J. Appl. Prob. **58,** 1114–1130 (2021)
*doi:*10.1017/jpr.2021.13

LARGE-DEVIATION ASYMPTOTICS OF CONDITION NUMBERS OF RANDOM MATRICES

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

Abstract

Let $\mathbf{X}$ be a $p \times n$ random matrix whose entries are independent and identically distributed real random variables with zero mean and unit variance. We study the limiting behaviors of the 2-normal condition number $k(p,n)$ of $\mathbf{X}$ in terms of large deviations for large n, with p being fixed or $p = p(n) \to \infty$ with $p(n) = o(n)$. We propose two main ingredients: (i) to relate the large-deviation probabilities of $k(p,n)$ to those involving n independent and identically distributed random variables, which enables us to consider a quite general distribution of the entries (namely the sub-Gaussian distribution), and (ii) to control, for standard normal entries, the upper tail of $k(p,n)$ using the upper tails of ratios of two independent χ^2 random variables, which enables us to establish an application in statistical inference.

Keywords: Condition numbers; large deviations; random matrices; Wishart matrices

2010 Mathematics Subject Classification: Primary 60B20
Secondary 60F10; 15A12

1. Introduction

1.1. Background

For any two positive integers $p, n \geq 2$, let us define a $p \times n$ random matrix $\mathbf{X}_{p\times n}$ whose entries X_{ij}, $1 \leq i \leq p$, $1 \leq j \leq n$, are independent and identically distributed (i.i.d.) real random variables satisfying

$$\mathbb{E}(X_{ij}) = 0, \qquad \mathbb{V}(X_{ij}) = 1, \tag{1}$$

where $\mathbb{E}$ denotes the expectation and $\mathbb{V}$ the variance. Then the 2-norm condition number $k(p,n)$ of $\mathbf{X}_{p\times n}$ is defined as $k(p, n) = \sigma_{\max}/\sigma_{\min}$, with $\sigma_{\max}$ and $\sigma_{\min}$ denoting the maximal and minimal singular values of $\mathbf{X}_{p\times n}$. The name '2-norm' comes from the fact that the maximal singular value $\sigma_{\max}$ coincides with the norm $||\mathbf{X}||_2 = \sup\left\{||\mathbf{X}x||_2 : x \in \mathbb{R}^n \text{ with } ||x||_2 = 1\right\}$, where the 2-norm of a vector $x \in \mathbb{R}^n$ is the Euclidean norm defined as $||x||_2 = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$ (for simplicity, $||x||_2$ will be written as $||x||$). In numerical linear algebra and the theory of probability in Banach spaces, condition numbers play an important role (cf. [17, 23]). In statistics, if we

Received 21 January 2020; revision received 22 February 2021.
* Postal address: Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden.
** Email address: martin.singull@liu.se
*** Email address: denise.uwamariya@liu.se
**** Email address: xiangfeng.yang@liu.se

define a square matrix $\mathbf{W}_{p\times p} := \mathbf{X}\mathbf{X}^\top/n$, then $\mathbf{W}$ is usually called a *sample covariance matrix* in the framework of estimating the population covariance matrix given vanishing population mean, where p denotes the dimension of the population and n is the sample size. In this setting the condition number has the equivalent form $k(p, n) = \sigma_{\max}/\sigma_{\min} = (\lambda_{\max}/\lambda_{\min})^{1/2}$, with $\lambda_{\max}$ and $\lambda_{\min}$ denoting the maximal and minimal eigenvalues of $\mathbf{W}_{p\times p}$.

One specific application of using condition numbers in statistics, which is also our main motivation, is to test the null hypothesis that the population covariance is a scalar multiple of identity. The *union-intersection test method* in [20, Section 7.4] suggests that the null hypothesis is rejected for large values of the condition number. To achieve this, we often need to study the null distribution of the sample condition number. For instance, in order to find the corresponding p-value, it is necessary to investigate the probability $\mathbb{P}(k(p, n) \geq c)$ with $c \geq 1$. Unfortunately, so far in the literature there is no efficient way to evaluate such a probability. In Section 5 we present an easy and efficient way to control such a probability in order to feasibly perform the hypotheses test using the union-intersection test method. Another interesting aspect of condition numbers is that they have one=to-one correspondence with the so-called *first anti-eigenvalues* defined as $2(\lambda_{\max}\lambda_{\min})^{1/2}/(\lambda_{\max} + \lambda_{\min})$, which can be found in many applications (cf. [9]). Since the condition number $k(p,n)$ of $\mathbf{X}_{p\times n}$ is invariant under matrix transpose, we shall in this paper always assume that $2 \leq p \leq n$. When the entries X_{ij} are i.i.d. real standard normal random variables, $\mathbf{W}_{p\times p}$ is called a real *central Wishart matrix* and denoted as $W_p(n, n^{-1}\mathbf{I})$, with $\mathbf{I} = \mathbf{I}_{p\times p}$ being an identity matrix.

From the matrix size point of view, the various studies of condition numbers of random matrices in the literature can be classified into two categories: for rectangular random matrices (i.e. $p < n$; see, for example, [1, 4, 7, 13]) and for square random matrices (i.e. $p = n$; see, for example, [6, 17, 21]). Results concerning lower/upper bounds of the minimal singular value $\sigma_{\min}$ can be found, for instance, in [14, 17–19], while results on limiting distributions of $\sigma_{\min}$ and $k(p,n)$ as n tends to infinity are contained in [6, Section 3] and [22], among others. From the distribution (of entries) point of view, we can also classify the results on condition numbers of random matrices in two categories: entries being standard normal random variables (see, for example, [1, 4, 6, 7]) and entries being other (especially discrete) random variables (see, for example, [13, 17, 21]). In general it is easier to study random matrices with standard normal entries (i.e. Wishart matrices $W_p(n, n^{-1}\mathbf{I})$) since there is an explicit (even though involved to some extent) joint distribution of the eigenvalues of $W_p(n, n^{-1}\mathbf{I})$, and it is not available for discrete random matrices. Also note that if the distribution of the entries is sub-Gaussian (see Section 1.2 for a detailed definition; cf. [13, 17, 24]), then all its moments, tail estimates, and moment-generating function can be controlled explicitly, which are in turn used to derive results on condition numbers.

The aim of this paper is to investigate limiting behaviors of the condition number $k(p,n)$ in terms of large deviations for large n (and possibly for large p at the same time as well), for rectangular random matrices (i.e. $p < n$) and for entries being i.i.d. sub-Gaussian random variables satisfying the conditions in (1). It turns out that such investigation heavily depends on the relation between p and n, and in this paper we will only focus on the case when p is fixed or $p = p(n) = o(n)$ as $n \to \infty$. Despite being in the framework of such a classical setting (i.e. one dimension of a random matrix is fixed or negligible with respect to the other one), large deviations of $k(p,n)$ have not appeared in the literature so far. Throughout the paper $f(n) = o(n)$ means $\lim_{n\to\infty} f(n)/n = 0$, and $f(n) = O(n)$ stands for $0 < c_1 \leq f(n)/n \leq c_2 < \infty$ for all n and some positive constants c_1 and c_2 independent of n.

Laws of large numbers of the extreme eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$ of the sample covariance matrices $\mathbf{W}_{p\times p}$ have been obtained in the forms $\lambda_{\max} \to (1+\kappa^{1/2})^2$ and $\lambda_{\min} \to (1-\kappa^{1/2})^2$ in probability under the assumption $p/n \to \kappa \in [0, 1]$ as $n \to \infty$ (see, for instance, [6, Lemma 4.1] for Wishart matrices and [2, 3] for general matrices). Therefore, when p is either fixed or $p(n) = o(n)$ (i.e. $\kappa = 0$), it always holds that $k(p, n) \to 1$ in probability. Then, a large-deviation probability of $k(p,n)$ takes the form $\mathbb{P}(k(p, n) \geq c)$ with $c \geq 1$. The specific aim of the paper is to study the limiting behaviors of $\mathbb{P}(k(p, n) \geq c)$ for large n.

The asymptotics of $\mathbb{P}(k(p, n) \geq c)$ with $c \geq 1$ as $n \to \infty$ cannot be readily obtained from the existing literature. To see this, we first note that for Wishart matrices an exact expression of the density function of the condition number $k(p,n)$ was derived in [1] for all $2 \leq p \leq n$. However, some complicated zonal polynomials appeared in the density function, which prevents us obtaining any useful asymptotics as $n \to \infty$. In [4, 7], lower and upper bounds of $\mathbb{P}(k(p, n) \geq c)$ were given (again only for Wishart matrices) for the purpose of studying tails of the condition number (i.e. for large c). Despite the tight bounds as $c \to \infty$ (with fixed p and n), the asymptotics with fixed c as $n \to \infty$ turn out to be very inaccurate. For general sample covariance matrices, large-deviation asymptotics for $\lambda_{\max}$ and $\lambda_{\min}$ (individually and jointly) as $n \to \infty$ were established in [8]. But the condition number cannot be precisely controlled by $\lambda_{\max}$ or/and $\lambda_{\min}$, and the contraction principle cannot be readily applied.

In this paper, while we employ the proof ideas in [8], we adopt several new strategies in order to improve certain restrictions and obtain non-asymptotic bounds. More specifically, in [8] the results were derived under the assumption $p = o(n/\ln\ln n)$, and in this paper we adopt the concentration inequality for the maximal eigenvalue (see Lemma 5) and have improved the assumption to $p = o(n)$; the employment of such concentration inequality also enables us to consider quite general distribution of the entries (namely sub-Gaussian distribution), improving the results in [8] where the entries are assumed to be symmetric and bounded (except for normal entries). Furthermore, the strategy of using two independent χ^2 random variables to control the condition number enables us to obtain non-asymptotic bounds for the distribution function of the condition number. To state the main result of the paper, let us introduce several notations/definitions.

1.2. 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, $1 \leq i \leq 3$, differing from each other by at most an absolute constant factor (cf. [24, Lemma 5.5]):

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

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

(iii) Super-exponential moment: $\mathbb{E}\exp\{X^2/K_3^2\} \leq \mathrm{e}$. If, further, $\mathbb{E}(X) = 0$, then (i)–(iii) are also equivalent to the following:

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

Furthermore, the *sub-Gaussian norm* of X is defined as $\sup_{p\geq 1} p^{-1/2}(\mathbb{E}|X|^p)^{1/p}$, namely the smallest K_2 in (ii). It is noted that normal (or Gaussian), Bernoulli, and bounded random variables are all sub-Gaussian.

1.3. Rate functions

To state large deviations, we need to introduce rate functions. In each Euclidean space $\mathbb{R}^p$ with $p \geq 2$, the Euclidean norm is, as before, written as $||x||$, and the inner product of x and y is written as $x \cdot y = x_1 y_1 + \cdots + x_p y_p$. For any α, $c \in \mathbb{R}$, we define

$$I_{p,\alpha}(c) = \inf_{x,y\in\mathbb{R}^p, ||x||=||y||=1, x\cdot y=0} \sup_{\theta\in\mathbb{R}} \left[\theta\alpha - \ln \mathbb{E} \exp\left\{\theta\left(S_{x,1}^2 - cS_{y,1}^2\right)\right\}\right],$$

where $S_{x,i} = \sum_{k=1}^{p} x_k X_{ki}$ for $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, $1 \leq i \leq n$. Notice that $I_{p,\alpha}(c)$ is non-increasing in p for each fixed α and c, and therefore the limit $I_{\infty,\alpha}(c) := \lim_{p\to\infty} I_{p,\alpha}(c)$ exists.

1.4. Main result

Theorem 1. *Suppose that the entries X_{ij}, $1 \leq i \leq p$, $1 \leq j \leq n$, are i.i.d. sub-Gaussian satisfying (1). Then, for any $c \geq 1$ we have, for fixed p,*

$$\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left(k^2(p, n) \geq c\right) = -I_{p,0}(c), \tag{2}$$

and, for $p = p(n) \to \infty$ with $p(n) = o(n)$,

$$\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left(k^2(p, n) \geq c\right) = -I_{\infty,0}(c). \tag{3}$$

Since the standard normal distribution is sub-Gaussian, a very special case of Theorem 1 is the real central Wishart matrix $W_p(n, n^{-1}\mathbf{I})$ for which the entries X_{ij}, $1 \leq i \leq p$, $1 \leq j \leq n$, are i.i.d. standard normal $N(0,1)$.

Corollary 1. *Suppose that the entries X_{ij}, $1 \leq i \leq p$, $1 \leq j \leq n$, are i.i.d. standard normal N(0,1). Then, for any $c \geq 1$, $\lim_{n\to\infty} n^{-1} \ln \mathbb{P}(k^2(p, n) \geq c) = -2^{-1} \ln [(c+1)^2/(4c)]$ when p is fixed or $p = p(n) \to \infty$ with $p(n) = o(n)$.*

In order to carry out a specific application of Corollary 1 in statistics (namely the union-intersection test method mentioned above), in Section 3 we prove a somewhat interesting result specified in Lemma 3: the upper tail of $k^2(p, n) = \lambda_{\max}/\lambda_{\min}$ can be controlled by the upper tails of ratios of two independent χ^2 random variables. This in turn yields an independent concise proof of Corollary 1 (under the assumption that p is fixed or $p(n) = o(n/\ln n)$). Note that individually $\lambda_{\max}$ and $\lambda_{\min}$ for Wishart matrices $W_p(n, n^{-1}\mathbf{I})$ are not χ^2 random variables (actually, the exact evaluation of the distributions of $\lambda_{\max}$ and $\lambda_{\min}$ is difficult, involving hypergeometric functions; see [15, Section 9.7]). Furthermore, it is clear that $\lambda_{\max}$ and $\lambda_{\min}$ are not independent. From this point of view, to control $k^2(p, n)$ using ratios of two independent χ^2 random variables is kind of unexpected. Such an application is formulated in detail in Section 5.

It should be mentioned in Theorem 1 that $I_{p,0}(1) = 0$, $I_{\infty,0}(1) = 0$, and both $0 < I_{p,0}(c) < \infty$ and $0 < I_{\infty,0}(c) < \infty$ for all $c > 1$ (see Lemma 7). To find explicit expressions for $I_{p,0}(c)$ is in general not easy because of the infimum; however, for some special cases it is feasible to compute $I_{p,0}(c)$. One particular example is, as we have seen, Corollary 1 for which $I_{p,0}(c)$ can be explicitly written (and turns out to be independent of p). Below is another example.

Example 1. Let $p = 2$ and the entries X_{ij} be $\mathbb{P}(X_{ij} = -1) = \mathbb{P}(X_{ij} = 1) = 1/2$. Then $I_{2,0}(c) = \ln (2c^{c/(c+1)}/(c+1))$. To see this, note that the matrix $\mathbf{W}_{2\times 2} := \mathbf{X}\mathbf{X}^\top/n$ can be explicitly

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

written as

$$\mathbf{W} = \frac{1}{n}\begin{bmatrix} n & \sum_{j=1}^{n} X_{1j}X_{2j} \\ \sum_{j=1}^{n} X_{1j}X_{2j} & n \end{bmatrix}.$$

Therefore, $\lambda_{\max} = \big(n + \big|\sum_{j=1}^{n} X_{1j}X_{2j}\big|\big)/n$ and $\lambda_{\min} = \big(n - \big|\sum_{j=1}^{n} X_{1j}X_{2j}\big|\big)/n$. Hence, $\mathbb{P}(k^2(p, n) \geq c) = \mathbb{P}\big(\big|\sum_{j=1}^{n} X_{1j}X_{2j}\big| \geq (c-1)n/(c+1)\big)$. Notice that $\{X_{1j}X_{2j},\ 1 \leq j \leq n\}$ are i.i.d. random variables with a common distribution $\mathbb{P}(X_{1j}X_{2j} = -1) = \mathbb{P}(X_{1j}X_{2j} = 1) = 1/2$; then Cramér's theorem yields

$$\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left(\left|\sum_{j=1}^{n} X_{1j}X_{2j}\right| \geq (c-1)n/(c+1)\right) = -\ln\left(2c^{c/(c+1)}/(c+1)\right),$$

so $I_{2,0}(c) = \ln\left(2c^{c/(c+1)}/(c+1)\right)$.

A full proof of Theorem 1 will be given in Section 4. It would be interesting to investigate explicit expressions of $I_{\infty,0}(c)$ as well. However, a major difficulty comes from the infimum over all $x, y \in \mathbb{R}^p$ with $||x|| = ||y|| = 1$ and $x \cdot y = 0$ as $p \to \infty$. If we can show that, under suitable additional assumptions,

$$I_{\infty,0}(c) = 2^{-1} \ln\left[(c+1)^2/(4c)\right], \tag{4}$$

then we prove an elegant universality result: the large-deviation asymptotics of the condition number of a sub-Gaussian random matrix coincide with those of a standard normal random matrix as $n \to \infty$. We note that (4) cannot hold in general only under the assumptions of Theorem 1. For instance, if $\mathbb{P}(X_{ij} = -1) = \mathbb{P}(X_{ij} = 1) = 1/2$, then $\mathbb{P}(\lambda_{\min} = 0) \geq \mathbb{P}(X_{ij} = 1, 1 \leq i \leq 2, 1 \leq j \leq n) = 2^{-2n}$. This implies that, for any $c \geq 1$,

$$\begin{aligned} I_{\infty,0}(c) &= -\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left(k^2(p, n) \geq c\right) \\ &\leq -\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left(\lambda_{\min} = 0\right) \leq 2\ln 2, \end{aligned}$$

and in this case (4) does not hold. Therefore, to show (4) it is likely that some regularity assumptions should be imposed on the distribution of X_{ij}. Furthermore, if we are able to manage to prove that for each $p = p(n)$ the minimizing couple $(x, y) = ((x_1, \ldots, x_p), (y_1, \ldots, y_p))$ in $I_{p,0}(c)$ satisfies the condition $\lim_{n\to\infty} \max(|x_1|, \ldots, |x_p|) = 0$ and $\lim_{n\to\infty} \max(|y_1|, \ldots, |y_p|) = 0$, then according to [12, Theorem 1] both $S_{x,1}$ and $S_{y,1}$ converge to $N(0,1)$ in distribution and (4) very likely holds. However, the minimizing couple seems to be very challenging to obtain explicitly, even for the simplest case described in Example 1 with $p = 2$.

At the end of this section, we make an observation. Because of the finiteness of the rate function at each $c \geq 1$ in Theorem 1, it follows that $\mathbb{P}(k^2(p, n) \geq c) > 0$ for n large enough (strictly positive), which is a non-trivial fact. On the other hand, because of the positivity of the rate function at each $c > 1$, $\mathbb{P}(k^2(p, n) \geq c)$ tends to zero exponentially fast as $n \to \infty$.

2. Proof outlines

Overall, the proof ideas are based on the ones in [8] which were used to study $\lambda_{\max}$ and $\lambda_{\min}$ individually and jointly. More precisely, for the lower bounds in (2) and (3), we relate the

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

condition number probability to that involving n i.i.d. random variables so that the classical Cramér's theorem can be used.

Lemma 1. *For any* $2 \le p \le n$ *and* $c \ge 1$*, the set* $\{k^2(p, n) \ge c\}$ *is equal to the set* $\{x, y \in \mathbb{R}^p \text{ s.t. } ||x|| = ||y|| = 1,\ x \cdot y = 0,\ \text{and } (x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) \ge c\}$.

Proof of Lemma 1. If $\omega \in \{\omega : k^2(p, n) \ge c\}$, then let us take x = the eigenvector of $\lambda_{\max}$ and y = the eigenvector of $\lambda_{\min}$. We can think of x and y as normalized vectors so that $||x|| = ||y|| = 1$ and $x \cdot y = 0$. Furthermore, $(x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) = k^2(p, n) \ge c$. To see the other direction, let us take $\omega \in \{(x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) \ge c\}$ for some $||x|| = ||y|| = 1$ and $x \cdot y = 0$. Since $\lambda_{\max} \ge z \cdot \mathbf{W}z$ and $\lambda_{\min} \le z \cdot \mathbf{W}z$ for all $z \in \mathbb{R}^p$ with $||z|| = 1$, it follows that $k^2(p, n) \ge (x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) \ge c$. □

With the help of Lemma 1, we can take any two fixed points $x, y \in \mathbb{R}^p$ with $||x|| = ||y|| = 1$ and $x \cdot y = 0$, and obtain

$$\mathbb{P}\left(k^2(p, n) \ge c\right) \ge \mathbb{P}\left((x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) \ge c\right) = \mathbb{P}\left(\sum_{i=1}^{n} (S_{x,i}^2 - cS_{y,i}^2)/n \ge 0\right). \tag{5}$$

Now the classical Cramér's theorem applied to the n i.i.d. random variables $(S_{x,i}^2 - cS_{y,i}^2)$, $1 \le i \le n$, gives the lower bounds.

The upper bounds in (2) and (3) are more complicated. In order to be able to still make use of Lemma 1, we need to divide the surface $S := \{x \in \mathbb{R}^p : ||x|| = 1\}$ of the p-dimensional sphere of unit radius into smaller pieces, and then take approximations.

Lemma 2. *For any* $0 < d < 1/2$*, let* N_d *denote the minimal number of spherical caps of chord* $2d\sqrt{1 - d^2/4}$ *needed in order to cover S, and* $\{x^{(i)}, 1 \le i \le N_d\}$ *be the centers of these spherical caps. Then, any two points* $x, y \in S$ *with* $x \cdot y = 0$ *can be approximated in the following way: there exist* $1 \le i, j \le N_d$ *such that* $||x - x^{(i)}|| \le d$, $||y - y^{(i,j)}|| \le 2d$, $||y^{(i,j)} - x^{(j)}|| \le d$*, and* $x^{(i)} \cdot y^{(i,j)} = 0$.

The proof of Lemma 2 is given in [8, p. 1056], and the total number of pairs $\{x^{(i)}, y^{(i,j)}\}_{i,j \ge 1}$ is bounded by N_d^2. Thanks to Lemma 2, the probability $\mathbb{P}\left(k^2(p, n) \ge c\right)$ essentially has an upper bound

$$\sum_{1 \le i,j \le N_d} \mathbb{P}\left((x^{(i)} \cdot \mathbf{W}x^{(i)})/(y^{(i,j)} \cdot \mathbf{W}y^{(i,j)}) \ge c_d\right)$$

in an appropriate form for some c_d depending on d (see estimates (9) below for a precise formulation of such an upper bound). However, estimating this upper bound requires subtle relations among N_d, p, and n, and it turns out that the desired upper bounds in (2) and (3) can be derived in this way only for fixed p or $p = p(n) \to \infty$ with $p(n) = o(n)$; see Section 4 for the proof details.

In order to achieve an application in statistics using Wishart matrices, some non-asymptotic estimates will be derived for the distribution function of the condition number using ratios of two independent χ^2 random variables; see Lemma 3 for a precise formulation. To this end, we decompose the joint probability density function of the ordered p eigenvalues.

The non-trivial proofs of the positivity $I_{p,0}(c) > 0$ and $I_{\infty,0}(c) > 0$ with $c > 1$ and of the finiteness $I_{p,0}(c) < \infty$ and $I_{\infty,0}(c) < \infty$ with $c \ge 1$ are given in Lemma 7.

3. Wishart matrices

In this section we restrict ourselves to real central Wishart matrices $\mathbf{W}_{p\times p} = \mathbf{X}\mathbf{X}^\top/n$. The condition number $k(p,n)$ of $\mathbf{W}$ does not depend on the scaling parameter $1/n$, so throughout this section for simplicity the non-negative real eigenvalues $\lambda_i \geq 0$, $1 \leq i \leq p$, of $\mathbf{X}\mathbf{X}^\top$ are considered. In this setting, let us denote $\lambda = (\lambda_1, \ldots, \lambda_p)$; then, the probability density function of λ can be written as (cf. [10])

$$f_{p,n}(\lambda) = c(p, n) \cdot \exp\left\{-\sum_{i=1}^{p} \lambda_i/2\right\} \cdot \prod_{1\leq i<j\leq p} |\lambda_i - \lambda_j| \cdot \prod_{i=1}^{p} \lambda_i^{(n-p-1)/2},$$

where $c(p, n) = (\pi^{p/2}/2^{(n+2)p/2}) \prod_{i=1}^{p} (\Gamma(1 + i/2)\Gamma((n - p + i)/2))^{-1}$. Hence, the probability density function $g_{p,n}(\lambda_{\text{ord}})$ of the ordered eigenvalues $\lambda_{\text{ord}} = (\lambda_{(1)}, \ldots, \lambda_{(p)})$ with $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)} \geq 0$ is $g_{p,n}(\lambda_{\text{ord}}) = p! f_{p,n}(\lambda_{\text{ord}})$ on $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)} \geq 0$. The main result of this section is to control the condition number using ratios of two independent χ^2 random variables.

Lemma 3. *For any $c \geq 1$ and $2 \leq p \leq n$,*

$$\mathbb{P}(U_1/U_2 \geq c) \leq \mathbb{P}(k^2(p, n) \geq c) \leq a(p, n) \cdot \mathbb{P}(U_3/U_4 \geq c),$$

where $U_1 \sim \chi^2(n)$, $U_2 \sim \chi^2(n)$, $U_3 \sim \chi^2(n + 3p - 5)$, and $U_4 \sim \chi^2(n - p + 1)$ are four independent χ^2 random variables, and

$$a(p, n) = \frac{\pi\,\Gamma((n + 3p - 5)/2)\Gamma((n - p + 1)/2)}{\Gamma(p/2)\Gamma(n/2)\Gamma((p - 1)/2)\Gamma((n - 1)/2)}.$$

Proof. The lower bound follows from (5). More specifically, in the summation $\sum_{i=1}^{n} (S_{x,i}^2 - cS_{y,i}^2)$, each $S_{x,i} = \sum_{k=1}^{p} x_k X_{ki} \sim N(0, 1)$ and $S_{y,i} = \sum_{k=1}^{p} y_k X_{ki} \sim N(0, 1)$. Furthermore, $S_{x,i}$ and $S_{y,j}$ are independent since $\mathbb{E}(S_{x,i}S_{y,j}) = x \cdot y = 0$. Therefore, $\sum_{i=1}^{n} S_{x,i}^2 \sim \chi^2(n)$, which is independent of $\sum_{i=1}^{n} S_{y,i}^2 \sim \chi^2(n)$.

For the upper bound, we decompose the density function $g_{p,n}(\lambda_{\text{ord}})$ of the ordered eigenvalues $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)} \geq 0$ as follows:

$$\begin{aligned} g_{p,n}(\lambda_{\text{ord}}) &= p! f_{p,n}(\lambda_{\text{ord}}) \\ &= p!c(p, n) \cdot \exp\left\{-\sum_{i=1}^{p} \lambda_{(i)}/2\right\} \cdot \prod_{1\leq i<j\leq p} (\lambda_{(i)} - \lambda_{(j)}) \cdot \prod_{i=1}^{p} \lambda_{(i)}^{(n-p-1)/2} \\ &= C(p, n) \cdot \Bigg[(\lambda_{(1)}\lambda_{(p)})^{(n-p-1)/2} \mathrm{e}^{-(\lambda_{(1)}+\lambda_{(p)})/2} \prod_{2\leq i\leq p} (\lambda_{(1)} - \lambda_{(i)}) \\ &\qquad \prod_{2\leq i\leq p-1} (\lambda_{(i)} - \lambda_{(p)})\Bigg] \cdot h_{p,n}(\lambda_{(2)}, \ldots, \lambda_{(p-1)}), \end{aligned}$$

where $C(p, n) = (p!c(p, n))/((p - 2)!c(p - 2, n - 2))$, and $h_{p,n}$ is another density function given by $h_{p,n}(\lambda_{(2)}, \ldots, \lambda_{(p-1)}) = (p - 2)!c(p - 2, n - 2) \cdot \exp\left\{-\sum_{i=2}^{p-1} \lambda_{(i)}/2\right\} \cdot \prod_{2\leq i<j\leq p-1} (\lambda_{(i)} - \lambda_{(j)}) \cdot \prod_{i=2}^{p-1} \lambda_{(i)}^{(n-p-1)/2}$. If one applies the bound $\prod_{2\leq i\leq p} (\lambda_{(1)} -$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

$\lambda_{(i)}) \prod_{2\le i\le p-1} (\lambda_{(i)} - \lambda_{(p)}) \le \lambda_{(1)}^{(p-1)+(p-2)}$ on $\lambda_{(1)} \ge \lambda_{(2)} \ge \cdots \ge \lambda_{(p)} \ge 0$, then $g_{p,n}(\lambda_{\mathrm{ord}}) \le C(p, n) \cdot \lambda_{(1)}^{(n+3p-7)/2} \cdot \lambda_{(p)}^{(n-p-1)/2} \cdot \mathrm{e}^{-(\lambda_{(1)}+\lambda_{(p)})/2} \cdot h_{p,n}(\lambda_{(2)}, \dots, \lambda_{(p-1)})$. Hence,

$$\begin{aligned}\mathbb{P}(k^2(p, n) \ge c) &= \int_{\lambda_{(1)}/\lambda_{(p)}\ge c, \lambda_{(1)}\ge\cdots\ge\lambda_{(p)}\ge 0} g_{p,n}(\lambda_{\mathrm{ord}})\, \mathrm{d}\lambda_{(1)} \cdots \mathrm{d}\lambda_{(p)} \\ &\le \int_{\lambda_{(1)}/\lambda_{(p)}\ge c} \int_{\lambda_{(2)}\ge\cdots\ge\lambda_{(p-1)}\ge 0} g_{p,n}(\lambda_{\mathrm{ord}})\, \mathrm{d}\lambda_{(1)} \cdots \mathrm{d}\lambda_{(p)} \\ &\le C(p, n) \int_{\lambda_{(1)}/\lambda_{(p)}\ge c} \lambda_{(1)}^{(n+3p-7)/2} \cdot \lambda_{(p)}^{(n-p-1)/2} \cdot \mathrm{e}^{-(\lambda_{(1)}+\lambda_{(p)})/2}\, \mathrm{d}\lambda_{(1)}\mathrm{d}\lambda_{(p)} \\ &= a(p, n) \cdot \mathbb{P}(U_3/U_4 \ge c),\end{aligned}$$

where the last equality comes from identifying the density functions of the corresponding two independent χ^2 random variables. We remark that similar decompositions were used for $\lambda_{\max}$ and $\lambda_{\min}$ individually in [11]. □

As mentioned earlier, the upper tail controls specified in Lemma 3 yield an independent concise proof of Corollary 1 (with $p = o(n/\ln n)$), which is presented here for the sake of completeness.

Proof of Corollary 1. It suffices to prove the upper bound for $c > 1$. Lemma 3 yields

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}\left(k^2(p, n) \ge c\right) \le \limsup_{n\to\infty} n^{-1} \ln a(p, n) + \limsup_{n\to\infty} n^{-1} \ln \mathbb{P}(U_3/U_4 \ge c).$$

We shall first prove $\lim_{n\to\infty} n^{-1} \ln a(p, n) = 0$ under $p = o(n/\ln n)$. If p is fixed, then, as $n \to \infty$,

$$\begin{aligned}\ln a(p, n) &= \ln \left[\frac{\pi\Gamma((n+3p-5)/2)\Gamma((n-p+1)/2)}{\Gamma(p/2)\Gamma(n/2)\Gamma((p-1)/2)\Gamma((n-1)/2)}\right] \\ &= \ln\Gamma((n+3p-5)/2) + \ln\Gamma((n-p+1)/2) \\ &\quad - \ln\Gamma(n/2) - \ln\Gamma((n-1)/2) + o(n).\end{aligned}$$

It follows from Stirling's approximation that $\ln\Gamma(x) = x\ln x - x + o(x)$ as $x \to \infty$, so $\ln a(p, n)$ becomes $(n/2)\ln((n+3p-5)/n) + (n/2)\ln((n-p+1)/(n-1)) + o(n) = o(n)$, which proves the limit $\lim_{n\to\infty} n^{-1}\ln a(p, n) = 0$. When $p = p(n) \to \infty$ with $p(n) = o(n/\ln n)$ as $n \to \infty$, similar arguments to those above yield $\ln a(p, n) = (n/2)\ln((n+3p-5)/n) + (n/2)\ln((n-p+1)/(n-1)) + O(1)p\ln n + o(n)$, and this again implies the limit $\lim_{n\to\infty} n^{-1}\ln a(p, n) = 0$.

Next, we shall prove the following estimate:

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}(U_3/U_4 \ge c) \le -2^{-1} \ln\left[(c+1)^2/(4c)\right]. \tag{6}$$

To this end, let us first rewrite the ratio of the two independent χ^2 random variables as summations of squares of independent standard normal random variables,

$$\mathbb{P}(U_3/U_4 \ge c) = \mathbb{P}\left(\sum_{i=1}^{n+3p-5} \xi_i^2 - c\sum_{i=1}^{n-p+1} \eta_i^2 \ge 0\right),$$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

where ξ_i and η_i are independent standard normal random variables. However, the two degrees of freedom $n+3p-5$ and $n-p+1$ are different, so Cramér's theorem cannot be readily applied. To achieve the large-deviation asymptotics, we make use of the Gärtner–Ellis theorem [5, Section 2.3]. To this end, let us define

$$Z_n = \left(\sum_{i=1}^{n+3p-5} \xi_i^2 - c \sum_{i=1}^{n-p+1} \eta_i^2 \right) / n,$$

and consider the logarithmic moment-generating function of Z_n defined as $\Lambda_n(\theta) = \ln \mathbb{E} \exp\{\theta Z_n\}$. By using the fact that $\mathbb{E} \exp\{\theta \xi_i^2\} = (1-2\theta)^{-1/2}$ for $\theta < 1/2$, one can establish the limit, for $-1/(2c) < \theta < 1/2$ and $p = o(n)$,

$$\Lambda(\theta) = \lim_{n\to\infty} n^{-1} \Lambda_n(n\theta) = \lim_{n\to\infty} n^{-1} \ln \mathbb{E} \exp\{n\theta Z_n\} = -2^{-1} \ln \left[(1-2\theta)(1+2c\theta)\right].$$

If one defines the Fenchel–Legendre transform of $\Lambda(\theta)$ as $\Lambda^*(\alpha) = \sup_{\theta\in\mathbb{R}} [\theta \cdot \alpha - \Lambda(\theta)]$, $\alpha \in \mathbb{R}$, then the Gärtner–Ellis theorem says that, for any closed set F,

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{P}\,(Z_n \in F) \le - \inf_{\alpha\in F} \Lambda^*(\alpha). \tag{7}$$

Now we take $F = [0, \infty)$, and we claim that, for any $\alpha \ge 0$, $\Lambda^*(\alpha) = \sup_{\theta\ge 0} [\theta \cdot \alpha - \Lambda(\theta)]$. This comes from the fact that $\Lambda(\theta) = \ln \mathbb{E} \exp\{\theta\xi_i^2 - c\theta\eta_i^2\} \ge \theta(1-c)$, thus implying $\theta \cdot \alpha - \Lambda(\theta) \le \theta(\alpha + (c-1)) \le 0$ for any $\theta \le 0$. Therefore, $\Lambda^*(\alpha)$ is non-decreasing in α for $\alpha \ge 0$. Hence, in (7) we have $\inf_{\alpha\in F} \Lambda^*(\alpha) = \Lambda^*(0) = 2^{-1} \ln \left[(c+1)^2/(4c)\right]$, so (6) now follows from (7). □

4. Random matrices with sub-Gaussian entries

In this section we consider sub-Gaussian random matrices $\mathbf{X}$ with i.i.d. entries X_{ij} being sub-Gaussian satisfying (1). To establish large-deviation asymptotics for the condition numbers in Theorem 1, we handle the two cases p being fixed and $p(n)$ being dependent on n separately since there are more subtle relations among N_d (which appeared in Lemma 2), d, p, and n when $p(n) \to \infty$. To avoid triviality let us focus on the case $c > 1$.

Since the lower bounds have already been proved in Section 2, we focus here on the upper bounds. In the spirit of Lemmas 1 and 2, the set $\{k^2(p, n) \ge c\}$ can be rewritten and estimated as follows (for notational simplicity, all points x, y, $x^{(i)}$, and $y^{(i,j)}$ below will be on S, and we will not write this explicitly):

$$\begin{aligned}\{k^2(p, n) \ge c\} &= \{x, y \text{ s.t. } x \cdot y = 0 \text{ and } (x \cdot \mathbf{W}x)/(y \cdot \mathbf{W}y) \ge c\} \\ &\subseteq \{x^{(i)}, y^{(i,j)} \text{ s.t. } x^{(i)} \cdot y^{(i,j)} = 0 \text{ and} \\ &\qquad x^{(i)} \cdot \mathbf{W}x^{(i)} - cy^{(i,j)} \cdot \mathbf{W}y^{(i,j)} \ge -2\lambda_{\max} d(2c+1)\},\end{aligned} \tag{8}$$

where the inclusion $\subseteq$ comes from Lemma 2 and the fact $|x \cdot \mathbf{W}x - x^{(i)} \cdot \mathbf{W}x^{(i)}| \le (||x|| + ||x^{(i)}||)||W||||x - x^{(i)}|| \le 2\lambda_{\max} d$.

4.1. Fixed dimension *p*

When p is fixed, the number N_d of spherical caps of chord $2d\sqrt{1-d^2/4}$ needed to cover S can be chosen fixed as well, and the exact expression of N_d is not important. To analyze

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

the probability of the latter set in (8), we further consider two disjoint sets $\{\lambda_{\max} \le pK\}$ and $\{\lambda_{\max} > pK\}$ for some large K which will be specified later. Since $\lambda_{\max} \le \operatorname{tr}(\mathbf{W})$, with $\operatorname{tr}(\mathbf{W})$ denoting the trace of $\mathbf{W}$, it then follows from (8) that

$$\begin{aligned}\mathbb{P}(k^2(p,n)\ge c) &\le \mathbb{P}\big(\text{there exists } x^{(i)}, y^{(i,j)} \text{ s.t. } x^{(i)}\cdot y^{(i,j)}=0 \text{ and}\\ &\qquad x^{(i)}\cdot \mathbf{W}x^{(i)} - cy^{(i,j)}\cdot\mathbf{W}y^{(i,j)} \ge -2pKd(2c+1)\big)\\ &\quad + \mathbb{P}\big(\operatorname{tr}(\mathbf{W})>pK\big)\\ &\le \sum_{1\le i,j\le N_d} \mathbb{P}\big(x^{(i)}\cdot\mathbf{W}x^{(i)} - cy^{(i,j)}\cdot\mathbf{W}y^{(i,j)} \ge -2pKd(2c+1)\big)\\ &\quad + \mathbb{P}\big(\operatorname{tr}(\mathbf{W})>pK\big).\end{aligned} \tag{9}$$

Therefore, denoting $\varepsilon := 2pKd(2c+1)$ with $-\varepsilon > 1-c$ for small d, Cramér's theorem yields

$$\begin{aligned}&\limsup_{n\to\infty} n^{-1}\ln \sum_{1\le i,j\le N_d} \mathbb{P}\big(x^{(i)}\cdot\mathbf{W}x^{(i)} - cy^{(i,j)}\cdot\mathbf{W}y^{(i,j)}\ge -\varepsilon\big)\\ &\quad= \max_{1\le i,j\le N_d}\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}\big(x^{(i)}\cdot\mathbf{W}x^{(i)} - cy^{(i,j)}\cdot\mathbf{W}y^{(i,j)}\ge -\varepsilon\big)\\ &\quad\le \max_{1\le i,j\le N_d} -\sup_{\theta\in\mathbb{R}}\Big[-\varepsilon\theta - \ln\mathbb{E}\exp\Big\{\theta\Big(S^2_{x^{(i)},1} - cS^2_{y^{(i,j)},1}\Big)\Big\}\Big]\\ &\quad\le -I_{p,-\varepsilon}(c).\end{aligned}$$

On the other hand, it follows from Cramér's theorem again that, with $K>1$,

$$\begin{aligned}\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}\big(\operatorname{tr}(\mathbf{W})>pK\big) &= \limsup_{n\to\infty} n^{-1}\ln\mathbb{P}\left(\sum_{i=1}^{p}\sum_{j=1}^{n} X_{ij}^2/(np) > K\right)\\ &\le -pI_{X^2}(K),\end{aligned}$$

where $I_{X^2}(K) = \sup_{\theta\in\mathbb{R}}\big[\theta K - \ln\mathbb{E}\exp\big\{\theta X_{11}^2\big\}\big]$. We now apply the following fact, whose proof will be presented in Section 6.

Lemma 4. *Under the assumptions of Theorem 1 with a fixed p, for any* $c>1$ *we have* $\lim_{\varepsilon\to 0^-} I_{p,\varepsilon}(c) = I_{p,0}(c) < \infty$ *and* $\lim_{K\to\infty} I_{X^2}(K) = \infty$.

Taking into account all these observations, we obtain, from (9),

$$\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}(k^2(p,n)\ge c) \le \max\left\{-I_{p,-\varepsilon}(c),\, -pI_{X^2}(K)\right\}.$$

By sending $d\to 0^+$ (equivalently $\varepsilon\to 0^+$), it follows that $I_{p,-\varepsilon}(c)\to I_{p,0}(c)$. Furthermore, by taking large K, it is clear that $I_{X^2}(K) > I_{p,0}(c)$. Therefore,

$$\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}(k^2(p,n)\ge c)\le -I_{p,0}(c).$$

4.2. High dimension $p = p(n)$

We note that the arguments in Section 4.1 do not go through when $p=p(n)\to\infty$ as $n\to\infty$ since, when we take the limit $n\to\infty$, the parameter $\varepsilon\to\infty$ as well. Furthermore, the number

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

N_d of spherical caps of chord $2\tilde{d} := 2d\sqrt{1-d^2/4}$ needed to cover S is important and an explicit expression for this in terms of d is needed. According to [16], $N_d = 4p(n)^{3/2}\tilde{d}^{-p(n)}(\ln p(n) + \ln\ln p(n) - \ln\tilde{d})(1 + O(1/\ln p(n)))$ for all $d < 1/2$ and large $p(n)$. The main ingredient of the proof in the high dimension setting is the following concentration inequality for the maximal singular value of the matrix $\mathbf{X}_{p\times n}$.

Lemma 5. [24, Theorem 5.39] *With $p \le n$, suppose that the entries X_{ij}, $1 \le i \le p$, $1 \le j \le n$ are i.i.d. sub-Gaussian. Then, for any $\gamma \ge 0$, $\mathbb{P}(\lambda_{\max} > (1+\kappa_1+\gamma)^2) \le 2\exp\{-\kappa_2\gamma^2 n\}$, where $\kappa_1, \kappa_2 > 0$ are two constants depending only on the sub-Gaussian norm of X_{ij}.*

With such a concentration inequality, we estimate as follows: $\mathbb{P}(k^2(p,n) \ge c) \le \mathbb{P}(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2) + \mathbb{P}(\lambda_{\max} > (1+\kappa_1+\gamma)^2)$. For the term $\mathbb{P}(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2)$ we can employ the idea used to prove (9) to obtain

$$\mathbb{P}\big(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2\big) \le N_d^2 \max_{1\le i,j\le N_d} \mathbb{P}\big(x^{(i)}\cdot \mathbf{W}x^{(i)} - c y^{(i,j)}\cdot \mathbf{W}y^{(i,j)} \ge -2(1+\kappa_1+\gamma)^2 d(2c+1)\big).$$

It now follows again from Cramér's theorem that, with $\varepsilon = 2(1+\kappa_1+\gamma)^2 d(2c+1)$,

$$\begin{aligned} n^{-1}\ln\mathbb{P}(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2) &\le (2/n)\ln N_d - \min_{1\le i,j\le N_d}\sup_{\theta\in\mathbb{R}}\big[-\varepsilon\theta - \ln\mathbb{E}\exp\big\{\theta\big(S^2_{x^{(i)},1} - cS^2_{y^{(i,j)},1}\big)\big\}\big] \\ &\le (2/n)\ln N_d - I_{p(n),-\varepsilon}(c) \\ &\le (2/n)\ln N_d - I_{\infty,-\varepsilon}(c), \end{aligned}$$

where the last inequality comes from the fact that $I_{p,-\varepsilon} \ge I_{\infty,-\varepsilon}(c)$ for small enough ε. The assumption $p = p(n) = o(n)$ implies that $\lim_{n\to\infty}(2/n)\ln N_d = 0$ for each fixed d. Therefore, $\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2) \le -I_{\infty,-\varepsilon}(c)$. Taking the limit $d \to 0^+$ implies that $\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}(k^2(p,n) \ge c, \lambda_{\max} \le (1+\kappa_1+\gamma)^2) \le -I_{\infty,0}(c)$. The limit $\lim_{\varepsilon\to 0^-} I_{\infty,\varepsilon}(c) = I_{\infty,0}(c)$ is established in Lemma 6. In summary, $\limsup_{n\to\infty} n^{-1}\ln\mathbb{P}(k^2(p,n) \ge c) \le \max(-I_{\infty,0}(c), -\kappa_2\gamma^2)$, and the proof is completed by sending $\gamma \to \infty$.

Lemma 6. *Under the assumptions of Theorem 1 with $p = p(n) \to \infty$ and $p(n) = o(n)$, we have, for any $c > 1$, $\lim_{\varepsilon\to 0^-} I_{\infty,\varepsilon}(c) = I_{\infty,0}(c) < \infty$.*

Proof. Since $I_{p,\varepsilon}(c)$ is non-increasing in p, the arguments leading to (10) still work in this case. Therefore, by taking $p(n) \to \infty$ we obtain $2(N+1)\varepsilon/\varepsilon_0 \le I_{\infty,\varepsilon}(c) - I_{\infty,0}(c) \le 0$, and complete the proof by sending $\varepsilon \to 0^-$. □

5. Application: The union-intersection test method

In this section, suppose that a population is p-variate normal with a zero mean vector and covariance matrix $\Sigma_{p\times p}$. The *sphericity* test deals with the hypotheses $H_0{:}\Sigma_{p\times p} = \sigma^2 I_{p\times p}$ for some $\sigma > 0$, $H_1{:}\Sigma_{p\times p} \ne \sigma^2 I_{p\times p}$. Among others, the union-intersection test method [20, Section 7.4] suggests that H_0 is rejected if $k^2(p,n) \ge c$, where c is determined from $\mathbb{P}(k^2(p,n) \ge c) = \alpha$ with a given significance level α. Unfortunately, so far in the literature there is no efficient way to evaluate the probability $\mathbb{P}(k^2(p,n) \ge c)$ under the hypothesis H_0. We remark here (again) that an exact expression for $\mathbb{P}(k^2(p,n) \ge c)$ was derived in [1] with

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

TABLE 1. Simulated powers when $p = 2$, for $\Sigma_{p\times p} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$, $p = 2$, and $\alpha = 0.05$.

n	5	10	20	50	100
$\hat{h}$	0.1114	0.2419	0.4923	0.9077	0.9981

TABLE 2. Simulated powers when $p = 4$, for $\Sigma_{p\times p} = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.4 \\ 0.25 & 1 & 0.1 & 0.2 \\ 0.25 & 0.1 & 1 & 0.3 \\ 0.4 & 0.2 & 0.3 & 1 \end{bmatrix}$, $p = 4$, and $\alpha = 0.05$.

n	10	20	60	100	160
$\hat{h}$	0.0033	0.0255	0.3752	0.7829	0.9818

$\Sigma_{p\times p} = I_{p\times p}$ for all $2 \le p \le n$ involving complicated zonal polynomials which prevents us efficiently obtaining the probability $\mathbb{P}(k^2(p, n) \ge c)$ (especially for large n). In this section we aim to control the probability $\mathbb{P}(k^2(p, n) \ge c)$ (under $\Sigma_{p\times p} = I_{p\times p}$ in which we choose, without loss of generality, $\sigma = 1$ since the condition number does not depend on σ) using the upper bound in Lemma 3 which is easy and efficient, and then apply it to test the above hypotheses using the union-intersection test method.

Recall the upper bound of $\mathbb{P}(k^2(p, n) \ge c)$ in Lemma 3: $\mathbb{P}(k^2(p, n) \ge c) \le a(p, n) \cdot \mathbb{P}(U_3/U_4 \ge c)$, where $U_3 \sim \chi^2(n + 3p - 5)$ and $U_4 \sim \chi^2(n - p + 1)$ are two independent χ^2 random variables. With such an upper bound, we are able to control the p-value for the above hypotheses test. More specifically, let $k^2_{\text{obs}}(p, n)$ denote the observed condition number for data simulated with $\Sigma_{p\times p} = I_{p\times p}$; then, the p-value has an upper bound p_{upp} defined as $p_{\text{upp}} = a(p, n) \cdot \mathbb{P}(U_3/U_4 \ge k^2_{\text{obs}}(p, n))$. The null hypothesis H_0 is then rejected if $p_{\text{upp}} < \alpha$. Based on these, the simulated powers $\hat{h}$ of the hypotheses are obtained for various $\Sigma_{p\times p}$ through computer simulations (performed $N = 10\,000$ times); see Tables 1 and 2.

These simulated powers suggest that when n is large, it is easy to make a correct conclusion that the population covariance matrix is not an identity matrix. Here we also note that if one uses the above procedures to simulate the type-I error α (under the setting that the population covariance matrix is an identity matrix), it can happen that the simulated type-I error is much smaller than α since an upper bound of the p-value is used during the procedures. For instance, with $p = 4$, $n = 100$, $N = 10\,000$, and $\alpha = 0.05$, the simulated type-I error is around 10^{-4}.

6. Other detailed proofs

This section contains proofs of the facts and the auxiliary lemmas used in previous sections.

Proof of Lemma 4. We first prove that $\lim_{K\to\infty} I_{X^2}(K) = \infty$. Since X_{ij} are sub-Gaussian, it follows from the super-exponential moment that $\mathbb{E}\exp\{tX_{11}^2\} < \infty$ for some $t > 0$ (actually, one can take $t = 1/K_3^2$). Therefore,

$$I_{X^2}(K) = \sup_{\theta\in\mathbb{R}} \left[\theta K - \ln \mathbb{E}\exp\{\theta X_{11}^2\}\right] \ge tK - \ln \mathbb{E}\exp\{tX_{11}^2\} \to \infty \text{ as } K \to \infty.$$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

We then prove that $I_{p,0}(c) < \infty$ for all $p \geq 2$ and $c \geq 1$. It suffices to just prove the case $p = 2$ since $0 \leq I_{p,0}(c) \leq I_{2,0}(c)$. To show this, we take $x^* = (x_1, x_2) = (1/\sqrt{2}, 1/\sqrt{2})$ and $y^* = (y_1, y_2) = (1/\sqrt{2}, -1/\sqrt{2})$. Hence,

$$\begin{aligned} S^2_{x^*,1} - cS^2_{y^*,1} &= \big(X_{11}/\sqrt{2} + X_{21}/\sqrt{2}\big)^2 - c\big(X_{11}/\sqrt{2} - X_{21}/\sqrt{2}\big)^2 \\ &= (1-c)\big(X_{11}^2 + X_{21}^2\big)/2 + (1+c)X_{11}X_{21}. \end{aligned}$$

If everything is restricted to the set $\{X_{21} + \varepsilon_1 \geq X_{11} \geq X_{21} \geq \varepsilon_2 \geq 0\}$ for some $\varepsilon_1, \varepsilon_2 \geq 0$, then

$$\begin{aligned} S^2_{x^*,1} - cS^2_{y^*,1} &\geq (1-c)X_{11}^2 + (1+c)X_{11}(X_{11} - \varepsilon_1) \\ &= X_{11}(2X_{11} - (1+c)\varepsilon_1) \\ &\geq X_{11}(2\varepsilon_2 - (1+c)\varepsilon_1) \\ &\geq 0 \quad \text{if } (1+c)\varepsilon_1 \leq 2\varepsilon_2. \end{aligned}$$

Therefore, for $(1+c)\varepsilon_1 \leq 2\varepsilon_2$,

$$\begin{aligned} I_{2,0}(c) &= \inf_{x,y\in\mathbb{R}^2, ||x||=||y||=1, x\cdot y=0} \sup_{\theta\in\mathbb{R}} \Big[-\ln \mathbb{E}\exp\{\theta(S^2_{x,1} - cS^2_{y,1})\}\Big] \\ &= \inf_{x,y\in\mathbb{R}^2, ||x||=||y||=1, x\cdot y=0} \sup_{\theta\geq 0} \Big[-\ln \mathbb{E}\exp\{\theta(S^2_{x,1} - cS^2_{y,1})\}\Big] \\ &\leq \sup_{\theta\geq 0} \Big[-\ln \mathbb{E}\exp\{\theta(S^2_{x^*,1} - cS^2_{y^*,1})\}\Big] \\ &\leq -\ln \mathbb{P}(X_{21} + \varepsilon_1 \geq X_{11} \geq X_{21} \geq \varepsilon_2 \geq 0). \end{aligned}$$

We now claim that, for any entries X_{ij} satisfying the assumptions of Theorem 1, there always exist ε_1 and ε_2 with $(1+c)\varepsilon_1 \leq 2\varepsilon_2$ such that $\mathbb{P}(X_{21} + \varepsilon_1 \geq X_{11} \geq X_{21} \geq \varepsilon_2 \geq 0) > 0$. Here we chose the right side of zero without loss of generality because $\mathbb{E}(X_{ij} = 0)$, but if needed the left side of zero can be chosen. To see this claim, let us first look at the case when the distribution of X_{ij} has a pure discrete part on $\mathbb{R}^+$, and in this case we can simply take $\varepsilon_1 = \varepsilon_2 = 0$ since $\mathbb{P}(X_{11} = X_{21} \geq 0) > 0$. If the distribution of X_{ij} does not contain a pure discrete part, then there must exist $0 < c_1 < c_2 < c_3$ with $c_3 = c_1 + 2c_1/(1+c)$ (here it is stressed again that, without loss of generality, we only consider the distribution on $\mathbb{R}^+$) such that $\mathbb{P}(X_{ij} \in (c_1, c_2)) > 0$ and $\mathbb{P}(X_{ij} \in (c_2, c_3)) > 0$. Setting $\varepsilon_1 = c_3 - c_1$ and $\varepsilon_2 = c_1$, it follows that $(1+c)\varepsilon_1 = 2\varepsilon_2$ and

$$\begin{aligned} \mathbb{P}(X_{21} + \varepsilon_1 \geq X_{11} \geq X_{21} \geq \varepsilon_2 \geq 0) &\geq \mathbb{P}(X_{21} \in (c_1, c_2), X_{11} \in (c_2, c_3)) \\ &\geq \mathbb{P}(X_{21} \in (c_1, c_2))\mathbb{P}(X_{11} \in (c_2, c_3)) > 0. \end{aligned}$$

This proves $I_{2,0}(c) < \infty$ for any $c \geq 1$. Next, we shall prove that, for $c > 1$, $\lim_{\varepsilon\to 0^+} I_{p,-\varepsilon}(c) = I_{p,0}(c)$. To achieve this limit, first note that we can actually prove a stronger finiteness result: $I_{2,\varepsilon_0}(c) < \infty$ for some small $\varepsilon_0 > 0$. ('Stronger' here refers to the fact that $I_{2,\alpha}(c)$ is non-decreasing in α for all $\alpha \geq 1 - c$.) By taking the same x^* and y^* as in the above arguments, we obtain

$$\mathbb{E}\exp\left\{\theta\big(S^2_{x^*,1} - cS^2_{y^*,1}\big)\right\} \geq \begin{cases} \exp\{\theta\varepsilon_2(2\varepsilon_2 - (1+c)\varepsilon_1)\}\mathbb{P}(X_{21} + \varepsilon_1 \geq X_{11} \geq X_{21} \geq \varepsilon_2 \geq 0), & \text{(I)} \\ \exp\{\theta 2\varepsilon_2^2\}\mathbb{P}(X_{11} = X_{21} \geq \varepsilon_2 \geq 0), & \text{(II)} \end{cases}$$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

where (I) represents the case when the distribution of X_{ij} has no pure discrete part, and (II) means that the distribution has a pure discrete part. If in the above arguments we take $c_3 = c_1 + 1.5c_1/(1+c)$ (everything else will be kept the same), then $(1+c)\varepsilon_1 = 1.5c_1 \le 2\varepsilon_2$. Hence, $I_{2,\varepsilon_0}(c) < \infty$ for any $\varepsilon_0 \le \min\{\varepsilon_2(2\varepsilon_2 - (1+c)\varepsilon_1), 2\varepsilon_2^2\}$, implying that $I_{p,\varepsilon_0}(c) < \infty$ for any $p \ge 2$.

If we denote $g_{x,y}(\alpha) = \sup_{\theta\in\mathbb{R}} \left[\theta\alpha - \ln \mathbb{E} \exp\left\{\theta\left(S_{x,1}^2 - cS_{y,1}^2\right)\right\}\right]$, then it is clear that $I_{p,\alpha}(c) = \inf_{x,y\in\mathbb{R}^p,||x||=||y||=1,x\cdot y=0} g_{x,y}(\alpha)$. Since $0 \le I_{p,\varepsilon}(c) \le I_{p,\varepsilon_0}(c) := N < \infty$ for all $1 - c \le \varepsilon \le \varepsilon_0$, it suffices to consider $A := \{(x, y) \in \mathbb{R}^{2p} : ||x|| = ||y|| = 1, x \cdot y = 0, g_{x,y}(\varepsilon_0) \le N + 1\}$, that is, $I_{p,\varepsilon}(c) = \inf_{x,y\in A} g_{x,y}(\varepsilon)$ for all $1 - c \le \varepsilon \le \varepsilon_0$. Then, for any $(x, y) \in A$, the convexity of $g_{x,y}(\varepsilon)$ in $\varepsilon \in (1 - c, 0)$ yields

$$(g_{x,y}(0) - g_{x,y}(\varepsilon))/(0-\varepsilon) \le (g_{x,y}(\varepsilon_0) - g_{x,y}(0))/(\varepsilon_0).$$

This gives us the following nice bounds (with $g_{x,y}(0) \ge 0$ in mind):

$$0 \le g_{x,y}(0) - g_{x,y}(\varepsilon) \le (N+1)(-\varepsilon)/\varepsilon_0.$$

Hence,

$$\begin{aligned}(N+1)\varepsilon/\varepsilon_0 &\le \inf_{(x,y)\in A} \left(g_{x,y}(\varepsilon) - g_{x,y}(0)\right)\\ &\le \inf_{(x,y)\in A} g_{x,y}(\varepsilon) - \inf_{(x,y)\in A} g_{x,y}(0) \qquad (10)\\ &= I_{p,\varepsilon}(c) - I_{p,0}(c) \le 0.\end{aligned}$$

Sending $\varepsilon \to 0^-$ in (10) completes the proof. □

Lemma 7. *The functions $I_{p,0}(c)$ and $I_{\infty,0}(c)$ in Theorem 1 satisfy $I_{p,0}(1) = 0$ and $0 < I_{p,0}(c) < \infty$ for $c > 1$, $I_{\infty,0}(1) = 0$ and $0 < I_{\infty,0}(c) < \infty$ for $c > 1$.*

Proof. The proof of the finiteness of $I_{p,0}(c)$ and $I_{\infty,0}(c)$ has been done separately in Lemmas 4 and 6. Here we prove the rest.

Jensen's inequality yields $\ln \mathbb{E} \exp\{\theta(S_{x,1}^2 - S_{y,1}^2)\} \ge \mathbb{E}(\theta(S_{x,1}^2 - S_{y,1}^2)) = 0$. Therefore, $\sup_{\theta\in\mathbb{R}} \left[-\ln \mathbb{E} \exp\left\{\theta\left(S_{x,1}^2 - S_{y,1}^2\right)\right\}\right] \le 0$, implying that $I_{p,0}(1) \le 0$. Taking $\theta = 0$ gives $I_{p,0}(1) = 0$. The term $I_{\infty,0}(1)$ as the limit of $I_{p,0}(1)$ must also satisfy $I_{\infty,0}(1) = 0$.

Now we prove that $I_{p,0}(c) > 0$ and $I_{\infty,0}(c) > 0$ for $c > 1$; this will be done in three steps.

Step 1 We claim that, for any $c \ge 1$, $p \ge 2$,

$$I_{p,0}(c) \ge 1/2 \cdot \inf_{x,y\in\mathbb{R}^p,||x||=||y||=1,x\cdot y=0} \sup_{\theta\ge 0} \left[-\ln\left(\mathbb{E} \exp\{\theta S_{x,1}^2\}\mathbb{E} \exp\{-\theta c S_{y,1}^2\}\right)\right]. \qquad (11)$$

To see (11), it suffices to notice that

$$\mathbb{E} \exp\{\theta(S_{x,1}^2 - cS_{y,1}^2)\} \le (\mathbb{E} \exp\{2\theta S_{x,1}^2\})^{1/2}(\mathbb{E} \exp\{-2\theta c S_{y,1}^2\})^{1/2}.$$

Step 2 For $c > 1$ and fixed $p \ge 2$, it will be proved that $I_{p,0}(c) > 0$. To this end, let us first note that, because of $\mathbb{E} \exp\{tX_{11}^2\} < \infty$ for some $t > 0$ (as remarked in the proof of Lemma 4),

$$\mathbb{E} \exp\{tS_{x,1}^2\} \le \mathbb{E} \exp\left\{t\sum_{i=1}^{p} X_{i1}^2\right\} = \left(\mathbb{E} \exp\{tX_{11}^2\}\right)^p < \infty. \qquad (12)$$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

Then, it follows from $e^x = 1 + x + x^2/2 + e^{\alpha x}x^3/6$ for some $\alpha \in [0, 1]$ that

$$\mathbb{E}\exp\{sS_{x,1}^2\} = 1 + s + s^2\mathbb{E}(S_{x,1}^4)/2 + s^3\mathbb{E}\left(\exp\{\alpha sS_{x,1}^2\}\cdot S_{x,1}^6\right)/6$$

for small $0 \le s \le \varepsilon_0$. The term $\mathbb{E}(S_{x,1}^4)$ can be bounded as

$$\mathbb{E}(S_{x,1}^4) = \mathbb{E}(X_{11}^4)\sum_{i=1}^{p} x_i^4 + 6\mathbb{E}(X_{11}^2)\sum_{1\le i<j\le p} x_i^2x_j^2 \le 16K_2^4 + 12pK_2^2,$$

and the term

$$\mathbb{E}\left(\exp\{\alpha sS_{x,1}^2\}\cdot S_{x,1}^6\right) \le \mathbb{E}\left(\exp\{\alpha sS_{x,1}^2\}\cdot(1+\exp\{\varepsilon S_{x,1}^2\})K\right) < \infty$$

for some $K > 0$ and ε with $\varepsilon_0 + \varepsilon \le t$, where the finiteness comes from (12). Therefore, we can rewrite, for small $0 \le s \le \varepsilon_0$,

$$\mathbb{E}\exp\{sS_{x,1}^2\} = 1 + s + s^2\cdot c_1 + s^3\cdot c_2(s),$$
$$\mathbb{E}\exp\{sS_{y,1}^2\} = 1 + s + s^2\cdot c_1^{'} + s^3\cdot c_2^{'}(s),$$

where $0 < c_1, c_1^{'} < \infty$ and $0 < c_2(s), c_2^{'}(s) < \infty$ for all $0 \le s \le \varepsilon_0$. Now we apply the estimate $\ln(1+x) \le x - x^2/4$ for all $|x| < 1$ and obtain

$$\ln\mathbb{E}\exp\{\theta S_{x,1}^2\} \le (\theta + \theta^2\cdot c_1 + \theta^3\cdot c_2(\theta)) - o(\theta),$$
$$\ln\mathbb{E}\exp\{-\theta cS_{y,1}^2\} \le (-\theta c + (\theta c)^2\cdot c_1^{'} - (\theta c)^3\cdot c_2^{'}(\theta)) - o(\theta)$$

for small θ. Hence, for small enough θ, uniform in x and y,

$$\ln\left(\mathbb{E}\exp\{\theta S_{x,1}^2\}\mathbb{E}\exp\{-\theta cS_{y,1}^2\}\right) \le \theta(1-c) + o(\theta).$$

Taking this inequality back to (11), we get, for small $\theta \le \epsilon$,

$$I_{p,0}(c) \ge 1/2 \cdot \inf_{x,y\in\mathbb{R}^p,||x||=||y||=1,x\cdot y=0}\ \sup_{\theta\le\epsilon}\left[-\ln\left(\mathbb{E}\exp\{\theta S_{x,1}^2\}\mathbb{E}\exp\{-\theta cS_{y,1}^2\}\right)\right]$$
$$\ge \sup_{\theta\le\epsilon}\left[\theta(c-1) + o(\theta)\right] > 0.$$

Step 3 For $c > 1$, we prove that $I_{\infty,0}(c) > 0$. The proof will be identical to the one in Step 2 if we one can show, instead of using (12), that $\mathbb{E}\exp\{tS_{x,1}^2\} \le C < \infty$ for some $t > 0$ and some constant $C > 0$ which is independent of p and x. To this end, we use the moment-generating function upper bound in the definition of a sub-Gaussian random variable:

$$\mathbb{E}\exp\{tX_{11}\} \le \exp\{t^2K_4^2\} \qquad \text{for all } t \in \mathbb{R}. \tag{13}$$

https://doi.org/10.1017/jpr.2021.13 Published online by Cambridge University Press

Now we apply (13) and the fact that $\mathbb{E}\exp\{tZ\} = \exp\{t^2/2\}$ for all $t \in \mathbb{R}$, where $Z \sim N(0, 1)$, which is independent of the entries, and derive, for any $0 \le t \le 1/(8K_4^2)$, that

$$\begin{aligned}
\mathbb{E}\exp\{tS_{x,1}^2\} &= \mathbb{E}\big(\mathbb{E}\big(\exp\{tS_{x,1}^2\} \mid S_{x,1}\big)\big) = \mathbb{E}\big(\mathbb{E}\big(\mathbb{E}\big(\exp\{\sqrt{2t}S_{x,1}Z\} \mid S_{x,1}\big) \mid S_{x,1}\big)\big) \\
&= \mathbb{E}\big(\mathbb{E}\big(\exp\{\sqrt{2t}S_{x,1}Z\} \mid S_{x,1}\big)\big) = \mathbb{E}\exp\{\sqrt{2t}S_{x,1}Z\} \\
&= \mathbb{E}\big(\mathbb{E}\big(\exp\{\sqrt{2t}S_{x,1}Z\} \mid Z\big)\big) \\
&= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=1}^{p}\mathbb{E}\exp\{\sqrt{2t}x_iX_{i1}Z\} \mid Z\right)\right) \le \mathbb{E}\left(\prod_{i=1}^{p}\exp\{2tx_i^2Z^2K_4^2\}\right) \\
&= \mathbb{E}\exp\{2tZ^2K_4^2\} = \big(1-4tK_4^2\big)^{-1/2} \le \sqrt{2}.
\end{aligned}$$

This completes the proof. □

7. Open problems and future work

In this paper we have only considered condition numbers for some special rectangular random matrices (i.e. $p < n$ with $p = o(n)$). It would be interesting and challenging to study large-deviation probabilities of condition numbers for square random matrices with $p = n$ (and also for almost square random matrices with $n = p + b$ for b fixed or $b = o(n)$). The main difficulty lies in the fact that the various estimates used throughout the paper become imprecise when $p = n$ (or $n = p + b$).

Another rectangular case we have not touched on here is when $p/n \to \kappa \in (0, 1)$, and in this case $k^2(p, n) \to (1+\kappa^{1/2})^2/(1-\kappa^{1/2})^2$ in probability. With this law of large numbers it is also natural to study the corresponding large-deviation asymptotics of $k(p,n)$ for large n.

The last open problem is about a universality result (mentioned in (4)): under what assumptions on the distribution of the entries does $I_{\infty,0}(c) = 2^{-1}\ln\big[(c+1)^2/(4c)\big]$? If the assumptions can be specified, then the large-deviation asymptotics of condition numbers for all such distributions coincide with those for Wishart matrices.

Acknowledgements

The authors are truly grateful to the two referees who carefully read the paper and provided constructive and essential comments which have led to an improved version of the paper, and to the Editor for help throughout the whole process.

References

[1] Anderson, W. and Wells, M. (2009). The exact distribution of the condition number of a Gaussian matrix. *SIAM J. Matrix Anal. Appl.* **31**, 1125–1130.

[2] Bai, Z., Silverstein, J. and Yin, Y. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivar. Anal.* **26**, 166–168.

[3] Bai, Z. and Yin, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Prob.* **21**, 1275–1294.

[4] Chen, Z. and Dongarra, J. (2005). Condition numbers of Gaussian random matrices. *SIAM J. Matrix Anal. Appl.* **27**, 603–620.

[5] Dembo, A. and Zeitouni, O. (2010). *Large Deviations Techniques and Applications*, corrected reprint of 2nd ed. Springer, Berlin.

[6] Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9**, 543–560.

[7] Edelman, A. and Sutton, B. (2005). Tails of condition number distributions. *SIAM J. Matrix Anal. Appl.* **27**, 547–560.

[8] 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.

[9] GUSTAFSON, K. (2012). *Antieigenvalue Analysis*, World Scientific, Hackensack, NJ.

[10] JAMES, A. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35**, 475–501.

[11] JIANG, T. AND LI, D. (2015). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J. Theoret. Prob.* **28**, 804–847.

[12] KEVEI, P. (2010). A note on asymptotics of linear combinations of iid random variables. *Period. Math. Hungar.* **60**, 25–36.

[13] LITVAK, A., PAJOR, A., RUDELSON, M. AND TOMCZAK-JAEGERMANN, N. (2005). Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195**, 491–523.

[14] LITVAK, A., TIKHOMIROV, K. AND TOMCZAK-JAEGERMANN, N. (2019). Small ball probability for the condition number of random matrices. In *Geometric Aspects of Functional Analysis*, ed. B. KLARTAG AND E. MILMAN, Vol. II, Springer, Berlin.

[15] MUIRHEAD, R. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley, New York.

[16] ROGERS, C. (1963). Covering a sphere with spheres. *Mathematika* **10**, 157–164.

[17] RUDELSON, M. (2008). Invertibility of random matrices: norm of the inverse. *Ann. Math.* **168**, 575–600.

[18] RUDELSON, M. AND VERSHYNIN, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218**, 600–633.

[19] RUDELSON, M. AND VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62**, 1707–1739.

[20] SRIVASTAVA, M. S. AND KHATRI, C. (1979). *An Introduction to Multivariate Statistics*. North-Holland, Amsterdam.

[21] TAO, T. AND VU, V. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. Math.* **169**, 595–632.

[22] TAO, T. AND VU, V. (2010). Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.* **20**, 260–297.

[23] TREFETHEN, L. AND BAU, D. (1997). *Numerical Linear Algebra*. SIAM, Philadelphia, PA.

[24] VERSHYNIN, R. (2012). *Introduction to the Non-Asymptotic Analysis of Random Matrices*. Cambridge University Press.