J. Appl. Probab. **60,** 1275–1280 (2023)
*doi:*10.1017/jpr.2022.130

LARGE DEVIATIONS OF EXTREMAL EIGENVALUES OF SAMPLE COVARIANCE MATRICES

DENISE UWAMARIYA,* ** AND
XIANGFENG YANG,* *** *Linköping University*

Abstract

Large deviations of the largest and smallest eigenvalues of $\mathbf{X}\mathbf{X}^\top/n$ are studied in this note, where $\mathbf{X}_{p\times n}$ is a $p \times n$ random matrix with independent and identically distributed (i.i.d.) sub-Gaussian entries. The assumption imposed on the dimension size p and the sample size n is $p = p(n) \to \infty$ with $p(n) = \mathrm{o}(n)$. This study generalizes one result obtained in [3].

Keywords: Large deviations; sample covariance matrices; extremal eigenvalues

2020 Mathematics Subject Classification: Primary 60B20
Secondary 60F10; 15A12

1. Introduction

For any two integers p, $n \geq 2$, let $\mathbf{X}_{p\times n}$ be a $p \times n$ random matrix with independent and identically distributed (i.i.d.) real entries. The matrix $\mathbf{W}$ defined by $\mathbf{W} = \mathbf{X}\mathbf{X}^\top/n$ (with $^\top$ standing for matrix transpose) is usually called a *sample covariance matrix* (see [1] and [11]), where p and n can be understood as dimension size and sample size respectively. When the entries are i.i.d. centered normal random variables, then $n\mathbf{W}$ is called a *Wishart matrix*. Sample covariance matrices appear naturally in many situations of multivariate statistical inference; in particular, many test statistics involve the extremal eigenvalues of $\mathbf{W}$. For instance, the union-intersection principle proposed in [8] suggests that one can use the largest eigenvalue of the sample covariance matrix to test whether or not the population covariance is identity. In the literature, weak convergence and law of large numbers of the extremal eigenvalues of $\mathbf{W}$ have been well studied; see [1], [5], [6], [11], and the references therein. In this note we study large deviations of the extremal eigenvalues of $\mathbf{W}$ as both p and n tend to infinity.

As the non-zero eigenvalues of $\mathbf{X}\mathbf{X}^\top$ are the same as those of $\mathbf{X}^\top\mathbf{X}$, it is without loss of generality to assume that $p \leq n$. Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalue of $\mathbf{W}$ respectively. It is assumed throughout the note that the i.i.d. entries $\{X_{ij}\}_{1\leq i\leq p, 1\leq j\leq n}$ of $\mathbf{X}$ have zero mean $\mathbb{E}(X_{ij}) = 0$ and unit variance $\mathbb{V}(X_{ij}) = 1$. Under the fourth finite moment assumption $EX_{ij}^4 < \infty$, Bai and Yin [1] proved that $\lambda_{\min} \to (1-\kappa^{1/2})^2$ and $\lambda_{\max} \to (1+\kappa^{1/2})^2$ almost surely as $n \to \infty$ and $p = p(n) \to \infty$ with $p(n)/n \to \kappa$. When $\kappa = 0$, the above results indicate that for large p and n the majority of $\lambda_{\min}$ lies in the region close to 1 from the left, and the majority of $\lambda_{\max}$ lies in the region close to 1 from the right. Therefore Fey *et al.*

Received 3 May 2022; revision received 22 November 2022.
* Postal address: Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden.
** Email address: denise.uwamariya@liu.se
*** Email address: xiangfeng.yang@liu.se

[3, Theorem 3.1] studied asymptotics on large deviation probabilities in the forms $\mathbb{P}(\lambda_{\min} \le c)$ with $0 \le c \le 1$ and $\mathbb{P}(\lambda_{\max} \ge c)$ with $c \ge 1$ for large p and n satisfying $p = o(n/\ln\ln n)$. They also noted [3, p. 1061] that the technical assumption $p = o(n/\ln\ln n)$ might be extended further by refining the arguments; however, this does not seem to be able to get rid of the logarithmic term.

The main result of this note (see Theorem 1 below) is an extension of [3, Theorem 3.1] in two respects: (a) the technical assumption is extended to $p = o(n)$, and (b) the i.i.d. entries are more general. To state our main result, let us recall the definition of sub-Gaussian distribution. A random variable X is said to be *sub-Gaussian* if it satisfies one of the following three equivalent properties, with parameters K_i, $1 \le i \le 3$ differing from each other by at most an absolute constant factor (see [10, Lemma 5.5]).

(i) Tails: $\mathbb{P}(|X| > t) \le \exp\left\{1 - t^2/K_1^2\right\}$ for all $t \ge 0$.

(ii) Moments: $(\mathbb{E}|X|^p)^{1/p} \le K_2\sqrt{p}$ for all $p \ge 1$.

(iii) Super-exponential moment: $\mathbb{E}\exp\left\{X^2/K_3^2\right\} \le e$.

If moreover $\mathbb{E}(X) = 0$, then (i)–(iii) are also equivalent to the following.

(iv) Moment generating function: $\mathbb{E}\exp\{tX\} \le \exp\left\{t^2K_4^2\right\}$ for all $t \in \mathbb{R}$ for some constant K_4.

Furthermore, the *sub-Gaussian norm* of X is defined as $\sup_{p\ge 1} p^{-1/2}(\mathbb{E}|X|^p)^{1/p}$, namely the smallest K_2 in (ii).

Theorem 1. *Suppose that the entries $\{X_{ij}\}_{1\le i\le p, 1\le j\le n}$ of $\mathbf{X}$ are i.i.d. sub-Gaussian with zero mean and unit variance. Then, for $p = p(n) \to \infty$ with $p(n) = o(n)$ as $n \to \infty$, we have the following.*

(i) *For any $c \ge 1$,*

$$\liminf_{n\to\infty} n^{-1} \ln \mathbb{P}(\lambda_{\max} \ge c) \ge -I(c), \tag{1}$$

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}(\lambda_{\max} \ge c) \le -\lim_{\epsilon\to 0} I(c-\epsilon). \tag{2}$$

(ii) *For any $0 \le c \le 1$,*

$$\liminf_{n\to\infty} n^{-1} \ln \mathbb{P}(\lambda_{\min} \le c) \ge -I(c), \tag{3}$$

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}(\lambda_{\min} \le c) \le -\lim_{\epsilon\to 0} I(c+\epsilon). \tag{4}$$

Here $I(c) := \lim_{p\to\infty} I_p(c)$ with

$$I_p(c) = \inf_{x\in\mathbb{R}^p, \|x\|=1} \sup_{\theta\in\mathbb{R}} \left[\theta c - \ln \mathbb{E}\exp\left\{\theta S_{x,1}^2\right\}\right],$$

$\|x\|$ being the Euclidean norm, and

$$S_{x,i} = \sum_{k=1}^{p} x_k X_{ki} \quad \textit{for } x = (x_1, \ldots, x_p) \in \mathbb{R}^p,\ 1 \le i \le n.$$

https://doi.org/10.1017/jpr.2022.130 Published online by Cambridge University Press

For standard normal entries, the results of Theorem 1 were proved in [3, Theorem 3.1] (assuming $p = \mathrm{o}(n/\ln\ln n)$), and in [4, Theorems 2 and 3] (under the assumption $p(n) = \mathrm{o}(n)$) where general β-Laguerre ensembles were considered (with $\beta = 1$ corresponding to entries being standard normal). From this point of view, Theorem 1 can also be regarded as an extension of [4, Theorems 2 and 3] from β-Laguerre ensembles to sub-Gaussian entries. The continuity of $I(c)$ is still largely unknown, as pointed out in [3]. However, with the arguments in [3, Theorem 3.2], $I(c)$ can be shown to be continuous on $[1, \infty)$ for some special sub-Gaussian entries; see Section 2.3 for more details. The proof of Theorem 1 makes use of a concentration inequality of the largest eigenvalue $\lambda_{\max}$ (see Section 2.1), which helps us to avoid refining the arguments in [3]. The same idea was employed in [9] for the study of condition numbers of sample covariance matrices.

2. Proof of Theorem 1

2.1. Concentration inequality for the largest eigenvalue

Vershynin [10, Theorem 5.39] considered a random matrix $A_{p\times n}$ whose columns A_j, $1 \leq j \leq n$ are independent sub-Gaussian isotropic random vectors in $\mathbb{R}^p$. Here we switched 'rows', which was originally written in [10, Theorem 5.39], to 'columns', as therein the largest singular value $s_{\max}(A)$ of A is defined as the largest eigenvalue of $(A^\top A)^{1/2}$, while in the current note we always consider the form $\mathbf{X}\mathbf{X}^\top$ because of the assumption $p \leq n$. If we now take $A = \mathbf{X}$, then the elements in each column are i.i.d. sub-Gaussian random variables, implying (based on [10, Lemma 5.24]) that the sub-Gaussian norm $\|A_j\|_{\psi_2}$ of each column A_j is finite, which is independent of p and n. As columns have the same distribution, it holds that $K := \|A_1\|_{\psi_2} = \cdots = \|A_n\|_{\psi_2}$. The concentration inequality in [10, Theorem 5.39] says that for any $t \geq 0$, and two absolute constants $\kappa_1, \kappa_2 > 0$ only dependent on K,

$$\mathbb{P}\big(s_{\max}(A) > \sqrt{n} + \kappa_1\sqrt{p} + t\big) \leq 2\mathrm{e}^{-\kappa_2 t^2}.$$

Note that $s^2_{\max}(A) = n\lambda_{\max}$ in the case $A = \mathbf{X}$, so the above non-asymptotic inequality reads

$$\mathbb{P}\big(\lambda_{\max} > \big(1 + \kappa_1\sqrt{p/n} + t/\sqrt{n}\big)^2\big) \leq 2\mathrm{e}^{-\kappa_2 t^2}.$$

With $\gamma := t/\sqrt{n}$ and the fact $p \leq n$, for any $\gamma \geq 0$ it becomes

$$\mathbb{P}\big(\lambda_{\max} > (1 + \kappa_1 + \gamma)^2\big) \leq 2\mathrm{e}^{-\kappa_2\gamma^2 n}. \tag{5}$$

2.2. Proof of the upper bounds

As suggested in [3], the fundamental first step of the proof is as follows:

$$\begin{aligned}\mathbb{P}(\lambda_{\max} \geq c) &= \mathbb{P}(\exists x \in \mathbb{R}^p \text{ with } \|x\| = 1 \text{ and } (x\cdot\mathbf{W}x) \geq c)\\ &= \mathbb{P}\Bigg(\exists x \in \mathbb{R}^p \text{ with } \|x\| = 1 \text{ such that } \sum_{i=1}^n S^2_{x,i}/n \geq c\Bigg),\end{aligned} \tag{6}$$

$$\begin{aligned}\mathbb{P}(\lambda_{\min} \leq c) &= \mathbb{P}(\exists x \in \mathbb{R}^p \text{ with } \|x\| = 1 \text{ and } (x\cdot\mathbf{W}x) \leq c)\\ &= \mathbb{P}\Bigg(\exists x \in \mathbb{R}^p \text{ with } \|x\| = 1 \text{ such that } \sum_{i=1}^n S^2_{x,i}/n \leq c\Bigg).\end{aligned} \tag{7}$$

https://doi.org/10.1017/jpr.2022.130 Published online by Cambridge University Press

Then the lower bounds (1) and (3) (for any $p \le n$) follow directly from Cramér's theorem for i.i.d. random variables $S_{x,i}$, $1 \le i \le n$. More specifically, we first fix an integer p and choose an x such that only the first p components are non-zero, then apply Cramér's theorem, and finally send p to infinity; see also the detailed arguments in [3, Section 3.2] leading to (3.8) therein. To prove the upper bounds in (2) and (4), as explained in [3] and [9], we shall use a finite number N_d of spherical caps of chord $2\tilde{d} := 2d\sqrt{1-d^2/4}$ with centers $x^{(j)}$ to cover the entire unit sphere S defined as $\|x\| = 1$, such that for any $x \in S$ there is some $x^{(j)} \in S$ close to x with $\|x - x^{(j)}\| \le d$. In this case,

$$|x \cdot \mathbf{W}x - x^{(j)} \cdot \mathbf{W}x^{(j)}| \le \big(\|x\| + \|x^{(j)}\|\big)\|\mathbf{W}\|\,\|x - x^{(j)}\| \le 2\lambda_{\max} d$$

(see [3, p. 1054]). For $p = p(n) \to \infty$ as $n \to \infty$, we need an explicit expression of N_d, which can be borrowed from [9] (see also [3] and [7]) as

$$N_d = 4\tilde{p}(n)^{3/2}\tilde{d}^{\,-\tilde{p}(n)}(\ln \tilde{p}(n) + \ln\ln \tilde{p}(n) - \ln \tilde{d})(1 + \mathrm{O}(1/\ln \tilde{p}(n)))$$

for all $d < 1/2$ and large $\tilde{p}(n) := p(n) - 1$. Then it is clear that for any fixed d, we have the limit $\lim_{n\to\infty} n^{-1} \ln N_d = 0$ with $p(n) = \mathrm{o}(n)$.

Thanks to the concentration inequality (5), the following upper estimates are used:

$$\mathbb{P}(\lambda_{\max} \ge c) \le \mathbb{P}\big(c \le \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) + \mathbb{P}\big(\lambda_{\max} > (1+\kappa_1+\gamma)^2\big), \tag{8}$$

$$\mathbb{P}(\lambda_{\min} \le c) \le \mathbb{P}\big(\lambda_{\min} \le c,\, \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) + \mathbb{P}\big(\lambda_{\max} > (1+\kappa_1+\gamma)^2\big). \tag{9}$$

To prove (2), applying (6) to (8) gives

$$\begin{aligned}
\mathbb{P}\big(c \le \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) &\le \mathbb{P}\big(\exists x^{(j)} : \big(x^{(j)} \cdot \mathbf{W}x^{(j)}\big) \ge c - 2d(1+\kappa_1+\gamma)^2\big) \\
&\le \sum_{1\le j\le N_d} \mathbb{P}\big(\big(x^{(j)} \cdot \mathbf{W}x^{(j)}\big) \ge c - 2d(1+\kappa_1+\gamma)^2\big) \\
&\le N_d \max_{1\le j\le N_d} \mathbb{P}\big(\big(x^{(j)} \cdot \mathbf{W}x^{(j)}\big) \ge c - 2d(1+\kappa_1+\gamma)^2\big) \\
&\le N_d \max_{1\le j\le N_d} \mathbb{P}\Bigg(\sum_{i=1}^{n} S^2_{x^{(j)},i}/n \ge c - 2d(1+\kappa_1+\gamma)^2\Bigg),
\end{aligned}$$

where the first inequality comes from the facts

$$\big|x \cdot \mathbf{W}x - x^{(j)} \cdot \mathbf{W}x^{(j)}\big| \le \big(\|x\| + \|x^{(j)}\|\big)\|\mathbf{W}\|\,\|x - x^{(j)}\| \le 2\lambda_{\max} d$$

and $\lambda_{\max} < (1+\kappa_1+\gamma)^2$. With $\epsilon := 2d(1+\kappa_1+\gamma)^2$, the Chernoff upper bound (see [2, remark (c) of Theorem 2.2.3]) implies

$$\begin{aligned}
&n^{-1} \ln \mathbb{P}\big(c \le \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) \\
&\quad\le n^{-1} \ln N_d - \min_{1\le j\le N_d} \sup_{\theta\in\mathbb{R}} \big[\theta(c-\epsilon) - \ln \mathbb{E}\exp\big\{\theta\big(S^2_{x^{(j)},1}\big)\big\}\big] \\
&\quad\le n^{-1} \ln N_d - I_{p(n)}(c-\epsilon) + \mathrm{o}(1).
\end{aligned}$$

With $p(n) = \mathrm{o}(n)$ and the fact $\lim_{n\to\infty} n^{-1} \ln N_d = 0$, it follows that

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}\big(c \le \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) \le -I(c-\epsilon).$$

https://doi.org/10.1017/jpr.2022.130 Published online by Cambridge University Press

Taking into account the concentration inequality (5), we obtain

$$\limsup_{n\to\infty} n^{-1}\mathbb{P}(\lambda_{\max} \ge c) \le \max\{-I(c-\epsilon),\, -\kappa_2\gamma^2\}.$$

Thus (2) is proved by first taking $d \to 0^+$ (implying that $\epsilon \to 0^+$) and then sending $\gamma \to \infty$.

In a very similar way (4) can be proved by applying (7) to (9) as follows:

$$\mathbb{P}\big(\lambda_{\min} \le c,\, \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) \le N_d \max_{1\le j\le N_d} \mathbb{P}\left(\sum_{i=1}^n S^2_{x^{(j)},i}/n \le c + 2d(1+\kappa_1+\gamma)^2\right).$$

Here we remark that the original proof in [3] is based on splitting the values of $\lambda_{\max}$ into two (or more) parts with the length of each part depending on n, which leads to the restrictive assumption $p = \mathrm{o}(n/\ln\ln n)$. Because of the uniform constant γ in the concentration inequality (5), it is thus possible to improve it as $p = \mathrm{o}(n)$.

2.3. Continuity of $I(c)$

It was remarked in [3] that the continuity of $I(c)$ is still largely unknown. Here we derive bounds for $I(c)$ using the ideas of [3, Theorem 3.2], and show that $I(c)$ is continuous on $[1, \infty)$ for sub-Gaussian entries satisfying the conditions in Theorem 1 and $K_4^2 = 1/2$ (recall that K_4^2 is given in the definition of sub-Gaussian distributions in Section 1).

Recall that $I(c) = \lim_{p\to\infty} I_p(c)$, where

$$I_p(c) = \inf_{x\in\mathbb{R}^p,\,\|x\|=1}\, \sup_{\theta\in\mathbb{R}} \big[\theta c - \ln \mathbb{E}\exp\{\theta S^2_{x,1}\}\big].$$

For $c \ge 1$, we have

$$\sup_{\theta\in\mathbb{R}}\big[\theta c - \ln \mathbb{E}\exp\{\theta S^2_{x,1}\}\big] = \sup_{\theta\ge 0}\big[\theta c - \ln \mathbb{E}\exp\{\theta S^2_{x,1}\}\big],$$

since

$$\theta c - \ln \mathbb{E}\exp\{\theta S^2_{x,1}\} \le \theta c - \mathbb{E}\big(\theta S^2_{x,1}\big) = \theta(c-1) \le 0 \quad \text{for } \theta < 0.$$

It was shown in [9] that

$$\mathbb{E}\exp\{\theta S^2_{x,1}\} \le \big(1 - 4\theta K_4^2\big)^{-1/2} \quad \text{for } \theta < 1/\big(4K_4^2\big).$$

Therefore, for $c \ge 1$,

$$\begin{aligned}
I_p(c) &\ge \inf_{x\in\mathbb{R}^p,\,\|x\|=1}\, \sup_{0\le\theta<1/(4K_4^2)} \big[\theta c - \ln \mathbb{E}\exp\{\theta S^2_{x,1}\}\big] \\
&\ge \inf_{x\in\mathbb{R}^p,\,\|x\|=1}\, \sup_{0\le\theta<1/(4K_4^2)} \Big[\theta c - \ln\Big(\big(1-4\theta K_4^2\big)^{-1/2}\Big)\Big] \\
&= \sup_{0\le\theta<1/(4K_4^2)} \Big[\theta c - \ln\Big(\big(1-4\theta K_4^2\big)^{-1/2}\Big)\Big] \\
&= c/\big(4K_4^2\big) - 1/2 + \big[\ln\big(2K_4^2/c\big)\big]/2 \quad \text{for } 1/2 \le K_4^2 \le c/2.
\end{aligned}$$

The restriction $1/2 \le K_4^2$ is from the assumptions that the entries X_{ij} have zero mean, unit variance, and

$$\mathbb{E}\exp\{tX_{ij}\} \le \exp\{t^2 K_4^2\} \quad \text{for all } t \in \mathbb{R}.$$

https://doi.org/10.1017/jpr.2022.130 Published online by Cambridge University Press

The other restriction $K_4^2 \leq c/2$ is from searching for the supremum. Therefore

$$I(c) \geq c/\big(4K_4^2\big) - 1/2 + \big[\ln\big(2K_4^2/c\big)\big]/2 \quad \text{for } 1/2 \leq K_4^2 \leq c/2.$$

On the other hand, Fey *et al.* [3, Theorem 3.2] proved that $I(c) \leq (c - 1 - \ln c)/2$ for $c \geq 1$. In summary, if one takes the entries X_{ij} as sub-Gaussian random variables satisfying the conditions in Theorem 1 and $K_4^2 = 1/2$, then $I(c) = (c - 1 - \ln c)/2$ for $c \geq 1$. As mentioned in [3], this is a kind of universality result as $(c - 1 - \ln c)/2$ is the corresponding rate function with i.i.d. standard normal entries. Furthermore, the condition $K_4^2 = 1/2$ is satisfied for at least three distributions: standard normal, Bernoulli ∓ 1 with equal probabilities, and uniform distribution on $[-\sqrt{3}, \sqrt{3}]$.

Acknowledgements

The authors are grateful to the referee and Editor for several constructive comments which have led to an improved version of the paper.

Funding information

There are no funding bodies to thank relating to this creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BAI, Z. AND YIN, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Prob.* **21**, 1275–1294.
[2] DEMBO, A. AND ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*, corrected reprint of 2nd (1998) edn. Springer, Berlin.
[3] FEY, A., VAN DER HOFSTAD, R. AND KLOK, M. (2008). Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Adv. Appl. Prob.* **40**, 1048–1071.
[4] JIANG, T. AND LI, D. (2015). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J. Theoret. Prob.* **28**, 804–847.
[5] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Commun. Math. Phys.* **209**, 437–476.
[6] JOHNSTONE, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295–327.
[7] ROGERS, C. (1963). Covering a sphere with spheres. *Mathematika* **10**, 157–164.
[8] ROY, S. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–238.
[9] SINGULL, M., UWAMARIYA, D. AND YANG, X. (2021). Large-deviation asymptotics of condition numbers of random matrices. *J. Appl. Prob.* **58**, 1114–1130.
[10] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, ed. Y. Eldar and G. Kutyniok, pp. 210–268. Cambridge University Press, Cambridge.
[11] YIN, Y. Q., BAI, Z. D. AND KRISHNAIAH, P. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Prob. Theory Relat. Fields* **78**, 509–521.