y\|_q = \|A_1^T y_0\|_q. \quad (3.24)$$

Now let $\tilde{x}_0 \in \mathbb{R}^n$ be the extended vector whose first m components are those of x and other components are zeroes, and $\tilde{\mathbb{R}}^m$ be the subspace of \mathbb{R}^n consists of the vectors whose first m components are zeroes, and take $V_0 := \text{span}(\tilde{x}_0) \oplus \tilde{\mathbb{R}}^m$. Since for any $\zeta \in \tilde{\mathbb{R}}^m$,

$$\frac{\|A(\tilde{x}_0 + \zeta)\|_p}{\|(\tilde{x}_0 + \zeta)\|_p} = \frac{\|A\tilde{x}_0\|_p}{\|\tilde{x}_0\|_p} \leq \frac{\|A\tilde{x}_0\|_p}{\|\tilde{x}_0\|_p} = \|A\tilde{x}_0\|_p = \|A_1\tilde{x}_0\|_p, \quad (3.25)$$

then $\sup_{x \in V_0, \|x\|_p=1} \|Ax\|_p = \|A_1\tilde{x}_0\|_p$. For any $V \subseteq \mathbb{R}^n$ with $\dim(V) = n - m + 1$, there is some $x' \in V$ with $\|x'\|_p = 1$ whose last $n - m$ components are zeroes, because of the dimensions, then we know

$$\sup_{x \in V, \|x\|_p=1} \|Ax\|_p \geq \|Ax'\|_p = \|A_1\bar{x}'\|_p \geq \|A_1\tilde{x}_0\|_p \quad (3.26)$$

where \bar{x}' is the vector in \mathbb{R}^m whose m components are the first m components of x' . Thus

$$s_m^{(p)}(A) = \inf_{V \subseteq \mathbb{R}^m, \dim(V)=n-m+1} \sup_{x \in V, \|x\|_p=1} \|Ax\|_p = \|A_1\tilde{x}_0\|_p. \quad (3.27)$$

On the other hand, $\|A^T y\|_q = \|A_1^T y\|_q$ for any $y \in \mathbb{R}^m$ with $\|y\|_q = 1$. Hence

$$s_m^{(q)}(A^T) = \inf_{y \in \mathbb{R}^m, \|y\|_q=1} \|A^T y\|_q = \inf_{y \in \mathbb{R}^m, \|y\|_q=1} \|A_1^T y\|_q = \|A_1^T y_0\|_q = \|A_1\tilde{x}_0\|_p, \quad (3.28)$$

by (3.24), and then the claim follows. □

For general rectangular matrices, we have the duality property for all 2-singular values.

Theorem 3.2.10 (Adjoint). *For any $m \times n$ matrix A ,*

$$s_k^{(2)}(A) = s_k^{(2)}(A^T) \quad (3.29)$$

for $k = 1, 2, \dots, \min(m, n)$.

Proof. Without loss of generality, we assume $m \leq n$.

If A is diagonal with $\lambda_1, \lambda_2, \dots, \lambda_m, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$, on its diagonal. Let $\{e_i : i = 1, 2, \dots, n\}$ be the standard basis for \mathbb{R}^n , and V_0 be the subspace spanned by $\{e_i : i = k, k+1, \dots, n\}$. For any $v := \sum_{i=k}^n t_i e_i \in V_0$ and $v \neq 0$,

$$\frac{\|Av\|_2}{\|v\|_2} = \sqrt{\frac{\sum_{i=k}^m \lambda_i^2 t_i^2}{\sum_{i=k}^n t_i^2}} \leq \sqrt{\frac{\sum_{i=k}^m \lambda_k^2 t_i^2}{\sum_{i=k}^n t_i^2}} = \lambda_k \sqrt{\frac{\sum_{i=k}^m t_i^2}{\sum_{i=k}^n t_i^2}} \leq \lambda_k. \quad (3.30)$$

Particularly, $\frac{\|Ae_k\|_2}{\|e_k\|_2} = \lambda_k$. So we have

$$\sup_{v \in V_0, \|v\|_2=1} \|Av\|_2 = \lambda_k. \quad (3.31)$$

Let V_1 be the subspace spanned by $\{e_i : i = 1, 2, \dots, k\}$, then for any $(n-k+1)$ -dimensional subspace V in \mathbb{R}^n , there exists some nonzero vector $v' \in V_1 \cap V$, since $\dim(V) + \dim(V_1) = n+1$. Thus there are $t_i \in \mathbb{R}, i = 1, 2, \dots, k$, which are not all zero, such that $v' = \sum_{i=1}^k t_i e_i$. Then we have

$$\sup_{v' \in V, \|v'\|_2=1} \|Av'\|_2 \geq \frac{\|Av'\|_2}{\|v'\|_2} = \sqrt{\frac{\sum_{i=1}^k \lambda_i^2 t_i^2}{\sum_{i=1}^k t_i^2}} \geq \sqrt{\frac{\sum_{i=1}^k \lambda_k^2 t_i^2}{\sum_{i=1}^k t_i^2}} = \lambda_k. \quad (3.32)$$

So $s_k^{(2)}(A) = \lambda_k$. Similarly, we have $s_k^{(2)}(A^T) = \lambda_k$, since A^T is also diagonal with $\lambda_1, \lambda_2, \dots, \lambda_m$ on its diagonal. Hence, $s_k^{(2)}(A) = s_k^{(2)}(A^T)$ for $k = 1, 2, \dots, \min(m, n)$ if A is diagonal.

For general matrix A of size $m \times n$, let $U\Lambda V^T := A$ be the singular value decomposition of A . Replace all the subspaces in the above argument by the image of the subspaces transformed by the orthogonal matrix V , and then the claim follows, since the orthogonal transformation U preserves the ℓ_2 -norm of every vector.

□

For general $p \geq 1$ and general rectangular matrices, we have the following duality theorem.

Theorem 3.2.11 (Duality). *For any $p \geq 1$ and $m \times n$ matrix A ,*

$$s_{\min(m,n)}^{(p)}(A) = s_{\min(m,n)}^{(q)}(A^T) \quad (3.33)$$

in which $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Without loss of generality, we assume $m < n$, since we have proved the case of $m = n$ in Lemma 3.2.8.

If $s_{\min(m,n)}^{(p)}(A) = 0$, then $s_{\min(m,n)}^{(q)}(A^T) = 0$ by Lemma 3.2.6.

If $s_{\min(m,n)}^{(p)}(A) \neq 0$, then A is of full rank, by Lemma 3.2.6. Assume $p > 1$ first. Since the spaces we are considering are finite dimensional, there exists some $w_0 \in \mathbb{R}^m$ with $\|w_0\|_q = 1$, such that

$$\|A^T w_0\|_q = \inf_{w \in \mathbb{R}^m, \|w\|_q=1} \|A^T w\|_q. \quad (3.34)$$

By Hölder's inequality,

$$\|A^T w_0\|_q \geq \langle A^T w_0, v \rangle \quad (3.35)$$

for all $v \in \mathbb{R}^n$ with $\|v\|_p = 1$, and there is some $v_0 \in \mathbb{R}^n$ with $\|v_0\|_p = 1$ such that

$$\|A^T w_0\|_q = \langle A^T w_0, v_0 \rangle. \quad (3.36)$$

Now take V_0 to be the direct sum of $\text{span}\{v_0\}$ and $\ker(A)$, that is an $(n - m + 1)$ -dimensional subspace in \mathbb{R}^n since $\ker(A)$ is $(n - m)$ -dimensional, as A has full rank. Thus it follows from (3.35) and (3.36) that for any $z \in \ker(A)$,

$$\langle A^T w_0, v_0 \rangle \geq \left\langle A^T w_0, \frac{v_0 + z}{\|v_0 + z\|_p} \right\rangle = \frac{1}{\|v_0 + z\|_p} \langle A^T w_0, v_0 + z \rangle. \quad (3.37)$$

But we also know

$$\langle A^T w_0, v_0 + z \rangle = \langle w_0, A(v_0 + z) \rangle = \langle w_0, Av_0 \rangle = \langle A^T w_0, v_0 \rangle, \quad (3.38)$$

thus

$$\langle A^T w_0, v_0 \rangle \geq \frac{1}{\|v_0 + z\|_p} \langle A^T w_0, v_0 \rangle, \quad (3.39)$$

in other words, $\|v_0 + z\|_p \geq 1$. Therefore we have

$$\frac{\|A(v_0 + z)\|_p}{\|v_0 + z\|_p} \leq \|Av_0\|_p \quad (3.40)$$

when considering the action of A on V_0 , and then $\sup_{v \in V_0, \|v\|_p=1} \|Av\|_q = \|Av_0\|_p$.

Next, we are going to show $\|Av_0\|_p = \|A^T w_0\|_q$. Since we know that $\langle w_0, Av_0 \rangle = \langle A^T w_0, v_0 \rangle = \|A^T w_0\|_q$ from (3.36), then it is sufficient to show that for any $w \in \mathbb{R}^m$ with $\|w\|_q = 1$, $\langle w_0, Av_0 \rangle \geq \langle w, Av_0 \rangle$, that is to show

$$\langle A^T w_0, v_0 \rangle \geq \langle A^T w, v_0 \rangle. \quad (3.41)$$

Let

$$\mathcal{S}_q(1) := \left\{ w \in \mathbb{R}^m : \|w\|_q = 1 \right\} \quad (3.42)$$

be the unit sphere in ℓ_q -norm in \mathbb{R}^m ,

$$\mathcal{S}_q\left(\|A^T w_0\|_q\right) := \left\{ v \in A^T(\mathbb{R}^m) : \|v\|_q = \|A^T w_0\|_q \right\} \quad (3.43)$$

be the sphere of radius $\|A^T w_0\|_q$ in ℓ_q -norm in $A^T(\mathbb{R}^m)$, and $A^T(\mathcal{S}_q(1))$ be the image of $\mathcal{S}_q(1)$ by map A^T . Since $\inf_{w \in \mathbb{R}^m, \|w\|_q=1} \|A^T w\|_q$ achieves its infimum at w_0 , then the surfaces $A^T(\mathcal{S}_q(1))$ and $\mathcal{S}_q\left(\|A^T w_0\|_q\right)$ are tangent at the point $A^T w_0$ in $A^T(\mathbb{R}^m)$ and let P be the same tangent plane of dimension $m-1$ to the surfaces $A^T(\mathcal{S}_q)$ and $\mathcal{S}_q\left(\|A^T w_0\|_q\right)$ at $A^T w_0$ in $A^T(\mathbb{R}^m)$. In fact, by (3.36), v_0 is the gradient of the function $f(u) := \|u\|_q$, $u \in \mathbb{R}^n$, at the point $u = A^T w_0$. Therefore, v_0 is orthogonal to P . Let $\mathcal{B}_q(1) := \left\{ w \in \mathbb{R}^m : \|w\|_q \leq 1 \right\}$ be the unit ball in ℓ_q -norm in \mathbb{R}^m , which is convex, as $q \geq 1$, then its image by the linear map A^T , $A^T(\mathcal{B}_q(1))$, is also convex. Therefore $A^T(\mathcal{B}_q(1))$ will be on the same side of P in $A^T(\mathbb{R}^m)$. Hence for any $w \in \mathbb{R}^m$ with $\|w\|_q = 1$, if $\langle A^T w, v_0 \rangle > 0$ then there is some $\lambda \geq 1$ such that $\lambda w \in P$, as v_0 is orthogonal to P , and it follows that

$$\langle A^T w, v_0 \rangle \leq \langle A^T(\lambda w), v_0 \rangle = \langle A^T w_0, v_0 \rangle + \langle A^T(\lambda w - w_0), v_0 \rangle = \langle A^T w_0, v_0 \rangle \quad (3.44)$$

as $\lambda w - w_0$ in P is orthogonal to v_0 ; if $\langle A^T w, v_0 \rangle \leq 0$ it is obvious that $\langle A^T w_0, v_0 \rangle \geq \langle A^T w, v_0 \rangle$, since $\langle A^T w_0, v_0 \rangle = \|A^T w_0\|_q \geq 0$. Thus (3.41) holds for all $w \in \mathbb{R}^m$ with $\|w\|_q = 1$. This proves that

$$s_m^{(p)}(A) \leq s_m^{(q)}(A^T). \quad (3.45)$$

On the other hand, there is a surface \mathcal{S} of dimension m in \mathbb{R}^n such that for any $z \in \ker(A)$ and $v \in \mathcal{S}$, $\|v + z\|_p \geq \|v\|_p$. We know that there is some $v_0 \in \mathcal{S}$ with $\|v_0\|_p = 1$ such that

$$\|Av_0\|_p = \inf_{v \in \mathcal{S}, \|v\|_p=1} \|Av\|_p, \quad (3.46)$$

and for any $(n - m + 1)$ -dimensional subspace V in \mathbb{R}^n , there is some $v' \in V \cap \mathcal{S}$ with $\|v'\|_p = 1$, because $\dim(V) + \dim(\mathcal{S}) = n + 1$ and \mathcal{S} is centrally symmetric, and then

$$\sup_{v \in V, \|v\|_p=1} \|Av\|_p \geq \|Av'\|_p \geq \|Av_0\|_p. \quad (3.47)$$

Thus $s_m^{(p)}(A) = \|Av_0\|_p$. Similar to the previous argument, by Hölder's inequality,

$$\|Av_0\|_p \geq \langle w, Av_0 \rangle \quad (3.48)$$

for all $w \in \mathbb{R}^n$ with $\|w\|_p = 1$, and there is some $w_0 \in \mathbb{R}^2$ with $\|w_0\|_q = 1$ such that

$$\|Av_0\|_p = \langle w_0, Av_0 \rangle. \quad (3.49)$$

Thus for the opposite direction of the inequality in (3.45), we just need to show $\|A^T w_0\|_q = \|Av_0\|_p$, that is to show

$$\langle w_0, Av_0 \rangle \geq \langle w_0, Av \rangle \quad (3.50)$$

for all $v \in \mathbb{R}^n$ with $\|v\|_p = 1$. But then we can use a very similar argument with the one for (3.41) to show this.

For $p = 1$, by the continuity of ℓ_p at $p = 1$ and as (3.33) holds for $p > 1$, we know that $s_{\min(m,n)}^{(1)}(A) = s_{\min(m,n)}^{(\infty)}(A^T)$. That completes the proof. □

Remark 3.2.12. For some computable examples on this theorem, see Appendix A.

3.3 THE SMALLEST q -SINGULAR VALUE OF RANDOM MATRICES

If $0 < q < 1$, we can still define the k -th singular value by (3.10), and Lemma 3.2.3 still holds for finite dimensional vector spaces. In [43] and [58], the estimates on the smallest singular value defined by the usual ℓ_2 -norm for square random matrix and the sum of a random matrix and a deterministic matrix were given. In this section, we study the probability estimates of the smallest q -singular value of rectangular random matrices and square random matrices.

For rectangular matrices, we are going to give the following estimate on the smallest q -singular value,

Theorem 3.3.1. *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $m \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $\varepsilon > 0$ there exist some $\gamma > 0$ and $c > 0$ and $r \in (0, 1)$ dependent on q and ε , such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}\right) < e^{-cm} \quad (3.51)$$

if $n < rm$.

To prove the above theorem, it is sufficient to show the following lemma, by considering the probability in union.

Lemma 3.3.2. *Given any $0 < q \leq 1$ and let A be an $m \times n$ pregaussian random matrix, then there exist some $c, \lambda \in (0, 1)$ such that*

$$\mathbb{P}\left(\|Av\|_q < \lambda_1 m^{\frac{1}{q}}\right) \leq c^m \quad (3.52)$$

for each $v \in \mathbb{S}_q^{m-1}$, where \mathbb{S}_q^{m-1} is the $(m-1)$ -dimensional unit sphere in ℓ_q -quasinorm.

First let us establish the following

Lemma 3.3.3. *Given any $0 < q \leq 1$, then for any ξ_1, \dots, ξ_n which are i.i.d. copies of a pregaussian random variable with variance 1, there exist $c, \lambda \in (0, 1)$ such that*

$$\mathbb{P}\left(\sum_{k=1}^n |\xi_k|^q < \lambda n\right) \leq c^n. \quad (3.53)$$

Proof. Same as the beginning of the proof for Theorem 2.3.15, for pregaussian random variables ξ_1, \dots, ξ_n with variance 1, we know that there exists some $\delta > 0$, such that

$$\varepsilon_0 := \mathbb{P}(|\xi_k| \leq \delta) < \frac{1}{2e} \quad (3.54)$$

for $k = 1, 2, \dots, n$. Then using the Riemann–Stieltjes integral for expectation, we have

$$\begin{aligned} \mathbb{E} \exp\left(-\frac{|\xi_k|^q}{\lambda}\right) &= \int_0^\infty \exp\left(-\frac{|\xi_k|^q}{\lambda}\right) d\mathbb{P}(|\xi_k| \leq t) \\ &\leq \int_0^\delta d\mathbb{P}(|\xi_k| \leq t) + \int_\delta^\infty \exp\left(-\frac{t^q}{\lambda}\right) d\mathbb{P}(|\xi_k| \leq t) \\ &= \varepsilon_0 + \int_\delta^\infty \exp\left(-\frac{t^q}{\lambda}\right) d\mathbb{P}(|\xi_k| \leq t). \end{aligned} \quad (3.55)$$

Choose $\lambda > 0$ to be small enough, such that

$$\exp\left(-\frac{t^q}{\lambda}\right) \leq \exp\left(-\frac{\delta^q}{\lambda}\right) < \frac{\varepsilon_0}{2(1-\varepsilon_0)} \quad (3.56)$$

for $t \geq \delta$. Therefore, it follows that

$$\mathbb{E} \exp\left(-\frac{|\xi_k|^q}{\lambda}\right) \leq \varepsilon_0 + \frac{\varepsilon_0}{2(1-\varepsilon_0)} \int_\delta^\infty d\mathbb{P}(|\xi_k| \leq t) \leq \varepsilon_0 + \frac{\varepsilon_0}{2} = \frac{3}{2}\varepsilon_0 < \frac{3}{4e}. \quad (3.57)$$

Finally, applying Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}\left(\sum_{k=1}^n |\xi_k|^q < \lambda n\right) &= \mathbb{P}\left(\exp\left(n - \frac{1}{\lambda} \sum_{k=1}^n |\xi_k|^q\right) > 1\right) \\ &\leq \mathbb{E}\left(\exp\left(n - \frac{1}{\lambda} \sum_{k=1}^n |\xi_k|^q\right)\right) \\ &= e^n \prod_{k=1}^n \mathbb{E} \exp\left(-\frac{|\xi_k|^q}{\lambda}\right) \\ &\leq c^n \end{aligned} \quad (3.58)$$

for $c := \frac{3}{2}\varepsilon_0 \cdot e < \frac{3}{4}$. □

The property of linear combination of pregaussian random variables will allow us to obtain the probabilistic estimate on $\|Av\|_q$ for the pregaussian ensemble A . Indeed, the combination of Lemma 3.3.3 and the lemma on linear combination of pregaussian random variables, Lemma 2.2.1 yields Lemma 3.3.2. Using the estimate on the covering number of the unit sphere in ℓ_q -quasinorm by ε -net ([21]) and the probabilistic union bound, we are going to prove Theorem 3.3.1.

Proof. Let the unit sphere in ℓ_q -quasinorm \mathbb{S}_q^n be covered by ε -net \mathcal{N} , we know from [21], that

$$|\mathcal{N}| \leq \left(1 + \frac{2}{\varepsilon}\right)^{\frac{n}{q}}. \quad (3.59)$$

By the probabilistic union bound,

$$\mathbb{P}\left(\exists v \in \mathcal{N} \text{ such that } \|Av\|_q < \lambda_1 m^{\frac{1}{q}}\right) \leq \left(1 + \frac{2}{\varepsilon}\right)^{\frac{n}{q}} c^m. \quad (3.60)$$

On the other hand, by Theorem 2.3.10, we know that

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right) \geq 1 - \exp(-C'm) \quad (3.61)$$

for some $C' > 0$.

Now assume the event that $\|Av'\|_q < \frac{1}{2}\lambda_1 m^{\frac{1}{q}}$ for some $v' \in \mathbb{S}_q^n$ occurs. Therefore, there exists some $v \in \mathcal{N}$ such that $\|v - v'\|_q^q \leq \varepsilon$, and by the triangular inequality,

$$\begin{aligned} \|Av\|_q^q &\leq \|A(v - v')\|_q^q + \|Av'\|_q^q \\ &\leq \left(s_1^{(q)}(A)\right)^q \|v - v'\|_q^q + \|Av'\|_q^q \\ &\leq K^q m \varepsilon + \frac{1}{2^q} \lambda_1^q m \\ &= \lambda_1^q m \end{aligned} \quad (3.62)$$

if we set $\varepsilon = \frac{\lambda_1}{2K}$. It follows that

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}} \text{ and } \|Av'\|_q < \frac{\lambda_1}{2} m^{\frac{1}{q}} \text{ for some } v' \in \mathbb{S}_q^n\right) \leq \mathbb{P}\left(\|Av\|_q < \lambda_1 m^{\frac{1}{q}}\right). \quad (3.63)$$

But the event $s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}$, where $\gamma := \frac{1}{2}\lambda_1$, implies $\|Av'\|_q < \frac{1}{2}\lambda_1 m^{\frac{1}{q}}$ for some $v' \in \mathbb{S}_q^n$, therefore

$$\begin{aligned} \mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}} \text{ and } s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}\right) &\leq \mathbb{P}\left(\|Av\|_q < \lambda_1 m^{\frac{1}{q}}\right) \\ &\leq \left(1 + \frac{2}{\varepsilon}\right)^{\frac{n}{q}} c^m \\ &= \left(\left(1 + \frac{2}{\varepsilon}\right)^{\frac{n}{qm}} c\right)^m \end{aligned} \quad (3.64)$$

by (3.60). Choosing an appropriate $r < 1$, we have that if $n < rm$,

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}} \text{ and } s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}\right) \leq e^{-c_1 m} \quad (3.65)$$

for some $c_1 > 0$.

Finally, the claim that

$$\mathbb{P}\left(s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}\right) \leq e^{-cm} \quad (3.66)$$

for some $c > 0$ follows from (3.65) and (3.61). \square

Now let us consider the square random matrices whose entries are independent and identically-distributed copies of a pregaussian random variable.

Theorem 3.3.4. *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $n \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $\varepsilon > 0$ and $0 < q \leq 1$, there exist some $K > 0$ and $c > 0$ dependent on q and ε , such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \varepsilon n^{-\frac{1}{q}}\right) < C\varepsilon + C\alpha^n + \mathbb{P}\left(\|A\| > Kn^{-\frac{1}{2}}\right). \quad (3.67)$$

where $\alpha \in (0, 1)$ and $C > 0$ depend only on the pregaussian random variable and K .

To prove the above theorem, that is to estimate $\|Av\|_q$ for all $v \in \mathbb{R}^n$, we decompose \mathbb{S}_q^{n-1} into the set of compressible vectors and the set of incompressible vectors, generalized from the approach used in [43], [58], for instance.

Definition 3.3.5 (Compressible and incompressible vectors in \mathbb{S}_q^{n-1}). Given any $\rho, \lambda \in (0, 1)$. Let $Comp_q(\lambda, \rho)$ be the set of vectors $v \in \mathbb{S}_q^{n-1}$ such that there is a vector v' with $\|v'\|_0 \leq \lambda n$ satisfying $\|v - v'\|_q \leq \rho$. The set of incompressible vector is defined as

$$Incomp_q(\lambda, \rho) := \mathbb{S}_q^{n-1} \setminus Comp_q(\lambda, \rho). \quad (3.68)$$

Now using the decomposition, we have

$$\begin{aligned} \mathbb{P}\left(s_n^{(q)}(A) < \varepsilon n^{-\frac{1}{q}}\right) &\leq \mathbb{P}\left(\inf_{v \in Comp_q(\lambda, \rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}}\right) \\ &\quad + \mathbb{P}\left(\inf_{v \in Incomp_q(\lambda, \rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}}\right), \end{aligned} \quad (3.69)$$

and in the following we are going to consider each term in the right hand side.

For the first term on compressible vectors, we can apply Lemma 3.3.2 and use the union bound to get the probabilistic estimate on $\|Av\|_q$ with $v \in \text{Comp}_q(\lambda, \rho)$, as

$$\mathbb{P} \left(\inf_{v \in \text{Comp}_q(\lambda, \rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}} \right) \leq \mathbb{P} \left(\inf_{v \in \text{Comp}_q(\lambda, \rho)} \|Av\|_q < \varepsilon n^{\frac{1}{q}} \right). \quad (3.70)$$

Therefore the first term actually decays exponentially for large n .

However, for incompressible vectors, we first consider $\text{dis}(X_j, H_j)$, which denotes the distance between column X_j of an $n \times n$ random matrix A and the span of other columns $H_j := \text{span}(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$, and obtain a generalized lemma for ℓ_q -quasinorm from the counterpart in [43], which allows us to transform the probabilistic estimate on $\|Av\|_q$ for $v \in \text{Incomp}_q(\lambda, \rho)$ to the probabilistic estimate on the average of the distances $\text{dist}(X_j, H_j)$, $j = 1, 2, \dots, n$.

Lemma 3.3.6. *Let A be an $n \times n$ random matrix with columns X_1, \dots, X_n , and $H_j := \text{span}(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$. Then for any $\rho, \lambda \in (0, 1)$ and $\varepsilon > 0$, one has*

$$\mathbb{P} \left(\inf_{v \in \text{Incomp}_q(\lambda, \rho)} \|Av\|_q < \varepsilon \rho n^{-\frac{1}{q}} \right) < \frac{1}{\lambda n} \sum_{j=1}^n \mathbb{P}(\text{dist}_q(X_j, H_j) < \varepsilon), \quad (3.71)$$

in which dist_q is the distance defined by the ℓ_q -quasinorm.

Proof. For every $v \in \text{Incomp}_q(\lambda, \rho)$, by Definition 3.3.5, there are at least λn components of v , v_j , satisfying $|v_j| \geq \rho n^{-\frac{1}{q}}$, because otherwise, v would be within ℓ_q -distance ρ of the sparse vector, the restriction of v on the components v_j satisfying $|v_j| \geq \rho n^{-\frac{1}{q}}$ with sparsity less than λn , and thus v would be compressible. Thus if we let $\mathcal{I}_1(v) := \left\{ j : |v_j| \geq \rho n^{-\frac{1}{q}} \right\}$, then $|\mathcal{I}_1(v)| \geq \lambda n$.

Next, let $\mathcal{I}_2(A) := \{j : \text{dist}_q(X_j, H_j) < \varepsilon\}$ and \mathcal{E} be the event that the cardinality of $\mathcal{I}_2(A)$, $|\mathcal{I}_2(A)| \geq \lambda n$. Applying Markov's inequality, we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &= \mathbb{P}(\{ \mathcal{I}_2(A) : |\mathcal{I}_2(A, \varepsilon)| \geq \lambda n \}) \\ &\leq \frac{1}{\lambda n} \mathbb{E} |\mathcal{I}_2(A)| \\ &= \frac{1}{\lambda n} \mathbb{E} \{ j : \text{dist}_q(X_j, H_j) < \varepsilon \} \\ &= \frac{1}{\lambda n} \sum_{j=1}^n \mathbb{P}(\text{dist}_q(X_j, H_j) < \varepsilon). \end{aligned} \quad (3.72)$$

Since \mathcal{E}^c is the event that

$$|\{j : \text{dist}_q(X_j, H_j) \geq \varepsilon\}| > (1 - \lambda)n \quad (3.73)$$

for random matrix A , thus if \mathcal{E}^c occurs, then for every $v \in \text{Incomp}_q(\lambda, \rho)$,

$$|\mathcal{I}_1(v)| + |\mathcal{I}_2(A)| > \lambda n + (1 - \lambda)n = n. \quad (3.74)$$

Hence there is some $j_0 \in \mathcal{I}_1(v) \cap \mathcal{I}_2(A)$. So we have

$$\|Av\|_q \geq \text{dist}_q(Av, H_{j_0}) = \text{dist}_q(v_{j_0}X_{j_0}, H_{j_0}) = |v_{j_0}| \text{dist}_q(X_{j_0}, H_{j_0}) \geq \varepsilon \rho n^{-\frac{1}{q}}. \quad (3.75)$$

Contra-positively, if the events $\|Av\|_q < \varepsilon \rho n^{-\frac{1}{q}}$ occurs then \mathcal{E} also occurs. Thus

$$\mathbb{P}\left(\inf_{v \in \text{Incomp}_q(\lambda, \rho)} \|Av\|_q < \varepsilon \rho n^{-\frac{1}{q}}\right) \leq \mathbb{P}(\mathcal{E}) \leq \frac{1}{\lambda n} \sum_{j=1}^n \mathbb{P}(\text{dist}_q(X_j, H_j) < \varepsilon). \quad (3.76)$$

□

Remark 3.3.7. dist_q defined by the ℓ_q -quasinorm is different from the usual dist defined by the ℓ_2 -quasinorm, because simply in a right triangle the hypotenuse as a vector can be shorter than the leg as a vector in ℓ_q -quasinorm.

Nevertheless, $\text{dist}_q(X_j, H_j) \geq \text{dist}(X_j, H_j)$ because $\|\cdot\|_q \geq \|\cdot\|_2$, therefore one can take the advantage of the estimate on $\mathbb{P}(\text{dist}(X_j, H_j) < \varepsilon)$ to obtain the estimate on $\mathbb{P}(\text{dist}_q(X_j, H_j) < \varepsilon)$.

Theorem 3.3.8 (Distance bound, [43]). *Let A be a random matrix whose entries are independent variables with variance at least 1 and fourth moment bounded by B . Let $K \geq 1$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\text{dist}(X_j, H_j) < \varepsilon \text{ and } \|A\| \leq Kn^{-\frac{1}{2}}\right) \leq C(\varepsilon + \alpha^n) \quad (3.77)$$

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

The above theorem implies that

$$\mathbb{P}(\text{dist}_q(X_j, H_j) < \varepsilon) \leq \mathbb{P}(\text{dist}(X_j, H_j) < \varepsilon) \leq C(\varepsilon + \alpha^n) + \mathbb{P}(\|A\| \leq Kn^{-\frac{1}{2}}). \quad (3.78)$$

Combining (3.69) and applying Lemma 3.3.6, we now reach the desired inequality in Theorem 3.3.4.

Furthermore, since A is pregaussian, using a standard concentration bound we know that for every $\varepsilon > 0$, there exists some $K > 0$ such that $\mathbb{P}(\|A\| \leq Kn^{-\frac{1}{2}}) < \varepsilon$. Thus, we have proved the following

Theorem 3.3.9 (Lower tail probabilistic estimate on the smallest q -singular value). *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $n \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $\varepsilon > 0$, there exists some $\gamma > 0$ such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \gamma n^{-\frac{1}{q}}\right) < \varepsilon, \quad (3.79)$$

where γ only depends on q , ε and the pregaussian random variable ξ .

3.4 UPPER TAIL PROBABILITY OF THE SMALLEST q -SINGULAR VALUE

In this section, we want to obtain the estimate on the upper tail probability of the smallest q -singular value of an $n \times n$ pregaussian random matrix.

Theorem 3.4.1 (Upper tail probabilistic estimate on the smallest q -singular value). *Given any $0 < q \leq 1$, and let ξ be the pregaussian random variable with variance 1 and A be an $n \times n$ matrix with i.i.d. copies of ξ in its entries. Then for any $K > e$, there exist some $C > 0$, $0 < c < 1$, and $\alpha > 0$ only dependent on pregaussian random variable ξ , q , such that*

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-\frac{1}{2}}\right) \leq \frac{C(\ln K)^\alpha}{K^\alpha} + c^n. \quad (3.80)$$

In particular, for any $\varepsilon > 0$, there exist some $K > 0$ and n_0 , such that

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-\frac{1}{2}}\right) < \varepsilon \quad (3.81)$$

for all $n \geq n_0$.

Proof. From the previous section and by Lemma 3.2.6, the $n \times n$ pregaussian random matrix A is invertible with very high probability. Therefore, we have

$$\mathbb{P}\left(s_n^{(q)}(A) \leq \frac{ut}{\varepsilon} \cdot n^{-\frac{1}{2}}\right) \geq \mathbb{P}\left(\|v\|_q \leq n^{\frac{1}{q}-\frac{1}{2}}u, \|A^{-1}v\|_q \geq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}} \text{ for some } v \in \mathbb{R}^n\right), \quad (3.82)$$

and thus it suffices to show

$$\mathbb{P}\left(\|v\|_q \leq n^{\frac{1}{q}-\frac{1}{2}}u, \|A^{-1}v\|_q \geq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}} \text{ for some } v \in \mathbb{R}^n\right) \geq 1 - \varepsilon. \quad (3.83)$$

Now we choose $v = X_j - \pi_j(X_j)$, where X_j is the j -th column vector of A and π_j is the projection onto the hyperplane $H_j := \text{span}(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$, to have a small probability for both the event $\|A^{-1}v\|_q \leq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}}$ and the event $\|v\|_q \geq u$ in (3.82).

Using the result established in [44], we can easily estimate the probability of the event that $\|A^{-1}v\|_q \leq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}}$ occurs. Since $\|A^{-1}v\|_q \geq \|A^{-1}v\|_2$, we know that

$$\begin{aligned} \mathbb{P}\left(\|A^{-1}v\|_q \leq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}}\right) &\leq \mathbb{P}\left(\|A^{-1}v\|_2 \leq \frac{\varepsilon}{t} \cdot n^{\frac{1}{q}}\right) \\ &= \mathbb{P}\left(\|A^{-1}v\|_2 \leq \frac{\varepsilon}{t} \cdot \left(n^{\frac{2}{q}}\right)^{\frac{1}{2}}\right) \\ &\leq 2p\left(4\varepsilon, t, n^{\frac{2}{q}}\right) \end{aligned} \quad (3.84)$$

where $p(\varepsilon, t, n) := C_1 \left(\varepsilon + e^{-c_1 t^2} + e^{-c_2 n}\right)$ for some $c_1, c_2, C_1 > 0$.

It remains to show that $\|v\|_q \geq u$ in (3.82) has a small probability, and that will be shown in the next lemma. \square

In fact, for the normal vector to the hyperplane H_j , $v = X_j - \pi_j(X_j)$, we have

Lemma 3.4.2. *For every $u > 0$, one has*

$$\mathbb{P}\left(\|X_j - \pi_j(X_j)\|_q \geq un^{\frac{1}{q}-\frac{1}{2}}\right) \leq C_2 e^{-c_3 u} + C_3 n^{-c_4} \quad (3.85)$$

for some $c_3, c_4, C_2, C_3 > 0$ and any $j = 1, 2, \dots, n$.

Proof. Without loss of generality, assume $j = 1$. Let $(a_1, a_2, \dots, a_n) := X_1 - \pi_1(X_1)$, normal to the hyperplane H_1 , and $(\xi_1, \xi_2, \dots, \xi_n) := X_1$.

Applying the Bessy-Esseen theorem (see for instance [54]), we know that

$$\mathbb{P}(\|X_j - \pi_j(X_j)\|_2 \geq u) = \mathbb{P}\left(\left|\frac{\sum_{i=1}^n a_i \xi_i}{\sqrt{\sum_{i=1}^n a_i^2}}\right| \geq u\right) = \mathbb{P}(|\mathbf{g}| \geq u) + O(n^{-c}) \quad (3.86)$$

for some $c > 0$, in which \mathbf{g} is a standard normal random variable.

By the discrete Hölder inequality,

$$\|X_j - \pi_j(X_j)\|_q \leq n^{\frac{1-q}{q}} \|X_j - \pi_j(X_j)\|_1 \leq n^{\frac{1}{q}-\frac{1}{2}} \|X_j - \pi_j(X_j)\|_2, \quad (3.87)$$

and then

$$\begin{aligned} \mathbb{P}\left(\|X_j - \pi_j(X_j)\|_q \geq n^{\frac{1}{q}-\frac{1}{2}} u\right) &\leq \mathbb{P}\left(n^{\frac{1}{q}-\frac{1}{2}} \|X_j - \pi_j(X_j)\|_2 \geq n^{\frac{1}{q}-\frac{1}{2}} u\right) \\ &= \mathbb{P}(\|X_j - \pi_j(X_j)\|_2 \geq u). \end{aligned} \quad (3.88)$$

Therefore it follows from (3.86) that

$$\begin{aligned} \mathbb{P}\left(\|X_j - \pi_j(X_j)\|_q \geq u n^{\frac{1}{q}-\frac{1}{2}}\right) &\leq \mathbb{P}(|\mathbf{g}| \geq u) + O(n^{-c}) \\ &= \frac{2}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{1}{2}x^2} dx + O(n^{-c}) \\ &\leq C_2 e^{-c_3 u} + C_3 n^{-c_4} \end{aligned} \quad (3.89)$$

for some $c_3, c_4, C_2, C_3 > 0$. □

Now we can choose $u = t = \sqrt{\ln M}$, where $M > 1$, and $\varepsilon = \frac{1}{M}$, then (3.82), (3.84), and (3.85) imply that

$$\mathbb{P}\left(s_n^{(q)}(A) > M \ln M \cdot n^{-\frac{1}{2}}\right) \leq \frac{C}{M^\alpha} + c^n \quad (3.90)$$

for some $C > 0$, $0 < c < 1$, and $\alpha > 0$. Then let $K := M \ln M$, thus we have

$$\mathbb{P}\left(s_n^{(q)}(A) > K n^{-\frac{1}{2}}\right) \leq \frac{C (\ln M)^\alpha}{K^\alpha} + c^n \leq \frac{C (\ln(M \ln M))^\alpha}{K^\alpha} + c^n = \frac{C (\ln K)^\alpha}{K^\alpha} + c^n \quad (3.91)$$

if $M \geq e$, that requires $K > e$.

Thus, we have completed the proof of Theorem 3.4.1.

Lastly, we make a table to summarize the results of the probabilistic estimates on the largest and smallest q -Singular values of pregaussian random square matrices for $0 < q \leq 1$ (see Table 3.1).

Table 3.1: Results obtained on the probabilistic estimates on the largest and smallest q -singular values of pregaussian random square matrices for $0 < q \leq 1$

| $n \times n$ matrix | Order of largest q -singular value | Order of smallest q -singular value |
|---------------------|--------------------------------------|--|
| Upper Tail | $n^{\frac{1}{q}}$ | $n^{-\frac{1}{2}}$ ($n^{-\frac{1}{q}}$ conjectured) |
| Lower Tail | $n^{\frac{1}{q}}$ | $n^{-\frac{1}{q}}$ |

CHAPTER 4

THE PROBABILISTIC ESTIMATES ON THE LARGEST p -SINGULAR VALUE OF PREGAUSSIAN RANDOM MATRICES FOR $p > 1$

4.1 INTRODUCTION

The largest and smallest q -singular values of pregaussian random matrices for $0 < q \leq 1$ have been studied in [32]. Similar to the q -singular value when $0 < q \leq 1$, we can define the largest p -singular value when $p > 1$.

Definition 4.1.1. For an $m \times N$ matrix A , the largest p -singular value of A denoted as $s_1^{(p)}(A)$ is defined as

$$s_1^{(p)}(A) := \sup \left\{ \frac{\|Ax\|_p}{\|x\|_p} : x \in \mathbb{R}^N \text{ with } x \neq 0 \right\} \quad (4.1)$$

for given $p > 1$.

4.2 LOWER TAIL PROBABILITY OF LARGEST p -SINGULAR VALUE FOR $p > 1$

In [32] we have established the following

Lemma 4.2.1 (Linear bound for partial binomial expansion). *For every positive integer n ,*

$$\sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} x^k (1-x)^{n-k} \leq 8x \quad (4.2)$$

for all $x \in [0, 1]$.

The above lemma can be applied to estimate probabilities.

Lemma 4.2.2. *Suppose $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d copies of a random variable ξ , then for any $\varepsilon > 0$,*

$$\mathbb{P} \left(\sum_{i=1}^n |\xi_i|^p \leq \frac{n\varepsilon}{2} \right) \leq 8\mathbb{P} (|\xi| \leq \varepsilon) \quad (4.3)$$

for any given $p > 1$.

Proof. Given $p > 1$, we have the relation on the probability events that

$$\left\{ (\xi_1, \dots, \xi_n) : \sum_{i=1}^n |\xi_i|^p \leq \frac{n\varepsilon}{2} \right\} \quad (4.4)$$

is contained in

$$\bigcup_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \{ (\xi_1, \dots, \xi_n) : |\xi_{i_1}|^p \leq \varepsilon, \dots, |\xi_{i_k}|^p \leq \varepsilon, |\xi_{i_{k+1}}|^p > \varepsilon, \dots, |\xi_{i_n}|^p > \varepsilon \} := \mathcal{E} \quad (4.5)$$

where $\{i_1, i_2, \dots, i_k\}$ is a subset of $\{1, 2, \dots, n\}$ and $\{i_{k+1}, \dots, i_n\}$ is its complement.

Let $x = \mathbb{P} (|\xi_1|^p \leq \varepsilon)$, then by the union probability,

$$\mathbb{P} (\mathcal{E}) = \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} x^k (1-x)^{n-k}, \quad (4.6)$$

and applying Lemma 4.2.1, we have

$$\mathbb{P} (\mathcal{E}) \leq 8x = 8\mathbb{P} (|\xi_1| \leq \varepsilon). \quad (4.7)$$

Since the event (4.4) is contained in the event (4.5),

$$\mathbb{P} \left(\sum_{i=1}^n |\xi_i|^p \leq \frac{n\varepsilon}{2} \right) \leq \mathbb{P} (\mathcal{E}) \leq 8\mathbb{P} (|\xi_1|^p \leq \varepsilon). \quad (4.8)$$

□

To estimate the lower tail probability of the largest p -singular value, we have

Theorem 4.2.3 (Lower tail probability of the largest p -singular value, $p > 1$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $p > 1$ and any $\varepsilon > 0$, there exists $\gamma > 0$ such that*

$$\mathbb{P} \left(s_1^{(p)}(A) \leq \gamma m^{\frac{1}{p}} \right) \leq \varepsilon \quad (4.9)$$

in which γ only depends on p, ε and the pregaussian random variable ξ .

Proof. Since a_{ij} is pregaussian with variance 1, then any $\varepsilon > 0$, there is some $\delta > 0$, such that

$$\mathbb{P}(|a_{ij}|^p \leq \delta) \leq \frac{\varepsilon}{8}. \quad (4.10)$$

But we know

$$s_1^{(p)}(A) \geq \left(\sum_{i=1}^m |a_{ij}|^p \right)^{\frac{1}{p}} \quad (4.11)$$

for all j , because by the definition of the largest p -singular value 4.1.1, choosing x to be the standard basis vectors of \mathbb{R}^N gives us $\max_j (\sum_{i=1}^m |a_{ij}|^p)^{\frac{1}{p}} \leq s_1^{(p)}(A)$. Therefore, by Lemma 4.2.2,

$$\mathbb{P} \left(s_1^{(p)}(A) \leq \left(\frac{\delta}{2} \right)^{\frac{1}{p}} m^{\frac{1}{p}} \right) \leq \mathbb{P} \left(\sum_{i=1}^m |a_{ij_0}|^p \leq \frac{m\delta}{2} \right) \leq 8\mathbb{P}(|a_{ij}|^p \leq \delta) \leq \varepsilon. \quad (4.12)$$

Thus let $\gamma = \left(\frac{\delta}{2} \right)^{\frac{1}{p}}$, then (4.9) follows. \square

4.3 UPPER TAIL PROBABILITY OF THE LARGEST p -SINGULAR VALUE FOR $p > 1$

For the upper tail probability of the largest p -singular value, $p > 1$, we can derive the following lemma first by using the Minkowski inequality and discrete Hölder inequality.

Lemma 4.3.1. *For $p \geq 1$, (2.17) defines a norm on the space of $m \times N$ matrices and*

$$\max_j \|a_j\|_p \leq s_1^{(p)}(A) \leq N^{\frac{p-1}{p}} \max_j \|a_j\|_p, \quad (4.13)$$

in which a_j , $j = 1, 2, \dots, N$, are the column vectors of A .

Applying the above lemma, an estimate we can derive easily is the following

Theorem 4.3.2 (Upper tail probability of the largest p -singular value of Bernoulli random matrices, $p > 1$). *Let ξ be a Bernoulli random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then*

$$m^{\frac{1}{p}} \leq s_1^{(p)}(A) \leq m^{\frac{1}{p}} N^{\frac{p-1}{p}} \quad (4.14)$$

For the more general rectangular matrices, we have

Theorem 4.3.3 (Upper tail probability of the largest p -singular value of rectangular matrices, $1 < p \leq 2$). *Let ξ be a pregaussian variable normalized to have variance 1 and A is an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $1 < p \leq 2$ and any $\varepsilon > 0$, there exists $K > 0$ such that*

$$\mathbb{P} \left(s_1^{(p)}(A) \geq K \left(m^{\frac{1}{p}} + m^{\frac{1}{p}-\frac{1}{2}} N^{\frac{1}{2}} \right) \right) \leq \varepsilon \quad (4.15)$$

where K only depends on p , ε and the pregaussian variable ξ .

Proof. By the discrete Hölder inequality and the definition of the largest p -singular value,

$$s_1^{(p)}(A) = \sup_{x \in \mathbb{R}^N, x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \leq \sup_{x \in \mathbb{R}^N, x \neq 0} \frac{m^{\frac{1}{p}-\frac{1}{2}} \|Ax\|_2}{\|x\|_2} = m^{\frac{1}{p}-\frac{1}{2}} s_1^{(2)}(A). \quad (4.16)$$

We also know that there exists $K > 0$ such that

$$\mathbb{P} \left(s_1^{(2)}(A) \geq K \left(m^{\frac{1}{2}} + N^{\frac{1}{2}} \right) \right) \leq \varepsilon. \quad (4.17)$$

Therefore, we have

$$\mathbb{P} \left(s_1^{(p)}(A) \geq K \left(m^{\frac{1}{p}} + m^{\frac{1}{p}-\frac{1}{2}} N^{\frac{1}{2}} \right) \right) \leq \mathbb{P} \left(s_1^{(2)}(A) \geq K \left(m^{\frac{1}{2}} + N^{\frac{1}{2}} \right) \right) \leq \varepsilon. \quad (4.18)$$

□

Using the duality lemma on the largest p -singular value, we have

Theorem 4.3.4 (Lower tail probability of the largest p -singular value of rectangular matrices, $p > 2$). *Let ξ be a pregaussian random variable normalized to have variance 1 and A be an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $p > 2$ and any $\varepsilon > 0$, there exists $\gamma > 0$ such that*

$$\mathbb{P} \left(s_1^{(p)}(A) \leq \gamma m^{\frac{p-1}{p}} \right) \leq \varepsilon \quad (4.19)$$

in which γ only depends on p , ε and the pregaussian random variable ξ .

Also, we have the upper tail probability of the largest p -singular value of rectangular matrices for $p > 2$.

Theorem 4.3.5 (Upper tail probability of the largest p -singular value of rectangular matrices, $p > 2$). *Let ξ be a pregaussian variable normalized to have variance 1 and A is an $m \times N$ matrix with i.i.d. copies of ξ in its entries, then for every $p > 2$ and any $\varepsilon > 0$, there exists $K > 0$ such that*

$$\mathbb{P} \left(s_1^{(p)}(A) \geq K \left(N^{\frac{p-1}{p}} + m^{\frac{1}{2}} N^{\frac{p-2}{2p}} \right) \right) \leq \varepsilon \quad (4.20)$$

where K only depends on p , ε and the pregaussian variable ξ .

Remark 4.3.6. In particular, for $p = \infty$, $s_1^{(\infty)}(A)$ is approximately $O(n)$. For some numerical experiments, see Appendix C.

CHAPTER 5

MODIFIED RESTRICTED ISOMETRY PROPERTY AND SPARSE RECOVERY

5.1 1-MODIFIED RESTRICTED ISOMETRY PROPERTY

For an integer $s \leq n$, the restricted isometry constant $\delta_s(A)$ is the smallest number δ which satisfies

$$(1 - \delta) \|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta) \|x\|^2$$

for all $x \in \mathbb{R}^n$, $|\text{support}(x)| \leq s$. Equivalently, the inequality

$$\sqrt{1 - \delta} \leq s_{\min}(A_S) \leq s_{\max}(A_S) \leq \sqrt{1 + \delta}$$

holds for any $m \times s$ submatrix A_S .

In an earlier version of [21], Foucart and Lai defined the so-called s -th q -modified restricted isometry property,

$$(1 - \delta) \|x\|_{f,q}^q \leq \|Ax\|_q^q \leq (1 + \delta) \|x\|_{f,q}^q \quad (5.1)$$

for all $x \in \mathbb{R}^N$ and $\|x\|_0 < s$. On the other hand, in [11], the q -restricted isometry property,

$$(1 - \delta_s) \|x\|_2^q \leq \|Ax\|_q^q \leq (1 + \delta_s) \|x\|_2^q \quad (5.2)$$

for all $x \in \mathbb{R}^N$ and $\|x\|_0 < s$ was defined. But in order to study the sparse recovery via ℓ_q -minimization, we want to use $\|x\|_q$ instead of $\|x\|_{f,q}$, so we define

Definition 5.1.1. An $m \times N$ matrix A is said to have the s -th q -restricted isometry property, if

$$(1 - \delta_s) \|x\|_q^q \leq \|Ax\|_q^q \leq (1 + \delta_s) \|x\|_q^q \quad (5.3)$$

for all $x \in \mathbb{R}^N$ and $\|x\|_0 < s$.

In an earlier version of [21] Foucart and Lai also introduced the new quasinorm

$$\|x\|_{f,q} := \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \quad (5.4)$$

for any pregaussian distribution probability density function f and $q > 0$, and showed that, for any $x \in \mathbb{R}^N$ and $0 < q \leq \frac{1}{3}$, the probability that x satisfies the q -modified restricted isometry property,

$$\mathbb{P} \left((1 - \varepsilon) m \|x\|_{f,q}^q \leq \|Ax\|_q^q \leq (1 + \varepsilon) m \|x\|_{f,q}^q \right) \geq 1 - 2e^{-\kappa m \varepsilon^2} \quad (5.5)$$

for some $\kappa > 0$ independent of q and f .

However, if we consider the case that $q = 1$ (for the sake of clearness, we will present the result for general $0 < q \leq 1$ in the later section), we have the following

Theorem 5.1.2. *Suppose that A is an $m \times N$ matrix whose entries are independent and identically-distributed copies of a symmetric pregaussian random variable with probability density function f . Then*

$$\mathbb{P} \left((1 - \varepsilon) m \|x\|_{f,1} \leq \|Ax\|_1 \leq (1 + \varepsilon) m \|x\|_{f,1} \right) \geq 1 - 2e^{-\kappa \frac{m}{\sqrt{N}} \varepsilon^2} \quad (5.6)$$

for any $0 < \varepsilon < 1$ and some $\kappa > 0$ dependent of f .

First of all, it turns out that the quasinorm (5.4) has the following properties

Lemma 5.1.3. *For any $x \in \mathbb{R}^N$, and even probability density function f ,*

1. $\|x\|_{f,2} = \sigma \|x\|_2$, in which σ^2 is the variance of the even probability density function f ;
2. $\|x\|_{f,q} \leq \|x\|_{f,q'}$ for any $0 < q \leq q'$;
3. $\|x\|_{f,q} \leq N^{\frac{1}{q}} \sigma_q^{\frac{1}{q}} \|x\|_p$, in which $\frac{1}{p} + \frac{1}{q} = 1$ for $p, q \geq 1$ and σ_q is the q -th absolute moment of the probability density function f ,

$$\sigma_q := \int_{-\infty}^{\infty} |t|^q f(t) dt; \quad (5.7)$$

4. $\|x\|_{f,q} \leq \|x\|_q \sigma_{q,p}^{\frac{1}{q}}$, in which

$$\sigma_{q,p} := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{i=1}^N (|t_1|^q + \cdots + |t_N|^q)^{\frac{q}{p}} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \quad (5.8)$$

and $\frac{1}{p} + \frac{1}{q} = 1$ for $p, q \geq 1$.

Proof. By the definition,

$$\begin{aligned} \|x\|_{f,2} &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^2 f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{2}} \\ &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sum_{i=1}^N x_i^2 t_i^2 \right. \\ &\quad \left. + \sum_{i,j=1, i \neq j}^N x_i x_j t_i t_j) f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{2}} \\ &= \left(\sum_{i=1}^N x_i^2 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_i^2 f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{2}} \\ &= \sigma \|x\|_2 \end{aligned} \quad (5.9)$$

since f is even, thus we obtain (1).

For (2), we can just use Hölder's inequality in the probability measure associated with the density function,

$$\begin{aligned} \|x\|_{f,q} &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\ &\leq \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^{q'} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q'}} \\ &\quad \cdot \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1 \cdot f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q} - \frac{1}{q'}} \\ &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^{q'} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q'}} \cdot 1 \\ &= \|x\|_{f,q'}. \end{aligned} \quad (5.10)$$

Again, by the discrete Hölder inequality,

$$\begin{aligned} \|x\|_{f,q} &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\ &\leq \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sum_{i=1}^N |t_i|^q) (\sum_{i=1}^N |x_i|^p)^{\frac{q}{p}} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\ &= \|x\|_p \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sum_{i=1}^N |t_i|^q) f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\ &= N^{\frac{1}{q}} \sigma_q^{\frac{1}{q}} \|x\|_p, \end{aligned} \quad (5.11)$$

which yields (3).

By the same token, (4) is obtained as follows,

$$\begin{aligned}
\|x\|_{f,q} &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&\leq \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sum_{i=1}^N |x_i|^q) (\sum_{i=1}^N |t_i|^p)^{\frac{q}{p}} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&= \|x\|_q \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\sum_{i=1}^N |t_i|^p)^{\frac{q}{p}} f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&= \|x\|_q \sigma_{q,p}^{\frac{1}{q}}.
\end{aligned} \tag{5.12}$$

□

In addition, the following useful lemmas were established in an earlier version of [21] by Foucart and Lai.

Lemma 5.1.4 ([21]). *For any $x \in \mathbb{R}^N$ and $q \geq 1$, $\|x\|_{f,q} \leq \|x\|_1 \sigma_q^{\frac{1}{q}}$.*

Proof. By Jensen's inequality and the convexity of the function $u \rightarrow u^q$ for $q \geq 1$,

$$\begin{aligned}
\|x\|_{f,q} &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&= \|x\|_1 \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i \frac{x_i}{\|x\|_1} \right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&\leq \|x\|_1 \left(\sum_{i=1}^N \frac{|x_i|}{\|x\|_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |t_i|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N \right)^{\frac{1}{q}} \\
&= \|x\|_1 \sigma_q^{\frac{1}{q}}.
\end{aligned} \tag{5.13}$$

□

Now, let us compare $\|x\|_{f,q}$ and $\|x\|_2$.

Lemma 5.1.5 (Comparison with ℓ_2 , [21]). *There exist $c_q, C_q > 0$ such that*

$$c_q \|x\|_2^q \leq \|x\|_{f,q}^q \leq C_q \|x\|_2^q \tag{5.14}$$

for all $x \in \mathbb{R}^N$.

Now let's prove Theorem 5.1.2.

Proof. [Proof of Theorem 5.1.2] Let $X_i := |\sum_{j=1}^N a_{ij} x_j|$ and $\xi_i := X_i - \mathbb{E}X_i$, then

$$\sum_{i=1}^m \xi_i = \|Ax\|_1 - m \|x\|_{f,1} \tag{5.15}$$

because

$$\|x\|_{f,1} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^N t_i x_i \right| f(t_1) \cdots f(t_N) dt_1 \cdots dt_N = \mathbb{E}X_i. \quad (5.16)$$

Obviously $\mathbb{E}\xi_i = 0$ and

$$\mathbb{E}\xi_i^2 = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = \|x\|_{f,2}^2 - \|x\|_{f,1}^2 := \nu^2 < \infty \quad (5.17)$$

since $\mathbb{E}|a_{ij}|^k \leq k!\lambda^k$ for some $\lambda > 0$. Moreover,

$$\mathbb{E}\xi_i^k = \mathbb{E}(X_i - \mathbb{E}X_i)^k = \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \mathbb{E}(X_i^j) \|x\|_{f,1}^{k-j}, \quad (5.18)$$

but by Lemma 5.1.4 we have

$$\mathbb{E}(X_i^j) \leq \|x\|_1^j \sigma_j \leq j!(\lambda \|x\|_1)^j. \quad (5.19)$$

Therefore

$$\begin{aligned} |\mathbb{E}\xi_i^k| &\leq \sum_{j=0}^k \binom{k}{j} j! (\lambda \|x\|_1)^j \|x\|_{f,1}^{k-j} \\ &\leq k! \sum_{j=0}^k \binom{k}{j} (\lambda \|x\|_1)^j (\|x\|_1 \sigma_1)^{k-j} \\ &\leq k! \sum_{j=0}^k \binom{k}{j} (\lambda \|x\|_1)^j (\lambda \|x\|_1)^{k-j} \\ &= k! (2\lambda \|x\|_1)^k \end{aligned} \quad (5.20)$$

by Lemma 5.1.4 again.

Now let $H := 2\lambda\|x\|_1$, then $|\mathbb{E}\xi_i^k| \leq k!H^k$. Applying the Bernstein inequality (see for instance [10] and [30]), we have

$$\mathbb{P}(|\|Ax\|_1 - m \|x\|_{f,1}| > t) \leq 2 \exp\left(-\frac{t^2}{2(tH + m\nu^2)}\right). \quad (5.21)$$

Then we can choose $t = \varepsilon m \|x\|_{f,1}$, thus

$$\begin{aligned} \mathbb{P}(|\|Ax\|_1 - m \|x\|_{f,1}| > \varepsilon m \|x\|_{f,1}) &\leq 2 \exp\left(-\frac{\varepsilon^2 m^2 \|x\|_{f,1}^2}{2(\varepsilon m \|x\|_{f,1} H + m\nu^2)}\right) \\ &= 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,1}^2}{2(\varepsilon \|x\|_{f,1} H + \nu^2)}\right). \end{aligned} \quad (5.22)$$

By Lemma 5.1.3, $\|x\|_{f,2} = \sigma\|x\|_2$ in which σ^2 is the variance of the even probability density function f , and $\|x\|_{f,2} \geq \|x\|_{f,1}$, and by Lemma 2.3 in [21],

$$\|x\|_{f,1} \geq c_1 \|x\|_2 = \frac{2^{\frac{1}{2}} \sigma_1}{\sqrt{\pi}} \min\left(\Gamma\left(\frac{3}{2}\right), \Gamma(1)\right) \|x\|_2 = \frac{\sqrt{2}}{2} \sigma_1 \|x\|_2, \quad (5.23)$$

in which σ_1 is the first absolute moment of the probability density function f , hence

$$\begin{aligned}
2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,1}^2}{2(\varepsilon \|x\|_{f,1} H + \nu^2)}\right) &\leq 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,1}^2}{2(\varepsilon \|x\|_{f,2} H + \|x\|_{f,2}^2)}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,1}^2}{2(\varepsilon \sigma \|x\|_2 H + \sigma^2 \|x\|_2^2)}\right) \\
&\leq 2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2 \|x\|_2^2}{4(\varepsilon \sigma \|x\|_2 H + \sigma^2 \|x\|_2^2)}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2 \|x\|_2}{4\sigma(\varepsilon H + \sigma \|x\|_2)}\right).
\end{aligned} \tag{5.24}$$

Furthermore

$$\begin{aligned}
2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2 \|x\|_2}{4\sigma(\varepsilon H + \sigma \|x\|_2)}\right) &= 2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2 \|x\|_2}{4\sigma(2\varepsilon \lambda \|x\|_1 + \sigma \|x\|_2)}\right) \\
&\leq 2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2 \|x\|_2}{4\sigma(2\varepsilon \lambda \sqrt{N} \|x\|_2 + \sigma \|x\|_2)}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m \sigma_1^2}{4\sigma(2\varepsilon \lambda \sqrt{N} + \sigma)}\right).
\end{aligned} \tag{5.25}$$

Thus, the combination of (5.22), (5.24) and (5.25) yields

$$\begin{aligned}
\mathbb{P}(\|Ax\|_1 - m \|x\|_{f,1} > \varepsilon m \|x\|_{f,1}) &\leq 2 \exp\left(-\frac{\sigma_1^2}{4\sigma(2\varepsilon \lambda + \sigma/\sqrt{N})} \cdot \frac{\varepsilon^2 m}{\sqrt{N}}\right) \\
&\leq 2 \exp\left(-\frac{\sigma_1^2}{4\sigma(2\lambda + \sigma)} \cdot \frac{\varepsilon^2 m}{\sqrt{N}}\right) \\
&= 2e^{-\kappa \frac{m}{\sqrt{N}} \varepsilon^2},
\end{aligned} \tag{5.26}$$

where $\kappa := \frac{\sigma_1^2}{4\sigma(2\lambda + \sigma)}$ dependent only of f , as desired. \square

We can see from (5.26) that the additional assumption $\frac{\sqrt{N}}{m} = o(1)$ will give us the exponential decay, so we have the following

Corollary 5.1.6. *If $N^{\frac{1}{2}} \ll m \ll N$, then we have the exponential decay for the tail probability*

$$\mathbb{P}(\|Ax\|_1 - m \|x\|_{f,1} > \varepsilon m \|x\|_{f,1}) \leq 2e^{-\kappa \varepsilon^2} \tag{5.27}$$

for some $\kappa > 0$ dependent of f .

and the following corollary on square matrices

Corollary 5.1.7. *If A is an $n \times n$ matrix whose entries are independent and identically-distributed copies of a symmetric pregaussian random variable with probability density function f . Then*

$$\mathbb{P}((1 - \varepsilon)n \ll x \ll_{f,1} \leq \|Ax\|_1 \leq (1 + \varepsilon)n \ll x \ll_{f,1}) \geq 1 - 2e^{-\kappa\sqrt{n}\varepsilon^2} \quad (5.28)$$

for any $0 < \varepsilon < 1$ and some $\kappa > 0$ dependent of f .

5.2 q -MODIFIED ISOMETRY PROPERTY

We can actually prove generally for $0 < q \leq 1$.

Theorem 5.2.1. *Suppose that A is an $m \times N$ matrix whose entries are independent and identically-distributed copies of a symmetric pregaussian random variable with probability density function f . Then*

$$\mathbb{P}\left((1 - \varepsilon)m \ll x \ll_{f,q}^q \leq \|Ax\|_q^q \leq (1 + \varepsilon)m \ll x \ll_{f,q}^q\right) \geq 1 - 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2} \quad (5.29)$$

for any $0 < \varepsilon < 1$ and $0 < q \leq 1$, and some $\kappa > 0$ dependent of f and q .

Proof. Let $X_i := \left|\sum_{j=1}^n a_{ij}x_j\right|^q$ and $\xi_i := X_i - \mathbb{E}X_i$. Since

$$\ll x \ll_{f,q}^q = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left|\sum_{i=1}^N t_i x_i\right|^q f(t_1) \cdots f(t_N) dt_1 \cdots dt_N = \mathbb{E}X_i. \quad (5.30)$$

then

$$\sum_{i=1}^m \xi_i = \|Ax\|_q^q - m \ll x \ll_{f,q}^q. \quad (5.31)$$

Obviously $\mathbb{E}\xi_i = 0$ and

$$\mathbb{E}\xi_i^2 = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = \ll x \ll_{f,2q}^{2q} - \ll x \ll_{f,q}^{2q} := \nu^2 < \infty \quad (5.32)$$

because $\mathbb{E}|a_{ij}|^k \leq k!\lambda^k$ for some $\lambda > 0$. Moreover,

$$\mathbb{E}\xi_i^k = \mathbb{E}(X_i - \mathbb{E}X_i)^k = \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \mathbb{E}(X_i^j) \ll x \ll_{f,q}^{q(k-j)}, \quad (5.33)$$

but by (2) of Lemma 5.1.3 and Lemma 5.1.4 we have

$$\mathbb{E}(X_i^j) = \|x\|_{f,q}^{qj} \leq \|x\|_{f,j}^{qj} \leq (\|x\|_1^j \sigma_j)^q \leq (j!)^q (\lambda \|x\|_1)^{qj} \leq j! (\lambda \|x\|_1)^{qj} \quad (5.34)$$

Therefore

$$\begin{aligned} |\mathbb{E}\xi_i^k| &\leq \sum_{j=0}^k \binom{k}{j} j! (\lambda \|x\|_1)^{qj} \|x\|_{f,q}^{q(k-j)} \\ &\leq k! \sum_{j=0}^k \binom{k}{j} (\lambda \|x\|_1)^{qj} \|x\|_{f,1}^{q(k-j)} \\ &= k! \sum_{j=0}^k \binom{k}{j} (\lambda \|x\|_1)^{qj} (\|x\|_1 \sigma_1)^{q(k-j)} \\ &\leq k! \sum_{j=0}^k \binom{k}{j} (\lambda \|x\|_1)^{qj} (\lambda \|x\|_1)^{q(k-j)} \\ &= k! (2(\lambda \|x\|_1)^q)^k \end{aligned} \quad (5.35)$$

by (2) of Lemma 5.1.3 and Lemma 5.1.4 again.

Now let $H := 2(\lambda \|x\|_1)^q$, then $|\mathbb{E}\xi_i^k| \leq k! H^k$. Applying the Bernstein inequality, we have

$$\mathbb{P}\left(\left|\|Ax\|_q^q - m \|x\|_{f,q}^q\right| > t\right) \leq 2 \exp\left(-\frac{t^2}{2(tH + m\nu^2)}\right). \quad (5.36)$$

Then we can choose $t = \varepsilon m \|x\|_{f,q}^q$, thus

$$\begin{aligned} \mathbb{P}\left(\left|\|Ax\|_q^q - m \|x\|_{f,q}^q\right| > \varepsilon m \|x\|_{f,q}^q\right) &\leq 2 \exp\left(-\frac{\varepsilon^2 m^2 \|x\|_{f,q}^{2q}}{2(\varepsilon m \|x\|_{f,q}^q H + m\nu^2)}\right) \\ &= 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,q}^{2q}}{2(\varepsilon \|x\|_{f,q}^q H + \nu^2)}\right). \end{aligned} \quad (5.37)$$

By Lemma 5.1.3, $\|x\|_{f,2} = \sigma \|x\|_2$ in which σ^2 is the variance of the even probability density function f , and $\|x\|_{f,2} \geq \|x\|_{f,1}$, and by Lemma 2.3 in [21],

$$c_q \|x\|_2^q \leq \|x\|_{f,q}^q \leq C_q \|x\|_2^q, \quad (5.38)$$

in which

$$c_q = \frac{2^{\frac{q}{2}} \sigma_q}{\sqrt{\pi}} \min\left(\Gamma\left(\frac{3}{2}\right), \Gamma\left(\frac{q+1}{2}\right)\right) = \frac{2^{\frac{q}{2}} \sigma_q}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 2^{\frac{q-2}{2}} \sigma_q \quad (5.39)$$

and

$$C_q = \frac{2^{\frac{q}{2}} \sigma_2^{\frac{q}{2}}}{\sqrt{\pi}} \max\left(\Gamma\left(\frac{3}{2}\right), \Gamma\left(\frac{q+1}{2}\right)\right) = \frac{(\sqrt{2}\sigma)^q}{\sqrt{\pi}} \Gamma\left(\frac{q+1}{2}\right) \quad (5.40)$$

since $\Gamma(\frac{3}{2}) < \Gamma(\frac{q+1}{2})$ for $0 < q \leq 1$ (see for example [35]). Hence by (1) and (2) of Lemma 5.1.3 again,

$$\begin{aligned}
2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,q}^{2q}}{2(\varepsilon \|x\|_{f,q}^q H + \nu^2)}\right) &\leq 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,q}^{2q}}{2(\varepsilon \|x\|_{f,q}^q H + \|x\|_{f,2q}^{2q})}\right) \\
&\leq 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,q}^{2q}}{2(\varepsilon \|x\|_{f,q}^q H + \|x\|_{f,2}^{2q})}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m \|x\|_{f,q}^{2q}}{2(\varepsilon \|x\|_{f,q}^q H + \sigma^{2q} \|x\|_2^{2q})}\right) \\
&\leq 2 \exp\left(-\frac{\varepsilon^2 m c_q^2 \|x\|_2^{2q}}{2(\varepsilon C_q \|x\|_2^q H + \sigma^{2q} \|x\|_2^{2q})}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m c_q^2 \|x\|_2^q}{2(\varepsilon C_q H + \sigma^{2q} \|x\|_2^q)}\right).
\end{aligned} \tag{5.41}$$

Furthermore,

$$\begin{aligned}
2 \exp\left(-\frac{\varepsilon^2 m c_q^2 \|x\|_2^q}{2(\varepsilon C_q H + \sigma^{2q} \|x\|_2^q)}\right) &= 2 \exp\left(-\frac{\varepsilon^2 m c_q^2 \|x\|_2^q}{2(2\varepsilon C_q (\lambda \|x\|_1)^q + \sigma^{2q} \|x\|_2^q)}\right) \\
&\leq 2 \exp\left(-\frac{\varepsilon^2 m c_q^2 \|x\|_2^q}{2(2\varepsilon C_q (\lambda \sqrt{N} \|x\|_2)^q + \sigma^{2q} \|x\|_2^q)}\right) \\
&= 2 \exp\left(-\frac{\varepsilon^2 m c_q^2}{2(2\varepsilon C_q (\lambda \sqrt{N})^q + \sigma^{2q})}\right).
\end{aligned} \tag{5.42}$$

Finally, the combination of (5.37), (5.41) and (5.42) yields

$$\begin{aligned}
\mathbb{P}\left(\left|\|Ax\|_q^q - m \|x\|_{f,q}^q\right| > \varepsilon m \|x\|_{f,q}^q\right) &\leq 2 \exp\left(-\frac{c_q^2 \cdot m N^{-\frac{q}{2}} \varepsilon^2}{2(2\varepsilon C_q \lambda^q + \sigma^{2q} / N^{\frac{q}{2}})}\right) \\
&\leq 2 \exp\left(-\frac{c_q^2}{2(2C_q \lambda^q + \sigma^{2q})} \cdot m N^{-\frac{q}{2}} \varepsilon^2\right) \\
&= 2e^{-\kappa m N^{-\frac{q}{2}} \varepsilon^2},
\end{aligned} \tag{5.43}$$

where $\kappa := \frac{c_q^2}{2(2C_q \lambda^q + \sigma^{2q})}$ dependent only of f , and the claim follows. \square

Also, we have

Corollary 5.2.2. *If $N^{\frac{q}{2}} \ll m \ll N$, $0 < q \leq 1$, then we have the exponential decay for the tail probability*

$$\mathbb{P}\left(\left|\|Ax\|_q^q - m \|x\|_{f,q}^q\right| > \varepsilon m \|x\|_{f,q}^q\right) 2e^{-\kappa \varepsilon^2} \tag{5.44}$$

for some $\kappa > 0$ dependent of f and q .

and the following corollary on square matrices

Corollary 5.2.3 (q-modified RIP for square matrices). *If A is an $n \times n$ matrix whose entries are independent and identically-distributed copies of a symmetric pregaussian random variable with probability density function f . Then*

$$\mathbb{P} \left((1 - \varepsilon) m \|x\|_{f,q}^q \leq \|Ax\|_q^q \leq (1 + \varepsilon) m \|x\|_{f,q}^q \right) \geq 1 - 2e^{-\kappa n^{\frac{2-q}{2}} \varepsilon^2} \quad (5.45)$$

for any $0 < \varepsilon < 1$ and some $\kappa > 0$ dependent of f and q , $0 < q \leq 1$.

5.3 ON THE SPARSE RECOVERY

Now, following the idea in [5] and that in [21] generalized by Foucart and Lai, which is using the estimate of cardinality of the finite cover of

$$\mathcal{S}_{f,q} := \{x \in \mathbb{R}^N : \|x\|_{f,q} = 1\}, \quad (5.46)$$

one can show

Theorem 5.3.1. *Suppose that the entries of the $m \times N$ matrix A are independent copies of a pregaussian random variable with an even probability density function f . Given $0 < q \leq 1$ and $0 < \varepsilon < 1$, there exist constants k_1 , k_2 , and k_3 depending on f and q such that*

$$\begin{aligned} \mathbb{P} \left(\left| \|Ax\|_q^q - m \|x\|_{f,q}^q \right| \leq \varepsilon m \|x\|_{f,q}^q \text{ for all } x \in \mathbb{R}^N \text{ with } \|x\| \leq s \right) \\ \geq 1 - 2e^{-\kappa_1 m N^{-\frac{q}{2}} \varepsilon^2} \end{aligned} \quad (5.47)$$

provided that $(k_2 \frac{s}{\varepsilon^2} + k_3 \frac{s}{\varepsilon^2} \ln(\frac{eN}{s})) N^{\frac{q}{2}} < m \ll N$.

Proof. We follow the proof of Theorem 4.1 of [21], with modified condition on m . By the cardinality estimate, union probability and Theorem 5.2.1, one has

$$\begin{aligned} \mathbb{P} \left(\left| \|Ax\|_q^q - m \|x\|_{f,q}^q \right| \geq \varepsilon m \|x\|_{f,q}^q \text{ for all } x \in \mathbb{R}^N \text{ with } \|x\| \leq s \right) \\ \leq \binom{N}{k} \left(1 + \frac{8}{\varepsilon}\right)^{\frac{t}{q}} 2e^{-k_1 m N^{-\frac{q}{2}} \varepsilon^2} \\ \leq 2 \binom{N}{k} \exp \left(-\frac{k_1 m N^{-\frac{q}{2}} \varepsilon^2}{4} + \frac{8s}{q\varepsilon} \right) \\ \leq 2 \exp \left(-\frac{k_1 m N^{-\frac{q}{2}} \varepsilon^2}{4} + \frac{8s}{q\varepsilon} + s \ln \left(\frac{eN}{s} \right) \right), \end{aligned} \quad (5.48)$$

and then the claim follows if $m > (k_2 \frac{s}{\varepsilon^2} + k_3 \frac{s}{\varepsilon^2} \ln(\frac{eN}{s})) N^{\frac{q}{2}}$ for $k_1 = \frac{\kappa}{8}$, $k_2 = \frac{64}{\kappa q}$ and $\kappa_1 =$

$\frac{8}{\kappa}$.

□

In compressed sensing, the ℓ_q -minimization problem with $0 < q \leq 1$ is the following:

$$\text{minimize}_{z \in \mathbb{R}^N} \|z\|_q \text{ subject to } Az = y, \quad (5.49)$$

that can be used to reconstruct a sparse vector that is the solution to the minimization

$$\text{minimize}_{z \in \mathbb{R}^N} \|z\|_0 \text{ subject to } Az = y. \quad (5.50)$$

Using Theorem 5.1 of [21], we are able to have the sparse recovery through pregaussian compression matrices.

Theorem 5.3.2. *Suppose that the entries of the $m \times N$ matrix A are independent copies of a pregaussian random variable with an even probability density function. If $0 < q \leq 1$, then the probability that every s -sparse vector $x \in \mathbb{R}^N$ is recovered as a solution to the optimization problem (5.49) exceeds $1 - \exp(-c_1 m N^{-\frac{q}{2}})$, provided that $c s \ln\left(\frac{N}{s}\right) N^{\frac{q}{2}} < m \ll N$, in which c and c_1 are dependent on the pregaussian distribution and q .*

CHAPTER 6

NULL SPACE PROPERTY FOR RECOVERY FROM MULTIPLE MEASUREMENTS VIA ℓ_q -MINIMIZATION

6.1 INTRODUCTION

In recent year, compressed sensing, a technique for recovering a sparse or compressible signals, attract much interest. The methods in compressed sensing including the convex ℓ_1 relaxation method and non-convex ℓ_q -method, $0 < q < 1$. On the ℓ_q -method, Foucart and Lai in [20] presented some numerical experimental results which indicates that the q -method performs better than other available methods and a sufficient condition on the matrix of an underdetermined linear system which guarantees that the solution of the system with minimal q -quasinorm.

The ℓ_q -method in unconstrained minimization can be also used to generate the sparse solution to underdetermined linear system. In data fitting, Tikhonov regularization is a method to get the regularized solution to underdetermined linear systems in least-square sense, which considers an unconstrained optimization problem instead of a constrained one. Recently, Lai and Wang in [33] gave an iterating algorithm to generate the sparse solution to underdetermined linear systems by defining the unconstrained ℓ_q -minimization

$$\arg \min_{x \in \mathbb{R}^N} \|x\|_q^q + \frac{1}{2\lambda} \|Ax - b\|_2^2 \quad (6.1)$$

for $0 < q \leq 1$ and using Γ -convergence to get the minimizer for the unconstrained ℓ_0 -minimization through the cluster points of minimizers for the unconstrained ℓ_q -minimization.

To get exact recovery from a single vector via ℓ_q -minimization, the sensing matrix needs to satisfy the ℓ_q null space property, as we have seen in Proposition 6.2.4. In multiple measurement problem (MMV), that is given a set of r measurements

$$Ax^{(k)} = b^{(k)} \text{ for } k = 1, \dots, r, \quad (6.2)$$

find the vectors $x^{(k)}$ which are jointly sparse, i.e., have nonzero entries at the same locations, and that arises in biomedical engineering such as neuromagnetic imaging, one may use the multiple-measurement-vector (MMV) non-convex optimization problem. It is

$$\text{minimize } \|X\|_{q,p} \text{ subject to } AX = B, \quad (6.3)$$

in which A , X and B are matrices, and

$$\|X\|_{q,p} := \left(\sum_{j=1}^N \|X^{j \rightarrow}\|_p^q \right)^{\frac{1}{q}}, \quad (6.4)$$

where $X^{j \rightarrow}$ is the j -th row of X , for $0 < q \leq 1$. We can find a condition for the exact recovery similar to the ℓ_q null space property for single measurement vector problem (SMV).

In the case of $r = 2$, $p = 2$ and $q = 1$, Foucart and Gribonval in [19] proved that the real null space and complex null space property are equivalent for the sparse recovery achieved by ℓ_1 -minimization. In the following sections, we will start with the case $r = 2$ and prove the equivalence for $p = 2$ and $0 < q \leq 1$ in this case, and then present our results for general r .

6.2 NULL SPACE PROPERTY FOR RECOVERY VIA ℓ_q MINIMIZATION

For minimization problem (5.50), the uniqueness of the solution is related to a parameter of the matrix A , $\text{spark}(A)$, which is defined as the cardinality of the smallest subset of linearly dependent columns of A (see [16]) and has the obvious property that $\text{spark}(A) \leq \text{rank}(A) + 1$. To be precise, if $\|z\|_0 < \frac{1}{2}\text{spark}(A)$ then z can be uniquely recovered by solving the ℓ_0 -minimization (5.50).

As for the minimization problem (5.49), one can introduce the so-called null space property of A . The null space property has also been used in quantifying the error of approximations, see e.g. [13]. However, the null space property which guarantees the sparse recovery is basically the following theorem in [28] by Gribonval and Nielsen, which is also generalized by them to general f -norm (but not necessarily a norm) on Hilbert space in [29].

Proposition 6.2.1 (Restricted null space property). *Let $S \subseteq \{1, 2, \dots, N\}$ be a fixed index set. Then a vector z with $\text{support}(z) \subseteq S$ can be uniquely recovered from $Az = b$ using ℓ_q -minimization (5.49) if for all non-zero v in the null space of A ,*

$$\|v_S\|_q < \|v_{S^c}\|_q, \quad (6.5)$$

in which S^c is the complement of S in $\{1, 2, \dots, N\}$.

For clarity we give a concise proof here.

Proof. We know for any non-zero x in the null space of A ,

$$\|z_S\|_q^q \leq \|x_S\|_q^q + \|(z+x)_S\|_q^q. \quad (6.6)$$

By the assumption $\|x_S\|_q < \|x_{S^c}\|_q$, we then have

$$\|z_S\|_q^q < \|(z+x)_S\|_q^q + \|x_{S^c}\|_q^q. \quad (6.7)$$

But we also know that z vanishes on S^c , thus

$$\|z\|_q^q = \|z_S\|_q^q < \|(z+x)_S\|_q^q + \|(z+x)_{S^c}\|_q^q = \|z+x\|_q^q, \quad (6.8)$$

and so $z \in \mathbb{R}^N$ is the unique solution to the minimization problem (5.49). \square

In [29], Gribonval and Nielsen also discussed the negated cases of (6.5) in general. Here we include a version in ℓ_q for our topic.

Proposition 6.2.2. *Let \mathcal{N} be the null space of A and S be any subset of $\{1, 2, \dots, N\}$.*

1. If $\|x_S\|_q > \|x_{S^c}\|_q$ for some $x \in \mathcal{N}$, then there exist some z and z' , such that $Az = Az'$, $\text{support}(z) \subseteq S$ and $\|z'\|_q < \|z\|_q$.
2. If $\|x_S\|_q = \|x_{S^c}\|_q$ for some non-zero $x \in \mathcal{N}$, then there exist some z and z' , $z \neq z'$, such that $Az = Az'$, $\text{support}(z) \subseteq S$ and $\|z'\|_q = \|z\|_q$.

Proof. For (1), without loss of generality, we can assume $S = \{1, 2, \dots, s\}$. Take z to be the vector whose components in S are the components of x restricted on S and whose other components are zeroes, and z' to be the vector whose components in S^c are the negative of the components of x restricted on S^c and whose other components are zeroes, then $z - z' = x$. It follows that

$$Az - Az' = A(z - z') = Ax = 0, \quad (6.9)$$

the support of z is in S , and

$$\|z'\|_q = \|x_{S^c}\|_q < \|x_S\|_q = \|z\|_q. \quad (6.10)$$

For (2), one can just change the “<” in (6.10) to “=” and other steps are the same with (1). □

Remark 6.2.3. This proposition tell us that z , that is supported on S , can not be uniquely recovered through solving the ℓ_q -minimization (5.49), if $\|x_S\|_q \geq \|x_{S^c}\|_q$ for some non-zero $x \in \mathcal{N}$. Therefore, if all the vectors supported on S can be uniquely recovered, then A satisfies the null space property on S .

Precisely, one has an equivalence as follows

Proposition 6.2.4. *Let $S \subseteq \{1, 2, \dots, N\}$ be a fixed index set. Then all the vectors supported on S can be uniquely recovered from $Az = b$ using ℓ_1 -minimization (5.49) if and only if for all non-zero v in the null space of A ,*

$$\|v_S\|_1 < \|v_{S^c}\|_1. \quad (6.11)$$

To give a concrete sense that $\|v_S\|_1 < \|v_{S^c}\|_1$ can fail if there is a vector supported on S can be uniquely recovered, we would like to give an example.

Example 6.2.5. Consider the ℓ_1 -minimization problem

$$\text{minimize}_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y, \quad (6.12)$$

in which

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}, \quad (6.13)$$

$N = 3$ and

$$y = \begin{pmatrix} 2 \\ 0 \end{pmatrix}. \quad (6.14)$$

To solve the minimization problem, we can do a linear programming which the ℓ_1 -minimization is equivalent to, see [7]. On the other hand, we can also obtain the solution directly. The solutions of $Az = y$ are

$$\left\{ \begin{pmatrix} t \\ 2-t \\ 1-t \end{pmatrix} : t \in \mathbb{R} \right\}, \quad (6.15)$$

and then the function $f(t) := |t| + |2-t| + |1-t|$, that is the ℓ_1 -norm of the solutions, achieves

its minimum at $t = 1$. So the minimization problem has a unique solution $z = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ with

the support $S = \{1, 2\}$. However, the non-zero null vector of A , $x := \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ does not

satisfy $\|x_S\|_1 < \|x_{S^c}\|_1$.

6.3 REAL VERSUS COMPLEX NULL SPACE PROPERTIES FOR ℓ_q

In this section, we consider the open problem whether for any $0 < q \leq 1$ the real null space property

$$\|x_S\|_q < \|x_{S^c}\|_q \quad (6.16)$$

in Proposition 6.2.1 or the so-called stable null space property

$$\|x_S\|_q \leq \rho \|x_{S^c}\|_q \text{ for some } 0 < \rho < 1, \quad (6.17)$$

for all $x \in \mathcal{N} \setminus \{0\}$ (in the following we use the notation “ \prec ” that was used in [19] to denote either “ $<$ ” or “ $\leq \rho$ ”) for the ℓ_q -minimization (5.49) is equivalent to the complex version

$$\sum_{j \in S} \left(\sqrt{v_j^2 + w_j^2} \right)^q \prec \sum_{j \in S^c} \left(\sqrt{v_j^2 + w_j^2} \right)^q, \quad (6.18)$$

in other words

$$\|z_S\|_q \prec \|z_{S^c}\|_q \quad (6.19)$$

for $z = v + \sqrt{-1}w$, for all $(v, w) \in \mathcal{N}^2 \setminus \{(0, 0)\}$, raised by Foucart and Gribonval in [19].

Remark 6.3.1. In finite dimensional vector spaces the null space property and the stable null space property are equivalent, because $\mathcal{N} \cap \mathcal{S}_q$ is compact and then the function $x \rightarrow \frac{\|x_S\|_q}{\|x_{S^c}\|_q}$ achieves its maximum which is strictly less than 1. But in infinite dimensional vector spaces, for instance Hilbert space, the null space property does not necessarily imply the stable null space property.

Theorem 6.3.2 (Comparison theorem). *Let $S \subset \{1, 2, \dots, N\}$ be an index set with $|S| = s$. Give $0 < q \leq 1$ and a matrix $B \in \mathbb{R}^{2 \times N}$ with columns $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N \in \mathbb{R}^2$, if*

$$\|(x, y)B_S\|_q \prec \|(x, y)B_{S^c}\|_q \quad (6.20)$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{0\}$ and some $S \subseteq \{1, 2, \dots, N\}$ with $|S| = s$. Then

$$\sum_{k \in S} \|\mathbf{c}_k\|_2^p \prec \sum_{k \in S^c} \|\mathbf{c}_k\|_2^p. \quad (6.21)$$

Proof. Let $B =: (b_{i,j})_{2 \times N}$, and without loss of generality we can assume $S := \{1, 2, \dots, s\}$.

We show (6.21) holds for $s = 1$ first and then for $s \geq 2$.

If $s = 1$, by the assumption,

$$|b_{1,1}x + b_{2,1}y|^q = \|(x, y)B_S\|_p^q \prec \|(x, y)B_{S^c}\|_p^q = \sum_{j=2}^N |b_{1,j}x + b_{2,j}y|^q \quad (6.22)$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. Choosing $x = \frac{b_{1,1}}{\sqrt{b_{1,1}^2 + b_{2,1}^2}}$ and $y = \frac{b_{2,1}}{\sqrt{b_{1,1}^2 + b_{2,1}^2}}$ and applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left(\sqrt{b_{1,1}^2 + b_{2,1}^2}\right)^q &\prec \sum_{j=2}^N \left| \frac{1}{\sqrt{b_{1,1}^2 + b_{2,1}^2}} (b_{1,j}b_{1,1} + b_{2,j}b_{2,1}) \right|^q \\ &\leq \sum_{j=2}^N \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^q \end{aligned} \quad (6.23)$$

from (6.22), thus the claim (6.21) for $s = 1$ follows.

For the case when $s \geq 2$, we have

$$\sum_{j=1}^s |b_{1,j}x + b_{2,j}y|^q \prec \sum_{j=s+1}^N |b_{1,j}x + b_{2,j}y|^q \quad (6.24)$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{0\}$ from the assumption. Let $v_j := \left(\frac{b_{1,j}}{\sqrt{b_{1,j}^2 + b_{2,j}^2}}, \frac{b_{2,j}}{\sqrt{b_{1,j}^2 + b_{2,j}^2}}\right) \in \mathbb{S}^1$ where \mathbb{S}^1 is the unit circle, $j = 1, 2, \dots, N$, then

$$\sum_{j=1}^s \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^q |\langle v_j, \xi \rangle|^q \prec \sum_{j=s+1}^N \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^q |\langle v_j, \xi \rangle|^q \quad (6.25)$$

for all unit vector $\xi \in \mathbb{S}^1$ particularly. Taking the integral of (6.25) on \mathbb{S}^1 , we have

$$\sum_{j=1}^s \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^q \int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^q d\xi \prec \sum_{j=s+1}^N \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^q \int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^q d\xi. \quad (6.26)$$

Note that $\int_{\mathbb{S}^1} |\langle \cdot, \xi \rangle|^q d\xi$ is a rotation invariant function from the perspective of integral geometry (cf. [1], [3] and [37]). That is, $\int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^q d\xi$ is constant independent of j . In fact we have

Lemma 6.3.3. *Given any integer $r \geq 2$. For any $q > 0$,*

$$\int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^q d\xi = C$$

for all $v \in \mathbb{S}^{r-1}$, where $C > 0$ is a constant dependent only on p .

Proof. Let U be an orthogonal transformation of \mathbb{R}^r . Then for any $v \in \mathbb{S}^{r-1}$, the sphere of the unit ball in \mathbb{R}^r , we have

$$\langle U(v), \xi \rangle = \langle v, U^{-1}(\xi) \rangle \quad (6.27)$$

for all $\xi \in \mathbb{S}^{r-1}$. Also, we know that

$$\mathbb{S}^{r-1} = \{U(v) : U \in O(r)\}, \quad (6.28)$$

where $O(r)$ denotes the set of all $r \times r$ orthogonal matrices. By change of variables and using the fact that $|\det(U^{-1})| = 1$, we get

$$\begin{aligned} \int_{\mathbb{S}^{r-1}} |\langle U(v), \xi \rangle|^q d\xi &= \int_{\mathbb{S}^{r-1}} |\langle v, U^{-1}(\xi) \rangle|^q d\xi \\ &= \int_{\mathbb{S}^{r-1}} |\langle v, U^{-1}(\xi) \rangle|^q dU^{-1}(\xi) \\ &= \int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^q d\xi \end{aligned} \quad (6.29)$$

for all $U \in O(r)$. Thus we see that $\int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^q d\xi \equiv C$ for some $C > 0$ and for all $v \in \mathbb{S}^{r-1}$. \square

Therefore

$$\sum_{j=1}^s \left(\sqrt{b_{1,j}^2 + b_{2,j}^2} \right)^q \prec \sum_{j=s+1}^N \left(\sqrt{b_{1,j}^2 + b_{2,j}^2} \right)^q. \quad (6.30)$$

So the claim follows. \square

An immediate consequence of the above theorem is the following theorem on the equivalence of real null space property and complex null space property.

Theorem 6.3.4. *The real null space property*

$$\|u_S\|_q \prec \|u_{S^c}\|_q \quad (6.31)$$

for all $u \in \ker(A) \setminus \{0\}$ is equivalent to the complex null space property

$$\sum_{j \in S} \left(\sqrt{v_j^2 + w_j^2} \right)^q \prec \sum_{j \in S^c} \left(\sqrt{v_j^2 + w_j^2} \right)^q, \quad (6.32)$$

for all $(v, w) \in (\ker(A))^2 \setminus \{(0, 0)\}$ and all $0 < q \leq 1$.

Proof. From (6.32) to (6.31), it is obvious.

For the converse, letting

$$B = \begin{pmatrix} v_1 & v_2 & \cdots & v_N \\ w_1 & w_2 & \cdots & w_N \end{pmatrix}, \quad (6.33)$$

we have

$$\|(x, y)B_S\|_q \prec \|(x, y)B_{S^c}\|_q \quad (6.34)$$

by (6.31). Then applying Theorem 6.3.2, we obtain that

$$\sum_{k \in S} \|\mathbf{c}_k\|_2^q \prec \sum_{k \in S^c} \|\mathbf{c}_k\|_2^q \quad (6.35)$$

where $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N$ are the columns of B . That yields (6.32). \square

Remark 6.3.5. By Remark B.0.16, we know that Corollary Theorem 6.3.4 also holds for $q > 1$.

The application of the result in Theorem 6.3.4 in compressed sensing is that we can recover a sparse vector in complex space via solving two real ℓ_q -minimization problems instead of a complex ℓ_q -minimization one, and an algorithm for the real ℓ_q -minimization problem (5.49) is given in [20].

6.4 ON RECOVERY FROM MULTIPLE MEASUREMENTS

We also want to consider another open problem that Foucart and Gribonval proposed in [19], which is whether the null space property (6.31) is equivalent to the so called mixed $\ell_{1,2}$ null space property

$$\sum_{j \in S} \sqrt{u_{1,j}^2 + \cdots + u_{n,j}^2} \prec \sum_{j \in S^c} \sqrt{u_{1,j}^2 + \cdots + u_{n,j}^2}, \quad (6.36)$$

for the joint recovery from multiple measurements via

$$\text{minimize } \sum_{i=1}^N \sqrt{z_{1,j}^2 + \cdots + z_{n,j}^2} \text{ subject to } Az_1 = Ax_1, \dots, Az_n = Ax_n \quad (6.37)$$

(see [8]), in which (x_1, \dots, x_n) is an n -tuple of vectors in \mathbb{R}^N . Based on the answer to the problem on complex null space above, we are able to solve the problem concerning n -tuples as well. More generally, we provide an affirmative answer to the mixed $\ell_{q,2}$ null space property

$$\sum_{j \in S} \left(\sqrt{u_{1,j}^2 + \dots + u_{n,j}^2} \right)^q \prec \sum_{j \in S^c} \left(\sqrt{u_{1,j}^2 + \dots + u_{n,j}^2} \right)^q, \quad (6.38)$$

for the joint recovery from multiple measurements via

$$\text{minimize } \sum_{i=1}^N \left(\sqrt{z_{1,i}^2 + \dots + z_{n,i}^2} \right)^q \text{ subject to } Az_1 = Ax_1, \dots, Az_n = Ax_n \quad (6.39)$$

(see [8]) for all $0 < q \leq 1$.

Theorem 6.4.1. *The null space property*

$$\|u_S\|_q \prec \|u_{S^c}\|_q \quad (6.40)$$

for all $u \in \ker(A) \setminus \{0\}$ is equivalent to the mixed $\ell_{q,2}$ null space property (6.38) for all $n \geq 3$ and all $0 < q \leq 1$.

Proof. It is trivial to deduce null space property (6.40) from mixed $\ell_{q,2}$ null space property (6.38).

For the converse, we can modify the proof of Theorem 6.3.2 to obtain a generalized comparison theorem for any $B \in \mathbb{R}^{n \times N}$, $n \geq 3$, and $0 < q \leq 1$. Specifically, we will have

$$\sum_{j \in S} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q |\langle v_j, \xi \rangle|^q \prec \sum_{j \in S^c} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q |\langle v_j, \xi \rangle|^q \quad (6.41)$$

for $v_j := \frac{1}{\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2}} (b_{1,j}, \dots, b_{n,j})$ and all vector ξ in the unit $n-1$ -sphere \mathbb{S}^{n-1} . Analogously to the case of $n = 2$, taking the integral of (6.41) on \mathbb{S}^{n-1} , we have

$$\begin{aligned} & \sum_{j \in S} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q \int_{\mathbb{S}^{n-1}} |\langle v_j, \xi \rangle|^q d\xi \\ & \prec \sum_{j \in S^c} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q \int_{\mathbb{S}^{n-1}} |\langle v_j, \xi \rangle|^q d\xi. \end{aligned} \quad (6.42)$$

Using the fact that $\int_{\mathbb{S}^{n-1}} |\langle \cdot, \xi \rangle|^q d\xi$ is a rotation invariant function from the perspective of integral geometry (see for instance [1], [3] and [37]), we get

$$\sum_{j \in S} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q \prec \sum_{j \in S^c} \left(\sqrt{b_{1,j}^2 + \dots + b_{n,j}^2} \right)^q. \quad (6.43)$$

□

For joint-sparse recovery from multiple measurements, it can also be achieved by mixed $\ell_{q,2}$ -minimization for $0 < q \leq 1$, for which one can see [20] and [8]. The above theorem may allow us to consider the ℓ_q -minimization instead, that will simplify the program for the mixed $\ell_{q,2}$ -minimization. Furthermore, multiple measurements have been used in many fields of technology, for instance, neuromagnetic imaging [15] and communication channels (see e.g. [14]), as pointed out in [8], so our results may have real applications in these fields.

6.5 THE NULL SPACE PROPERTY FOR RECOVERY FROM MULTIPLE MEASUREMENTS

In this section, we want to study the MMV non-convex optimization problem, which is

$$\text{minimize } \|X\|_{q,p} \quad \text{subject to } AX = B \quad \text{with } 0 < q \leq 1. \quad (6.44)$$

In particular, when $p = 2$, (6.44) becomes (6.39). On the null space property of the matrix A for joint recovery from multiple measurements via (6.39), in which $p = 2$ and $0 < q \leq 1$, we have

Theorem 6.5.1. *Let A be an $m \times N$ matrix and $S \subset \{1, 2, \dots, N\}$ be an index set. Then for any $0 < q \leq 1$, all $X_0 \in \mathbb{R}^{N \times r}$ with the support of the rows of X_0 contained in S can be uniquely recovered from $AX = B$ using (6.44) for $p = 2$ if and only if for all Z with columns $Z^{\downarrow k} \in \ker(A) \setminus \{0\}$, $k = 1, 2, \dots, r$,*

$$\|Z_S\|_{q,2} < \|Z_{S^c}\|_{q,2} \quad (6.45)$$

in which $\|Z_S\|_{q,2} := \left(\sum_{j \in S} \|Z^{j \rightarrow}\|_2^q \right)^{\frac{1}{q}}$ and $\|Z_{S^c}\|_{q,2} := \left(\sum_{j \in S^c} \|Z^{j \rightarrow}\|_2^q \right)^{\frac{1}{q}}$.

Proof. Assume that (6.45) holds for all Z with columns $Z^{\downarrow k} \in \ker(A) \setminus \{0\}$ and $X \in \mathbb{R}^{N \times r}$ with the support of the rows of X contained in S is an solution to $AX = B$. Then for Z with columns $Z^{\downarrow k} \in \ker(A) \setminus \{0\}$, for $0 < q \leq 1$,

$$\sum_{j \in S} \|X^{j \rightarrow}\|_2^q \leq \sum_{j \in S} \|Z^{j \rightarrow}\|_2^q + \sum_{j \in S} \left\| (X + Z)^{j \rightarrow} \right\|_2^q. \quad (6.46)$$

From the assumption, $\sum_{j \in S} \|Z^{j \rightarrow}\|_2^q < \sum_{j \in S^c} \|Z^{j \rightarrow}\|_2^q$, we thus have

$$\sum_{j \in S} \|X^{j \rightarrow}\|_2^q < \sum_{j \in S^c} \|Z^{j \rightarrow}\|_2^q + \sum_{j \in S} \|(X + Z)^{j \rightarrow}\|_2^q. \quad (6.47)$$

But the support of the rows of X are contained in S , hence

$$\begin{aligned} \sum_{j=1}^N \|X^{j \rightarrow}\|_2^q &= \sum_{j \in S} \|X^{j \rightarrow}\|_2^q \\ &< \sum_{j \in S^c} \|(X + Z)^{j \rightarrow}\|_2^q + \sum_{j \in S} \|(X + Z)^{j \rightarrow}\|_2^q \\ &= \sum_{j=1}^N \|(X + Z)^{j \rightarrow}\|_2^q, \end{aligned} \quad (6.48)$$

and so $X \in \mathbb{R}^{N \times r}$ is the unique solution to the minimization problem (6.44) for $p = 2$ and $0 < q \leq 1$.

For the converse, assume that there is some Z with columns $Z^{\downarrow k} \in \ker(A) \setminus \{0\}$ such that $\|Z_S\|_{q,2} \geq \|Z_{S^c}\|_{q,2}$. We can choose $X \in \mathbb{R}^{N \times r}$ such that the rows of X restricted on S equal to those of Z , and the remaining rows of X are zeros. Therefore,

$$AX = AX - AZ = A(X - Z) \quad (6.49)$$

and $\|X\|_{q,2} = \|X_S\|_{q,2} = \|Z_S\|_{q,2} \geq \|Z_{S^c}\|_{q,2} = \|X - Z\|_{q,2}$, which contradicts with the uniqueness of the recovery. \square

Corollary 6.5.2. *Let A be an $m \times N$ matrix and $S \subset \{1, 2, \dots, N\}$ be an index set. Then for any $p \geq 1$ and $0 < q \leq 1$, all $X_0 \in \mathbb{R}^{N \times r}$ with the support of the rows of X_0 contained in S can be uniquely recovered from $AX = B$ using (6.44) if and only if for all Z with columns $Z^{\downarrow k} \in \ker(A) \setminus \{0\}$, $k = 1, 2, \dots, r$,*

$$\|Z_S\|_{q,p} < \|Z_{S^c}\|_{q,p}. \quad (6.50)$$

Proof. One can just change the $\|\cdot\|_2$ -norm to $\|\cdot\|_p$ -norm for $p \geq 1$ in the proof of Theorem 6.5.1, and that will give us a proof for the claim. \square

Furthermore, we have

Theorem 6.5.3. *Let A be an $m \times N$ matrix and $S \subset \{1, 2, \dots, N\}$ be an index set. Then for any $0 < q \leq 1$, all $X_0 \in \mathbb{R}^{N \times r}$ with the support of the rows of X_0 contained in S can be uniquely recovered via solving the MMV non-convex optimization problem (6.44) for $p = 2$ if and only if*

$$\|u_S\|_q \prec \|u_{S^c}\|_q \quad (6.51)$$

for all $u \in \ker(A) \setminus \{0\}$.

Proof. This is a consequence of the combination of Theorem 6.5.1 and Theorem 6.4.1. \square

BIBLIOGRAPHY

- [1] Alesker, S., Continuous rotation invariant valuations on convex sets, *Annals of Mathematics*, 149 (1999), 977-1005.
- [3] Alesker, S. and J. Bernstein, Range characterization of the cosine transform on higher Grassmannians, *Advances in Mathematics*, Volume 184, Issue 2, 1 June 2004, Pages 367-379.
- [2] Antoulas, A. C., Approximation of large-scale dynamical systems, Series: *Advances in Design and Control* (No. 6), Society for Industrial and Applied Mathematic, 2005.
- [4] Bai, Z. D., Circular law, *Ann. Probab.* Volume 25, Number 1 (1997), 494-529.
- [5] Baraniuk, R., M. Davenport, and R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constructive Approximation*, 2008, 253-263.
- [8] Berg, E. and M. P. Friedlander, Joint-sparse recovery from multiple measurements, *arXiv:0904.2051v1*, 2009.
- [7] Bloomfield, P., W. L. Steiger, *Least absolute deviations: theory, applications, and algorithms*, Volume 6 of (*Progress in probability and statistics*), Birkhäuser, 1983.
- [9] Buldygin, V. V., Sub-Gaussian random variables, *Ukrainian Mathematical Journal*, Volume 32, Number 6, November, 1980, 483-489.
- [10] Buldygin, V. V. and I. U. V. Kozachenko, *Metric characterization of random variables and random processes*, *Translations of mathematical monographs*, Vol. 188, Published by AMS Bookstore, 2000.

- [6] Candes, E. and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, Vol. 51, No. 12. (December 2005), pp. 4203-4215.
- [11] Chartrand, R. and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Problems*, Volume 24, Number 3, 2008, doi: 10.1088/0266-5611/24/3/035020.
- [12] Chen, J. and X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, *IEEE Transactions on Signal Processing* , 2006, 54 (12) : 4634-4643.
- [13] Cohen, A. and W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, *J. Amer. Math. Soc.* 22 (2009), 211-231.
- [14] Cotter, S. F. and B. D. Rao, Sparse channel estimation via matching pursuit with application to equalization, *IEEE Transactions on Communications*, 50(3), March 2002, 374-377.
- [15] Cotter, S. F., B. D. Rao, K. Engang, and K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, *IEEE Transactions on Signal Processing*, 53:2477-2488, July 2005.
- [16] Donoho, D. L. and Elad, M., Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proceedings of the National Academy of Sciences*, 2003, doi: 10.1073/pnas.0437847100.
- [17] Edelman, A., Eigenvalues and condition numbers of random matrices, *SIAM J. Matrix Anal. Appl.* 9 (1988), 543-560.
- [18] Fisher, A., C. Dickson, and W. Boynge, *The mathematical theory of probabilities and its application to frequency curves and statistical methods*, New York, Macmillan Co., 1922.

- [19] Foucart, S. and R. Gribonval, Real vs. complex null space properties for sparse vector recovery, Submitted, 2009.
- [20] Foucart, S. and M. J. Lai, Sparsest Solutions of Underdetermined Linear Systems via ℓ_q minimization for $0 < q \leq 1$, Applied and Computational Harmonic Analysis, Volume 26, Issue 3, May 2009, Pages 395-407.
- [21] Foucart, S. and M. J. Lai, Sparse Recovery with Pre-Gaussian Random Matrices, accepted for publication by Stud. Math., 2010.
- [22] Gorodnitskya, I. F., J. S. Georgeb and B. D. Raoa, Neuromagnetic source imaging with FOCUS: a recursive weighted minimum norm algorithm, Electroencephalography and Clinical Neurophysiology Volume 95, Issue 4, October 1995, Pages 231-251.
- [23] Garling, D. J. H., Inequalities : a journey into linear analysis, Cambridge University Press, 2007.
- [24] Gribonval, R. and M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, Appl. Comput. Harmon. Anal. 22 (2007) 335–355.
- [25] Golub, G. H. and C. F. V. Loan, Matrix computations (Johns Hopkins Studies in Mathematical Sciences), The Johns Hopkins University Press, 1996.
- [26] Geman, S., A limit theorem for the norm of random matrices, Ann. Probab. 8 (1980), 252–261.
- [27] Girko, V. L., "Circular Law." Theory Probab. Appl. 29, 694-706, 1984.
- [28] Gribonval, R. and M. Nielsen, Sparse decompositions in unions of bases, IEEE Trans. Inform. Theory 49 (12) (2003) 3320–3325.

- [29] Gribonval, R. and M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measurestar, Applied and Computational Harmonic Analysis Volume 22, Issue 3, May 2007, Pages 335-355.
- [30] Lai, M. J., On sparse solution of underdetermined linear systems, Journal of Concrete and Applicable Mathematics, 8(2010), 296–327.
- [31] Lai, M. J. and Y. Liu, The Null Space Property for Sparse Recovery from Multiple Measurement Vectors, submitted to Applied and Computational Harmonic Analysis, 2010.
- [32] Lai, M. J. and Y. Liu, The Probabilistic Estimates on the Largest and Smallest q -Singular Values of Pre-Gaussian Random Matrices, submitted to Advances in Mathematics, 2010.
- [33] Lai, M. J. and J. Wang, An unconstrained ℓ_q -minimization for sparse solution of under determined linear systems, submitted, 2009.
- [34] Latala, R., Some estimates of norms of random matrices, Proc. Amer. Math. Soc. 133 (2005), 1273-1282.
- [35] Li, X. and C. P. Chen, Inequalities for the Gamma Function, vol. 8, iss. 1, Journal of Inequalities in Pure and Applied Mathematics, 2007.
- [36] Li, W. V., On the lower tail of Gaussian measures on \mathbb{R}^p , Progress in Probability, Vol. 30, 106-117, (1992).
- [37] Lonke, Y., Derivatives of the L^p -cosine transform, Advances in Mathematics, Volume 176, Issue 2, 25 June 2003, Pages 175-186.
- [38] Malioutov, D., M. Cetin, and A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Transactions on Signal Processing, 53(8):3010–3022, August 2005.

- [39] Mishali, M. and Y. C. Eldar, The ReMBo Algorithm: Accelerated Recovery of Jointly Sparse Vectors, 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008.
- [40] Mishali M. and Y. C. Eldar, Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *IEEE Transactions on Signal Processing*, 56(10):4692-4702, October 2008.
- [41] Pisier, G., *The volume of convex bodies and Banach space geometry*, Cambridge University Press, 1999.
- [42] Robbins, H. A., Remark of Stirling's Formula, *Amer. Math. Monthly* 62, 26-29, 1955.
- [43] Rudelson, M. and R. Vershynin, The Littlewood–Offord problem and invertibility of random matrices, *Advances in Mathematics*, Volume 218, Issue 2, 1 June 2008, Pages 600-633.
- [44] Rudelson, M. and R. Vershynin, The least singular value of a random square matrix is $o(n^{-1/2})$, *Comptes rendus-Mathématique*, Volume 346, Issues 15-16, August 2008, Pages 893-896.
- [45] Rudelson, M. and R. Vershynin, Smallest singular value of a random rectangular matrix, *Communications on Pure and Applied Mathematics*, 2009 *Communications on Pure and Applied Mathematics*, Volume 62 Issue 12, Pages 1707 - 1739, 2009.
- [46] Rudelson, M. and R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, *Proceedings of the International Congress of Mathematicians*, Hyderabad, India, 2010.
- [47] Seginer, Y., The expected norm of random matrices, *Combin. Probab. Comput.* 9 (2000), 149– 166.
- [48] Shen, J., On the singular values of Gaussian random matrices, *Linear Algebra and its Applications*, Volume 326, Issues 1-3, 15 March 2001, Pages 1-14.

- [49] Siahhaan, H. B., S. Weiland, and A. A. Stoorvogel, Optimal approximation of linear operators: a singular value decomposition approach, In D.S. Gilliam, J. Rosenthal (Eds.), Proceedings 15th International Symposium on Mathematical Theory of Networks and Systems (MNTS'02, Notre Dame), Session FM3. (pp. 580-590), Univ. Notre Dame, 2002.
- [50] Soshnikov, A., A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *J. Statist. Phys.* 108 (2002), no. 5-6, 1033–1056.
- [51] Soshnikov, A. and Y. V. Fyodorov, On the largest singular values of random matrices with independent Cauchy entries, *J. Math. Phys.* 46, 033302 (2005); doi:10.1063/1.1855932 (15 pages).
- [52] Spielman, D. and S. H. Teng, Smoothed analysis of algorithms. Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002), 597–606, Higher Ed. Press, Beijing, 2002.
- [54] Stroock, D., Probability theory. An analytic view, Cambridge University Press, Cambridge, 1993.
- [53] Szarek, S. J., Spaces with large distance to ℓ_∞^n and random matrices, *American Journal of Mathematics* 112 (1990), no. 6, 899–942.
- [55] Tao, T. and V. Vu, On the singularity probability of random Bernoulli matrices, *J. Amer. Math. Soc.* 20, 603-628 (2007).
- [56] Tao, T. and V. Vu, Random Matrices: The circular Law, *Communications in Contemporary Mathematics*, 10 (2008), 261-307.
- [57] Tao, T. and V. Vu, On the permanent of random Bernoulli matrices, *Advances in Mathematics*, Volume 220, Issue 3, 15 February 2009, Pages 657-669.

- [58] Tao, T. and V. Vu, Smooth Analysis of the Condition Number and the Least Singular Value, Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, 2009, 714-737, DOI: 10.1007/978-3-642-03685-9.
- [59] Tao, T. and V. Vu, Random matrices: the distribution of smallest singular values, Geometric And Functional Analysis, Volume 20, Number 1 / June, 2010, 260-297.
- [60] Tao, T., V. Vu, and M. Krishnapur, Random matrices: Universality of ESDs and the circular law, preprint arXiv:0807.4898 [math.PR], to appear in the Annals of Probability, 2010.
- [61] Tropp, J. A., Algorithms for simultaneous sparse approximation: Part II: Convex relaxation. Signal Processing, 86:589–602, 2006.
- [62] Wigner, E., "On the Distribution of the Roots of Certain Symmetric Matrices." Ann. of Math. 67, 325-328, 1958.
- [63] Wojtaszczyk, P., Banach Spaces For Analysts, Cambridge University Press, 1996.
- [64] Ye, J. C., S. Tak, and Y. Han, H. W. Park, Projection reconstruction MR imaging using FOCUSS, Magn Reson Med. 2007 Apr;57(4):764-75.
- [65] Yin, Y. Q., Z. D. Bai, and P. R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probab. Theory Related Fields 78 (1988), 509–521.

APPENDIX A

EXAMPLES ON THE DUALITY

Example A.0.4. $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $m = 2$, $n = 3$, and $p \geq 1$.

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then $A^T y = \begin{pmatrix} y_1 \\ y_2 \\ y_1 + y_2 \end{pmatrix}$. Therefore,

$$\begin{aligned} \|A^T y\|_q &= \left\| \begin{pmatrix} y_1 \\ y_2 \\ y_1 + y_2 \end{pmatrix} \right\|_q \\ &\geq \|y\|_q. \end{aligned} \tag{A.1}$$

and equality in (A.1) holds when $y_1 = -y_2$. So we have

$$\begin{aligned} s_2^{(q)}(A^T) &= \inf_{V \subseteq \mathbb{R}^2, \dim(V)=1} \sup_{y \in V, \|y\|_q=1} \|A^T y\|_q \\ &= \inf_{y \in \mathbb{R}^2, \|y\|_q=1} \|A^T y\|_q \\ &= 1. \end{aligned} \tag{A.2}$$

On the other hand, let $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, then $Ax = \begin{pmatrix} x_1 + x_3 \\ x_2 + x_3 \end{pmatrix}$. For any $V \subseteq \mathbb{R}^3$ with

$\dim(V) = 2$, there is a vector $\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$ in V whose third coordinate is zero and $\left\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\|_p =$

1. Therefore

$$\|A|_V\|_p \geq \left\| A \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \right\|_p = \left\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\|_p = 1. \quad (\text{A.3})$$

Choose particular $V_0 := \text{span} \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$, then $\|A|_{V_0}\|_p = 1$. So

$$\begin{aligned} s_2^{(1)}(A) &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, \|x\|_1=1} \|Ax\|_1 \\ &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \|A|_V\|_p \\ &= 1. \end{aligned} \quad (\text{A.4})$$

Thus, we conclude that for any $p \geq 1$,

$$s_{\min(m,n)}^{(p)}(A) = s_{\min(m,n)}^{(q)}(A^T), \quad (\text{A.5})$$

where $\frac{1}{p} + \frac{1}{q} = 1$, in this example.

Example A.0.5. $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \end{pmatrix}$, $m = 2$, $n = 3$, and $p = 1$.

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then $A^T y = \begin{pmatrix} y_1 \\ y_2 + y_2 \\ y_1 - 2y_2 \end{pmatrix}$. Therefore

$$\begin{aligned} s_2^{(\infty)}(A^T) &= \inf_{V \subseteq \mathbb{R}^2, \dim(V)=1} \sup_{y \in V, \|y\|_\infty=1} \|A^T y\|_\infty \\ &= \inf_{y \in \mathbb{R}^2, \|y\|_\infty=1} \|A^T y\|_\infty \\ &= \inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|). \end{aligned} \quad (\text{A.6})$$

Now discuss it in two cases:

1. If $|y_1| = 1$ and $|y_2| \leq 1$, then

$$\max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) \geq |y_1| = 1. \quad (\text{A.7})$$

2. If $|y_2| = 1$ and $|y_1| \leq 1$, then

$$\max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) \geq |y_1 - 2y_2| \geq ||y_1| - 2|y_2|| = ||y_1| - 2| \geq 1. \quad (\text{A.8})$$

Thus $\max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) \geq 1$ for all $(y_1, y_2) \in \mathbb{R}^2$ with $\max(|y_1|, |y_2|) = 1$. Hence

$$\inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|) = 1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) \geq 1. \quad (\text{A.9})$$

Particularly at $(y_1, y_2) = (1, 0)$, we have

$$[\max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|)]_{y_1=1, y_2=0} = 1. \quad (\text{A.10})$$

Thus

$$\inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|) = 1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) = 1,$$

in other words, $s_2^{(\infty)}(A^T) = 1$.

On the other hand, let $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, then $Ax = \begin{pmatrix} x_1 + x_2 + x_3 \\ x_2 - 2x_3 \end{pmatrix}$. Therefore,

$$\begin{aligned} s_2^{(1)}(A) &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, \|x\|_1=1} \|Ax\|_1 \\ &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, |x_1|+|x_2|+|x_3|=1} |x_1 + x_2 + x_3| + |x_2 - 2x_3| \quad (\text{A.11}) \\ &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2 + x_3| + |x_2 - 2x_3|}{|x_1| + |x_2| + |x_3|}. \end{aligned}$$

Let $F(V) := \sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2 + x_3| + |x_2 - 2x_3|}{|x_1| + |x_2| + |x_3|}$, then $s_2^{(1)}(A) = \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} F(V)$. For partic-

ular $V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_2 = 2x_3 \right\}$, we know

$$\sup_{x \in V_0 \setminus \{0\}} \frac{|x_1 + x_2 + x_3| + |x_2 - 2x_3|}{|x_1| + |x_2| + |x_3|} = \sup_{x \in V_0 \setminus \{0\}} \frac{|x_1 + x_2 + x_3|}{|x_1| + |x_2| + |x_3|} = 1, \quad (\text{A.12})$$

and it follows that $s_2^{(1)}(A) \leq 1$. For generic V , that is actually a plane passing through the origin in \mathbb{R}^3 , we can assume

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : ax_1 + bx_2 + cx_3 = 0 \right\} \quad (\text{A.13})$$

where a , b and c are not all zeroes. Now we discuss (A.11) in three cases:

1. If $c \neq 0$, then $x_3 = \frac{a}{c}x_1 + \frac{b}{c}x_2 = \alpha x_1 + \beta x_2$ for $\alpha := \frac{a}{c}$ and $\beta := \frac{b}{c}$. Thus we have

$$\begin{aligned} F(V) &= \sup_{(x_1, x_2) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1+\alpha)x_1 + (1+\beta)x_2| + |(1-2\beta)x_2 - 2\alpha x_1|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2|} \\ &\geq \left[\frac{|(1+\alpha)x_1 + (1+\beta)x_2| + |(1-2\beta)x_2 - 2\alpha x_1|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2|} \right]_{x_1=1, x_2=0} \\ &= \frac{|1+\alpha| + |2\alpha|}{1+|\alpha|}. \end{aligned} \quad (\text{A.14})$$

It is not hard to see that $\frac{|1+\alpha| + |2\alpha|}{1+|\alpha|} \geq 1$ for all $\alpha \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $c \neq 0$.

2. If $b \neq 0$, then $x_2 = \frac{a}{b}x_1 + \frac{c}{b}x_3 = \alpha x_1 + \gamma x_3$ for $\alpha := \frac{a}{b}$ and $\gamma := \frac{c}{b}$. Therefore,

$$\begin{aligned} F(V) &= \sup_{(x_1, x_3) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1+\alpha)x_1 + (1+\gamma)x_3| + |\alpha x_1 + (\gamma-2)x_3|}{|x_1| + |\alpha x_1 + \gamma x_3| + |x_3|} \\ &\geq \left[\frac{|(1+\alpha)x_1 + (1+\gamma)x_3| + |\alpha x_1 + (\gamma-2)x_3|}{|x_1| + |\alpha x_1 + \gamma x_3| + |x_3|} \right]_{x_1=0, x_3=1} \\ &= \frac{|1+\gamma| + |\gamma-2|}{1+|\gamma|}. \end{aligned} \quad (\text{A.15})$$

It is not hard to see that $\frac{|1+\gamma| + |\gamma-2|}{1+|\gamma|} \geq 1$ for all $\gamma \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $b \neq 0$.

3. If $a \neq 0$, then $x_1 = \frac{b}{a}x_2 + \frac{c}{a}x_3 = \beta x_2 + \gamma x_3$ for $\beta := \frac{b}{a}$ and $\gamma := \frac{c}{a}$. Therefore,

$$\begin{aligned} F(V) &= \sup_{(x_2, x_3) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1+\beta)x_2 + (1+\gamma)x_3| + |x_2 - 2x_3|}{|\beta x_2 + \gamma x_3| + |x_2| + |x_3|} \\ &\geq \left[\frac{|(1+\beta)x_2 + (1+\gamma)x_3| + |x_2 - 2x_3|}{|\beta x_2 + \gamma x_3| + |x_2| + |x_3|} \right]_{x_2=0, x_3=1} \\ &= \frac{|1+\gamma| + 2}{|\gamma| + 1}. \end{aligned} \quad (\text{A.16})$$

It is not hard to see that $\frac{|1+\gamma| + 2}{|\gamma| + 1} \geq 1$ for all $\gamma \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $a \neq 0$.

So we have shown that

$$\sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2 + x_3| + |x_2 - 2x_3|}{|x_1| + |x_2| + |x_3|} \geq 1 \quad (\text{A.17})$$

for all $V \subseteq \mathbb{R}^3$ with $\dim(V) = 2$. Hence $s_2^{(1)}(A) = 1$ in (A.11).

Finally, we conclude that

$$s_{\min(m,n)}^{(1)}(A) = s_{\min(m,n)}^{(\infty)}(A^T) \quad (\text{A.18})$$

in this example.

Remark A.0.6. For $p = 2$, using singular value decomposition, we get $s_{\min(m,n)}^{(2)}(A) = s_{\min(m,n)}^{(2)}(A^T) = \sqrt{4 - \sqrt{2}}$ for this example.

Let us give another example.

Example A.0.7. $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $m = 2$, $n = 3$, and $p = 1$.

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then $A^T y = \begin{pmatrix} y_1 \\ y_1 + y_2 \\ 0 \end{pmatrix}$. Therefore

$$\begin{aligned} s_2^{(\infty)}(A^T) &= \inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|) = 1} \max(|y_1|, |y_1 + y_2|, 0) \\ &= \inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|) = 1} \max(|y_1|, |y_1 + y_2|). \end{aligned} \quad (\text{A.19})$$

For any $(y_1, y_2) \in \mathbb{R}^2$ with $\max(|y_1|, |y_2|) = 1$, $|y_1 + y_2|$ is the distance between the origin and the y_1 -intercept of the line passing through (y_1, y_2) with slope -1 and obviously, $|y_1|$ is the horizontal (y_1 -directional) distance between the origin and (y_1, y_2) , and see Figure A.1. Minimizing the maximum of these two distances, we can see from Figure A.1 that $\max(|y_1|, |y_1 + y_2|)$ achieves its minimum at $(y_1, y_2) = (-\frac{1}{2}, 1)$ on the boundary of the square, $\{(y_1, y_2) \in \mathbb{R}^2 : \max(|y_1|, |y_2|) = 1\}$. Thus we have

$$s_2^{(\infty)}(A^T) = [\max(|y_1|, |y_1 + y_2|)]_{y_1 = -\frac{1}{2}, y_2 = 1} = \frac{1}{2}. \quad (\text{A.20})$$

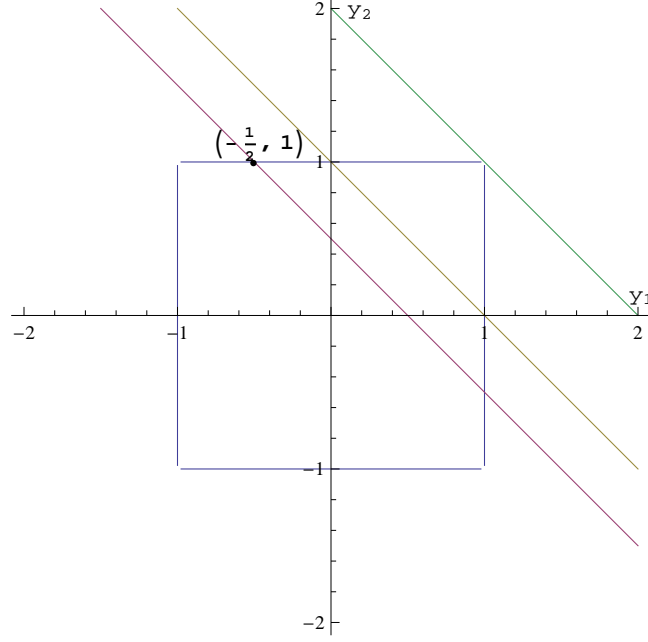


Figure A.1: Minimization of the maximum of $|y_1|$ and $|y_1 + y_2|$

On the other hand, let $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, then $Ax = \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix}$. Therefore,

$$\begin{aligned}
 s_2^{(1)}(A) &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, \|x\|_1=1} \|Ax\|_1 \\
 &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, |x_1|+|x_2|+|x_3|=1} |x_1 + x_2| + |x_2| \\
 &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V \setminus \{0\}} \frac{|x_1+x_2|+|x_2|}{|x_1|+|x_2|+|x_3|}.
 \end{aligned} \tag{A.21}$$

Let $F(V) := \sup_{x \in V \setminus \{0\}} \frac{|x_1+x_2|+|x_2|}{|x_1|+|x_2|+|x_3|}$, then $s_2^{(1)}(A) = \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} f(V)$. For generic $V \subseteq \mathbb{R}^3$ with $\dim(V) = 2$, we assume

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : ax_1 + bx_2 + cx_3 = 0 \right\} \tag{A.22}$$

where a , b and c are not all zeroes. Now we discuss (A.21) in three cases:

1. If $c \neq 0$, then $x_3 = \frac{a}{c}x_1 + \frac{b}{c}x_2 = \alpha x_1 + \beta x_2$ for $\alpha := \frac{a}{c}$ and $\beta := \frac{b}{c}$. Thus when α and β are not both zeroes, we have

$$\begin{aligned}
F(V) &= \sup_{(x_1, x_2) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2|} \\
&\geq \left[\frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2|} \right]_{x_1 = \beta, x_2 = -\alpha} \\
&= \frac{|\beta - \alpha| + |\alpha|}{|\alpha| + |\beta|} \\
&= \frac{(|\beta - \alpha| + |\alpha|) + |\beta - \alpha| + |\alpha|}{2(|\alpha| + |\beta|)} \\
&\geq \frac{1}{2}
\end{aligned} \tag{A.23}$$

and when α and β are both zeroes, obviously

$$F(V) = \sup_{(x_1, x_2) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2|} \geq \left[\frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2|} \right]_{x_2=0} = 1. \tag{A.24}$$

So $F(V) \geq \frac{1}{2}$ for all V in which $c \neq 0$.

2. If $b \neq 0$, then $x_2 = \frac{a}{b}x_1 + \frac{c}{b}x_3 = \alpha x_1 + \gamma x_3$ for $\alpha := \frac{a}{b}$ and $\gamma := \frac{c}{b}$, and then $F(V)$ can be written as

$$F(V) = \sup_{(x_1, x_3) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1 + \alpha)x_1 + \gamma x_3| + |\alpha x_1 + \gamma x_3|}{|x_1| + |\alpha x_1 + \gamma x_3| + |x_3|}. \tag{A.25}$$

By the triangle inequality,

$$|(1 + \alpha)x_1 + \gamma x_3| + |\alpha x_1 + \gamma x_3| \geq \frac{1}{2} (|x_1| + |(1 + \alpha)x_1 + \gamma x_3| + |\alpha x_1 + \gamma x_3|), \tag{A.26}$$

and therefore,

$$F(V) \geq \sup_{(x_1, x_3) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|x_1| + |(1 + \alpha)x_1 + \gamma x_3| + |\alpha x_1 + \gamma x_3|}{2(|x_1| + |\alpha x_1 + \gamma x_3| + |x_3|)}. \tag{A.27}$$

When $x_1 = 1 + |\gamma|$ and $x_3 = |1 + \alpha|$, using the triangle inequality,

$$\begin{aligned}
|(1 + \alpha)x_1 + \gamma x_3| &\geq ||1 + \alpha|x_1 - |\gamma||x_3|| \\
&= |1 + \alpha|(1 + |\gamma|) - |\gamma| \\
&\geq |1 + \alpha| \\
&= |x_3|.
\end{aligned} \tag{A.28}$$

Finally, from (A.27) we have

$$F(V) \geq \left[\frac{|x_1| + |(1+\alpha)x_1 + \gamma x_3| + |\alpha x_1 + \gamma x_3|}{2(|x_1| + |\alpha x_1 + \gamma x_3| + |x_3|)} \right]_{x_1=1+|\gamma|, x_3=|1+\alpha|} \geq \frac{1}{2} \quad (\text{A.29})$$

for all V in which $b \neq 0$.

3. If $a \neq 0$, then $x_1 = \frac{b}{a}x_2 + \frac{c}{a}x_3 = \beta x_2 + \gamma x_3$ for $\beta := \frac{b}{a}$ and $\gamma := \frac{c}{a}$. Therefore,

$$\begin{aligned} F(V) &= \sup_{(x_1, x_3) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1+\beta)x_2 + \gamma x_3| + |x_2|}{|\beta x_2 + \gamma x_3| + |x_2| + |x_3|} \\ &\geq \left[\frac{|(1+\beta)x_2 + \gamma x_3| + |x_2|}{|\beta x_2 + \gamma x_3| + |x_2| + |x_3|} \right]_{x_3=0} \\ &= \frac{|1+\beta|+1}{|\beta|+1} \\ &= \frac{(|1+\beta|+1)+|1+\beta|+1}{2(|\beta|+1)} \\ &\geq \frac{1}{2} \end{aligned} \quad (\text{A.30})$$

for all V in which $a \neq 0$.

So we have shown that

$$\sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2| + |x_3|} \geq \frac{1}{2} \quad (\text{A.31})$$

for all $V \subseteq \mathbb{R}^3$ with $\dim(V) = 2$. However, for particular

$$V_0 := \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : x_1 = -x_2 \right\}, \quad (\text{A.32})$$

we know

$$\sup_{x \in V_0 \setminus \{0\}} \frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2| + |x_3|} = \frac{1}{2}, \quad (\text{A.33})$$

and thus $F(V) = \sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2| + |x_2|}{|x_1| + |x_2| + |x_3|}$ attains its infimum at V_0 . Hence $s_2^{(1)}(A) = \frac{1}{2}$.

Finally, we conclude that

$$s_{\min(m,n)}^{(1)}(A) = s_{\min(m,n)}^{(\infty)}(A^T) \quad (\text{A.34})$$

holds for this example.

Lastly, let us also see an example of 2 by 4 matrix.

Remark A.0.8. For $p = 2$, using singular value decomposition, we get $s_{\min(m,n)}^{(2)}(A) = s_{\min(m,n)}^{(2)}(A^T) = \frac{\sqrt{5}-1}{2}$ for this example.

Example A.0.9. $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -1 \end{pmatrix}$, $m = 2$, $n = 4$, and $p = 1$.

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then $A^T y = \begin{pmatrix} y_1 \\ y_2 \\ y_1 - 2y_2 \\ y_1 - y_2 \end{pmatrix}$. Therefore

$$\begin{aligned} s_2^{(\infty)}(A^T) &= \inf_{V \subseteq \mathbb{R}^2, \dim(V)=1} \sup_{y \in V, \|y\|_\infty=1} \|A^T y\|_\infty \\ &= \inf_{y \in \mathbb{R}^2, \|y\|_\infty=1} \|A^T y\|_\infty \\ &= \inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|, |y_1 - y_2|) \\ &\geq \inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) \end{aligned} \quad (\text{A.35})$$

By Example A.0.5,

$$\inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|) = 1 \quad (\text{A.36})$$

Thus

$$\inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|, |y_1 - y_2|) \geq 1. \quad (\text{A.37})$$

Particularly at $(y_1, y_2) = (1, 0)$, we have

$$[\max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|, |y_1 - y_2|)]_{y_1=1, y_2=0} = 1. \quad (\text{A.38})$$

Thus

$$\inf_{(y_1, y_2) \in \mathbb{R}^2, \max(|y_1|, |y_2|)=1} \max(|y_1|, |y_1 + y_2|, |y_1 - 2y_2|, |y_1 - y_2|) = 1,$$

in other words, $s_2^{(\infty)}(A^T) = 1$.

On the other hand, let $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$, then $Ax = \begin{pmatrix} x_1 + x_2 + x_3 + x_4 \\ x_2 - 2x_3 - x_4 \end{pmatrix}$. Therefore,

$$\begin{aligned} s_2^{(1)}(A) &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, \|x\|_1=1} \|Ax\|_1 \\ &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V, |x_1|+|x_2|+|x_3|+|x_4|=1} |x_1 + x_2 + x_3 + x_4| + |x_2 - 2x_3 - x_4| \\ &= \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} \sup_{x \in V \setminus \{0\}} \frac{|x_1+x_2+x_3+x_4|+|x_2-2x_3-x_4|}{|x_1|+|x_2|+|x_3|+|x_4|}. \end{aligned} \quad (\text{A.39})$$

Let $F(V) := \sup_{x \in V \setminus \{0\}} \frac{|x_1+x_2+x_3+x_4|+|x_2-2x_3-x_4|}{|x_1|+|x_2|+|x_3|+|x_4|}$, then $s_2^{(1)}(A) = \inf_{V \subseteq \mathbb{R}^3, \dim(V)=2} F(V)$. For

particular $V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_2 = 2x_3 + x_4 \right\}$, we know

$$\sup_{x \in V_0 \setminus \{0\}} \frac{|x_1 + x_2 + x_3 + x_4| + |x_2 - 2x_3 - x_4|}{|x_1| + |x_2| + |x_3| + |x_4|} = \sup_{x \in V_0 \setminus \{0\}} \frac{|x_1 + x_2 + x_3 + x_4|}{|x_1| + |x_2| + |x_3| + |x_4|} = 1, \quad (\text{A.40})$$

and it follows that $s_2^{(1)}(A) \leq 1$. For generic V , that is actually a 3-dimensionanl subspace in \mathbb{R}^4 , we can assume

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^3 : ax_1 + bx_2 + cx_3 + dx_4 = 0 \right\} \quad (\text{A.41})$$

where a, b, c and d are not all zeroes. Now we discuss (A.39) in four cases:

1. If $a \neq 0$, then $x_1 = \frac{b}{a}x_2 + \frac{c}{a}x_3 + \frac{d}{a}x_4 = \beta x_2 + \gamma x_3 + \delta x_4$ for $\beta := \frac{b}{a}$, $\gamma := \frac{c}{a}$ and $\delta := \frac{d}{a}$.

Therefore,

$$\begin{aligned} F(V) &= \sup_{(x_2, x_3, x_4) \in \mathbb{R}^3 \setminus \{(0,0,0)\}} \frac{|(1+\beta)x_2 + (1+\gamma)x_3 + (1+\delta)x_4| + |x_2 - 2x_3 - x_4|}{|\beta x_2 + \gamma x_3 + \delta x_4| + |x_2| + |x_3| + |x_4|} \\ &\geq \left[\frac{|(1+\beta)x_2 + (1+\gamma)x_3 + (1+\delta)x_4| + |x_2 - 2x_3 - x_4|}{|\beta x_2 + \gamma x_3 + \delta x_4| + |x_2| + |x_3| + |x_4|} \right]_{x_2=0, x_3=1, x_4=0} \\ &= \frac{|1+\gamma|+2}{|\gamma|+1}. \end{aligned} \quad (\text{A.42})$$

It is not hard to see that $\frac{|1+\gamma|+2}{|\gamma|+1} \geq 1$ for all $\gamma \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $a \neq 0$.

2. If $b \neq 0$, then $x_2 = \frac{a}{b}x_1 + \frac{c}{b}x_3 + \frac{d}{b}x_4 = \alpha x_1 + \gamma x_3 + \delta x_4$ for $\alpha := \frac{a}{b}$, $\gamma := \frac{c}{b}$ and $\delta := \frac{d}{b}$.

Therefore,

$$\begin{aligned} F(V) &= \sup_{(x_1, x_3, x_4) \in \mathbb{R}^3 \setminus \{(0,0,0)\}} \frac{|(1+\alpha)x_1 + (1+\gamma)x_3 + (1+\delta)x_4| + |\alpha x_1 + (\gamma-2)x_3 - x_4|}{|x_1| + |\alpha x_1 + \gamma x_3| + |x_3| + |x_4|} \\ &\geq \left[\frac{|(1+\alpha)x_1 + (1+\gamma)x_3 + (1+\delta)x_4| + |\alpha x_1 + (\gamma-2)x_3 - x_4|}{|x_1| + |\alpha x_1 + \gamma x_3| + |x_3| + |x_4|} \right]_{x_1=0, x_3=1, x_4=0} \\ &= \frac{|1+\gamma|+|\gamma-2|}{1+|\gamma|}. \end{aligned} \quad (\text{A.43})$$

It is not hard to see that $\frac{|1+\gamma|+|\gamma-2|}{1+|\gamma|} \geq 1$ for all $\gamma \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $b \neq 0$.

3. If $c \neq 0$, then $x_3 = \frac{a}{c}x_1 + \frac{b}{c}x_2 + \frac{d}{c}x_4 = \alpha x_1 + \beta x_2 + \delta x_4$ for $\alpha := \frac{a}{c}$, $\beta := \frac{b}{c}$ and $\delta := \frac{d}{c}$.

Thus we have

$$\begin{aligned} F(V) &= \sup_{(x_1, x_2, x_4) \in \mathbb{R}^3 \setminus \{(0,0,0)\}} \frac{|(1+\alpha)x_1 + (1+\beta)x_2 + (1+\delta)x_4| + |(1-2\beta)x_2 - 2\alpha x_1 - x_4|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2| + |x_4|} \\ &\geq \left[\frac{|(1+\alpha)x_1 + (1+\beta)x_2 + (1+\delta)x_4| + |(1-2\beta)x_2 - 2\alpha x_1 - x_4|}{|x_1| + |x_2| + |\alpha x_1 + \beta x_2| + |x_4|} \right]_{x_1=1, x_2=0, x_4=0} \\ &= \frac{|1+\alpha|+|2\alpha|}{1+|\alpha|}. \end{aligned} \quad (\text{A.44})$$

It is not hard to see that $\frac{|1+\alpha|+|2\alpha|}{1+|\alpha|} \geq 1$ for all $\alpha \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $c \neq 0$.

4. If $d \neq 0$, then $x_4 = \frac{a}{d}x_1 + \frac{b}{d}x_2 + \frac{c}{d}x_3 = \alpha x_1 + \beta x_2 + \gamma x_3$ for $\alpha := \frac{a}{d}$, $\beta := \frac{b}{d}$ and $\gamma := \frac{c}{d}$.

Thus we have

$$\begin{aligned} F(V) &= \sup_{(x_1, x_2) \in \mathbb{R}^2 \setminus \{(0,0)\}} \frac{|(1+\alpha)x_1 + (1+\beta)x_2 + (1+\gamma)x_3| + |(1-\beta)x_2 - \alpha x_1 - (2+\gamma)x_3|}{|x_1| + |x_2| + |x_3| + |\alpha x_1 + \beta x_2 + \gamma x_3|} \\ &\geq \left[\frac{|(1+\alpha)x_1 + (1+\beta)x_2 + (1+\gamma)x_3| + |(1-\beta)x_2 - \alpha x_1 - (2+\gamma)x_3|}{|x_1| + |x_2| + |x_3| + |\alpha x_1 + \beta x_2 + \gamma x_3|} \right]_{x_1=0, x_2=1, x_3=0} \\ &= \frac{|1+\beta|+|1-\beta|}{1+|\beta|}. \end{aligned} \quad (\text{A.45})$$

It is not hard to see that $\frac{|1+\beta|+|1-\beta|}{1+|\beta|} \geq 1$ for all $\beta \in \mathbb{R}$. So $F(V) \geq 1$ for all V in which $d \neq 0$.

So we have shown that

$$\sup_{x \in V \setminus \{0\}} \frac{|x_1 + x_2 + x_3 + x_4| + |x_2 - 2x_3 - x_4|}{|x_1| + |x_2| + |x_3| + |x_4|} \geq 1 \quad (\text{A.46})$$

for all $V \subseteq \mathbb{R}^4$ with $\dim(V) = 3$. Hence $s_2^{(1)}(A) = 1$ in (A.39).

Finally, we conclude that

$$s_{\min(m,n)}^{(1)}(A) = s_{\min(m,n)}^{(\infty)}(A^T) \quad (\text{A.47})$$

in this example.

Remark A.0.10. For $p = 2$, using singular value decomposition, we get $s_{\min(m,n)}^{(2)}(A) = s_{\min(m,n)}^{(2)}(A^T) = \sqrt{5 - \sqrt{5}}$ for this example.

Remark A.0.11. The above are some computable examples for the duality. However, if we change $p = 1$ to some finite number $p > 1$ and $p \neq 2$ for a general rectangular matrix A , to get the exact values of the smallest p -singular value of A and the smallest q -singular value of A^T can be computationally hard, due to the potential difficulty in finding the exact subspace at which the infimum is achieved, but one may show they turn out to be equal in an indirect way.

APPENDIX B

NON-CONVEX GEOMETRY UNDER LINEAR MAP

In this appendix, we would like to present some work on non-convex geometry under linear map, as it is closely related to the ℓ_q -null space property of matrices for $0 < q < 1$.

First of all, let us introduce the generalized q -perimeter of a polygon.

Definition B.0.12 (q -perimeter). The q -perimeter of a polygon \mathcal{P} with edges e_1, e_2, \dots, e_m , denoted by $\text{Perimeter}_q(\mathcal{P})$, is defined as

$$\text{Perimeter}_q(\mathcal{P}) := \sum_{k=1}^m (\text{length}(e_k))^q. \quad (\text{B.1})$$

We have the following lemma for the generalized q -perimeter of polygon under linear map.

Lemma B.0.13 (q -perimeter of polygon under linear map). *For any $2 \times n$ matrix M with columns $\mathbf{c}_1, \mathbf{c}_1, \dots, \mathbf{c}_n \in \mathbb{R}^2$, we have*

$$\text{Perimeter}_q(M[-1, 1]^n) = 2^{q+1} \sum_{k=1}^n \|\mathbf{c}_k\|_2^q. \quad (\text{B.2})$$

Proof. By the singular value decomposition, there exist a 2×2 orthogonal matrix U , an $n \times n$ orthogonal matrix V and a $2 \times n$ diagonal matrix Λ with non-negative entries λ_1 and λ_2 on its main diagonal, such that $M = U\Lambda V^T$.

Let $M' := \Lambda V^T$, and let's first show that if the claim holds for M' then it is true for $M = UM'$. Now suppose

$$\text{Perimeter}_q(M'[-1, 1]^n) = 2^{q+1} \sum_{k=1}^n \|\mathbf{c}'_k\|_2^q \quad (\text{B.3})$$

where $\mathbf{c}'_1, \mathbf{c}'_2, \dots, \mathbf{c}'_n$ are the columns of M' . Since an orthogonal transformation preserves the geometric features including the length of the edges of a polygon, then

$$\text{Perimeter}_q(M[-1, 1]^n) = \text{Perimeter}_q(M'[-1, 1]^n). \quad (\text{B.4})$$

Moreover, we also have $\|\mathbf{c}'_k\|_2 = \|U\mathbf{c}_k\|_2 = \|\mathbf{c}_k\|_2$. So (B.2) holds for $M' = \Lambda V^T$ if it is true for M' .

Next, we'll show that (B.3) is true for $M' = \Lambda V^T$. For the orthogonal matrix $V^T = (V_{i,j})_{n \times n}$, we know that V^T is a composition of permutations, reflections, and (or) rotations, hence $V^T[-1, 1]^n$ is also an n -cube obtained by a rigid body motion of the n -cube $[-1, 1]^n$. Let $V^T = (V_1, V_2, \dots, V_n)$ in which V_i is the i -th column of V for $i = 1, 2, \dots, n$. To get the q -perimeter of a polygon, we just need to find its vertices. The set of vertices of the n -cube $[-1, 1]^n$ is

$$\{(\epsilon_1, \epsilon_1, \dots, \epsilon_n) \in \mathbb{R}^n : \epsilon_j = \pm 1, j = 1, 2, \dots, n\}, \quad (\text{B.5})$$

and then that of the n -cube $V^T[-1, 1]^n$ is

$$\left\{ \left(\sum_{j=1}^n \epsilon_j V_j \right)^T \in \mathbb{R}^n : \epsilon_j = \pm 1, j = 1, 2, \dots, n \right\}. \quad (\text{B.6})$$

Therefore, $\Lambda V^T[-1, 1]^n$ is a polygon with vertices

$$\left\{ \left(\lambda_1 \sum_{j=1}^n \epsilon_j V_{1,j}, \lambda_2 \sum_{j=1}^n \epsilon_j V_{2,j} \right) \in \mathbb{R}^2 : \epsilon_j = \pm 1, j = 1, 2, \dots, n \right\}. \quad (\text{B.7})$$

Summing the q -length of the edges between the adjacent vertices, in which only one component of the ϵ 's differs, we have

$$\begin{aligned} \text{Perimeter}_q(\Lambda V^T[-1, 1]^n) &= 2 \sum_{j=1}^n \left(\sqrt{(2\lambda_1 V_{1,j})^2 + (2\lambda_2 V_{2,j})^2} \right)^q \\ &= 2^{q+1} \sum_{j=1}^n \left(\sqrt{(\lambda_1 V_{1,j})^2 + (\lambda_2 V_{2,j})^2} \right)^q. \end{aligned} \quad (\text{B.8})$$

On the other hand, if $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ are the columns of ΛV^T , then

$$\sum_{k=1}^n \|\mathbf{c}_k\|_2^q = \left\| \begin{pmatrix} \lambda_1 V_{1,j} \\ \lambda_2 V_{2,j} \end{pmatrix} \right\|_2^q = \sum_{j=1}^n \left(\sqrt{(\lambda_1 V_{1,j})^2 + (\lambda_2 V_{2,j})^2} \right)^q. \quad (\text{B.9})$$

Thus we have finished the proof. \square

Remark B.0.14. Lemma B.0.13 also holds for $q > 1$. In the case when $n = 1$, $M[-1, 1]^n$ is a line segment, but we view it as a digon, which is degenerate in \mathbb{R}^2 , and then $\text{Perimeter}_q(M[-1, 1]^n)$ is 2 times the q -th power of the length of the line segment.

This lemma and Theorem 6.3.2 imply the following

Corollary B.0.15. *Given $0 < q \leq 1$ and $B \in \mathbb{R}^{2 \times N}$,*

$$\|(x, y)B_S\|_q \prec \|(x, y)B_{S^c}\|_q \quad (\text{B.10})$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{0\}$ and some $S \subseteq \{1, 2, \dots, N\}$ with $|S| = s$. Then

$$\text{Perimeter}_q(B_S([-1, 1]^s)) \prec \text{Perimeter}_q(B_{S^c}([-1, 1]^{N-s})). \quad (\text{B.11})$$

Remark B.0.16. In [19], it was shown that

$$B_S([-1, 1]^s) \subseteq B_{S^c}([-1, 1]^{N-s}) \quad (\text{B.12})$$

holds if

$$\|(x, y)B_S\|_1 \prec \|(x, y)B_{S^c}\|_1 \quad (\text{B.13})$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, for which dual spaces and the Hahn-Banach theorem were used in the proof. In fact, more generally, (B.12) holds if

$$\|(x, y)B_S\|_q \prec \|(x, y)B_{S^c}\|_q \quad (\text{B.14})$$

for some $q \geq 1$ and all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, because of the convexity of the function $t \rightarrow t^q$ when $q \geq 1$. Then the containing relation (B.12) on convex polygons allows one to compare their perimeters. However, this method becomes infeasible for the case of $0 < q < 1$ because of the following reasons.

Firstly, (B.12) may not hold if (B.14) holds for some $0 < q < 1$ and all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. An example is $q = \frac{1}{2}$,

$$B = \begin{pmatrix} 1601 & 500 & 1100 \\ 1600 & 720 & 900 \end{pmatrix} \quad (\text{B.15})$$

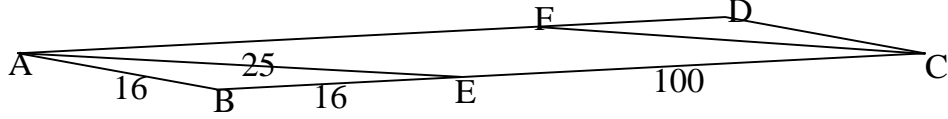


Figure B.1: Parallelograms

and $S = \{1\}$. In this example,

$$\|(x, y)B_S\|_{\frac{1}{2}} < \|(x, y)B_{S^c}\|_{\frac{1}{2}} \quad (\text{B.16})$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, that will be verified later, but the point $B_S(\{1\}) = \{(1601, 1600)\}$ in $B_S([-1, 1])$ is not contained in the parallelogram $B_{S^c}([-1, 1]^2)$.

Secondly, for a given $0 < q < 1$, (6.21) may not hold even if (B.12) holds, due to the non-convexity of the function $t \rightarrow t^q$ when $0 < q < 1$. An example for this is the following. Let B be the 2×4 matrix such that $B_{S^c}([-1, 1]^2)$ is parallelogram $ABCD$ and $B_S([-1, 1]^2)$ is parallelogram $AECF$ with the lengths of the edges prescribed in Figure B.1, and $q = \frac{1}{2}$. The $\frac{1}{2}$ -perimeter of parallelogram $AECF$ is 30 and the $\frac{1}{2}$ -perimeter of parallelogram $ABCD$ is $8 + 4\sqrt{29}$, so the q -perimeter of parallelogram $AECF$ is larger than the q -perimeter of parallelogram $ABCD$ for $q = \frac{1}{2}$, although parallelogram $AECF$ is contained in parallelogram $ABCD$.

Now let us verify the claim that if

$$B = \begin{pmatrix} 1601 & 500 & 1100 \\ 1600 & 720 & 900 \end{pmatrix}, \quad (\text{B.17})$$

then

$$\|(x, y)B_S\|_{\frac{1}{2}} < \|(x, y)B_{S^c}\|_{\frac{1}{2}} \quad (\text{B.18})$$

for $S = \{1\}$ and all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$.

It suffices to show

$$|1601x + 1600y|^{\frac{1}{2}} < |500x + 720y|^{\frac{1}{2}} + |1100x + 900y|^{\frac{1}{2}} \quad (\text{B.19})$$

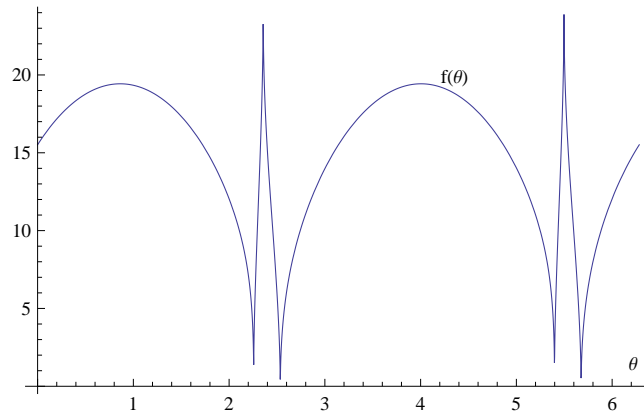


Figure B.2: The graph of difference function $f(\theta)$

all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, that can be written as

$$|1601 \cos \theta + 1600 \sin \theta|^{\frac{1}{2}} < |500 \cos \theta + 720 \sin \theta|^{\frac{1}{2}} + |1100 \cos \theta + 900 \sin \theta|^{\frac{1}{2}} \quad (\text{B.20})$$

for all $\theta \in [0, 2\pi]$.

Indeed, (B.20) holds for all $\theta \in [0, 2\pi]$, because graphing the difference function

$$f(\theta) := \sqrt{|500 \cos \theta + 720 \sin \theta|} + \sqrt{|1100 \cos \theta + 900 \sin \theta|} - \sqrt{|1601 \cos \theta + 1600 \sin \theta|} \quad (\text{B.21})$$

by using Mathematica, we can see from Figure B.2 that $f(\theta)$ is always positive for $\theta \in [0, 2\pi]$.

So (B.18) holds for $S = \{1\}$ and all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$.

APPENDIX C

SOME NUMERICAL EXPERIMENTS ON p -SINGULAR VALUE FOR $p > 1$ AND q -SINGULAR VALUE FOR $0 < q \leq 1$ OF RANDOM MATRICES

In this appendix, we show the results from some numerical experiments on the p -singular value for $p > 1$ and q -singular value for $0 < q \leq 1$ of random matrices.

For $p = 2$, we plot the largest 2-singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 100. See Figure C.1. This graph shows that the 2-singular value is $O(\sqrt{n})$.

For $p = 1$, in the first numerical experiment we plot the largest 1-singular value of Gaussian random of size $n \times n$, where n runs from 1 through 100. See Figure C.2. The graph shows that the largest 1-singular value is $O(n)$, as estimated in Theorem 2.3.10, Theorem 2.3.16, and Theorem 4.2.3 as well.

In the second numerical experiment for $p = 1$, we plot the largest 1-singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 200. See Figure C.3.

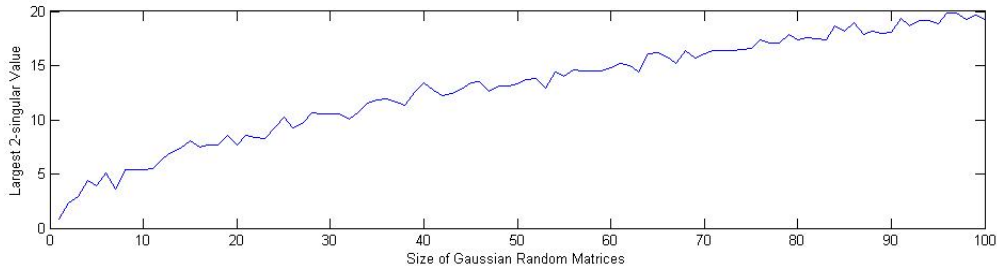


Figure C.1: Largest 2-singular value of Gaussian random matrices

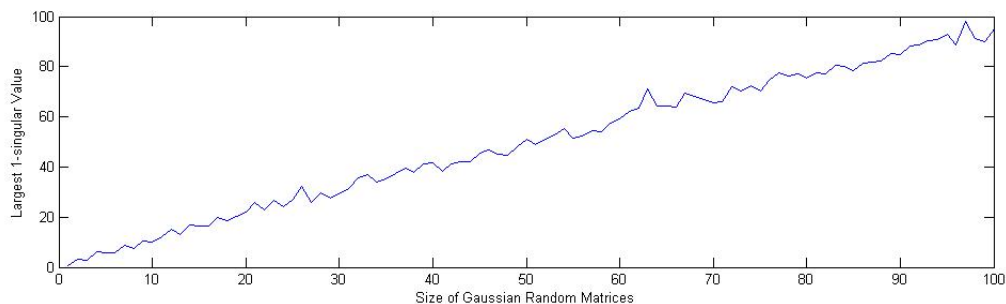


Figure C.2: Largest 1-singular value of Gaussian random matrices: Experiment 1

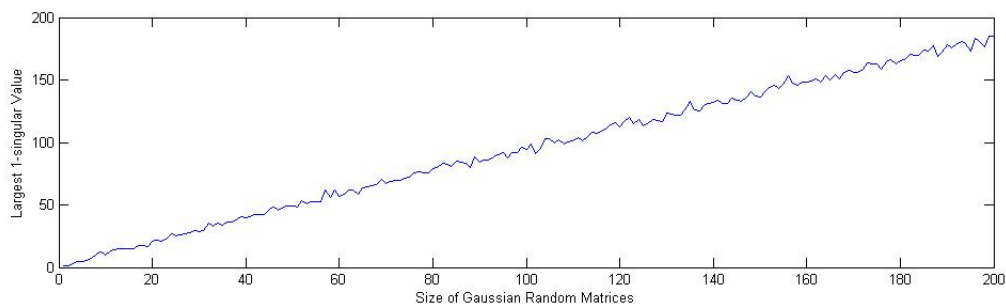


Figure C.3: Largest 1-singular value of Gaussian random matrices: Experiment 2

The graph shows that the largest 1-singular value is $O(n)$, as estimated in Theorem 2.3.10, Theorem 2.3.16, and Theorem 4.2.3 as well.

In the third experiment for $p = 1$, we plot the largest 1-singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 400. See Figure C.4. The graph shows that the largest 1-singular value is $O(n)$, as estimated in Theorem 2.3.10, Theorem 2.3.16, and Theorem 4.2.3 as well.

For $p = \infty$, we plot the largest ∞ -singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 500. See Figure C.5. This graph shows that the ∞ -singular value is $O(n)$, as estimated in Theorem 2.3.10 and Theorem 2.3.16.

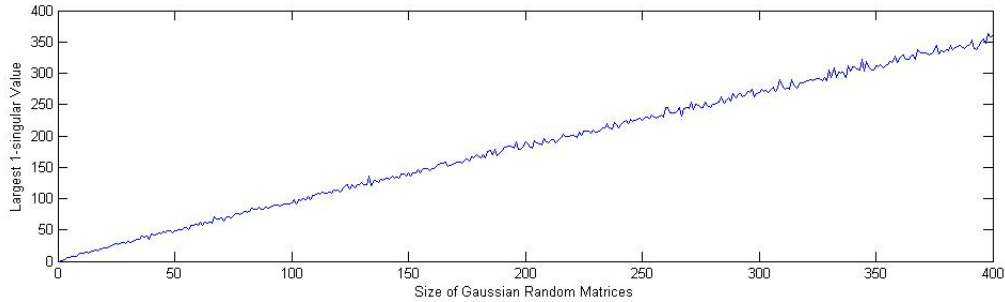


Figure C.4: Largest 1-singular value of Gaussian random matrices: Experiment 3

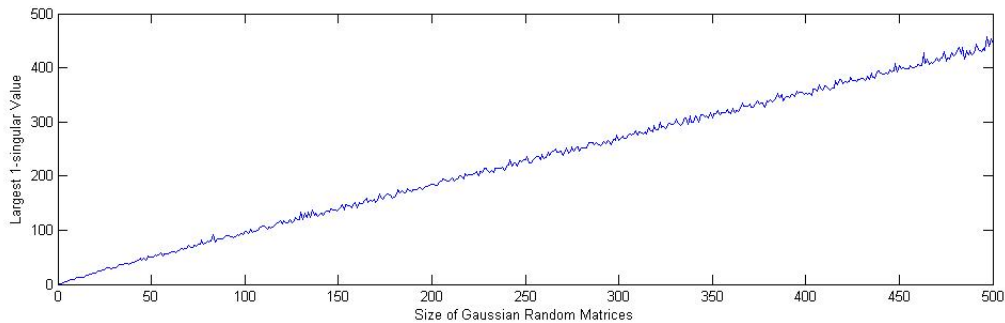


Figure C.5: Largest ∞ -singular value of Gaussian random matrices

For $p = \frac{1}{3}$, we plot the largest $\frac{1}{3}$ -singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 500. See Figure C.6. This graph shows that the ∞ -singular value is approximately $O(n^3)$, as estimated in Theorem 2.3.10 and Theorem 2.3.16.

For $p = \frac{1}{4}$, we plot the largest $\frac{1}{4}$ -singular value of Gaussian random matrices of size $n \times n$, where n runs from 1 through 300. See Figure C.7. This graph shows that the $\frac{1}{4}$ -singular value is approximately $O(n^4)$, as estimated in Theorem 2.3.10 and Theorem 2.3.16.

For rectangular matrices, we also plot the largest $\frac{1}{4}$ -singular value of Gaussian random matrices of size $m \times n$, where m and n run from 1 through 100. See Figure C.8. This graph

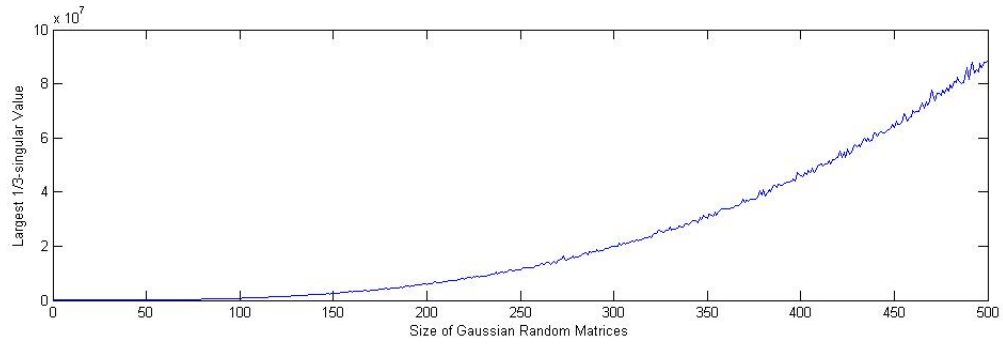


Figure C.6: Largest $\frac{1}{3}$ -singular value of Gaussian random matrices

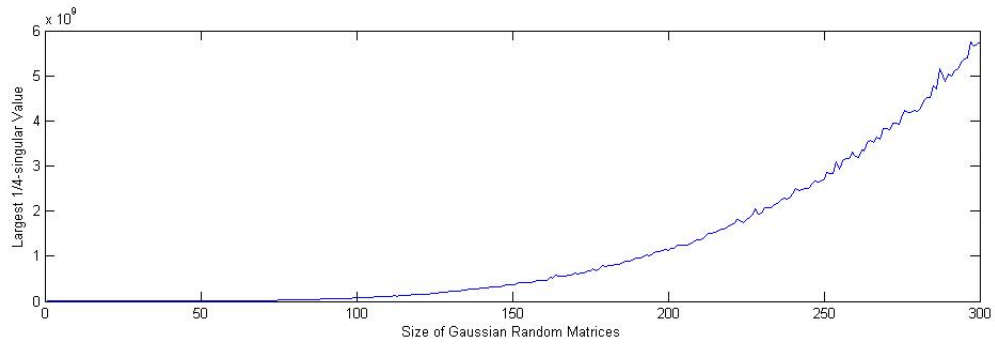


Figure C.7: Largest $\frac{1}{4}$ -singular value of Gaussian random matrices

shows that the $\frac{1}{4}$ -singular value is approximately $O(m^4)$, as estimated in Theorem 2.3.10 and Theorem 2.3.16.

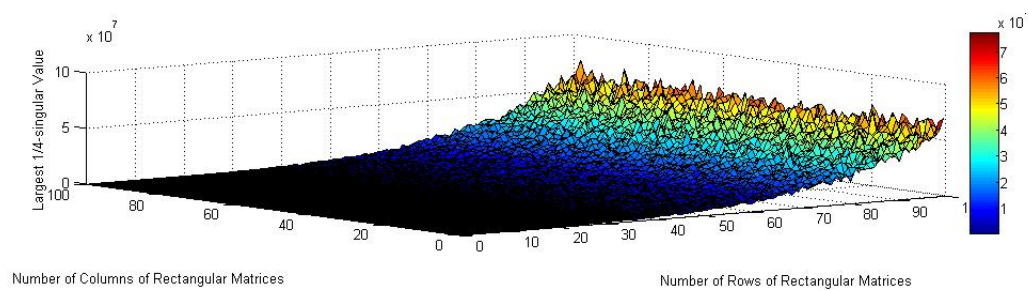


Figure C.8: Largest $\frac{1}{4}$ -singular value of rectangular Gaussian random matrices

INDEX

Bernoulli random variable, 15, 23, 55

Bernstein inequality, 16, 23, 62, 65

ℓ_q -minimization problem, 58, 68, 71, 74

Laplacian random variable, 16, 17

Largest q singular value, 18, 25, 31

MMV, 9, 79, 81

Modified restricted isometry property, 58, 59,
67

Null space property, 71, 74

Pregaussian random variable, 14, 15, 31

Smallest q -singular value, 35, 37, 49