On the singular values of random matrices

Shahar Mendelson [1,3,5]   Grigoris Paouris[2,4,6]

July 5, 2012

Abstract

We present an approach that allows one to bound the largest and smallest singular values of an $N \times n$ random matrix with iid rows, distributed according to a measure on $\mathbb{R}^n$ that is supported in a relatively small ball and linear functionals are uniformly bounded in L_p for some $p > 8$, in a quantitative (non-asymptotic) fashion. Among the outcomes of this approach are optimal estimates of $1 \pm c\sqrt{n/N}$ not only in the case of the above mentioned measure, but also when the measure is log-concave or when it a product measure of iid random variables with "heavy tails".

1 Introduction

The question of estimating the extremal singular value of a random matrix of the form $\Gamma = N^{-1/2} \sum_{i=1}^{N} \langle X_i, \cdot \rangle e_i$, that is, of an $N \times n$ matrix with iid rows, distributed according to a probability measure μ on $\mathbb{R}^n$, has attracted much attention in recent years. As a part of the non-asymptotic approach to the theory of random matrices, obtaining sharp quantitative bounds has many important applications, for example, in Asymptotic Geometric Analysis and

[1]Department of Mathematics, Technion, I.I.T, Haifa 32000, Israel.

[2]Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, U.S.A.

[3]Part of this research was supported by the Centre for Mathematics and its Applications, The Australian National University, Canberra, ACT 0200, Australia. Additional support was given by an Australian Research Council Discovery grant DP0986563, the European Community's Seventh Framework Programme (FP7/2007-2013) under ERC grant agreement 203134, and by the Israel Science Foundation grant 900/10.

[4]Part of this research was supported by the A. Sloan Foundation and the US National Science Foundation grant DMS-0906150.

[5]Email: shahar@tx.technion.ac.il

[6]Email: grigoris@math.tamu.edu

in Statistics. Instead of listing some of those applications, we refer the reader to [8, 17, 10, 5, 1, 2, 3, 19, 21] and references therein for more details on the history of the problem and its significance. General surveys on the non-asymptotic theory of random matrices may be found in [18, 20].

Our main motivation is to identify assumptions on the measure μ that allow one to obtain the typical behavior of the extremal singular values of Γ, i.e., assumptions that ensure that for $N \geq n$, with high probability,

$$1 - c\sqrt{\frac{n}{N}} \leq s_{\min}(\Gamma) \leq s_{\max}(\Gamma) \leq 1 + c\sqrt{\frac{n}{N}},$$

where c is an absolute constant.

Two particularly interesting cases are when μ is an isotropic, log-concave measure [8, 17, 10, 5, 11, 12, 1, 2, 3], and when μ is some natural extension of the situation in the asymptotic Bai-Yin theorem [21, 19, 13], formulated below.

Theorem 1.1 *[7] Let $A = A_{N,n}$ be an $N \times n$ random matrix with independent entries, distributed according to a random variable ξ, for which*

$$\mathbb{E}\xi = 0, \quad \mathbb{E}\xi^2 = 1 \quad \text{and} \quad \mathbb{E}\xi^4 < \infty.$$

If $N, n \to \infty$ and the aspect ratio n/N converges to $\beta \in (0, 1]$, then

$$\frac{1}{\sqrt{N}} s_{\min}(A) \to 1 - \sqrt{\beta}, \quad \frac{1}{\sqrt{N}} s_{\max}(A) \to 1 + \sqrt{\beta}$$

almost surely. Also, without the fourth moment assumption, $s_{\max}(A)/\sqrt{N}$ is almost surely unbounded.

In a more general setting we assume that the n-dimensional rows X_i, $1 \leq i \leq N$, of the matrix Γ are independent and distributed according to an isotropic probability measure μ, (that is, for every $t \in S^{n-1}$, $\mathbb{E}\big\langle X, t \big\rangle = 0$ and $\mathbb{E}|\big\langle X, t\big\rangle|^2 = 1$), and that every linear functional has bounded p moments, i.e. that $\sup_{t \in S^{n-1}} \|\big\langle X, t\big\rangle\|_p \leq \kappa_1$ (or in the "ψ_1-case", that $\sup_{t \in S^{n-1}} \|\big\langle X, t\big\rangle\|_{\psi_1} \leq \kappa_2$, where $\|\langle X, t\rangle\|_{\psi_1} = \inf\{s > 0 : \mathbb{E}\exp(\frac{|\langle X,t\rangle|}{s}) \leqslant 2\}$). Note that obtaining the desired bound is equivalent to showing that with high probability,

$$\sup_{t \in B_2^n} |\sum_{i=1}^{N} (\big\langle X_i, t\big\rangle^2 - 1)| \leq c\sqrt{Nn}, \tag{1.1}$$

where c is a constant that depends only on p and κ_1 (or just on κ_2 in the ψ_1 case), and B_2^n is the Euclidean unit ball in $\mathbb{R}^n$. Since we are interested in CLT-type rates, with a decay of $\sim 1/\sqrt{N}$, we will focus on the case $p > 4$, because for $p < 4$, CLT rates are false. Such rates in the non-asymptotic Bai-Yin estimate have recently been established in [13] for $X = (\xi_i)_{i=1}^n$, where the ξ_i's are iid, mean-zero, variance 1 random variables that belong to some L_p space for $p > 4$ (while different rates have been proved there for $2 < p \leq 4$).

The common threads linking the log-concave case and the "heavy tails" one are that in both, the random vector X satisfies that with high probability, the Euclidean norm $\|X\|$ is of the order of $\sqrt{n}$, and that linear functionals $\langle X, t\rangle$ are well behaved: for a log-concave measure $\sup_{t\in S^{n-1}} \|\langle X,t\rangle\|_{\psi_1} \leq \kappa_2$, and in the "heavy tails" case, $\sup_{t\in S^{n-1}} \|\langle X,t\rangle\|_{L_p} \leq \kappa_1(p)$.

Having this in mind, the goal of this note is to present a proof of the following result:

Theorem 1.2 *Let μ be an isotropic probability measure on $\mathbb{R}^n$, set $N \geq n$ and assume that $\max_{i\leq N} \|X_i\| \leq C_0(Nn)^{1/4}$. Let $\kappa_1 \geq 1$ and set k_0 to be the first integer which satisfies that $k_0 \log(eN/k_0) \geq n$.*
If $p > 8$, $\sup_{t\in B_2^n} \|\langle t, \cdot\rangle\|_{L_p} \leq \kappa_1$ and $1 \leq \beta \leq c_1 k_0$, then with μ^N-probability at least

$$1 - c_2\left(\frac{1}{N^\beta} + \exp(-c_3 n)\right),$$

$$\sup_{t\in S^{n-1}} |\sum_{i=1}^N \langle X_i, t\rangle^2 - 1| \leq c_4\sqrt{nN},$$

where c_1, c_2, c_3 and c_4 depend only on β, p, C_0 and κ_1.

Following the proof of Theorem 1.2, one can establish the same result in the ψ_1-case but with a better estimate on the probability. The following theorem has already appeared in [1, 2, 4] and recently M. Talagrand found a shorter proof of the same fact [16]. Instead of essentially repeating the proof of Theorem 1.2 we will state at each step the corresponding result in the ψ_1 case and only sketch the changes required in the proof.

Theorem 1.3 *Let μ be an isotropic probability measure on $\mathbb{R}^n$, set $N \geq n$ and assume that $\max_{i\leq N} \|X_i\| \leq C_0(Nn)^{1/4}$. If $\sup_{t\in B_2^n} \|\langle t, \cdot\rangle\|_{\psi_1} \leq \kappa_2$, then with μ^N-probability at least*

$$1 - 2\left(\exp(-c_1(Nn)^{1/4}) + \exp(-c_1 n)\right),$$

$$\sup_{t \in S^{n-1}} \left| \sum_{i=1}^{N} \left\langle X_i, t \right\rangle^2 - 1 \right| \leq c_2 \sqrt{nN},$$

where c_1 and c_2 are constants that depend only on C_0 and κ_2.

As will be explained later, the probability estimate of $\exp(-cn)$ that appears in Theorems 1.2 and 1.3 is the correct one when N is larger than $\exp(c_p n)$ and $\exp(cn)$ respectively.

The two theorems lead to the desired estimates on the singular values of Γ by a standard argument which we will not present in full. It is well understood that one may replace the L_∞ condition on $\|X\|$ with the assumption that $Pr(\max_{i \leq N} \|X_i\| \geq t(Nn)^{1/4})$ is well behaved, and the modifications needed in the proofs are minimal. Moreover, in all the examples mentioned above the probability $Pr(\max_{i \leq N} \|X_i\| \geq t(Nn)^{1/4})$ is well behaved. Indeed, if μ is log-concave then it follows from [15] that $Pr(\max_{i \leq N} \|X_i\| \geq t(Nn)^{1/4}) \leq 2\exp(-ct(Nn)^{1/4})$; and if $\xi \in L_p$ for $p > 4$ and $X = (\xi_i)_{i=1}^n$, one may show that $Pr(\max_{i \leq N} \|X_i\| \geq t(Nn)^{1/4}) \leq c_p (n/N)^{p/4-1} t^{-p}$. Since adapting the proof from the L_∞ assumption to the tail-based one is standard and has appeared in many places, we will not repeat it here.

Theorem 1.2 extends the recent result from [13] beyond the case in which X has iid coordinate, distributed according to $\xi \in L_p$ for some $p > 4$, and with a considerably easier proof than the original one (although it does not cover the range $4 < p \leq 8$, nor can it be extended to a more general context than the case of the Euclidean ball as in (1.1).

Theorem 1.3 was established in [1, 2], but with a weaker probability estimate of $1 - 2\exp(-c\sqrt{n})$. Very recently the original proof from [1, 2] was simplified in [4] and [16], and with the same probability estimate as we obtain here. In fact, several ideas used in both these proofs are essential in ours as well, although we believe that our proof is simpler. Moreover, the proofs from [1, 2] and [4], [16] use the ψ_1 assumption in an essential way and cannot be extended to the "heavy tails" case.

Throughout, we will denote absolute constants by $c_1, c_2, \ldots$. Their value may change from line to line. We write $A \lesssim B$ if there is an absolute constant c_1 for which $A \leq c_1 B$. $A \sim B$ means that $c_1 A \leq B \leq c_2 A$ for absolute constants c_1 and c_2. If the constants depend on some parameter r we will write $A \lesssim_r B$ or $A \sim_r B$. We will denote the Euclidean norm by $\| \; \|$. Finally, if (a_n) is a sequence, set (a_n^*) to be the non-increasing rearrangement of $(|a_n|)$.

2 The Proof

We begin with the following simple observation on a monotone rearrangement of iid random variables. Recall that k_0 satisfies that $k_0 \log(eN/k_0) \sim n$ if $\log(eN) \lesssim n$, and $k_0 = 1$ otherwise.

Lemma 2.1 *Let $Z_1, ..., Z_N$ be iid random variables, distributed according to Z.*

1. *If $p > 4$ and $C_0, \beta > 0$, there exist constants c_0, c_1, c_2 and c_3 that depend only on p, C_0 and β for which the following hold. If $\|Z\|_{L_\infty} \leq C_0(Nn)^{1/4}$ and $u \geq c_0$, then $\sum_{i \leq uk_0} (Z_i^*)^2 \leq c_1(1 + u\|Z\|_{L_p}^2)(Nn)^{1/2}$ with probability at least $1 - c_2 N^{-\beta}$ and $\sum_{i=uk_0+1}^{N} (Z_i^*)^4 \leq c_1 \|Z\|_{L_p}^4 N$ with probability at least $1 - 2\exp(-c_3 un)$.*
2. *There exist absolute constants $c_4, ..., c_7$ for which the following hold. If $Z \in L_{\psi_1}$, then with probability at least $1 - 2\exp(-c_4(Nn)^{1/4})$, $\sum_{i \leq k_0} (Z_i^*)^2 \leq c_5 \|Z\|_{\psi_1}^2 (Nn)^{1/2}$. Also, for $u \geq c_6$, with probability at least $1 - 2\exp(-c_7 un)$, $\sum_{i=k_0+1}^{N} (Z_i^*)^4 \leq c_5 u^4 \|Z\|_{\psi_1}^4 N$.*

Proof. The fact that $Pr(Z_{2^s}^* \geq t) \leq \binom{N}{2^s} \left(Pr(|Z| \geq t)\right)^{2^s}$ is the main ingredient in the proof. We will also assume that $k_0 > 1$, and in particular, that $k_0 \log(eN/k_0) \sim n$. If $k_0 = 1$, the modifications required are minimal and we will omit the proof in that case.

First, consider the L_p case. Fix $\varepsilon = p/4 - 1$, let $\beta \geq 1$ and set s_2 which depends only on β and p and will be named later. For $2^{s_2} \leq 2^s \leq k_0$ put $\alpha_s = (eN/2^s)^{(1+\varepsilon)/p}/(Nn)^{1/4} = 2^{s/4}/n^{1/4}$. Since $Pr(|Z| \geq \|Z\|_{L_p} t) \leq t^{-p}$, then in that range, $Pr(Z_{2^s}^* \geq \|Z\|_{L_p} \alpha_s (Nn)^{1/4}) \leq (eN/2^s)^{-\varepsilon 2^s}$. Hence, for a right choice of $s_2(\beta, p)$, and since $4(1+\varepsilon)/p = 1$, then with probability at least $1 - (eN/2^{s_2})^{-c\varepsilon 2^{s_2}} \geq 1 - c_0 N^{-\beta}$,

$$\sum_{2^s \leq uk_0} 2^s (Z_{2^s}^*)^2 \leq \|Z\|_{L_\infty}^2 2^{s_2} + \|Z\|_{L_p}^2 (Nn)^{1/2} \sum_{2^{s_2} \leq 2^s \leq uk_0} 2^s \alpha_s^2$$

$$\lesssim_{C_0} 2^{s_2}(Nn)^{1/2} + \|Z\|_{L_p}^2 N^{1/2} \sum_{2^{s_2} \leq 2^s \leq un} 2^{s/2} \lesssim_{\beta,p} (1 + u^{1/2}\|Z\|_{L_p}^2)(Nn)^{1/2}.$$

For the second part, take $t_s = \|Z\|_{L_p}(eN/2^s)^{(1+\varepsilon)/p} = \|Z\|_{L_p}(eN/2^s)^{1/4}$ and let $\max\{2/\varepsilon, 1\} < u \lesssim (N/k_0)^{1/2}$. Hence, with probability at least

$$1 - \sum_{uk_0 \leq 2^s \leq N} \exp(-\varepsilon 2^s \log(eN/2^s)) \geq 1 - \exp(-c_1 \varepsilon u k_0 \log(eN/k_0))$$

$$\geq 1 - \exp(-c_2 \varepsilon un),$$

$$\sum_{uk_0 \le 2^s \le N} 2^s (Z^*_{2^s})^4 \lesssim \|Z\|_{L_p}^4 \sum_{uk_0 \le 2^s \le N} 2^s (eN/2^s)^{4(1+\varepsilon)/p} \lesssim_p \|Z\|_{L_p}^4 N.$$

Next, consider the ψ_1 case. Set s_1 to be the first integer for which $2^s \log(eN/2^s) \ge (Nn)^{1/4}$ and assume without loss of generality that $2^{s_1} \le k_0$. Put $\alpha_s \sim 1/2^s$ for $s \le s_1$ and let $\alpha_s \sim \log(eN/2^s)/(Nn)^{1/4}$ for $2^{s_1} \le 2^s \le k_0$. Note that if $s \le s_1$ then

$$\begin{aligned} Pr(Z^*_{2^s} \ge \|Z\|_{\psi_1} \alpha_s (Nn)^{1/4}) &\le \exp(2^s \log(eN/2^s) - c_1 (Nn)^{1/4}) \\ &\le \exp(-c_2 (Nn)^{1/4}), \end{aligned}$$

and if $2^{s_1} \le 2^s \le k_0$ then

$$\Pr(Z^*_{2^s} \ge \|Z\|_{\psi_1} \alpha_s (Nn)^{1/4}) \le \exp(-c_3 2^s \log(eN/2^s)).$$

Since $k_0 \log^2(eN/k_0) \lesssim n \log(eN/n)$ then

$$\sum_{2^s \le k_0} 2^s \alpha_s^2 \le \sum_{2^s \le k_0} 2^{-s} + \frac{2^s \log^2(eN/2^s)}{(Nn)^{1/2}} \lesssim 1 + \left(\frac{n}{N}\right)^{1/2} \log\left(\frac{eN}{n}\right) \lesssim c_4.$$

Summing the probabilities, it follows that with probability at least $1 - 2\exp(-c_5 (Nn)^{1/4})$,

$$\sum_{i=1}^{k_0} (Z_i^*)^2 \lesssim \sum_{2^s \le k_0} 2^s (Z^*_{2^s})^2 \lesssim \|Z\|_{\psi_1}^2 \sqrt{Nn},$$

which proves our first claim in the ψ_1 case.

Turning to the second part, fix $u \ge 2$ and consider $t_s = u\|Z\|_{\psi_1} \log(eN/2^s)$. Since $Pr(Z^*_{2^s} \ge t_s) \le \exp(-(u-1)2^s \log(eN/2^s))$ and $k_0 \log(eN/k_0) \sim n$, then by summing the probabilities, it is evident that

$$\sum_{k_0 \le 2^s \le N} 2^s (Z^*_{2^s})^4 \le u^4 \|Z\|_{\psi_1}^4 \sum_{k_0 \le 2^s \le N} 2^s \log^4(eN/2^s) \lesssim u^4 \|Z\|_{\psi_1}^4 N$$

with probability at least $1 - 2\exp(-c_6 un)$. ■

The following corollary uses the same idea as in Lemma 2.1 and we will need it only when $k_0 > 1$. To formulate it, fix $0 < \gamma < 1$ and κ_3 to be named later, let $k_\ell = \gamma^\ell k_0$ and set ℓ_0 to be the first integer satisfying that $k_{\ell_0} \log(eN/k_{\ell_0}) \le \kappa_3 (Nn)^{1/4}$. The constants γ and κ_3 will depend only on p and their value will be specified in the proof of Lemma 2.3 below.

Corollary 2.2 *There exist a constant c_1 such that for every γ there exist constant $c_2 = c_2(\gamma)$ for which the following holds. Let $p > 4$ and $\varepsilon = p/4 - 1$, set $\ell_1 > 0$ to be any integer for which $k_{\ell_1} \geq 1$, and let $Z_1, ..., Z_N$ be iid random variables, distributed according to Z with $\|Z\|_p < \infty$. Then, for every $0 \leq \ell < \ell_1$, with probability at least $1-(eN/k_{\ell+1})^{-\varepsilon k_{\ell+1}}$, $(\sum_{j=k_{\ell+1}}^{k_\ell} (Z_j^*)^2)^{1/2} \leq c_1 \|Z\|_p \eta_\ell$, where $\eta_\ell \sim (Nk_\ell)^{\frac{1}{4}}$. In particular we have that $\sum_{\ell=0}^{\ell_1 - 1} \eta_\ell \leq c_2 (Nn)^{1/4}$.*

Moreover, if $Z_1, ..., Z_N$ are iid random variables, distributed according to Z with $\|Z\|_{\psi_1} < \infty$, there exist absolute constants c_3, c_4 and c_5 for which the following holds. Let $\gamma = 1/2$, and for every $0 \leq \ell < \ell_0$ and $u \geq c_3$, with probability at least $1 - 2\exp(-c_4 u k_\ell \log(eN/k_\ell))$, $(\sum_{j=k_{\ell+1}}^{k_\ell} (Z_j^*)^2)^{1/2} \leq c_5 u \|Z\|_{\psi_1} \bar{\eta}_\ell$, where $\sum_{\ell=0}^{\ell_0-1} \bar{\eta}_\ell \leq c_5 (Nn)^{1/4}$.

The proof of Corollary 2.2 follows from the same argument used in the second parts of the L_p and ψ_1 cases in Lemma 2.1, with the choice of $t_s = (eN/k_\ell)^{(1+\varepsilon)/p} = (eN/k_\ell)^{1/4}$ in the L_p case and $t_s = u \log(eN/k_\ell)$ in the ψ_1 one, combined with a straightforward calculation.

Next, let us turn to the main ingredient of the proof. Consider $U_k = \{x \in S^{N-1} : |\text{supp}(x)| \leq k\}$ and set $A_k = \sup_{a \in U_k} \| \sum_{i=1}^N a_i X_i \|$. The motivation for studying this quantity is that for every $k \leq N$, $A_k = \sup_{t \in B_2^n} \left(\sum_{i=1}^k (\langle X_i, t \rangle^*)^2 \right)^{1/2}$, but for reasons that will become clear later, we only need to bound A_{k_0}.

For every k, let δ_k be determined later and set $\mathcal{N}_k$ a subset of B_2^N satisfying that for every $x \in \mathbb{R}^N$,

$$\sup_{y \in \mathcal{N}_k} \langle y, x \rangle \geq (1 - \delta_k) \sup_{z \in U_k} \langle y, x \rangle.$$

It is standard to verify that there is a set $\mathcal{N}_k$ as above of cardinality at most $\exp(k \log(eN/k\delta_k))$.

The main application of Corollary 2.2 is the following Lemma.

Lemma 2.3 *For every $p > 8$, C_0, κ_1 and $\beta > 0$ as in Theorem 1.2 , there exist constants c_1 and c_2 that depend only on p, C_0, κ_1 and β and for which the following holds. If $I \subset \{1, ..., N\}$, then in the L_p case, with μ^N-probability at least $1 - c_1/N^\beta$,*

$$\sup_{a \in U_{k_0}} \sup_{b \in U_{k_0}} \left\langle \sum_{i \in I} a_i X_i, \sum_{i \in I^c} b_i X_i \right\rangle \leq c_2 (Nn)^{1/4} A_{k_0}.$$

Also, in the ψ_1 case, there are constants c_3 and c_4 that depend only on C_0 and κ_2, for which, with μ^N-probability at least $1-2\exp(-c_3(Nn)^{1/4})$,

$$\sup_{a\in U_{k_0}}\sup_{b\in U_{k_0}}\big\langle\sum_{i\in I}a_iX_i,\sum_{i\in I^c}b_iX_i\big\rangle\leq c_4(Nn)^{1/4}A_{k_0}.$$

Again, we will restrict ourselves to the case when $k_0>1$, since the modifications needed for the case $k_0=1$ are minor.

Proof. Let us begin with the L_p case. Consider the sets U_{k_ℓ} as above and let

$$B_{k_\ell}=\sup_{a\in U_{k_\ell}}\sup_{b\in U_{k_\ell}}\big\langle\sum_{i\in I}a_iX_i,\sum_{i\in I^c}b_iX_i\big\rangle.$$

The main observation is that for every $0\leq\ell\leq\ell_1$,

$$\begin{aligned}\rho_{k_\ell}B_{k_\ell}\leq&B_{k_{\ell+1}}+\sup_{b\in\mathcal{N}_{k_\ell}}\Big(\sum_{i=k_{\ell+1}+1}^{k_\ell}\big(\big\langle\sum_{j\in I^c}b_jX_j,X_i\big\rangle^*\big)^2\Big)^{1/2}\\&+\sup_{a\in\mathcal{N}_{k_{\ell+1}}}\Big(\sum_{i=k_{\ell+1}+1}^{k_\ell}\big(\big\langle\sum_{i\in I}a_iX_i,X_j\big\rangle^*\big)^2\Big)^{1/2},\end{aligned}\tag{2.1}$$

where $\rho_{k_\ell}=(1-\delta_{k_\ell})(1-\delta_{k_{\ell+1}})$ and ℓ_1 will be defined later.

Indeed, fix $a\in U_{k_\ell}$ and let $Z_{a,j}=\big\langle\sum_{i\in I}a_iX_i,X_j\big\rangle$. By the definition of $\mathcal{N}_{k_\ell}$

$$\sup_{b\in U_{k_\ell}}\sum_{j\in I^c}b_jZ_{a,j}\leq(1-\delta_{k_\ell})^{-1}\sup_{b\in\mathcal{N}_{k_\ell}}\sum_{j\in I^c}b_jZ_{a,j}.$$

Note that

$$\sup_{a\in U_{k_\ell}}\sup_{b\in\mathcal{N}_{k_\ell}}\sum_{j\in I^c}b_jZ_{a,j}=\sup_{b\in\mathcal{N}_{k_\ell}}\sup_{a\in U_{k_\ell}}\sum_{i\in I}a_i\big\langle X_i,\sum_{j\in I^c}b_jX_j\big\rangle=(*),$$

and setting $W_{b,i}=\big\langle X_i,\sum_{j\in I^c}b_jX_j\big\rangle$ for $i\in I$, it is evident that

$$\begin{aligned}(*)\leq&\sup_{b\in\mathcal{N}_{k_\ell}}\Big(\sum_{i=1}^{k_\ell}(W_{b,i}^*)^2\Big)^{1/2}\\\leq&\sup_{a\in U_{k_{\ell+1}}}\sup_{b\in\mathcal{N}_{k_\ell}}\big\langle\sum_{i\in I}a_iX_i,\sum_{j\in I^c}b_jX_j\big\rangle+\sup_{b\in\mathcal{N}_{k_\ell}}\Big(\sum_{i=k_{\ell+1}+1}^{k_\ell}(W_{b,i}^*)^2\Big)^{1/2}.\end{aligned}$$

Replacing $U_{k_{\ell+1}}$ by $\mathcal{N}_{k_{\ell+1}}$ and repeating the argument used above for the first term (while reversing the roles of a and b) proves (2.1).

Since $|N_{k_\ell}| \leq \exp(k_\ell \log(eN/k_\ell \delta_{k_\ell}))$, and using the independence of $(X_j)_{j\in I^c}$ and $(X_i)_{i\in I}$, a straightforward application of Corollary 2.2, shows that with probability at least

$$1 - 2\exp(-(p/4-1)k_{\ell+1}\log(eN/k_{\ell+1}) + k_\ell \log(eN/k_\ell \delta_{k_\ell})) = (**),$$

for every $b \in \mathcal{N}_{k_\ell}$ and every $a \in \mathcal{N}_{k_{\ell+1}}$,

$$\left(\sum_{i=k_{\ell+1}+1}^{k_\ell} (\langle \sum_{j\in I^c} b_j X_j, X_i \rangle^*)^2 \right)^{1/2} \leq (cNk_\ell)^{1/4} A_{k_0}$$

and

$$\left(\sum_{i=k_{\ell+1}+1}^{k_\ell} (\langle \sum_{i\in I} a_i X_i, X_j \rangle^*)^2 \right)^{1/2} \leq (cNk_\ell)^{1/4} A_{k_0}.$$

Since $p/4 > 2$, there is $\gamma < 1$ for which $(p/4-1)\gamma > 1$. Thus, for $p > 8$ there are γ, c_1 and c_2 that depend only on p, and for which one may take $\delta_{k_\ell} = (k_\ell/N)^{c_1}$, satisfying that

$$(**) \geq 1 - 2\exp(-c_2 k_{\ell+1}\log(eN/k_{\ell+1})).$$

Now set ℓ_1 to be the largest integer ℓ for which both $k_\ell - k_{\ell+1} > 1$ and

$$\sum_{j=0}^{\ell} \exp(-c_2 k_{j+1} \log(eN/k_{j+1})) \leq N^{-\beta}.$$

Therefore, ℓ_1 is the first integer satisfying $(p/4-2)k_\ell \log(eN/k_\ell) \leq \kappa_3 \beta \log N$ for an appropriate choice of κ_3.

Observe that there is a constant c_3 that depends only on p for which $\prod_{\ell=0}^{\ell_1}(1-\delta_{k_\ell})^2 \geq c_3$. Hence, repeating this dimension reduction procedure up to $\ell = \ell_1$ and then applying the "large coordinates" estimate from Lemma 2.1 for $B_{k_{\ell_1}}$, (while observing that $k_{\ell_1} \leqslant k_0$), concludes the proof.

The proof in the ψ_1 case is similar - only with a different termination point for the dimension reduction process: k_{ℓ_0} instead of k_{ℓ_1}. We omit the details of this case. ■

Observe that in the proof of the previous Lemma we needed that $\frac{p}{4} - 1 > 1$. This is the only point in our proof where the fact $p > 8$ is required.

Theorem 2.4 *Under the assumptions of Theorem 1.2, there are constants c_1 and c_2 that depend only β, p, C_0 and κ_1, for which with probability at least $1 - c_1 N^{-\beta}$, $A_{k_0} \leq c_2 (Nn)^{1/4}$.*

Under the assumptions of Theorem 1.3, with probability at least $1 - 2\exp(-c_3 (Nn)^{1/4})$, $A_{k_0} \leq c_4 (Nn)^{1/4}$, where c_3, c_4 depend only on C_0 and κ_2.

Proof. We will only present a proof in the L_p case, as the ψ_1 one has an almost identical proof. Clearly, for every $a \in U_{k_0}$, $\|\sum_{i=1}^N a_i X_i\|^2 = \sum_{i \neq j} a_i a_j \langle X_i, X_j \rangle + \sum_{i=1}^N a_i^2 \|X_i\|^2$, and since $\|a\| \leq 1$, the second term is at most $\max_{i \leq N} \|X_i\|^2 \leq C_0^2 (Nn)^{1/2}$.

To bound the first term, let $(\varepsilon_i)_{i=1}^N$ be independent Bernoulli random variables. Note that

$$\mathbb{E}_\varepsilon \sum_{i \neq j} (1+\varepsilon_i)(1-\varepsilon_j) a_i a_j \langle X_i, X_j \rangle = \sum_{i \neq j} a_i a_j \langle X_i, X_j \rangle,$$

and thus it suffices to control

$$\begin{aligned} &\sup_{a \in U_{k_0}} \mathbb{E}_\varepsilon \sum_{i \neq j} (1+\varepsilon_i)(1-\varepsilon_j) a_i a_j \langle X_i, X_j \rangle \\ \leq &\mathbb{E}_\varepsilon \sup_{a \in U_{k_0}} \sum_{i \neq j} (1+\varepsilon_i)(1-\varepsilon_j) a_i a_j \langle X_i, X_j \rangle \equiv \mathbb{E}_\varepsilon H((\varepsilon_i)_{i=1}^N, (X_i)_{i=1}^N). \end{aligned}$$

Observe that if $I_\varepsilon = \{i : \varepsilon_i = 1\}$ then

$$H((\varepsilon_i)_{i=1}^N, (X_i)_{i=1}^N) = 4 \sup_{a \in U_{k_0}} \Big\langle \sum_{i \in I_\varepsilon} a_i X_i, \sum_{j \in I_\varepsilon^c} a_j X_j \Big\rangle$$

for every realization of $(\varepsilon_i)_{i=1}^N$.

Fix $(\varepsilon_i)_{i=1}^N$, then

$$H((\varepsilon_i)_{i=1}^N, (X_i)_{i=1}^N) \lesssim \sup_{a \in U_{k_0}} \sup_{b \in U_{k_0}} \Big\langle \sum_{i \in I_\varepsilon} a_i X_i, \sum_{j \in I_\varepsilon^c} b_j X_j \Big\rangle.$$

Applying Lemma 2.3, if $p > 8$, with μ^N-probability at least $1 - cN^{-\beta}$, $H((\varepsilon_i)_{i=1}^N, (X_i)_{i=1}^N) \lesssim_p (Nn)^{1/4} A_{k_0}$. Thus, by a Fubini argument, there exists a set $\mathcal{B} \subset \Omega^N$ of μ^N-probability at least $1 - c_1 N^{-\beta/2}$, on which, with μ_ε^N-probability at least $1 - c_2 N^{-\beta/2}$, $H((\varepsilon_i)_{i=1}^N, (X_i)_{i=1}^N) \lesssim_p (Nn)^{1/4} A_{k_0}$.

Hence, for every $(X_i)_{i=1}^N \in \mathcal{B}$,

$$\begin{aligned}
&\mathbb{E}_\varepsilon H((\varepsilon_i)_{i=1}^N,(X_i)_{i=1}^N) \lesssim_p \\
&A_{k_0}(Nn)^{1/4} + N^{-\beta/2} \sup_{a \in U_{k_0}} |\sum_{i \neq j} a_i a_j \langle X_i, X_j \rangle| \\
\lesssim_{p,C_0} &A_{k_0}(Nn)^{1/4} + N^{-\beta/2}(A_{k_0}^2 + (Nn)^{1/2}),
\end{aligned} \tag{2.2}$$

where the last inequality follows from the Cauchy-Scwarz inequality and the definition of A_{k_0}. Therefore, on $\mathcal{B}$, if $\beta > 0$ and N is large enough, then $A_{k_0}^2 \lesssim_{p,\beta,C_0} A_{k_0}(Nn)^{1/4} + (Nn)^{1/2}$ and the claim follows. ■

The final observation we need is a straightforward application of Lemma 2.1 to the random variables $Z_t = \langle X, t \rangle$, for vectors t in a 1/2-net in B_2^n.

Lemma 2.5 *Under the assumptions of Theorem 1.2 there exist absolute constants c_1, c_2 and c_3 depending only on κ_1 for which the following holds. If $\mathcal{N}$ is a maximal 1/2-separated subset of B_2^n then with probability at least $1 - 2\exp(-c_1 n)$, $\sup_{t \in \mathcal{N}} \left(\sum_{i=c_3k_0+1}^N (\langle X_i, t \rangle^*)^4\right)^{1/2} \leq c_2\sqrt{N}$.*

Moreover under the assumptions of Theorem 1.3 there exist absolute constants c_4 and c_5 depending only on κ_2 for which with probability at least $1 - 2\exp(-c_1 n)$, $\sup_{t \in \mathcal{N}} \left(\sum_{i=k_0+1}^N (\langle X_i, t \rangle^*)^4\right)^{1/2} \leq c_2\sqrt{N}$.

Proof of Theorem 1.2. Let $\mathcal{N}$ be a maximal 1/2-separated subset of B_2^n and let $\mathcal{C}$ be the intersection of the events from Theorem 2.4 and Lemma 2.5. Note that on $\mathcal{C}$, with μ_ε^N-probability at least $1 - 2\exp(-c_1 n)$,

$$\sup_{t \in B_2^n} |\sum_{i=1}^N \varepsilon_i \langle X_i, t \rangle^2| \lesssim_{C_0,p} \sqrt{Nn}.$$

Indeed, let c_3 be the constant from Lemma 2.5, fix $t, t' \in \mathcal{N}$ and let J be the union of the sets of the largest c_3k_0 coordinates of $(|\langle X_i, t \rangle|)_{i=1}^N$ and $(|\langle X_i, t' \rangle|)_{i=1}^N$. By Höffding's inequality, for every $v > 0$, with μ_ε^N-probability at least $1-2\exp(-c_4v^2)$,

$$|\sum_{i=1}^N \varepsilon_i \langle X_i,t\rangle\langle X_i,t'\rangle| \lesssim \sum_{i\in J}|\langle X_i,t\rangle\langle X_i,t'\rangle| + v(\sum_{i\in J^c}\langle X_i,t\rangle^2\langle X_i,t'\rangle^2)^{1/2}$$

$$\leq 2c_3(\sum_{i=1}^{k_0}(\langle X_i,t\rangle^*)^2)^{1/2}(\sum_{i=1}^{k_0}(\langle X_i,t'\rangle^*)^2)^{1/2}$$

$$+v(\sum_{i=c_3k_0+1}^{N}(\langle X_i,t\rangle^*)^4)^{1/4}(\sum_{i=c_3k_0+1}^{N}(\langle X_i,t'\rangle^*)^4)^{1/4}$$

$$\lesssim A_{k_0}^2 + v\sqrt{N}. \tag{2.3}$$

Let $v\sim\sqrt{n}$, and since $|\mathcal{N}|\leq 5^n$, there is a set $\mathcal{D}\subset\{-1,1\}^N$ of μ_ε^N-probability at least $1-2\exp(-c_5n)$ on which (2.3) holds for any pair t,t' taken from $\mathcal{N}\times\mathcal{N}$. Since each $t\in B_2^n$ can be written as $\sum_{i=1}^\infty \beta_i t_i$ with $0\leq\beta_i\lesssim 2^{-i}$ and $t_i\in\mathcal{N}$, then on $\mathcal{D}$,

$$\sup_{t\in B_2^n}|\sum_{i=1}^N \varepsilon_i\langle X_i,t\rangle^2| \lesssim (Nn)^{1/2}\sum_{i,j=1}^\infty 2^{-i}2^{-j}\lesssim (Nn)^{1/2},$$

with constants that depends on κ_0, C_0, p and β. The assertion now follows from a standard application of a variation of the Giné-Zinn symmetrization theorem [9] (see also §5.3 in [13]). ■

The proof of 1.3 follows the same line and we will not present the details. Finally, let us point out that the estimate on the probability in Theorem 1.2 (and in Theorem 1.3 as well) is of the right order when $N\geqslant e^{c_pn}$, where $c_p>0$ is a constant that depends only on p; observe that in that range, the dominant term in the probability estimate is e^{-cn}.

Indeed, set $A=\sup_{t\in B_2^n}|N^{-1/2}\sum_{i=1}^N(\langle X_i,t\rangle^2-1)|$, and note that for any fixed $t\in S^{n-1}$, $Pr(A>cn^{1/2})\geq Pr(|N^{-1/2}\sum_{i=1}^N(\langle X_i,t\rangle^2-1)|>cn^{1/2})$. By a variant of Berry-Esseen theorem (see [14], Theorem 2.2) it follows that

$$\left|Pr((|N^{-1/2}\sum_{i=1}^N(\langle X_i,t\rangle^2-1)|>cn^{1/2})-Pr(|g|>cn^{1/2})\right|\lesssim\frac{1}{N^\alpha},$$

where α depends only on p (and is positive for any $p>4$), and g is a standard gaussian variable. Hence, under our assumptions and for those very large values of N, it is evident that $Pr(A>cn^{1/2})>(1/2)\exp(-c_1n)$.

2.1 Final Remarks

Many of the ideas used in the proof of Theorem 1.2 can actually be traced back to Bourgain [8], who studied the log-concave case and obtained estimates on the random variables $\max_{|I| \leq m} \| \sum_{i \in I} X_i \|$ using a combination of self-bounding and decoupling arguments. This led to a bound on the non-increasing rearrangement of vectors $(\langle X_i, t \rangle)_{i=1}^N$, uniformly for $t \in B_2^n$.

In [11], similar uniform bounds were obtained in the more general, empirical processes setup, and under a ψ_1-tail assumption; that is, estimates on $\sup_{f \in F} \max_{|I|=m} | \sum_{i \in I} f(X_i)|$ for a general class of functions F with a bounded diameter in L_{ψ_1}. In both cases, the quantity that was estimated was not the right one for the problem at hand, and thus the approach resulted in slightly suboptimal estimates on $\sup_{f \in F} | \sum_{i=1}^N f^2(X_i) - \mathbb{E} f^2|$.

Bourgain's method was extended and improved in [1, 2], in which the parameters A_m were introduced. This, combined with the correct level of truncation ($(Nn)^{1/4}$ rather than $n^{1/2}$) were the main ingredients in the solution of the log-concave case, though only with the probability estimate of $1 - 2\exp(-c\sqrt{n})$.

At the same time, it was noted in [12] that one may use a chaining argument to control $\sup_{f \in F} \max_{|I|=m} (\sum_{i \in I} f^2(X_i))^{1/2}$ for a general class of functions F that has a bounded diameter in L_{ψ_1}. Of course, when considering $F = \{\langle t, \cdot \rangle : t \in B_2^n\}$, this quantity is just A_m. This approach was extended further in [13], allowing one to control the empirical process $\sup_{f \in F} | \sum_{i=1}^N f^2(X_i) - \mathbb{E} f^2|$ for classes that are only bounded in L_p rather than in L_{ψ_1}.

To see why our proof follows the same ideas as [12, 13], one should observe that the key point in [12, 13] was to study the fine structure of the random coordinate projection $V = \{(f(X_i))_{i=1}^N : f \in F\}$, and then use this structure to handle the Bernoulli process indexed by V^2 (without reverting to the gaussian process indexed by the same set!). To that end, one obtains information on the monotone rearrangement of each "link" $((\pi_{s+1} f - \pi_s f)(X_i))_{i=1}^N$ in the chain given by the admissible sequence (F_s), where at each step, one balances the cardinality of the set of links and $\binom{N}{k}$. In this way, one may obtain uniform information on the k largest coordinates of $((\pi_{s+1} f - \pi_s f)(X_i))_{i=1}^N$ for that value of k. Moreover, these k largest coordinates are controlled in terms of a "global" notion of complexity of F (e.g. the γ_2 functional), while the smaller coordinates are estimated in the same way we did here – using tail estimates on each random variable $(\pi_s f - \pi_{s+1} f)(X)$.

Unlike the general case, here, the structure is rather simple because B_2^n

is both large and very regular. In particular, one should not expect chaining to have any advantage over the union bound – which can be viewed as "one-step chaining", or alternatively, chaining that starts at a set of cardinality $\exp(cn)$. Having this in mind, our proof follows the path mentioned above: the balance should be between the "cardinality" of B_2^n - i.e. $\exp(cn)$, and $\binom{N}{k}$, which is precisely the definition of k_0. What happens on the "large" k_0 coordinates (i.e. A_{k_0}) depends on a "global" property – $\max_{i \leq N} \|X_i\|$ (Theorem 2.4), while the "small" coordinates are estimated using only individual tail estimates (Lemma 2.5).

$$< \sum a_i X_i, \sum b_j X_j > \leq A_{k_0} < \sum a_i X_i, v > \leq A_{k_0} \| \sum a_i X_i \| \|v\| =$$

$$A_{k_0} \| \sum a_i X_i \| \leq A_{k_0} \sum \|X_i\|$$

References

[1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles, J. Amer. Math. Soc. 23 535-561, 2010.

[2] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Sharp bounds on the rate of convergence of empirical covariance matrix, C.R. Math. Acad. Sci. Paris, 349, 195–200, 2011.

[3] R. Adamczak, R. Latała, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Chevet type inequality and norms of submatrices, preprint.

[4] R. Adamczak, R. Latała, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Tail estimates for norms of sums of log-concave random vectors, preprint.

[5] G. Aubrun, Sampling convex bodies: a random matrix approach, Proc. Amer. Math. Soc. 135, 1293-1303, 2007.

[6] S. Alesker, *ψ_2-estimate for the Euclidean norm on a convex body in isotropic position*, Geom. Aspects of Funct. Analysis (Lindenstrauss-Milman eds.), Oper. Theory Adv. Appl. **77** (1995), 1–4.

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

[8] J. Bourgain, Random points in isotropic convex bodies, in Convex Geometric Analysis (Berkeley, CA, 1996) Math. Sci. Res. Inst. Publ. 34 (1999), 53-58.

[9] E. Giné and J. Zinn, Some limit theorems for empirical processes, Ann. Probab. 12(4), 929-989, 1984.

[10] O. Guédon, M. Rudelson, L_p moments of random vectors via majorizing measures, Adv. Math. 208(2), 798-823, 2007.

[11] S. Mendelson, On weakly bounded empirical processes, Math. Annalen, 340(2), 293-314, 2008.

[12] S. Mendelson, Empirical processes with a bounded ψ_1 diameter, Geometric and Functional Analysis, 20(4) 988-1027, 2010.

[13] S. Mendelson, G. Paouris, On generic chaining and the smallest singular values of random matrices with heavy tails, Journal of Functional Analysis, to appear (41 pages), DOI10.1016/j.jfa.2012.01.027.

[14] E. Mossel, R. O'Donnell, K. Oleszkiewicz, Noise stability of functions with low influences: invariance and optimality. Annals of Mathematics 171(1), pp. 295-341 (2010).

[15] G. Paouris, Concentration of mass on convex bodies, Geometric and Functional Analysis, 16(5), 1021-1049, 2006.

[16] M. Talagrand, private communication.

[17] M. Rudelson, Random vectors in the isotropic position, J. Funct. Anal. 164, 60-72, 1999.

[18] M. Rudelson, R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, Proceedings of the International Congress of Mathematicians, Hyderabad, India, 2010, to appear.

[19] N. Srivastava, R. Vershynin, Covariance estimation for distributions with $2 + \epsilon$ moments. arXiv:1106.2775.

[20] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices. In: *Compressed Sensing: Theory and Applications*, Yonina Eldar and Gitta Kutyniok (eds), Cambridge University Press, to appear.

[21] R. Vershynin, How close is the sample covariance matrix to the actual covariance matrix? Journal of Theoretical Probability, to appear.