MATHEMATICS OF COMPUTATION
Volume 84, Number 294, July 2015, Pages 1775–1794
S 0025-5718(2014)02895-0
Article electronically published on October 30, 2014

THE PROBABILISTIC ESTIMATES ON THE LARGEST AND SMALLEST q-SINGULAR VALUES OF RANDOM MATRICES

MING-JUN LAI AND YANG LIU

ABSTRACT. We study the q-singular values of random matrices with pre-Gaussian entries defined in terms of the ℓ_q-quasinorm with $0 < q \le 1$. In this paper, we mainly consider the decay of the lower and upper tail probabilities of the largest q-singular value $s_1^{(q)}$, when the number of rows of the matrices becomes very large. Based on the results in probabilistic estimates on the largest q-singular value, we also give probabilistic estimates on the smallest q-singular value for pre-Gaussian random matrices.

1. INTRODUCTION

The extremal spectrum or the largest and smallest singular values of random matrices have been of interest to many research communities including numerical analysis and multivariate statistics. For example, the condition numbers of random matrices were of interest as early as in von Neumann and Goldstein'1947, [28] and Smale'1985, [19], and distribution of the largest and smallest eigenvalues of Wishart matrices was studied in Wishart'1928, [30]. Some estimates for the probability distribution of the norm of a random matrix transformation were obtained in Bennett, Goodman and Newman'1975, [2]. In 1988, Edelman presented a comprehensive study on the distribution of the condition numbers of Gaussian random matrices together with many numerical experiments (cf. [5]). In particular, Edelman explained several interesting applications of eigenvalues of random matrices in graph theory, the zeros of Riemann zeta functions, as well as in nuclear physics (cf. [6]). Indeed, the well-known semi-circle law (cf. Wigner'1962, [29]) states that the histogram for the eigenvalues of a large random matrix is roughly a semi-circle. To be more precise, let A be a Gaussian random matrix and $M(x)$ denote the proportion of eigenvalues of the Gaussian orthogonal ensemble $(A + A^T)/(2\sqrt{n})$ (the symmetric part of $A/\sqrt{n}$) that are less than x. Then the semi-circle law asserts that

$$\frac{d}{dx}M(x) \to \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{if } x \in [-1,1], \\ 0, & \text{otherwise.} \end{cases}$$

This interesting property has made a long lasting impact and attracted many researchers to extend and generalize the semi-circle law. See recent papers of

Received by the editor November 26, 2012 and, in revised form September 23, 2013.

2010 *Mathematics Subject Classification.* Primary 60B20; Secondary 60F10, 60G50, 60G42.

Key words and phrases. Random matrices, probability, pre-Gaussian random variable, generalized singular values.

The first author was partly supported by the National Science Foundation under grant DMS-0713807.

The second author was partially supported by the Air Force Office of Scientific Research under grant AFOSR 9550-12-1-0455.

Tao and Vu'2008, [24] and Rudelson and Vershynin'2010, [17] for new results and surveys and the references therein. It is known that the largest eigenvalue of $M_s = \frac{1}{s} V_{n\times s}(V_{n\times s})^T$ converges to $(1+\sqrt{y})^2$ almost surely (cf. Geman'1980, [10]) and the smallest eigenvalue converges to $(1-\sqrt{y})^2$ almost surely (cf. Silverstein'1985, [18]), where $V_{n\times s}$ is a Gaussian random matrix of size $n \times s$ with $n/s \to y \in (0,1]$ and $V_{n\times s}(V_{n\times s})^T$ is called a Wishart matrix. The behavior of the largest singular value of random matrices A with i.i.d. entries is well studied. If a random variable ξ has a bounded fourth moment, then the largest eigenvalue $s_1(A)$ of an $n \times n$ random matrix A with i.i.d. copies of ξ satisfies the following property:

$$\lim_{n\to\infty} \frac{s_1(A)}{\sqrt{n}} = 2\sqrt{\mathbb{E}\xi^2}$$

almost surely. See, e.g., Yin, Bai, Krishnaiah'1988, [31] and Bai, Silverstein and Yin'1988, [1]. The bounded fourth moment is necessary and sufficient in this case. However, the behavior of the smallest singular value for general random matrices has been much less known. Although Edelman showed that for every $\epsilon > 0$, the smallest eigenvalue $s_n(A)$ of Gaussian random matrix A of size $n \times n$ has

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

for any $\epsilon > 0$, the probability estimates for $s_n(A)$ for general random matrix A were not known until the results in Rudelson and Vershynin'2008, [14]. In fact, Rudelson in [16] presented a less accurate probability estimate for $s_n(A)$, and soon both Rudelson and Vershynin found a simpler proof of much accurate estimate in [15]. More precisely, Rudelson and Vershynin first showed (cf. [15]) the following results:

Theorem 1.1. *If A is a matrix of size $n \times n$ whose entries are independent random variables with variance* 1 *and bounded fourth moment, then*

$$\lim_{\epsilon\to 0_+} \limsup_{n\to\infty} \mathbb{P}\left(s_n(A) \le \frac{\epsilon}{\sqrt{n}}\right) = 0.$$

Furthermore, in Rudelson and Vershynin'2008, [14], they presented a proof of the following

Theorem 1.2. *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*

$$\lim_{K\to+\infty} \limsup_{n\to\infty} \mathbb{P}\left(s_n(A) \ge \frac{K}{\sqrt{n}}\right) = 0.$$

These two results settled down a conjecture by Smale in [18] (the results on the Gaussian case were established by Edelman and Szarek; see [6] and [22]). More precise estimates for largest and smallest eigenvalues are given for sub-Gaussian random matrices, Bernoulli matrices, covariance matrices, and general random matrices of the form $M + A$ with deterministic matrix M and random matrix A in the last ten years. See, e.g. [25], [20], [14], [26], [23] and the references in [17].

In this paper, we extend these studies on the probability estimate of the largest and smallest singular values of random matrices in the ℓ_2-norm and give estimates for these extremal spectra in the setting of the ℓ_q-quasinorm for $0 < q \le 1$. Not only is it interesting to know if the probability estimates for largest and smallest singular values of random matrices in the ℓ_2-norm can be extended to the setting of the ℓ_q-quasinorm, there are also some definite advantages of using the general ℓ_q-quasinorm when studying the restricted isometry property of random matrices as suggested in Chartrand and Steneva'2008, [4], Foucart and Lai'2009, [8] and Foucart and Lai'2010, [9]. In addition to Gaussian and sub-Gaussian random matrices, we would like to study the probability estimates for pre-Gaussian random matrices. A random variable ξ is pre-Gaussian if ξ has mean zero and the moment growth condition $\mathbb{E}(|\xi|^k) \leq k!\lambda^k/2$, i.e. $(\mathbb{E}(|\xi|^k))^{1/k} \leq C\lambda k$ for $k \geq 1$ (cf. Buldygin and Kozachenko'2000, [3]). Note that the moment growth condition for a sub-Gaussian random variable η is $\left(\mathbb{E}\left(|\eta|^k\right)\right)^{1/k} \leq BC\sqrt{k}$.

To be precise on what we are going to study in this paper, for any vector $\mathbf{x} = (x_1, \cdots, x_n)^T$ in $\mathbb{R}^n$, let

$$\|\mathbf{x}\|_q^q = \sum_{i=1}^n |x_i|^q$$

for $q \in (0, \infty)$. It is known that for $q \geq 1$, $\|\cdot\|_q$ is a norm for $\mathbb{R}^n$ and $\|\cdot\|_q^q$ is a quasinorm for $\mathbb{R}^n$ for $q \in (0,1)$ that satisfies all the properties for a norm except the triangle inequality. Let $A = (a_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ be a matrix. The standard largest q-singular value is defined by

$$s_1^{(q)}(A) := \sup\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0\right\}. \tag{1.1}$$

It is known that for $q \geq 1$, the equation in (1.1) defines a norm on the space of $m \times n$ matrices. In addition, we know

$$\max_j \|a_j\|_q \leq s_1^{(q)}(A) \leq n^{\frac{q-1}{q}} \max_j \|a_j\|_q\,, \tag{1.2}$$

where a_j, $j = 1, 2, \cdots, n$, are the column vectors of A. We refer to any book on matrix theory for the properties of the largest singular value $s_1^q(A)$ when $q \geq 1$, for example, [11]. However, for $q \in (0,1)$, the properties of $s_1^q(A)$ are not well-known. For convenience, we shall explain some useful properties in the Preliminaries section.

The purpose of this paper is to study the matrix spectrum, e.g. $s_1^q(A)$ for random matrix A with pre-Gaussian entries. Two sets of our main results are the following

Theorem 1.3 (Upper tail probability of the largest q-singular value)**.** *Let ξ be a pre-Gaussian 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* $0 < q < 1$,

$$\mathbb{P}\left(s_1^{(q)}(A) \geq Cm^{\frac{1}{q}}\right) \leq \exp\left(-C'm\right) \tag{1.3}$$

for some C, $C' > 0$ only dependent on the pre-Gaussian variable ξ.

Theorem 1.4 (Lower tail probability of the largest q-singular value)**.** *Let ξ be a pre-Gaussian 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 there exists a constant $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right) \leq c^m \tag{1.4}$$

for some $0 < c < 1$, where K only depends on the pre-Gaussian variable ξ.

These results have their counterparts in papers by Yin, Bai, Krishnaiah'1988, [31], Bai, Silverstein and Yin'1988, [1] and Sosnikov'2002, [20] for the ℓ_2-norm. It is interesting to know if the above results hold for general random matrices whose entries are i.i.d. copies of a random variable of the bounded fourth moment.

Next we would like to study the smallest singular values. In general we can define the k-th q-singular value as follows.

Definition 1.1. The k-th q-singular value of an $m \times n$ matrix A is defined by

$$s_k^{(q)}(A) := \inf\left\{\sup\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in V \setminus \{0\}\right\} : V \subseteq \mathbb{R}^n,\, \dim(V) \geq n-k+1\right\}. \tag{1.5}$$

It is easy to see that

$$s_1^{(q)}(A) \geq s_2^{(q)}(A) \geq \ldots \geq s_{\min(m,n)}^{(q)}(A) \geq 0. \tag{1.6}$$

The smallest singular value $s^q_{\min(m,n)}$ is also of special interest in various studies. In the lower tail probability estimate, we divide the study in two cases when $m > n$ (tall matrices) and $m = n$ (square matrices) under the assumption that A is of full rank. The study is heavily dependent on the known results on the compressible and incompressible vectors. In the upper tail probability estimate, we use the known estimates on the projection in the ℓ_2-norm. Another set of main results is as follows. For tall random matrices, we have

Theorem 1.5 (Lower tail probability on the smallest q-singular value)**.** *Let us fix $0 < q \leq 1$. Let ξ be the pre-Gaussian random variable with mean 0 and variance 1. Suppose that A is an $m \times n$ matrix with i.i.d. copies of ξ in its entries with $m > n$. Then there exist some $\varepsilon > 0, c > 0$ and $\lambda \in (0,1)$ dependent on q and ε such that*

$$\mathbb{P}\left(s_m^{(q)}(A) \leq \varepsilon m^{1/q}\right) < e^{-cm} \tag{1.7}$$

when $n \leq \lambda m$.

For square random matrices, we have

Theorem 1.6 (Lower tail probability on the smallest q-singular value)**.** *Let us fix $0 < q \leq 1$. Let ξ be the pre-Gaussian 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 $\varepsilon > 0$, one has*

$$\mathbb{P}\left(s_n^{(q)}(A) \leq \gamma n^{-1/q}\right) < \varepsilon, \tag{1.8}$$

where $\gamma > 0$ depends only on the pre-Gaussian variable ξ.

The above theorem is an extension of Theorem 1.1. Finally we have

Theorem 1.7 (Upper tail probability on the smallest q-singular value)**.** *Given any $0 < q \leq 1$, let ξ be a pre-Gaussian 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 $\alpha > 0$ which are only dependent on the pre-Gaussian variable ξ such that*

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-1/2}\right) \leq \frac{C(\ln K)^\alpha}{K^\alpha} + c^n. \tag{1.9}$$

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

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-1/2}\right) < \varepsilon \tag{1.10}$$

for all $n \geq n_0$.

The above theorem is an extension of Theorem 1.2. Note that we are not able to prove

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-1/q}\right) < \varepsilon \tag{1.11}$$

under the assumptions in Theorem 1.7. However, we strongly believe that the above inequality holds. We leave it as a conjecture.

The remainder of the paper is devoted to the proof of these five theorems which give a good understanding of the spectrum of pre-Gaussian random matrices in ℓ_q-quasinorm with $0 < q \leq 1$. We shall present the analysis in four separate sections after the Preliminaries section.

2. Preliminaries

First of all, one can easily derive the following

Lemma 2.1. *For $0 < q < 1$, the equation in (1.1)) defines a quasinorm on the space of $m \times N$ matrices. In particular, we have*

$$\left(s_1^{(q)}(A+B)\right)^q \leq \left(s_1^{(q)}(A)\right)^q + \left(s_1^{(q)}(B)\right)^q$$

for any $m \times N$ matrices A and B. Moreover,

$$s_1^{(q)}(A) = \max_j \|a_j\|_q \tag{2.1}$$

for $0 < q \leq 1$, where a_j, $j = 1, \ldots, N$, are the columns of matrix A.

Proof. It is straightforward and not hard to show that $s_1^{(q)}(A)$, $q \leq 1$, defines a quasinorm on matrices by using the quasi-norm properties of $\|\mathbf{x}\|_q$, the ℓ_q-quasinorm on vectors.

To prove equation (2.1), on one hand, we have

$$\|Ax\|_q^q \leq \sum_{j=1}^N |x_j|^q \cdot \|a_j\|_q^q \leq \|x\|_q^q \max_j \|a_j\|_q^q \tag{2.2}$$

for $0 < q \leq 1$, which implies

$$s_1^{(q)}(A) \leq \max_j ||a_j||_q. \tag{2.3}$$

On the other hand, by (1.1), we have

$$s_1^{(q)}(A) = \sup_{x \in \mathbb{R}^N, \|x\|_q = 1} \|Ax\|_q \geq \|Ae_j\|_q = ||a_j||_q \tag{2.4}$$

for every j, where e_j is the j-th standard basis vector of $\mathbb{R}^N$, and then it follows that

$$s_1^{(q)}(A) \geq \max_j ||a_j||_q. \tag{2.5}$$

Thus, combined with (2.3), we obtain the equation (2.1) for $0 < q \leq 1$ as desired. □

Next we need the following elementary estimate. Mainly we need a linear bound for partial binomial expansion.

Lemma 2.2 (Linear bound for partial binomial expansion)**.** *For every positive integer n,*

$$\sum_{k=\left\lfloor \frac{n}{2}\right\rfloor+1}^{n} \binom{n}{k} x^k (1-x)^{n-k} \le 8x$$

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

Proof. Let us start with an even integer. For every $x \in \left[\frac{1}{8}, 1\right]$, we have

$$\sum_{k=n+1}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} \le \sum_{k=0}^{2n} \binom{2n}{k} x^k (1-x)^{2n-k} = 1 \le 8x. \tag{2.6}$$

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

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

By De Moivre-Stirling's formula (see e.g. [7]) and furthermore the estimate in [13],

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

where $\frac{1}{12n+1} < \lambda_n < \frac{1}{12n}$. We have

$$\binom{2n}{n} = \frac{\sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n} e^{\lambda_{2n}}}{\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\lambda_n}\right)^2} = \frac{4^n}{\sqrt{\pi n}} e^{\lambda_{2n}-2\lambda_n} \le \frac{4^n}{\sqrt{\pi n}}. \tag{2.7}$$

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

$$f(x) \le \sum_{k=n+1}^{2n} \binom{2n}{n} x^k (1-x)^{2n-k} \le \sum_{k=n+1}^{2n} \binom{2n}{n} x^k \le n \binom{2n}{n} x^{n+1} \tag{2.8}$$

for all $x \in [0,1]$. Using (2.7), we have

$$f(x) \le 4^n \sqrt{\frac{n}{\pi}} x^{n+1}. \tag{2.9}$$

Letting $g(x) = 4^n \sqrt{\frac{n}{\pi}} x^n$, we have

$$\ln(g(x)) = n\ln(4x) + \frac{1}{2}\ln n - \frac{1}{2}\ln \pi \le -n\ln 2 + \frac{1}{2}\ln n - \frac{1}{2}\ln\pi \le 0$$

for $x \in [0, 1/8]$. Thus we have $f(x) \le x \le 8x$. Also, we can have a similar estimate for odd integers. These complete the proof. ☐

Remark 2.1. The coefficient on the right-hand side can be improved by Markov's inequality, but the estimate obtained by the analytic technique above is actually good enough for the purposes of this paper.

Next we review the smallest q-singular values. Without loss of generality, we consider $m \ge n$. Then the n-th q-singular value is the smallest q-singular value which can also be expressed in another way.

Lemma 2.3. *Let A be an $m\times n$ matrix with $m \geq n$. Then the smallest q-singular value*

$$s_n^{(q)}(A) = \inf\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0\right\}. \tag{2.10}$$

Proof. By the definition,

$$\begin{aligned} s_n^{(q)}(A) &= \inf\left\{\sup\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in V \setminus \{0\}\right\} : V \subseteq \mathbb{R}^n, \dim(V) \geq 1\right\} \\ &\leq \inf\left\{\sup\left\{\frac{\|Av\|_q}{\|v\|_q} : v \in V \setminus \{0\}\right\} : V = \operatorname{span}(x) : x \in \mathbb{R}^n \setminus \{0\}\right\} \\ &= \inf\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0\right\}. \end{aligned} \tag{2.11}$$

We also know the infimum can be achieved by considering the unit $\mathcal{S}_q$-sphere in the finite-dimensional space, and so the claim follows. □

In particular, if A is an $n \times n$ matrix, we know

$$\begin{aligned} s_n^{(q)}(A) &= \inf\left\{\frac{\|Ax\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0\right\} \\ &= \frac{1}{\sup\left\{\frac{\|A^{-1}x\|_q}{\|x\|_q} : x \in \mathbb{R}^n \text{ with } x \neq 0\right\}} \\ &= \frac{1}{s_1^{(q)}(A^{-1})}. \end{aligned} \tag{2.12}$$

The estimate of the largest q-singular value can be used to estimate the smallest q-singular values based on this relation.

As we see, the q-singular value is defined by the ℓ_q-quasinorm, as opposed to the ℓ_2-norm, but using a similar proof for the relationship between the rank of a matrix and its smallest singular value in ℓ_2, one has the following relationship between the rank of a matrix and its smallest q-singular value.

Lemma 2.4. *For any positive integer m and n, an $m \times n$ matrix A is of full rank if and only if $s_{\min(m,n)}^{(q)}(A) > 0$.*

Remark 1. One could also derive this lemma by the properties of singular values defined by the ℓ_2-norm and by using the inequalities on the relations between the ℓ_2-norm and the ℓ_q-quasinorm.

We shall need the following result to estimate the smallest q-singular values.

Lemma 2.5. *Let A be a matrix of size $m \times N$. Suppose that $m \geq N$. Then*

$$s_{\min(m,N)}^{(q)}(A) \leq \min_j \|a_j\|_q.$$

Proof. Choose e_{j_0} to be a standard basis vector of $\mathbb{R}^N$ such that $\|Ae_{j_0}\|_q = \min_j \|a_j\|_q$ and use the definition of $s_{\min(m,N)}^{(q)}(A)$ for $m \geq N$. □

The following generalization of Lemma 4.10 in Pisier'1999, [12] will be used in a later section.

Lemma 2.6. *For $0 < q \le 1$, let $\mathcal{S}_q := \{x \in \mathbb{R}^n : |x|_q = 1\}$ denote the unit sphere of $\mathbb{R}^n$ in the ℓ_q-quasinorm. For any $\delta > 0$, there exists a finite set $\mathcal{U}_q \subseteq \mathcal{S}_q$ with*

$$\min_{\mathbf{u}\in\mathcal{U}_q} \|x-u\|_q^q \le \delta \quad \textit{for all } x \in \mathcal{S}_q \qquad \textit{and} \qquad \mathrm{card}(\mathcal{U}_q) \le \left(1+\frac{2}{\delta}\right)^{n/q}.$$

Proof. Let $(u_1, \ldots, u_k)$ be a set of k points on the sphere $\mathcal{S}_q$ such that $|u_i -_j|_q^q > \delta$ for all $i \neq j$. We choose k as large as possible. Thus, it is clear that

$$\min_{1\le i\le k} |x-u_i|_q^q \le \delta \quad \text{for all } x \in \mathcal{S}_q.$$

Let $\mathcal{B}_q := \{x \in \mathbb{R}^n : |x|_q \le 1\}$ be the unit ball of $\mathbb{R}^n$ relative to the quasinorm $|\cdot|_q$. It is easy to see that the $(\delta/2)$-balls centered at $\mathbf{u}_i$,

$$u_i + \left(\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q, \quad 1 \le i \le k,$$

are disjoint. Indeed, if x would belong to the $(\delta/2)$-ball centered at x_i as well as the $(\delta/2)$-ball centered at x_j, we would have

$$|u_i - u_j|_q^q \le |u_i - x|_q^q + |u_j - x|_q^q \le \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

which is a contradiction. Besides, it is easy to see that

$$u_i + \left(\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q \subseteq \left(1+\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q, \quad 1 \le i \le k.$$

By comparison of volumes, we get

$$k\mathrm{Vol}\left(\left(\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q\right) = \sum_{i=1}^k \mathrm{Vol}\left(\mathbf{u}_i + \left(\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q\right) \le \mathrm{Vol}\left(\left(1+\frac{\delta}{2}\right)^{1/q} \mathcal{B}_q\right).$$

Then, by homogeneity of the volumes, we have

$$k\left(\frac{\delta}{2}\right)^{n/q} \mathrm{Vol}\left(\mathcal{B}_q\right) \le \left(1+\frac{\delta}{2}\right)^{n/q} \mathrm{Vol}\left(\mathcal{B}_q\right),$$

which implies that $k \le \left(1+\frac{2}{\delta}\right)^{n/q}$. This completes the proof. ☐

3. The upper tail probability of the largest q-singular value

We begin with the following

Theorem 3.1 (Upper tail probability of the largest 1-singular value)**.** *Let ξ be a pre-Gaussian 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*

$$\mathbb{P}\left(s_1^{(1)}(A) \ge Cm\right) \le \exp\left(-C'm\right) \tag{3.1}$$

for some C, $C' > 0$ only dependent on the pre-Gaussian variable ξ.

Proof. Since a_{ij} are i.i.d. copies of the pre-Gaussian variable ξ, $\mathbb{E}a_{ij} = 0$, and there exist some $\lambda > 0$ such that $\mathbb{E}\,|a_{ij}|^k \le k!\lambda^k$ for all k. Without loss of generality, we may assume that $\lambda \ge 1$. With the variance $\mathbb{E}a_{ij}^2 = 1$, we have

$$\mathbb{E}\,|a_{ij}|^k \le \frac{\mathbb{E}a_{ij}^2}{2} H^{k-2} k!$$

for $H := 2\lambda^3$ and all $k \geq 2$. By the Bernstein inequality (cf. Theorem 5.2 in [3]), we know that

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

for all $t > 0$ and for each $j = 1, \cdots, N$. In particular, when $t = Cm$,

$$\mathbb{P}\left(\left|\sum_{j=1}^{m} a_{ij}\right| \geq Cm\right) \leq 2\exp\left(-\frac{C^2 m}{4C\lambda^3+2}\right). \tag{3.2}$$

Here a condition on C will be determined later.

On the other hand, by Lemma 2.1,

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

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

$$\mathbb{P}\left(\sum_{i=1}^{m} |a_{ij}| \geq t\right) \leq \sum_{(\epsilon_1,\ldots,\epsilon_m)\in\{-1,1\}^m} \mathbb{P}\left(\sum_{i=1}^{m} \epsilon_i a_{ij} \geq t\right). \tag{3.3}$$

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

$$\begin{aligned}
\mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right) &\leq m \quad \mathbb{P}\left(\sum_{i=1}^{m} |a_{ij}| \geq Cm\right) \\
&\leq \quad 2^m m\mathbb{P}\left(\left|\sum_{i=1}^{m} a_{ij}\right| \geq Cm\right) \\
&\leq \quad 2^m m\exp\left(-\tfrac{C^2 m}{4C\lambda^3+2}\right) \\
&\leq \quad \exp\left(-\left(\tfrac{C^2}{4C\lambda^3+2} - \ln 2 - 1\right) m\right).
\end{aligned} \tag{3.4}$$

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

$$C > 2\lambda^3 + 2\lambda^3\ln 2 + \sqrt{2 + 2\ln 2 + 4\lambda^6 + 8\lambda^6\ln 2 + 4\lambda^6\ln^2 2}. \tag{3.5}$$

That is, choosing $C' = \frac{C^2}{4C\lambda^3+2} - \ln 2 - 1$, we get (3.1). □

The previous theorem allows us to estimate the largest q-singular value for $0 < q < 1$. The estimate can follow easily from Theorem 3.1, but it is one of the tail probabilistic estimates we wanted to obtain, so let us state it as a theorem, which is Theorem 1.3.

Proof of Theorem 1.3. By Hölder's inequality, we have $\|a_j\|_q \leq m^{\frac{1}{q}-1}\|a_j\|_1$ for $0 < q < 1$. It follows from Lemma 2.1 that

$$s_1^{(q)}(A) = \max_j \|a_j\|_q \leq m^{\frac{1}{q}-1} s_1^{(1)}(A). \tag{3.6}$$

From (3.1), we have

$$\begin{aligned}\mathbb{P}\left(s_1^{(q)}(A) \geq Cm^{\frac{1}{q}}\right) &\leq \mathbb{P}\left(m^{\frac{1}{q}-1}s_1^{(1)}(A) \geq Cm^{\frac{1}{q}}\right)\\ &= \mathbb{P}\left(s_1^{(1)}(A) \geq Cm\right)\\ &\leq \exp\left(-C'm\right)\end{aligned} \tag{3.7}$$

for some $C,\, C' > 0$. ∎

4. The lower tail probability of the largest q-singular value

Let us use the result in Lemma 2.2 to give estimates on the lower tail probabilities of the largest q-singular value.

Lemma 4.1. *Suppose $\xi_1,\, \xi_2,\, \cdots,\, \xi_n$ are i.i.d. copies of a random variable ξ. Then for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^{n} |\xi_i| \leq \frac{n\varepsilon}{2}\right) \leq 8\mathbb{P}\left(|\xi| \leq \varepsilon\right). \tag{4.1}$$

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

$$\left\{(\xi_1,\, \dots,\xi_n) : \sum_{i=1}^{n} |\xi_i| \leq \frac{n\varepsilon}{2}\right\} \tag{4.2}$$

is contained in

$$\bigcup_{k=\lfloor \frac{n}{2}\rfloor+1}^{n} \;\bigcup_{\substack{\{i_1,\dots,i_k\}\\ \subset\{1,\dots,n\}}} \left\{(\xi_1,\dots,\xi_n) : |\xi_{i_1}| \leq \varepsilon,\dots,|\xi_{i_k}| \leq \varepsilon, \left|\xi_{i_{k+1}}\right| > \varepsilon,\dots,|\xi_{i_n}| > \varepsilon\right\}, \tag{4.3}$$

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, and let us denote the set (4.3) by $\mathcal{E}$.

Let $x = \mathbb{P}\left(|\xi_1| \leq \varepsilon\right)$. Then by the union probability,

$$\mathbb{P}\left(\mathcal{E}\right) = \sum_{k=\lfloor \frac{n}{2}\rfloor+1}^{n} \binom{n}{k} x^k \left(1-x\right)^{n-k}, \tag{4.4}$$

and applying Lemma 2.2, we have

$$\mathbb{P}\left(\mathcal{E}\right) \leq 8x = 8\mathbb{P}\left(|\xi_1| \leq \varepsilon\right). \tag{4.5}$$

Since the event (4.2) is contained in the event (4.3), we have

$$\mathbb{P}\left(\sum_{i=1}^{n} |\xi_i| \leq \frac{n\varepsilon}{2}\right) \leq \mathbb{P}\left(\mathcal{E}\right) \leq 8\mathbb{P}\left(|\xi_1| \leq \varepsilon\right). \tag{4.6}$$

∎

We start with a lower tail probability for the 1-singular values.

Theorem 4.1 (Lower tail probability of the largest 1-singular value)**.** *Let ξ be a pre-Gaussian 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 there exists a constant $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(1)}(A) \leq Km\right) \leq c^m \tag{4.7}$$

for some $0 < c < 1$, *where* K *only depends on the pre-Gaussian variable* ξ.

Proof. Since a_{ij} has variance 1, there exists $\delta > 0$ and $0 \leq \beta < 1$ such that

$$\mathbb{P}\left(|a_{ij}| \leq \delta\right) = \beta. \tag{4.8}$$

Let B_j be the number of variables in $\{a_{ij}\}_{i=1}^{m}$ that are less than or equal to δ. Then if $\sum_{i=1}^{m} |a_{ij}| \leq \delta \cdot \lambda m$ for $0 < \lambda < 1$, then $B_j \geq (1-\lambda)\, m$, because otherwise $\sum_{i=1}^{m} |a_{ij}| > \delta \cdot \lambda m$. It follows that

$$\mathbb{P}\left(\sum_{i=1}^{m} |a_{ij}| \leq \delta \cdot \lambda m\right) \leq \mathbb{P}\left(B_j \geq (1-\lambda)\, m\right). \tag{4.9}$$

By Markov's inequality,

$$\mathbb{P}\left(B_j \geq (1-\lambda)\, m\right) \leq \frac{\mathbb{E}B_i}{(1-\lambda)\, m}, \tag{4.10}$$

but B_j satisfies a binomial distribution of m independent experiments, each of which yields success with probability β; therefore

$$\mathbb{P}\left(B_j \geq (1-\lambda)\, m\right) \leq \frac{\beta}{1-\lambda}. \tag{4.11}$$

By choosing suitable λ, we can make $0 < \frac{\beta}{1-\lambda} < 1$. Thus

$$\mathbb{P}\left(\sum_{i=1}^{m} |a_{ij}| \leq \delta \cdot \lambda m\right) \leq c \tag{4.12}$$

for some $0 < c < 1$. It follows that

$$\begin{aligned} \mathbb{P}\left(s_1^{(1)}(A) \leq \lambda\delta m\right) &= \mathbb{P}\left(\max_{1\leq j\leq N}\left(\textstyle\sum_{i=1}^{m} |a_{ij}|\right) \leq \lambda\delta m\right) \\ &= \textstyle\prod_{j=1}^{m} \mathbb{P}\left(\left(\sum_{i=1}^{m} |a_{ij}|\right) \leq \lambda\delta m\right) \\ &\leq c^m. \end{aligned} \tag{4.13}$$

Thus letting $K = \lambda\delta$, we obtain (3.1).

□

For general $0 < q < 1$, we have Theorem 1.4.

Proof of Theorem 1.4. We can use the same method as in the proof of Theorem 4.1. Since a_{ij} has nonzero variance, there exists $\delta > 0$ and $0 \leq \beta < 1$ such that

$$\mathbb{P}\left(|a_{ij}|^q \leq \delta\right) = \beta. \tag{4.14}$$

Then by Lemma 4.1 and substituting a_{ij} in the proof of Theorem 4.1 by $|a_{ij}|^q$,

$$\begin{aligned} \mathbb{P}\left(s_1^{(q)}(A) \leq (\lambda\delta)^{\frac{1}{q}}\, m^{\frac{1}{q}}\right) &= \mathbb{P}\left(\max_{1\leq j\leq N}\left(\textstyle\sum_{i=1}^{m} |a_{ij}|^q\right) \leq \lambda\delta m\right) \\ &= \textstyle\prod_{j=1}^{m} \mathbb{P}\left(\left(\sum_{i=1}^{m} |a_{ij}|^q\right) \leq \lambda\delta m\right) \\ &\leq c^m \end{aligned} \tag{4.15}$$

for some $0 < c < 1$. Thus letting $K = (\lambda\delta)^{\frac{1}{q}}$, (1.4) follows. □

Remark 2. If one uses the quasinorm comparison inequality $s_1^{(q)}(A) \leq s_1^{(1)}(A)$ for $0 < q \leq 1$, one can get

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km\right) \leq c^m \tag{4.16}$$

for $0 < q \leq 1$, but with a loss of the estimate on $\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right)$.

Since the bounded moment growth condition for pre-Gaussian variables is not needed in the proof of Theorem 4.1, the above proofs also show that the theorem holds for any random variable with nonzero variance. Therefore, more generally, we have

Theorem 4.2. *Let ξ be a random variable with non-zero variance and A be an $m \times m$ matrix with i.i.d. copies of ξ in its entries. Then there exists a constant $K > 0$ such that*

$$\mathbb{P}\left(s_1^{(q)}(A) \leq Km^{\frac{1}{q}}\right) \leq c^m \tag{4.17}$$

for some $0 < c < 1$, where K only depends on ε and the random variable ξ.

5. The lower tail probability of the smallest q-singular value

In this section, we first study the probability estimates of the smallest q-singular value of rectangular random matrices with $m > n$. Then we give some estimates for square random matrices.

5.1. The tall random matrix case.

In this subsection, we assume that $n \leq \lambda m$ with $\lambda \in (0, 1)$ and consider the smallest q-singular value of random matrices of size $m \times n$.

Theorem 5.1. *Given any $0 < q \leq 1$, let ξ be the pre-Gaussian random variable with variance 1 and A be an $m \times n$ matrix with i.i.d. copies of ξ in its entries. Then there exist some $\gamma > 0, b > 0$ and $\nu \in (0, 1)$ dependent on the pre-Gaussian random variable ξ such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \gamma m^{1/q}\right) < e^{-bm} \tag{5.1}$$

with $n \leq \nu m$.

To prove this result, we need to establish a few lemmas.

Lemma 5.1. *Fix any $0 < q \leq 1$. For any $\xi_1, \cdots, \xi_m$ that are i.i.d. copies of a pre-Gaussian variable with non-zero variance, for any $c \in (0, 1)$ there exists $\lambda \in (0, 1)$, that does not depend on m, such that*

$$\mathbb{P}\left(\sum_{k=1}^{m} |\xi_k|^q < \lambda m\right) \leq c^m. \tag{5.2}$$

Proof. For any $\xi_1, \cdots, \xi_m$ that are i.i.d. copies of a pre-Gaussian variable with non-zero variance, we know that there exists some $\delta > 0$ such that

$$\varepsilon_0 := \mathbb{P}\left(|\xi_k| \leq \delta\right) < 1 \tag{5.3}$$

for $k = 1, 2, \cdots, m$, because otherwise the pre-Gaussian variable would have a zero variance. Then using the Riemann–Stieltjes integral for expectation, we have

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

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

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

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

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

Finally, applying Markov's inequality, we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^m |\xi_k|^q < \lambda m\right) &= \mathbb{P}\left(\exp\left(m - \frac{1}{\lambda}\sum_{k=1}^m |\xi_k|^q\right) > 1\right) \\ &\le \mathbb{E}\left(\exp\left(m - \frac{1}{\lambda}\sum_{k=1}^m |\xi_k|^q\right)\right) \\ &= e^m \prod_{k=1}^m \mathbb{E}\exp\left(-\frac{|\xi_k|^q}{\lambda}\right). \\ &\le (3e\epsilon_0/2)^m.\end{aligned}$$

For any $c \in (0, 1)$, we choose ϵ_0 such that $3e\epsilon_0/2 = c$. This completes the proof. ☐

The following lemma is a property of the linear combination of pre-Gaussian variables, which allows us to obtain the probabilistic estimate on $\|Av\|_q$ for the pre-Gaussian ensemble A.

Lemma 5.2 (Linear combination of pre-Gaussian variables). *Let a_{ij}, $i = 1, 2, \cdots, m$ and $j = 1, 2, \ldots, n$ be pre-Gaussian variables and $\eta_i = \sum_{j=1}^n a_{ij} x_j$. Then η_i are pre-Gaussian variables for $i = 1, 2, \ldots, m$.*

Proof. Since a_{ij} are pre-Gaussian variables, $\mathbb{E}a_{ij} = 0$, and there are constants $\lambda_{ij} > 0$ such that $\mathbb{E}\left|a_{ij}\right|^k \le k!\lambda_{ij}^k$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, N$. It is easy to see

$$\mathbb{E}\eta_i = \sum_{j=1}^N x_j \mathbb{E}a_{ij} = 0.$$

Letting $\|x\|_1 = \sum_{i=1}^N |x_j|$, we use the convexity to have

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

for all integers $k \geq 1$. Thus, η_k is a pre-Gaussian random variable. □

Combining two lemmas above, we obtain the following

Lemma 5.3. *Given any $0 < q \leq 1$ and letting A be an $m \times n$ pre-Gaussian matrix, for any $c \in (0,1)$ there exists $\lambda \in (0,1)$ such that*

$$\mathbb{P}\left(\|Av\|_q < \lambda^{1/q} m^{1/q}\right) \leq c^m \tag{5.4}$$

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

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. By using Lemma 2.6, for any $\delta > 0$ there exists a δ-net $\mathcal{U}_q$ in unit sphere $\mathbb{S}_q$ such that

$$\min_{u \in \mathcal{U}_q} \|x - u\|_q^q \leq \delta \quad \text{for all } x \in \mathcal{S}_q \qquad \text{and} \qquad \text{card}(\mathcal{U}_q) \leq \left(1 + \frac{2}{\delta}\right)^{n/q}.$$

By Lemma 5.3, for all $v \in \mathcal{U}_q$ we have

$$\mathbb{P}\left(\|Av\|_q^q < \lambda m, \text{ for all } v \in \mathcal{U}_q\right) \leq \left(1 + \frac{2}{\delta}\right)^{n/q} c^m. \tag{5.5}$$

Since the event $s_n^{(q)}(A) < \gamma m^{\frac{1}{q}}$ implies $\|Av'\|_q < 2\gamma m^{\frac{1}{q}}$ for some $v' \in \mathbb{S}_q$,

$$\mathbb{P}(s_n^{(q)}(A) < \gamma m^{1/q}) \leq \mathbb{P}\left(\|Av\|_q < 2\gamma m^{1/q} \text{ for some } v \in \mathbb{S}_q\right).$$

If $v \in \mathcal{U}_q$, we use (5.5) with $2\gamma < \lambda^{1/q}$ to have

$$\mathbb{P}(s_n^{(q)}(A) < \gamma m^{1/q}) \leq \left(1 + \frac{2}{\delta}\right)^{n/q} c^m. \tag{5.6}$$

If $v \notin \mathcal{U}_q$, we use Theorem 1.3 to have

$$\begin{aligned}&\mathbb{P}\left(\|Av\|_q < 2\gamma m^{1/q} \text{ with } v \in \mathbb{S}_q \backslash \mathcal{U}_q\right) \\ &\leq e^{-c_1 m} + \mathbb{P}\left(s_1^{(q)}(A) \leq K m^{1/q} \text{ and } \|Av\|_q < 2\gamma m^{1/q} \text{ with } v \in \mathbb{S}_q \backslash \mathcal{U}_q\right).\end{aligned}$$

When $v \in \mathbb{S}_q \backslash \mathcal{U}_q$ in the event that $s_1^{(q)}(A) \leq K m^{1/q}$ and $\|Av\|_q < 2\gamma m^{1/q}$, there exists a $u \in \mathcal{U}_q$ within a q-distance δ such that

$$\begin{aligned}\|Au\|_q^q &\leq \|A(v-u)\|_q^q + \|Av\|_q^q \\ &\leq \left(s_1^{(q)}(A)\right)^q \|v - u\|_q^q + \|Av\|_q^q \\ &\leq K^q m\delta + (2\gamma)^q m \\ &< \lambda^q m\end{aligned}$$

if $\delta < \dfrac{\lambda^q - (2\gamma)^2}{K^q}$. We can use (5.5) again to conclude

(5.7)
$$\mathbb{P}\left(s_1^{(q)}(A) \le Km^{1/q} \text{ and } \|Av\|_q < 2\gamma m^{1/q} \text{ for some } v \in \mathbb{S}_q \backslash \mathcal{U}_q\right) \le \left(1 + \frac{2}{\delta}\right)^{n/q} c^m.$$

If we choose ν and c small enough in Lemma 5.1 with $n = \nu m$ such that

$$c_2 := \left(1 + \frac{2}{\delta}\right)^{\nu/q} c < 1,$$

we have thus completed the proof by choosing $b > 0$ such that $e^{-c_1 m} + e^{-c_2 m} \le e^{-bm}$. □

5.2. **The square random matrix case.** Now let us consider the square random matrices with pre-Gaussian entries.

Theorem 5.2. *Given any $0 < q \le 1$, let ξ be the pre-Gaussian 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 $0 < q \le 1$, there exist some $K > 0$ and $c > 0$ dependent on ε and the pre-Gaussian random variable ξ such that*

$$\mathbb{P}\left(s_n^{(q)}(A) < \varepsilon n^{-\frac{1}{q}}\right) < C\varepsilon + C\alpha^n + \mathbb{P}\left(\|A\| > Kn^{-\frac{1}{2}}\right), \tag{5.8}$$

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

To prove the above theorem, we generalize the ideas in Rudelson and Vershynin'2008, [15] to the setting of the ℓ_q-quasinorm. We first decompose $\mathbb{S}_q^{n-1}$ into the set of compressible vectors and the set of incompressible vectors. The concepts of compressible and incompressible vectors in $\mathbb{S}_2^{n-1}$ were introduced in [15]. See also Tao and Vu'2009, [27]. We shall use a generalized version of these concepts. Recall that $\|x\|_0$ denotes the number of nonzero entries of the vector $x \in \mathbb{R}^n$.

Definition 5.1 (Compressible and incompressible vectors in $\mathbb{S}_q^{n-1}$)**.** Fix $\rho, \lambda \in (0,1)$. 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 $\|v'\|_0 \le \lambda n$ satisfying $\|v - v'\|_q \le \rho$. The set of incompressible vectors is defined as

$$Incomp_q(\lambda, \rho) := \mathbb{S}_q^{n-1} \setminus Comp_q(\lambda, \rho). \tag{5.9}$$

Now using the decomposition in Definition 5.1, we have

$$\begin{aligned}\mathbb{P}\left(s_n^{(q)}(A) < \varepsilon n^{-\frac{1}{q}}\right) \quad &\le \quad \mathbb{P}\left(\inf_{v \in Comp_q(\lambda,\rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}}\right) \\ &\quad + \mathbb{P}\left(\inf_{v \in Incomp_q(\lambda,\rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}}\right).\end{aligned} \tag{5.10}$$

In the following we are going to consider each of the two terms on the right hand side of the above equation. For the first term on compressible vectors, we can apply Lemma 5.3 since

$$\mathbb{P}\left(\inf_{v \in Incomp_q(\lambda,\rho)} \|Av\|_q < \varepsilon n^{-\frac{1}{q}}\right) \le \mathbb{P}\left(\inf_{v \in Incomp_q(\lambda,\rho)} \|Av\|_q < \nu n^{\frac{1}{q}}\right), \tag{5.11}$$

to conclude that it actually decays exponentially for $n > 1$, where $\nu = \lambda^{1/q}$ as in Lemma 5.3.

However, for incompressible vectors, we first consider $dist_q(X_j, H_j)$, which denotes the distance between column X_j of an $n \times n$ random matrix A and the span of other columns $H_j := span\left(X_1, \cdots, X_{j-1}, X_{j+1}, \ldots, X_n\right)$ in the ℓ_q-quasinorm. We generalize a result on the probability estimate of the distance in the ℓ_2-norm in [15] to the ℓ_q-quasinorm setting. This allows us to transform the probabilistic estimate on $\|Av\|_q$ for $v \in Incomp_q(\lambda, \rho)$ to the probabilistic estimate on the average of the distances $dist_q(X_j, H_j)$, $j = 1, 2, \ldots, n$.

Lemma 5.4. *Let A be an $n \times n$ random matrix with columns $X_1, \ldots, X_n$, and let*

$$H_j := span\left(X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_n\right).$$

Then for any $\rho, \lambda \in (0,1)$ and $\varepsilon > 0$, one has

$$\mathbb{P}\left(\inf_{v \in Incomp_q(\lambda,\rho)} \|Av\|_q < \varepsilon \rho n^{-\frac{1}{q}}\right) < \frac{1}{\lambda n} \sum_{j=1}^{n} \mathbb{P}\left(dist_q(X_j, H_j) < \varepsilon\right), \tag{5.12}$$

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

Proof. For every $v \in Incomp_q(\lambda, \rho)$, by Definition 5.1, 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 $\mathcal{I}_2(A) := \{j : dist_q(X_j, H_j) < \varepsilon\}$ and $\mathcal{E}$ be the event such that for the cardinality of $\mathcal{I}_2(A)$, $|\mathcal{I}_2(A)| \geq \lambda n$. Applying Markov's inequality, we have

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

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

$$|\{j : dist_q(X_j, H_j) \geq \varepsilon\}| > (1 - \lambda)\, n$$

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

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

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

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

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

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

These complete the proof. ☐

Note that $dist_q(X_j, H_j) \geq dist(X_j, H_j)$ because $\|\cdot\|_q \geq \|\cdot\|_2$. Thus we can take the advantage of the estimate on $\mathbb{P}(dist(X_j, H_j) < \varepsilon)$ given in [15] to obtain the estimate on $\mathbb{P}(dist_q(X_j, H_j) < \varepsilon)$.

Theorem 5.3 (Distance bound (cf. [15]))**.** *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 $\varepsilon > 0$,*

$$\mathbb{P}\left(dist\left(X_j, H_j\right) < \varepsilon \text{ and } \|A\| \leq Kn^{-\frac{1}{2}}\right) \leq C\left(\varepsilon + \alpha^n\right), \tag{5.13}$$

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

The above theorem implies that

$$\mathbb{P}\left(dist_q\left(X_j, H_j\right) < \varepsilon\right) \leq \mathbb{P}\left(dist\left(X_j, H_j\right) < \varepsilon\right) \leq C\left(\varepsilon + \alpha^n\right) + \mathbb{P}\left(\|A\| \leq Kn^{-\frac{1}{2}}\right). \tag{5.14}$$

Combining (5.10) and applying Lemma 5.4, we now reach the desired inequality in Theorem 5.2.

Furthermore, since A is pre-Gaussian, using a standard concentration bound we know that for every $\varepsilon > 0$ there exists some $K > 0$ such that $\mathbb{P}\left(\|A\| \leq Kn^{-\frac{1}{2}}\right) < \varepsilon$. Thus, we have proved Theorem 1.6.

6. The upper tail probability of the smallest q-singular value

In this section, we continue to study the estimate of the upper tail probability of the smallest q-singular value of an $n \times n$ pre-Gaussian matrix. Mainly we are going to prove Theorem 1.7. To do so, we need some preparation.

Let X_j be the j-th column vector of A and π_j be the projection onto the subspace $H_j := span\left(X_1, \ldots, X_{j-1}, X_{j+1}, \cdots, X_n\right)$. We first have

Lemma 6.1. *For every $\alpha > 0$, one has*

$$\mathbb{P}\left(\left\|X_j - \pi_j\left(X_j\right)\right\|_q \geq \alpha n^{\frac{1}{q}-\frac{1}{2}}\right) \leq c_1 e^{-c_2\alpha} + c_3 n^{-c_4} \tag{6.1}$$

for each $j = 1, 2, \ldots, n$, where $c_1, c_2, c_3, c_4 > 0$ are constants independent of j, n, and q.

Proof. Without loss of generality, assume $j = 1$. Write $(a_1, a_2, \ldots, a_n) := X_1 - \pi_1\left(X_1\right)$. Applying the Bessy-Esseen theorem (see for instance [21]), we know that

$$\mathbb{P}\left(\left\|X_j - \pi_j\left(X_j\right)\right\|_2 \geq \alpha\right) = \mathbb{P}\left(\left|\frac{\sum_{i=1}^n a_i\xi_i}{\sqrt{\sum_{i=1}^n a_i^2}}\right| \geq \alpha\right) = \mathbb{P}\left(|\mathbf{g}| \geq \alpha\right) + O\left(n^{-c}\right) \tag{6.2}$$

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

By the Hölder inequality,

$$\left\|X_j - \pi_j\left(X_j\right)\right\|_q \leq n^{\frac{1-q}{q}}\left\|X_j - \pi_j\left(X_j\right)\right\|_1 \leq n^{\frac{1}{q}-\frac{1}{2}}\left\|X_j - \pi_j\left(X_j\right)\right\|_2.$$

It follows that

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

Therefore it follows from (6.2) that

$$\begin{aligned}\mathbb{P}\left(\left\|X_j - \pi_j\left(X_j\right)\right\|_q \geq \alpha n^{\frac{1}{q}-\frac{1}{2}}\right) &\leq \mathbb{P}\left(|\mathbf{g}| \geq \alpha\right) + O\left(n^{-c}\right) \\ &= \tfrac{2}{\sqrt{2\pi}}\int_\alpha^\infty e^{-\frac{1}{2}x^2}dx + O\left(n^{-c}\right) \\ &\leq c_1 e^{-c_2\alpha} + c_3 n^{-c_4}\end{aligned}$$

for some positive constants c_1, c_2, c_3, c_4. $\square$

We are now ready to prove Theorem 1.7.

Proof of Theorem 1.7. From Section 5.2 and by Lemma 2.4, we know that the $n\times n$ pre-Gaussian matrix A is invertible with very high probability. Thus, we have

$$\mathbb{P}\left(s_n^{(q)}(A) \le \frac{\alpha t}{\varepsilon}\cdot n^{-1/q}\right) \ge \mathbb{P}\left(\|v\|_q \le \alpha, \left\|A^{-1}v\right\|_q \ge \frac{\varepsilon}{t}\cdot n^{1/q} \text{ for some } v\in\mathbb{R}^n\right). \tag{6.3}$$

Thus it suffices to show that

$$\mathbb{P}\left(\|v\|_q \le \alpha, \left\|A^{-1}v\right\|_q \ge \frac{\varepsilon}{t}\cdot n^{1/q}\right) \ge 1-\varepsilon \tag{6.4}$$

for some vector $v \neq 0$.

Using the result established in Rudelson and Vershynin'2008, [14], we can easily get the desired probability of the event that $\left\|A^{-1}v\right\|_q \le \frac{\varepsilon}{t}\cdot n^{\frac{1}{q}}$ occurs. Indeed, since $\left\|A^{-1}v\right\|_q \ge \left\|A^{-1}v\right\|_2$, we know that

$$\begin{aligned} \mathbb{P}\left(\left\|A^{-1}v\right\|_q \le \tfrac{\varepsilon}{t}\cdot n^{-1/q}\right) &\le \mathbb{P}\left(\left\|A^{-1}v\right\|_2 \le \tfrac{\varepsilon}{t}\cdot n^{1/q}\right)\\ &= \mathbb{P}\left(\left\|A^{-1}v\right\|_2 \le \tfrac{\varepsilon}{t}\cdot \left(n^{2/q}\right)^{1/2}\right)\\ &\le 2p\left(4\varepsilon, t, n^{2/q}\right), \end{aligned} \tag{6.5}$$

where $p(\varepsilon,t,n) := c_5\left(\varepsilon + e^{-c_6t^2} + e^{-c_7n}\right)$ for some positive constants c_5, c_6, c_7.

Next let us choose $v = X_1 - \pi_1(X_1)$. Lemma 6.1 together with the estimate in (6.5) yield (6.4). Indeed, letting $u = t = \sqrt{\ln M}$ with $M > 1$ and $\varepsilon = \frac{1}{M}$, we have

$$\mathbb{P}\left(s_n^{(q)}(A) > M\ln M\cdot n^{-1/2}\right) \le \frac{C}{M^\alpha} + c^n \tag{6.6}$$

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

$$\mathbb{P}\left(s_n^{(q)}(A) > Kn^{-1/2}\right) \le \frac{C(\ln M)^\alpha}{K^\alpha} + c^n \le \frac{C\left(\ln(M\ln M)\right)^\alpha}{K^\alpha} + c^n = \frac{C(\ln K)^\alpha}{K^\alpha} + c^n \tag{6.7}$$

if $M \ge e$, which requires $K > e$. These complete the proof. ☐

Acknowledgment

The authors would like to thank the referees for useful comments.

References

[1] Z. D. Bai, Jack W. Silverstein, and Y. Q. Yin, *A note on the largest eigenvalue of a large-dimensional sample covariance matrix*, J. Multivariate Anal. **26** (1988), no. 2, 166–168, DOI 10.1016/0047-259X(88)90078-4. MR963829 (89i:62083)

[2] G. Bennett, V. Goodman, and C. M. Newman, *Norms of random matrices*, Pacific J. Math. **59** (1975), no. 2, 359–365. MR0393085 (52 #13896)

[3] V. V. Buldygin and Yu. V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, Translations of Mathematical Monographs, vol. 188, American Mathematical Society, Providence, RI, 2000. Translated from the 1998 Russian original by V. Zaiats. MR1743716 (2001g:60089)

[4] Rick Chartrand and Valentina Staneva, *Restricted isometry properties and nonconvex compressive sensing*, Inverse Problems **24** (2008), no. 3, 035020, 14, DOI 10.1088/0266-5611/24/3/035020. MR2421974 (2009d:94027)

[5] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560, DOI 10.1137/0609045. MR964668 (89j:15039)

[6] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Massachusetts Institute of Technology, 1989. PRoQuest LLC. MR2941174

[7] A. Fisher, *The Mathematical Theory of Probabilities and its Application to Frequency Curves and Statistical Methods*, vol. 1, The Macmillan Company, 1922.

[8] Simon Foucart and Ming-Jun Lai, *Sparsest solutions of underdetermined linear systems via l_q-minimization for* $0 < q \leq 1$, Appl. Comput. Harmon. Anal. **26** (2009), no. 3, 395–407, DOI 10.1016/j.acha.2008.09.001. MR2503311 (2011b:65045)

[9] Simon Foucart and Ming-Jun Lai, *Sparse recovery with pre-Gaussian random matrices*, Studia Math. **200** (2010), no. 1, 91–102, DOI 10.4064/sm200-1-6. MR2720209 (2011g:15061)

[10] Stuart Geman, *A limit theorem for the norm of random matrices*, Ann. Probab. **8** (1980), no. 2, 252–261. MR566592 (81m:60046)

[11] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996. MR1417720 (97g:65006)

[12] Gilles Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge Tracts in Mathematics, vol. 94, Cambridge University Press, Cambridge, 1989. MR1036275 (91d:52005)

[13] Herbert Robbins, *A remark on Stirling's formula*, Amer. Math. Monthly **62** (1955), 26–29. MR0069328 (16,1020e)

[14] Mark Rudelson and Roman Vershynin, *The least singular value of a random square matrix is* $O(n^{-1/2})$ (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **346** (2008), no. 15-16, 893–896, DOI 10.1016/j.crma.2008.07.009. MR2441928 (2009i:60104)

[15] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633, DOI 10.1016/j.aim.2008.01.010. MR2407948 (2010g:60048)

[16] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), no. 2, 575–600, DOI 10.4007/annals.2008.168.575. MR2434885 (2010f:46021)

[17] Mark Rudelson and Roman Vershynin, *Non-asymptotic Theory of Random Matrices: Extreme Singular Values*, Proceedings of the International Congress of Mathematicians. Volume III, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602. MR2827856 (2012g:60016)

[18] Jack W. Silverstein, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab. **13** (1985), no. 4, 1364–1368. MR806232 (87b:60050)

[19] Steve Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), no. 2, 87–121, DOI 10.1090/S0273-0979-1985-15391-1. MR799791 (86m:65061)

[20] Alexander Soshnikov, *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, DOI 10.1023/A:1019739414239. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. MR1933444 (2003h:62108)

[21] Daniel W. Stroock, *Probability Theory*, 2nd ed., Cambridge University Press, Cambridge, 2011. An analytic view. MR2760872 (2012a:60003)

[22] Stanisław J. Szarek, *Condition numbers of random matrices*, J. Complexity **7** (1991), no. 2, 131–149, DOI 10.1016/0885-064X(91)90002-F. MR1108773 (92i:65086)

[23] Terence Tao and Van Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628, DOI 10.1090/S0894-0347-07-00555-3. MR2291914 (2008h:60027)

[24] Terence Tao and Van Vu, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307, DOI 10.1142/S0219199708002788. MR2409368 (2009d:60091)

[25] Terence Tao and Van Vu, *On the permanent of random Bernoulli matrices*, Adv. Math. **220** (2009), no. 3, 657–669, DOI 10.1016/j.aim.2008.09.006. MR2483225 (2010b:15014)

[26] Terence Tao and Van Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. **20** (2010), no. 1, 260–297, DOI 10.1007/s00039-010-0057-8. MR2647142 (2011m:60020)

[27] Terence Tao and Van Vu, *Smooth analysis of the condition number and the least singular value*, Math. Comp. **79** (2010), no. 272, 2333–2352, DOI 10.1090/S0025-5718-2010-02396-8. MR2684367 (2011k:65065)

[28] John von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. **53** (1947), 1021–1099. MR0024235 (9,471b)

[29] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327. MR0095527 (20 #2029)

[30] J. Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika **20** (1928), no. 1/2, 32–52.

[31] Y. Q. Yin, 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), no. 4, 509–521, DOI 10.1007/BF00353874. MR950344 (89g:60117)

Department of Mathematics, The University of Georgia, Athens, Georgia 30602
E-mail address: mjlai@math.uga.edu

Department of Mathematics, Michigan State University, East Lansing, Michigan 488244-1027
E-mail address: yliu@math.msu.edu