Probab. Theory Relat. Fields (2009) 144:221–246
DOI 10.1007/s00440-008-0146-x

Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles

Tiefeng Jiang

Received: 15 April 2007 / Revised: 3 January 2008 / Published online: 11 March 2008

Abstract We develop a tool to approximate the entries of a large dimensional complex Jacobi ensemble with independent complex Gaussian random variables. Based on this and the author's earlier work in this direction, we obtain the Tracy–Widom law of the largest singular values of the Jacobi emsemble. Moreover, the circular law, the Marchenko–Pastur law, the central limit theorem, and the laws of large numbers for the spectral norms are also obtained.

Keywords Haar measure · Eigenvalue · Random matrix · Largest eigenvalue · Empirical distribution · Limiting distribution

Mathematics Subject Classification (2000) 15A33 · 15A52 · 60F05 · 60F15

1 Introduction and main results

There are two purposes in this paper. The first is approximating a part of Haar distributed unitary matrices by independent complex normals. The second is to prove, by using the previous result, that the limiting distribution of the largest singular value of a Jacobi ensemble follows the Tracy–Widom distribution. Besides, for the squared singular values of the Jacobi ensembles, we prove that the empirical distributions converge to the Marchenko–Pastur law; the central limit theorem holds; the law of large numbers of the largest eigenvalues holds. We also prove that the empirical distribution of the (complex) eigenvalues of a Jacobi ensemble converges to the circular

Supported in part by NSF #DMS-0449365.

T. Jiang (✉)
School of Statistics, University of Minnesota, 224 Church Street, Minneapolis, MN 55455, USA
e-mail: tjiang@stat.umn.edu

law. Before stating our results, we review some central themes in Random Matrix Theory.

Random Matrix Theory mainly concerns on the eigenvalues of matrices with random entries. These matrices include, among others, Gaussian symmetric matrices, Wishart matrices, Jacobi ensembles, Haar invariant matrices on compact groups. A recent work by Bryc et al. [13] studied the Toeplitz, Hankel and Markov matrices, which are different than the matrices investigated before. About eigenvalues, early work of random matrices by statisticians such as Hsu, Girshick and Wilks are on density functions of eigenvalues of real and complex Wishart matrices, see, e.g., Anderson [2] and Muirhead [55]; the celebrated semicircular law was established by physicist Wigner [68]; the limiting distributions of the largest eigenvalues of Gaussian orthogonal, unitary and symplectic ensembles were obtained by Tracy and Widom [63–65].

Random matrix theory has a vast literature, one can check, for instance, Anderson [2], Bai [3,5], Diaconis [21], Eaton [27] and Muirhead [55] for the interest of statistics. For some connections between the random matrix theory and Engineering, one can see Tulino and Verdu [66]. A good reference for random matrix theory and applications in Physics are Beenakker [11], Bohigas [12], Forrester [30], Guhr [35] and Mehta [54]. For some connections between random matrix theory and other fields of mathematics, particularly the number theory, one can see, e.g., Conrey et al. [18], Deift [23], Katz and Sanark [48], and Mezzardi and Snaith [53].

Investigating the entries of large dimensional random matrices is another interest in Random Matrix Theory. Compared with the studies of eigenvalues, there are few literatures about this. For the research in this direction, one can see, for example, D'Aristotle et al. [20], Diaconis et al. [22], and Jiang [42–44], which are based on statistical testing problems, and the image analysis. In the forthcoming paper, by using the properties of the entries of Haar distributed matrices, Jiang [41] obtain the exact formula of the variance for a quantum conductance studied in [10,11,58].

In the first part of this paper, we will develop a tool to understand the entries of a truncated part of a Haar distributed unitary matrices, which has the same distribution as that of a Jacobian ensemble. We will then apply this tool to study the asymptotic properties of the eigenvalues of the Jacobi ensembles.

The classical definition of a Jacobi ensemble is matrix $J = (C^*C + D^*D)^{-1/2} C(C^*C + D^*D)^{-1/2}$, where C and D are $n_1 \times m$ and $n_2 \times m$ matrices, respectively, $n_1 \geq m$ and $n_2 \geq m$, and the total $(n_1+n_2)m$ entries are i.i.d. standard real (complex) Gaussian random variables. See, e.g., Anderson [2], Constantine [19], Eaton [27], Muirhead [55], Collins [17] and Forrester [30] for full details. The density of the eigenvalues of J is that

$$f(\lambda_1, \ldots, \lambda_m) = C \cdot \prod_{i=1}^{m} \lambda_i^{a\beta/2} \cdot \prod_{i=1}^{m} (1-\lambda_i)^{b\beta/2} \cdot \prod_{1\leq i<j\leq m} |\lambda_i - \lambda_j|^{\beta} \cdot I_A, \tag{1.1}$$

where C is a normalizing constant, $A = \{0 \leq \lambda_i \leq 1;\ 1 \leq i \leq m\}$, $\beta = 1$ corresponds to the real case, $\beta = 2$ corresponds to the complex case, $a = n_1 - m + 1 - 2/\beta$ and $b = n_2 - m + 1 - 2/\beta$, see, e.g., (3.15) and (4.1) from [29].

Table 1 Dyson's "threefold way" of classical random matrices

	Real, $\beta = 1$	Complex, $\beta = 2$	Quaternion, $\beta = 4$
Hermite	GOE	GUE	GSE
Laguerre	Real Wishart	Complex Wishart	Quaternion Wishart
Jacobi	Real MANOVA	Complex MANOVA	Quaternion MANOVA

It is shown in [16,29] that J has the same distribution as that of U^*U, where U is the $p \times q$ upper-left corner of an $n \times n$ Haar orthogonal (unitary) matrix, where $p = n_1, q = m$ and $n = n_1 + n_2$ (see (3.14) and (4.1) from [29]). Therefore, the eigenvalues of J, a Jacobi ensemble, is the *squared* singular values of U. For this reason, we will only consider the eigenvalues of U in this paper.

Now, to better understand what our limiting results are in the general context of Random Matrix Theory, let's look at Dyson's "threefold way" of classical random matrices, see Table 1 (see also [25,26]), on which much research of the Random Matrix Theory focuses.

For the first two rows of matrices in Table 1 and for $\beta = 1, 2, 4$, the circular law, the semi-circular law, the Marchenko–Pastur law of eigenvalues, the strong law of largest eigenvalues, the Tracy–Widom distribution for the largest eigenvalues had been derived, respectively, by many authors since Wigner [68]. In this paper we prove that if p and q are of order $o(\sqrt{n})$, then all these results also hold for the third row of matrices for $\beta = 1$ and 2, that is, the real and complex Jacobi ensemble of matrices. To the author's knowledge, these results are new. While stating our main results below, some relevant references will also be given.

Let μ and ν be two probability measures on $(\mathbb{R}^m, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra. The variation distance between μ and ν, denoted by $\|\mu - \nu\|$, is equal to

$$\|\mu - \nu\| = 2 \cdot \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)| = \int_{\mathbb{R}^m} |f(x) - g(x)|\, dx_1 dx_2 \ldots dx_m \tag{1.2}$$

provided μ and ν have density functions $f(x)$ and $g(x)$ with respect to the Lesbegue measure, respectively. This is one of the strongest probability distances between two probability measures. In fact, four probability distances will be used in this paper. We use $\mathcal{L}(X)$ to denote the probability distribution of random matrix X. When we say "X is the upper-left $p \times q$ block of matrix A", we mean X is the matrix formed by the first p rows and first q columns of A. Our first result is as follows.

Theorem 1 (Normal approximation) *Let $\{p_n;\ n \geq 1\}$ and $\{q_n;\ n \geq 1\}$ be two sequences of positive integers such that $p_n = o(\sqrt{n})$ and $q_n = o(\sqrt{n})$. Let U_n be the upper-left $p_n \times q_n$ block of an $n \times n$ Haar invariant unitary matrix. Let X_n be a $p_n \times q_n$ matrix whose $p_n q_n$ entries are i.i.d. standard complex normal random variables. Then* $\lim_{n\to\infty} \|\mathcal{L}(\sqrt{n}U_n) - \mathcal{L}(X_n)\| = 0$.

The next result shows that above result is also sharp.

Theorem 2 (Sharp order) *Given $x > 0$, $y > 0$ and integer n. Set $p_n = [x\sqrt{n}]$ and $q_n = [y\sqrt{n}]$ for $n \geq 1$. Let U_n be the upper-left $p_n \times q_n$ block of an $n \times n$ Haar invariant unitary matrix. Let X_n be a $p_n \times q_n$ matrix whose $p_n q_n$ entries are i.i.d. standard complex normal random variables. Then*

$$\liminf_{n\to\infty} \|\mathcal{L}(\sqrt{n}U_n) - \mathcal{L}(X_n)\| \geq 2\Phi\left(\frac{xy}{2}\right) - 1,$$

where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt$ for $x \in \mathbb{R}$.

Theorems 1 and 2 also hold for the Haar orthogonal matrices. They are proved in Jiang [42]. In the applications next, we will need the following corollary. As usual, $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ stand for the eigenvalues of an $n \times n$ Hermitian matrix A.

Corollary 1.1 *Let the condition in Theorem 1 hold for U_n being the truncated part of a Haar orthogonal or unitary matrix. Given integer $k \geq 1$, let $f_n(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}^k$ be measurable functions for all $n \geq 1$. If for some Borel set $F \subset \mathbb{R}^k$ such that $P(f_n(\lambda_1(X_n^* X_n), \ldots, \lambda_n(X_n^* X_n)) \in F) \to C$ for some constant C, then this also holds if $X_n^* X_n$ is replaced by $nU_n^* U_n$ for each $n \geq 1$.*

In the following we will investigate the law of large numbers for spectral radius, the circular law, the semicircular law for the Jacobi ensemble.

The fact that a scaled spectral radius of a matrix of i.i.d. standard normals as entries converges was first proved by Geman [31]. Later Bai et al. [7], Yin et al. [71] generalized it to the matrices with entries being i.i.d. (arbitrary) random variables with the finite fourth moment. Yin et al. [70] and Silverstein [60,61] are the first to obtain some properties of smallest eigenvalues. Silverstein [59] proved the strong convergence of the smallest eigenvalue of Wishart matrices. Bai and Yin [7] generalized it to the covariance matrices whose entries are not necessarily Gaussian random variables.

For a $p \times q$ matrix $U_{p,q}$, denote $\lambda_{\max}$ the largest eigenvalue of $U_{p,q}^* U_{p,q}$, and

$$\lambda_{\min} = \begin{cases} \text{the smallest eigenvalue of } U_{p,q}^* U_{p,q}, & \text{if } p \geq q; \\ \text{the } (q-p+1)\text{th smallest eigenvalue of } U_{p,q}^* U_{p,q}, & \text{if } p < q. \end{cases} \tag{1.3}$$

The reason for this definition is that the $q - p$ smallest eigenvalues of $q \times q$ matrix $U_{p,q}^* U_{p,q}$ are all zero if $q > p$. For the Jacobi ensemble, we have the following result.

Theorem 3 (Spectral limit) *Suppose Γ_n is an $n \times n$ Haar invariant orthogonal or unitary matrix, and $U_{p,q}$ be its upper-left $p \times q$ sub-matrix. Let $\lambda_{\max}$ be the largest eigenvalues of $U_{p,q}^* U_{p,q}$, and $\lambda_{\min}$ be as in (1.3). If $p \to \infty$, $p = o(\sqrt{n})$ and $q/p \to c \in (0, +\infty)$, then*

$$\left(\frac{n}{p}\right)\lambda_{\max} \to (1+\sqrt{c})^2 \quad \textit{and} \quad \left(\frac{n}{p}\right)\lambda_{\min} \to (1-\sqrt{c})^2 \tag{1.4}$$

in probability as $n \to \infty$.

The empirical distribution of the eigenvalues of the covariance matrices was first obtained by Marchenko and Pastur [52] (see also [8,34,45,67,69] for some of the many studies of this and related questions). To state the result in the Jacobi case, we need some notation first. Let Γ_n be an $n \times n$ Haar invariant matrix, and $U = U_{p,q}$ be its upper-left $p \times q$ sub-matrix. Let $F_{p,q}$ be the empirical distribution of $(n/p)\lambda_i$, where $\lambda_i,\ 1 \le i \le q$, are the eigenvalues of U^*U, that is,

$$F_{p,q}(x) = \frac{1}{q}\sum_{i=1}^{q} I\left\{(np^{-1})\lambda_i \le x\right\}, \quad x \in \mathbb{R}. \tag{1.5}$$

The Marchenko–Pastur distribution is the probability distribution with density

$$p(x) = \begin{cases} \frac{1}{2\pi c x}\sqrt{(b-x)(x-a)}, & \text{if } a < x < b; \\ 0, & \text{otherwise,} \end{cases} \tag{1.6}$$

and a point mass $1-c^{-1}$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$ for some $c \in (0, 1]$.

Hiai and Petz [36] showed that the large deviations for $F_{p,q}$ holds when $p/n \to \gamma \in (0, \infty)$, and the rate function has an unique zero. This means that $F_{p,q}$ converges weakly to a probability distribution. The exact limiting distribution for complex case was given explicitly in [14,51].

Considering here a smaller value of p_n, we complement their result by explicitly characterizing the limiting distribution and proving that the convergence also holds for the Kolmogorov–Smirnov distance.

Theorem 4 (Marchenko–Pastur limit) *Let Γ_n be an $n \times n$ Haar invariant orthogonal or unitary matrix. If $p \to +\infty,\ q/p \to c > 0$ and $p = o(n/(\log n))$, then $\|F_{p,q} - F\| := \sup_{x\in\mathbb{R}} |F_{p,q}(x) - F(x)| \to 0$ in probability, where $F(x) = \int_{-\infty}^{x} p(t)\,dt$ for $x \in \mathbb{R}$, and $p(x)$ is as in (1.6).*

Next we study the circular law. Let μ and ν be two probability measures on $\mathbb{C}$ (or $\mathbb{R}^2$). Define

$$\rho(\mu, \nu) = \sup_{\|f\|_L \le 1} \left| \int_{\mathbb{C}} f(x)\,\mu(dx) - \int_{\mathbb{C}} f(x)\,\nu(dx) \right|, \tag{1.7}$$

where f above is a bounded Lipschitz function defined on $\mathbb{C}$ with $\|f\| = \sup_{x\in\mathbb{C}} |f(x)|$, and $\|f\|_L = \|f\| + \sup_{x\neq y} |f(x) - f(y)|/|x-y|$. This metric generates the topology of the weak convergence of probability measures on $\mathbb{C}$ (see, e.g., Chap. 11 from [24]), that is, μ_n converges to μ weakly if and only if $\rho(\mu_n, \mu) \to 0$ as $n \to \infty$.

Girko [32,33] found the Circular Law, that is, the empirical distribution of the (complex) eigenvalues of a non-symmetric square matrix of i.i.d. entries goes to the uniform distribution on the unit disc. Silverstein proved the Circular Law for the complex normal case in his unpublished notes, which was reported in Hwang

[39]. Edelman [28] proved that the expected value of the empirical distribution converges. Bai [4] is the first one who rigorously proved the Circular Law when the entries of the matrices are i.i.d. random variables (not necessarily normally distributed).

For an upper-left $p \times p$ corner of an $n \times n$ Haar unitary matrix, Życzkowski and Sommers [72] simulated the empirical measure of its eigenvalues for small n and p. It is very interesting to see from Petz and Réffy [56] that the above empirical measure converges to a non-uniform distribution on the unit disc in the complex plane when $p/n \to \gamma \in (0, \infty)$ as $n \to \infty$. They proved the result by using the large deviations. We complement their results by showing the weak convergence to a uniform distribution on the unit disk for $p_n = o(\sqrt{n})$.

Theorem 5 (Circular law) *Suppose Γ_n is an $n \times n$ Haar invariant orthogonal or unitary matrix, and U_p is its upper-left $p \times p$ sub-matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of U_p. Denote by μ_n the empirical measure of $\sqrt{n/p}\lambda_i$'s, and μ the uniform distribution on $\{z \in \mathbb{C};\ |z| \le 1\}$. If $p \to +\infty$ and $p = o(\sqrt{n})$, then $\rho(\mu_n, \mu) \to 0$ in probability as $n \to \infty$.*

The central limit theorems for the traces of kth power of n-dimensional sample covariance matrices was proved by Jonsson [45] when k is fixed and $n \to \infty$. Sinai and Soshnikov [62] extended this CLT to $k = k(n)$ which grows sufficiently slow in n. Bai and Silverstein [6] proved a CLT for linear spectral statistics of large-dimensional generalized sample covariance matrices. The CLT for Jacobi ensembles is given next.

Theorem 6 (Central limit theorem) *Let Γ_n be an $n \times n$ Haar invariant matrix, and $U_{p,q}$ be its upper-left $p \times q$ sub-matrix. Assume that $p \to +\infty$ and $q/p \to c \in (0, \infty)$ and $p = o(\sqrt{n})$. Let $\lambda_1, \lambda_2, \ldots, \lambda_q$ be the eigenvalues of $(n/p)U_{p,q}^* U_{p,q}$, and $f_1, f_2, \ldots, f_k$ be functions on $\mathbb{R}$ analytic on an open interval containing $[(1-c)^2, (1+c)^2]$.*

(i) *If Γ_n is a Haar invariant orthogonal matrix, then $\left(\sum_{i=1}^q f_1(\lambda_i), \ldots, \sum_{i=1}^q f_k(\lambda_i)\right) - c_n$ converges weakly to a k-dimensional normal distribution $N_k(\mu, \Sigma)$, where μ, c_n and Σ are deterministic (see Theorem 1.1 from [6] for further details about the expressions of μ, c_n and Σ).*

(ii) *If Γ_n is a Haar invariant unitary matrix, then $\left(\sum_{i=1}^q f_1(\lambda_i), \ldots, \sum_{i=1}^q f_k(\lambda_i)\right) - c_n$ converges weakly to $N_k\left(0, \frac{1}{2}\Sigma\right)$, where c_n and Σ are as in (i).*

Recently, Killip [49] obtained the CLT for $F_{p,q}(x)$ as in (1.5) for fixed x and its muti-dimensional analogue, which is different than Theorem 6.

Now we study the asymptotic distributions of the largest eigenvalues of the Jacobi ensembles. Let $q(x)$ be the solution of the Painléve II (non-linear) differential equation

$$
\begin{aligned}
&q''(x) = xq(x) + 2q(x)^3 \text{ with boundary condition,}\\
&q(x) \sim \mathrm{Ai}(x) \text{ as } x \to +\infty,
\end{aligned}
$$

where $\mathrm{Ai}(x)$ denotes the Airy function. The Tracy–Widom distributions are as follows.

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x) + (x-s)q(x)^2\, dx\right), \tag{1.8}$$

$$F_2(s) = \exp\left(-\int_s^\infty (x-s)q(x)^2\, dx\right), \quad s \in \mathbb{R}. \tag{1.9}$$

Let Γ_n be an $n \times n$ matrix, and $U_{p,q}$ be its upper-left $p \times q$ sub-matrix. Denote

$$\mu_{np} = \left(\sqrt{p} + \sqrt{q}\right)^2,$$
$$\sigma_{np} = \left(\sqrt{p} + \sqrt{q}\right)\left(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}}\right)^{1/3}.$$

Tracy and Widom [63–65] established the limiting distributions of the largest eigenvalues of Gaussian orthogonal ensembles (GOE), Gaussian unitary ensembles (GUE) and Gaussian symplectic ensembles (GSE). These distributions are called the Tracy–Widom distribution subsequently. Johansson [46] proved that the largest eigenvalues of complex Wishart matrices converge to F_2. Johnstone [47] later showed that this is also true for the real Wishart matrices with limiting distribution F_1.

For the third row of Table 1, Constantine [19] in 1963 established the exact distribution of the largest eigenvalues of the real Jacobi ensembles. Koev and Dumitriu [50] recently generalized it to the β-Jacobi matrices. In particular, their results hold for Gaussian orthogonal ($\beta = 1$), Gaussian unitary ($\beta = 2$) and Gaussian symplectic ($\beta = 4$) cases. Considering the largest principal angles between random subspaces, Absil et al. [1] obtained some formulas similar to those in [50]. Their results in [1,50] are based on an infinite series of terms involving with the Jack functions. We complement these works by proving the weak convergence to a Tracy–Widom distribution.

Theorem 7 (Tracy–Widom limit law) *Suppose Γ_n is an $n \times n$ Haar invariant matrix, $U_{p,q}$ is its upper-left $p\times q$ sub-matrix, and $\lambda_{\max}$ is the largest eigenvalue of $U_{p,q}^* U_{p,q}$. Assume $p \to \infty,\ p = o(\sqrt{n})$ and $q/p \to c \in (0, +\infty)$.*

- (i) *If Γ_n is a Haar orthogonal matrix, then $(n\lambda_{\max} - \mu_{n,p})/\sigma_{n,p}$ converges weakly to F_1 as in (1.8).*
- (ii) *If Γ_n is a Haar unitary matrix, then $(n\lambda_{\max} - \mu_{n,p})/\sigma_{n,p}$ converges weakly to F_2 as in (1.9).*

The idea of the proofs of Theorems 1 and 2 is essentially to approximate the density function in (1.1) by that of a Wishart matrix, which has form

$$f(\lambda_1, \ldots, \lambda_m) = C' \cdot \prod_{i=1}^m \lambda_i^{a\beta/2} \cdot e^{-b\beta/2 \sum_{i=1}^m \lambda_i} \cdot \prod_{1 \le i < j \le m} |\lambda_i - \lambda_j|^\beta \tag{1.10}$$

by using the heuristic that $1 - x \sim e^{-x}$ as $x \to 0$, where C' is a numerical constant.

Future problems. In the above theorems, conclusions hold under two different orders of p_n and q_n. Recall the definition of variation distance in (1.2), Theorem 1 is sharp for the approximation order that $p = o(\sqrt{n})$ and $q = o(\sqrt{n})$, this does not mean that the orders of p and q in Theorems 3, 5, 6 and 7 cannot be improved. The reason is that the variation distance is the maximum of the differences of two probabilities over *all* Borel sets, whereas in those theorems we only need that the difference of two probabilities of a *specific* set goes to zero. The same argument applies to Theorems 4 in terms of Theorems 8 and 9 stated in Sect. 3. It will be interesting to see how much the order can be improved.

The organization of this paper is as follows: the main results are stated in this section; we prove Theorems 1, 2 and Corollary 1.1 in Sect. 2; we prove Theorems 3–7 in Sect. 3.

2 Proofs of Theorems 1, 2 and Corollary 1.1

Certain steps of the proofs in this section are similar to those of Theorems 1 and 2 from [42]. However, the proofs here are more friendly.

The general form of the probability density function of the Jacobian ensemble is known, see, e.g., [17,29]. We will first calculate the normalizing constant to make it to be a density function, which will be used later in the proofs of Theorems 1 and 2.

The first part of the following lemma belongs to Hsu [38]. The second part is simply the Jacobian determinant for the transform from a positive definite Hermitian matrix to its eigenvalues, see, e.g., [40]. The abbreviation "p.d.f." next means "probability density function".

Lemma 2.1 *Let Z be a m by n ($m \le n$) matrix of complex entries with p.d.f. $g(ZZ^*)$. Then the p.d.f. of $R = ZZ^*$ is given by $\pi^{mn}(\Gamma_m(n))^{-1}(\det(R))^{n-m}g(R)$. Moreover, the p.d.f. of the eigenvalues $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_m)$ of R is given by*

$$\frac{\pi^{m(n+m-1)}}{\Gamma_m(n)\Gamma_m(m)} g(\Lambda) \cdot \left(\prod_{i=1}^{m} \lambda_i\right)^{n-m} \prod_{1\le i<j\le m} (\lambda_i - \lambda_j)^2,$$

if $g(U\Lambda U^*) = g(\Lambda)$ for any unitary matrix U, where $\Gamma_m(n) = \pi^{m(m-1)/2}\prod_{i=1}^{m}\Gamma(n-i+1)$.

The following well-known formula can be found in many places, e.g., Mehta [54] and Forrester [30].

Lemma 2.2 (Selberg integral) *Let $N \ge 2$ be an integer, and α, β and γ be positive numbers. Then*

$$\int_0^1\int_0^1 \cdots \int_0^1 \prod_{i=1}^{N} x_i^{\alpha-1}(1-x_i)^{\beta-1} \cdot \prod_{1\le j<k\le N} |x_j - x_k|^{2\gamma}\, dx_1\, dx_2 \dots dx_N$$

$$= \prod_{l=0}^{N-1} \frac{\Gamma(1+\gamma+l\gamma)\Gamma(\alpha+l\gamma)\Gamma(\beta+l\gamma)}{\Gamma(1+\gamma)\Gamma(\alpha+\beta+(N+l-1)\gamma)}.$$

The next lemma is a special case of (2.15) and (2.16) from [3].

Lemma 2.3 *Let $\{p_n; n \geq 1\}$ and $\{q_n; n \geq 1\}$ be two sequences of positive integers such that $p_n \to \infty$ and $p_n/q_n \to \eta \in (0, \infty)$. Let X_n be a $p_n \times q_n$ random matrix whose entries are i.i.d. complex normals with mean zero and variance one. The following two statements hold. For each integer $k \geq 1$,*

$$\text{(i)} \quad E\left(tr(X_n^* X_n)^k\right) \sim p_n^k q_n \sum_{r=0}^{k-1} \frac{1}{r+1} \left(\frac{q_n}{p_n}\right)^r \binom{k}{r}\binom{k-1}{r}$$

as $n \to \infty$.
(ii)

$$\frac{tr\left((X_n^* X_n)^k\right)}{q_n^{k+1}} \to \sum_{r=0}^{k-1} \frac{\eta^{k-r}}{r+1}\binom{k}{r}\binom{k-1}{r}$$

in probability as $n \to \infty$.

The following lemma is quite similar to Lemma 2.4 from [42]. The only difference is that the fourth moment of a standard real normal is 3 and that of the absolute value of a standard complex normal is 2.

Lemma 2.4 *Given $\epsilon \in (0, 1)$. Let $\{p_n;\ n \geq 1\}$ and $\{q_n;\ n \geq 1\}$ be two sequences of positive integers such that $\epsilon < p_n/q_n < \epsilon^{-1}$ for all $n \geq 1$, and $p_n \to \infty$. For each n, let $X_n = (x_{ij})$ be a $p_n \times q_n$ matrix whose entries are i.i.d. standard complex normals. Then, as $n \to \infty$,*

$$(i)\ \mathrm{Var}(tr((X^*X)^2)) \sim p_n^2 q_n^2 + 4p_n q_n (p_n + q_n)^2,$$
$$(ii)\ \mathrm{Cov}(tr(X^*X)^2, tr(X^*X)) \sim 2p_n q_n (p_n + q_n).$$

Proof For convenience, we simply write $p = p_n$ and $q = q_n$ in what follows. For a standard complex normal ξ, we know $|\xi|^2 \sim Exp(1)$. Therefore

$$E|\xi|^2 = 1,\ \ E|\xi|^4 = 2 \text{ and } \ E|\xi|^6 = 6. \tag{2.1}$$

Use identity $tr(UV) = tr(VU)$ to obtain that

$$tr((X^*X)^2) = tr(XX^*XX^*) = \sum_{1 \leq i,k \leq p_n, 1 \leq j,l \leq q_n} x_{ij}\bar{x}_{kj}x_{kl}\bar{x}_{il}.$$

Classifying the sum according to that $i=k$ or $j=l$, we have that

$$
\begin{aligned}
tr(X^*X) &= \sum_{j=1}^{q}\sum_{i=1}^{p}|x_{ij}|^2,\\
tr((X^*X)^2)&\\
&= \sum_{j=1}^{q}\sum_{i=1}^{p}|x_{ij}|^4+\sum_{j=1}^{q}\sum_{i\neq l=1}^{p}|x_{ij}|^2|x_{lj}|^2+\sum_{i=1}^{p}\sum_{j\neq k=1}^{q}|x_{ij}|^2|x_{ik}|^2\\
&\quad+\sum_{i\neq k,\, j\neq l} x_{ij}\bar{x}_{kj}x_{kl}\bar{x}_{il}.
\end{aligned}
\tag{2.2}
$$

Set

$$
\begin{aligned}
B_1 &= \sum_{j=1}^{q}\sum_{i=1}^{p}(|x_{ij}|^4-2),\quad B_2=\sum_{j=1}^{q}\sum_{i\neq l=1}^{p}(|x_{ij}|^2-1)(|x_{lj}|^2-1),\\
B_3 &= \sum_{i=1}^{p}\sum_{j\neq k=1}^{q}(|x_{ij}|^2-1)(|x_{ik}|^2-1),\quad B_4=\sum_{i\neq k,\, j\neq l} x_{ij}\bar{x}_{kj}x_{kl}\bar{x}_{il}.
\end{aligned}
$$

It is easy to check that

$$
tr((X^*X)^2)=\left(\sum_{i=1}^{4}B_i\right)+2(p+q-2)(tr(X^*X)-pq)+C_{p,q}, \tag{2.3}
$$

where $C_{p,q}$ is a constant on p and q. Moreover, it is not difficult to see that $EB_i=0$ for $1\le i\le 4$, $\mathrm{Cov}(B_i,B_j)=0$ for all $1\le i\neq j\le 4$, and $\mathrm{Cov}(B_i,tr(X^*X))=0$ for $i=2,3,4$. Also, each B_i is a sum of uncorrelated random variables. Therefore,

$$
\begin{aligned}
\mathrm{Var}\left(tr((X^*X)^2)\right) &= \left(\sum_{i=1}^{4}\mathrm{Var}(B_i)\right)+4(p+q-2)^2\mathrm{Var}(tr(X^*X))\\
&\quad+4(p+q-2)\mathrm{Cov}(B_1,tr(X^*X)).
\end{aligned}
$$

Now it is easy to verify that $\mathrm{Cov}(B_1,tr(X^*X))=O(p^2)$ and $\mathrm{Var}(B_i)=O(p^3)$ for $i=1,2,3$ as $p\to\infty$. Moreover, from (2.2), we know B_4 is a real number, so $\mathrm{Var}(B_4)=E|B_4|^2=pq(p-1)(q-1)$ and $\mathrm{Var}(tr(X^*X))=pq\mathrm{Var}(|x_{11}|^2)=pq$ by (2.1). Combining these quantities together, we obtain (i).

(ii) By (2.3) again,

$$
\begin{aligned}
\mathrm{Cov}(tr((X^*X)^2),\ tr(X^*X)) &= \mathrm{Cov}(tr(X^*X),\ B_1)+2(p+q-2)\cdot\mathrm{Var}(tr(X^*X))\\
&\sim 2pq(p+q)
\end{aligned}
$$

as $n\to\infty$. □

To get the uniform integrability for a certain sequence of random variables, we need the following lemma.

Lemma 2.5 *Let the notation be as in Lemma 2.4 except that x_{11} be any random variable with $Ex_{11} = 0$, $E|x_{11}|^2 = 1$ and $E|x_{11}|^{16} < \infty$. Set $Z_{n,i} = p_n^{-i}\big(tr(X_n^*X_n)^i - Etr(X_n^*X_n)^i\big)$ for $i = 1, 2$. Then $sup_{n\geq 1} E|Z_{n,i}|^4 < \infty$ for $i = 1, 2$.*

Proof As in the proof of Lemma 2.4, write $p = p_n$ and $q = q_n$. Recall (2.2), $pZ_{n,1}$ is a sum of pq i.i.d. random variables with mean zero. Then $E|pZ_{n,1}|^4 = O((pq)^2)$ as $n \to \infty$. So the conclusion for $i = 1$ holds.

Recall (2.3). By the convex property of function $h(x) = x^4$, we have that

$$E(p^2 Z_{ni})^4 \leq C\left(\sum_{i=1}^{4} E|B_i|^4 + p^4 E(tr(X_n^*X_n) - Etr(X_n^*X_n))^4\right).$$

The last term in the parenthesis is of order $p^8 E|Z_{n,1}|^4 = O(p^8)$. Again, B_1 is a sum of pq independent random variables with mean zero. Then $E(B_1^4) = O((pq)^2) = O(p^4)$. Notice that B_4 is a sum of uncorrelated random variables, and each term is a product of four independent and centered random variables sitting in the corners of a non-degenerate rectangle in the matrix. Then EB_4^4 is the same as the fourth moment of the sum of $pq(p-1)(q-1)$ independent random variables with mean zero. So, $E(B_4)^4 = O(p^8)$.

Set $y_{ij} = |x_{ij}|^2 - 1$ for all i, j. Then $B_2 = \sum_{j=1}^{q}\sum_{i\neq l=1}^{p} y_{ij}y_{lj}$. Then

$$B_2 = \sum_{j=1}^{q}\left(\left(\sum_{i=1}^{p} y_{ij}\right)^2 - p\right) - \sum_{j=1}^{q}\sum_{i=1}^{p}(y_{ij}^2 - 1).$$

The second term is a sum of pq i.i.d. r.v.'s. Then its fourth moment is $O((pq)^2)$. The fourth moment of the first term is bounded by

$$Cq^2 E\left(\left(\sum_{i=1}^{p} y_{ij}\right)^2 - p\right)^4 \leq C'q^2\left(p^4 + E\left(\sum_{i=1}^{p} y_{i1}\right)^8\right) = O(p^6)$$

for some universal constants $C > 0$, $C' > 0$, since $E(\sum_{i=1}^{p} y_{i1})^8 = O(p^4)$. Thus, $E(B_2)^4 = O(p^6)$. Similarly, $E(B_3)^4 = O(p^6)$. In summary, $p^8 E|Z_{n2}|^4 = E(p_n^2 Z_{n2})^4 = O(p^8)$ as $n \to \infty$. The conclusion then follows for $i = 2$. The proof is complete. □

The lower bound in Theorem 2 actually comes from the following.

Lemma 2.6 *Let $Z \sim N(0,1)$. Then, $E|1 - e^{-t^2+tZ}| = 2\Phi(t) - 1 \in (0,1)$ for any $t > 0$, where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt$.*

Proof Observe that

$$\begin{aligned}
&E(1-e^{-t^2+tZ})I(Z\le t)+E(e^{-t^2+tZ}-1)I(Z>t)\\
&=E(1-e^{-t^2+tZ})+2E(e^{-t^2+tZ}-1)I(Z>t)\\
&=1-e^{-t^2/2}+\frac{2}{\sqrt{2\pi}}\int_t^{\infty}(e^{-t^2+tx}-1)e^{-x^2/2}\,dx.
\end{aligned}$$

Now, note that $(x-t)^2/2=t^2/2-tx+x^2/2$ the last term is equal to

$$\begin{aligned}
&\frac{2}{\sqrt{2\pi}}\int_t^{\infty}e^{-t^2+tx-x^2/2}\,dx-\frac{2}{\sqrt{2\pi}}\int_t^{\infty}e^{-x^2/2}\,dx\\
&=e^{-t^2/2}\left(\frac{2}{\sqrt{2\pi}}\int_0^{\infty}e^{-x^2/2}\,dx\right)-2P(Z>t)\\
&=e^{-t^2/2}+2\Phi(t)-2.
\end{aligned}$$

The two identities prove the lemma. □

Now we state the density function of the complex Jacobi ensemble as follows.

Proposition 2.1 *Let Γ_n be an $n\times n$ Haar invariant unitary matrix. Let p and q be two positive integers such that $p\ge q$ and $p+q\le n$. Let also U be the $p\times q$ upper-left block of Γ_n. Then the density of U is*

$$f(U)=C_{p,q}\cdot(\det(I-U^*U))^{n-p-q}I(\lambda_{\max}(U^*U)\le 1), \tag{2.4}$$

where

$$\frac{1}{C_{p,q}}=\pi^{pq}\cdot\prod_{i=1}^{q}\frac{(n-i-p)!}{(n-i)!}.$$

Proof By Theorem 5.1 from [17] or (3.10) from [29], we know (2.4) holds with some unknown constant $C_{p,q}>0$. We compute it next. Define

$$g(A)=C_{p,q}\cdot(\det(I-A))^{n-p-q}I(\lambda_{\max}(A)\le 1),$$

where A is a $q\times q$ non-negative definite matrix. Then $g(W^*\Lambda W)=g(\Lambda)$ for any unitary matrix W, where $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_q)$ and $\lambda_1,\ldots,\lambda_q$ are eigenvalues of A.

Taking $Z = U^*$, by Lemma 2.1, the density of eigenvalues of ZZ^* is

$$C_{p,q}\frac{\pi^{q(p+q-1)}}{\Gamma_q(p)\Gamma_q(q)}\prod_{i=1}^{q}\lambda_i^{p-q}(1-\lambda_i)^{n-p-q} \times \prod_{1\le i<j\le q}(\lambda_i-\lambda_j)^2 I(0\le\lambda_1<\ldots,\lambda_q\le 1), \tag{2.5}$$

where $\Gamma_q(p) = \pi^{q(q-1)/2}\prod_{i=1}^{q}\Gamma(p-i+1)$. Obviously,

$$\int\limits_{0\le\lambda_1<\cdots<\lambda_q\le 1} h(\lambda_1,\ldots,\lambda_q)\,d\lambda_1\ldots d\lambda_q = \frac{1}{q!}\int_0^1\cdots\int_0^1 h(\lambda_1,\ldots,\lambda_q)\,d\lambda_1\ldots\lambda_q$$

for any symmetric function $g(\lambda_1,\ldots,\lambda_q)$. Taking $\alpha = p-q+1$, $\beta = n-p-q+1$ and $\gamma = 1$ in Lemma 2.2, and integrating the function in (2.5) over $\mathbb{R}^q$, we have that

$$\begin{aligned}
1 &= \frac{1}{q!}\cdot C_{p,q}\frac{\pi^{q(p+q-1)}}{\Gamma_q(p)\Gamma_q(q)}\cdot\prod_{l=0}^{q-1}\frac{\Gamma(l+2)\Gamma(l+p-q+1)\Gamma(n+l-p-q+1)}{\Gamma(n-q+l+1)}\\
&= \frac{1}{q!}\cdot C_{p,q}\pi^{pq}\cdot\left(\frac{\prod_{i=1}^{q}\Gamma(i+1)}{\prod_{i=1}^{q}\Gamma(q-i+1)}\right)\cdot\left(\frac{\prod_{i=1}^{q}\Gamma(i+p-q)}{\prod_{i=1}^{q}\Gamma(p-i+1)}\right)\\
&\quad\times\left(\prod_{i=1}^{q}\frac{\Gamma(n+i-p-q)}{\Gamma(n-q+i)}\right)\\
&= \frac{1}{q!}\cdot C_{p,q}\pi^{pq}\Gamma(q+1)\cdot\prod_{i=1}^{q}\frac{\Gamma(n-p-i+1)}{\Gamma(n-q+i)},
\end{aligned}$$

where we set $i = l+1$ in the second equality, and use in the third equality the fact that the first parenthesis is equal to $\Gamma(q+1)$, the second is identical to 1. Note that $\prod_{i=1}^{q}\Gamma(n-q+i) = \prod_{i=1}^{q}\Gamma(n-i+1)$. The conclusion then follows from the fact that $\Gamma(m+1) = m!$ for any integer $m\ge 0$. □

Lemma 2.7 *Let U be the upper-left $p\times q$ block of an $n\times n$ Haar invariant unitary matrix. Assume $q\le p$ and $p+q\le n$. Let $X_{p,q}$ be a $p\times q$ matrix whose pq entries are i.i.d. standard complex normal random variables. Then the variation distance between the distribution of $\sqrt{n}U$ and that of $X_{p,q}$ is equal to $E|K_nL_n-1|$, where*

$$K_n = \frac{1}{n^{pq}}\prod_{i=1}^{q}\frac{(n-i)!}{(n-i-p)!},$$

$$L_n = e^{\sum_{i=1}^{q}\lambda_i}\left(\prod_{i=1}^{q}\left(1-\frac{\lambda_i}{n}\right)\right)^{n-p-q} I\left(\max_{1\le i\le q}\lambda_i\le n\right),$$

and $\lambda_1, \ldots, \lambda_q$ are eigenvalues of Z^*Z with Z being a $p \times q$ matrix whose entries are independent and standard complex normals.

Proof Let $f_{\sqrt{n}U}(x)$ be the probability density function of $\sqrt{n}U$, and $f_Y(x)$ be the probability density function of Y. Recalling that $\sqrt{n}U$ has $2pq$ variables, then

$$\begin{aligned} f_{\sqrt{n}U}(x) &= \frac{1}{n^{pq}} f\left(\frac{X}{\sqrt{n}}\right) \\ &= \frac{C_{p,q}}{n^{pq}} \left(\det\left(I - \frac{X^*X}{n}\right)\right)^{n-p-q} I(\lambda_{\max}(X^*X) \le 1), \end{aligned}$$

and $f_Y(x) = (\sqrt{\pi})^{-2pq} e^{-tr(X^*X)}$, where $C_{p,q}$ is as in Proposition 2.1. Let $\|\mathcal{L}(\sqrt{n}U) - \mathcal{L}(Y)\|$ denote the variation distance between $\sqrt{n}U$ and Y. Then

$$\begin{aligned} \|\mathcal{L}(\sqrt{n}U) - \mathcal{L}(Y)\| &= \int_{\mathbb{R}^{2pq}} \left| f_{\sqrt{n}U}(x) - f_Y(x) \right| dx \\ &= \int_{\mathbb{R}^{2pq}} \left| \frac{f_{\sqrt{n}U}(x)}{f_Y(x)} - 1 \right| f_Y(x)\, dx \\ &= E\left| \frac{f_{\sqrt{n}U}(Z)}{f_Y(Z)} - 1 \right|, \end{aligned}$$

where Z is a $p \times q$ matrix with entries being i.i.d. standard complex normals. The conclusion then follows from the facts that $\det(I - Z^*Z/n)^{n-p-q} = (\prod_{i=1}^q (1 - n^{-1}\lambda_i))^{n-p-q}$, and $tr(Z^*Z) = \sum_{i=1}^q \lambda_i$, where λ_i's are eigenvalues of Z^*Z. □

The next two lemmas analyze the precise behaviors of K_n and L_n appearing in Lemma 2.7.

Lemma 2.8 *Given $x > 0$ and $y > 0$, let $p = [xn^{1/2}]$ and $q = [yn^{1/2}]$. Set*

$$K_n = \frac{1}{n^{pq}} \prod_{j=1}^{q} \frac{(n-j)!}{(n-j-p)!}.$$

Then

$$K_n = \exp\left\{ -\frac{p^2q + pq^2}{2n} - \frac{2x^3y + 2xy^3 + 3x^2y^2}{12} + O\left(\frac{1}{\sqrt{n}}\right) \right\} \tag{2.6}$$

as n is sufficiently large.

Proof Write $K_n = n^{-pq} \prod_{j=1}^q \prod_{i=-1}^{p-2} (n - i - j) = \prod_{j=1}^q \prod_{i=0}^{p-1} (1 - (i+j)/n)$. Then

$$\log K_n = \sum_{j=1}^{q} \sum_{i=0}^{p-1} \log\left(1 - \frac{i+j}{n}\right).$$

Note that $|\log(1-x)+x+(x^2/2)| \le x^3$ for x small enough, and $(p+q)/n \to 0$ as $n \to \infty$ since $p \sim x\sqrt{n}$ and $q \sim y\sqrt{n}$, we have that

$$\begin{aligned} &\left|\log K_n + \frac{1}{n}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i+j) + \frac{1}{2n^2}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i+j)^2\right| \\ &\quad \le \frac{1}{n^3}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i+j)^3 \end{aligned}$$

as n is sufficiently large. Let

$$H_n = \frac{1}{n}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i+j) \text{ and } I_n = \frac{1}{2n^2}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i+j)^2$$

for $n \ge 1,\ p \ge 1$ and $q \ge 1$. Since $(i+j)^3 \le (p+q)(i+j)^2$ for $0 \le i \le p$ and $1 \le j \le q$, we obtain

$$|\log K_n + H_n + I_n| \le \frac{C}{\sqrt{n}} I_n \tag{2.7}$$

for some constant $C = C(x, y) > 0$ as n is sufficiently large. Now we evaluate H_n and I_n. Recall

$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1) \ \text{ and } \ \sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$$

for $n \ge 1$. Then

$$H_n = \frac{q}{n}\sum_{i=0}^{p-1} i + \frac{p}{n}\sum_{j=1}^{q} j = \frac{pq(p-1)}{2n} + \frac{pq(q+1)}{2n} = \frac{pq(p+q)}{2n}. \tag{2.8}$$

Now

$$\begin{aligned} I_n &= \frac{1}{2n^2}\sum_{j=1}^{q}\sum_{i=0}^{p-1}(i^2+j^2+2ij) \\ &= \frac{q}{2n^2}\frac{(p-1)p(2p-1)}{6} + \frac{p}{2n^2}\frac{q(q+1)(2q+1)}{6} + \frac{2}{2n^2}\frac{(p-1)p}{2}\cdot\frac{q(q+1)}{2} \\ &= \frac{2x^3y + 2xy^3 + 3x^2y^2}{12} + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

as $n \to \infty$, where we sum i^2, j^2 and ij separately in the second equality above, and the equalities $p = x\sqrt{n} + O(1)$ and $q = y\sqrt{n} + O(1)$ are used in the last step. This together with (2.7) and (2.8) proves the lemma. □

Lemma 2.9 *Let $x > 0$ and $y > 0$ be constants, and $p = p_n = [x\sqrt{n}]$ and $q = q_n = [y\sqrt{n}]$. Recall function L_n in Lemma 2.7. Then, $e^{-a_n}L_n$ converges weakly to the distribution of $e^{\sigma\xi}$ where $\xi \sim N(0, 1)$, and*

$$a_n = \frac{p^2q + pq^2}{2n} + \frac{x^3y + xy^3}{6} \text{ and } \sigma = \frac{xy}{2}.$$

Proof Set $X_n = X_{p,q}$ and

$$f(x) = \begin{cases} x + (n - p - q)\log\left(1 - \frac{x}{n}\right), & \text{if } 0 \le x < n; \\ -\infty, & \text{otherwise.} \end{cases} \tag{2.9}$$

Then, $L_n = \exp(\sum_{i=1}^q f(\lambda_i))$. For any $x \in (0, n)$, by the Taylor expansion, there exists $\xi = \xi_x \in (0, x)$ such that

$$\log\left(1 - \frac{x}{n}\right) = -\frac{x}{n} - \frac{x^2}{2n^2} - \frac{x^3}{3n^3} - \frac{x^4}{4}\cdot\frac{1}{(\xi - n)^4}.$$

Then

$$f(x) = \frac{p+q}{n}x - \frac{n-p-q}{2n^2}x^2 - \frac{n-p-q}{3n^3}x^3 + g_n(\xi)\frac{x^4}{n^3}, \quad x \in (0, n), \tag{2.10}$$

where $g_n(x) = -n^3(n-p-q)/(4(x-n)^4)$. It is trivial to see that $\sup_{0\le x\le \alpha n}|g_n(x)| \le (1-\alpha)^{-4}$ for any $\alpha \in (0, 1)$. Recall that $\lambda_1, \lambda_2, \ldots, \lambda_q$ are eigenvalues of $X_n^*X_n = X_{p,q}^*X_{p,q}$, where the entries of the $p \times q$ matrix X_n are independent standard complex normals. Recall that $p \sim x\sqrt{n}$ and $q \sim y\sqrt{n}$. By Theorem 2.16 from [3], there exists a constant $c(x, y) \in (0, \infty)$ such that

$$\frac{\max_{1\le i\le q}\lambda_i}{p} \to c(x, y) \tag{2.11}$$

in probability as $n \to \infty$. Define $\Omega_n := \{\max_{1\le i\le q}\lambda_i \le (c(x, y) + 1)p\}$. Then

$$P(\Omega_n^c) \to 0 \tag{2.12}$$

as $n \to \infty$. Now on Ω_n, by (2.10),

$$\begin{aligned}\sum_{i=1}^q f(\lambda_i) = {} & \frac{p+q}{n}tr\left(X_n^*X_n\right) - \frac{n-p-q}{2n^2}tr\left((X_n^*X_n)^2\right) \\ & - \frac{n-p-q}{3n^3}tr\left((X_n^*X_n)^3\right) + \tilde{g}_n\frac{tr\left((X_n^*X_n)^4\right)}{n^3},\end{aligned} \tag{2.13}$$

where $|\tilde{g}_n| \in [0, 2)$ as n is sufficiently large.

By (ii) of Lemma 2.3,

$$\frac{tr((X^*X)^4)}{n^3} = \frac{tr((X^*X)^4) - Etr((X^*X)^4)}{n^3} + O\left(\frac{q^5}{n^3}\right) \to 0 \tag{2.14}$$

in probability as $n \to \infty$. Taking $\eta = x/y$ in (ii) of Lemma 2.3, we obtain

$$\frac{p+q}{2n^2} tr\left((X_n^*X_n)^2\right) = \frac{(p+q)q^3}{2n^2} \cdot \frac{tr\left((X_n^*X_n)^2\right)}{q^3} \to \frac{xy(x+y)^2}{2}, \tag{2.15}$$

$$\frac{n-p-q}{3n^3} tr\left((X_n^*X_n)^3\right) \sim \frac{q^4}{3n^2} \cdot \frac{tr\left((X_n^*X_n)^3\right)}{q^4} \to \frac{xy(x^2+y^2+3xy)}{3} \tag{2.16}$$

and

$$\frac{tr((X^*X)^4)}{n^3} = \frac{q^5}{n^3} \cdot \frac{tr((X^*X)^4)}{q^5} \to 0$$

in probability as $n \to \infty$ since $p \sim x\sqrt{n}$ and $q \sim y\sqrt{n}$. It is easy to check that the term on left hand side of (2.15) minus that of (2.16) converges to $(x^3y + xy^3)/6$ in probability as $n \to \infty$. Define

$$R_n = \frac{p+q}{n} tr\left(X_n^*X_n\right) - \frac{1}{2n} tr\left((X_n^*X_n)^2\right). \tag{2.17}$$

By (2.12) and (2.13), to prove the lemma, it is enough to show

$$R_n - \frac{p^2q + pq^2}{2n} \text{ converges weakly to } N\left(0, \frac{x^2y^2}{4}\right) \tag{2.18}$$

as $n \to \infty$. Reviewing (2.2), we have that

$$\begin{aligned} tr((X_n^*X_n)^2) = {} & \sum_{j=1}^{q}\sum_{i=1}^{p}|x_{ij}|^4 + \sum_{j=1}^{q}\sum_{i\neq l=1}^{p}|x_{ij}|^2|x_{lj}|^2 + \sum_{i=1}^{p}\sum_{j\neq k=1}^{q}|x_{ij}|^2|x_{ik}|^2 \\ & + \sum_{i\neq k,\ j\neq l} x_{ij}\bar{x}_{kj}x_{kl}\bar{x}_{il}. \end{aligned} \tag{2.19}$$

Then, by (2.1),

$$Etr\left((X_n^*X_n)^2\right) = pq(E|x_{11}|^4) + qp(p-1) + pq(q-1) = pq(p+q).$$

Easily, $E\left(tr(X_n^*X_n)\right) = pq$. Thus

$$ER_n = \frac{pq(p+q)}{n} - \frac{pq(p+q)}{2n} = \frac{pq(p+q)}{2n}.$$

Now set $h_i = tr(X_n^* X_n)^i - E\left(tr(X_n^* X_n)^i\right)$ for $i = 1, 2$. From (2.18) and the above, to prove the lemma, it suffices to show that

$$W_n := \frac{p+q}{n} h_1 - \frac{1}{2n} h_2 \text{ converges to } N(0, \sigma^2) \text{ weakly,} \tag{2.20}$$

where σ is as in the statement of the lemma. Since, $tr(X_n^* X_n) = \sum_{i,j} |x_{ij}|^2$, which is a sum of i.i.d. random variables, $\mathrm{Var}(h_1) = \mathrm{Var}(tr(X_n^* X_n)) = pq$. By Lemma 2.4, $\mathrm{Var}(h_2)/n^2$ converges to a positive constant. By (ii) of Theorem 1.1 from [6], $(h_1/\sqrt{\mathrm{Var}(h_1)},\ h_2/\sqrt{\mathrm{Var}(h_2)})$ converges weakly to a normal distribution with mean zero. It follows that W_n converges weakly to a normal distribution with mean zero and variance σ^2. By Lemma 2.5, $\{W_n^2;\ n \geq 1\}$ is uniformly integrable, therefore $\sigma^2 = \lim_{n\to\infty} E W_n^2$. Now,

$$\begin{aligned}\mathrm{Var}(W_n) = {} & \frac{(p+q)^2}{n^2} \mathrm{Var}\left(tr(X_n^* X_n)\right) + \frac{1}{4n^2} \mathrm{Var}\left(tr\left((X_n^* X)^2\right)\right) \\ & - \frac{p+q}{n^2} \cdot \mathrm{Cov}\left(tr(X_n^* X_n), tr\left((X_n^* X_n)^2\right)\right).\end{aligned}$$

Since $\mathrm{Var}(tr(X_n^* X_n)) = pq$ as calculated earlier, by Lemma 2.4 again, the above yields

$$\mathrm{Var}(W_n) \to \frac{x^2 y^2}{4}$$

as $n \to \infty$. Therefore, $\sigma^2 = x^2 y^2/4$. The proof is completed. □

Now we are ready to prove Theorems 1 and 2.

Proof of Theorem 2 Lemma 2.7 says that $\|\mathcal{L}(U_n) - \mathcal{L}(X_n)\| = E|K_n L_n - 1|$. Lemmas 2.8 and 2.9 imply that $K_n L_n$ converges to $e^{\mu + \sigma\xi}$ in distribution, where $\mu = -x^2 y^2/4$ and $\sigma = xy/2$. Then the conclusion follows from the Fatou lemma and Lemma 2.6. □

Proof of Theorem 1 Heuristically, the proof here corresponds to the case $x = y = 0$ in the proof of Theorem 2. We now make it rigorous.

As discussed in the proof of Theorem 1 in [42], we assume, without loss of generality, $p_n \to \infty$ and $p_n = q_n = o(\sqrt{n})$ as $n \to \infty$.

Let K_n and L_n be as in Lemma 2.7. Following the proof of Lemma 2.8, it is easy to see that

$$K_n \sim e^{-(p^2 q + p q^2)/2n} \tag{2.21}$$

as $n \to \infty$. Now we claim that

$$e^{-(p^2 q + p q^2)/2n} L_n \to 1 \tag{2.22}$$

in probability as $n \to \infty$. Recall the proof of Lemma 2.9, every step follows from the beginning to (2.18) with $x = y = 0$. We need to show that

$$R_n - \frac{p^2q + pq^2}{2n} \to 0$$

in probability, where R_n is as in (2.17). As checked before assertion (2.20), $ER_n = (p^2q + pq^2)/(2n)$. Therefore, it suffices to prove that

$$\frac{p+q}{n} h_1 \to 0 \ \text{ and } \ \frac{1}{n} h_2 \to 0 \tag{2.23}$$

in probability as $n \to \infty$, where $h_i = tr(X_n^* X_n)^i - E\left(tr(X_n^* X_n)^i\right)$ for $i = 1, 2$. Recall the arguments immediately after (2.20), we know that $\text{Var}(h_1) = pq$. Thus, $\text{Var}((p+q)h_1/n) = (p+q)^2 pq/n^2 \to 0$ since $p = q = o(\sqrt{n})$. Then the first assertion in (2.23) holds by Chebyshev's inequality. Finally, $\text{Var}(h_2/n) \sim 17p^4/n^2 \to 0$ by (i) of Lemma 2.4. The second conclusion in (2.23) follows.

Now by Lemma 2.7, $\|\mathcal{L}(U_n) - \mathcal{L}(X_n)\| = E|K_n L_n - 1|$. From (2.21) and (2.22), we know that $0 \le K_n L_n \to 1$ in probability. Also, $E(K_n L_n) = \int f_{\sqrt{n}U}(x)\, dx = 1$, where $f_{\sqrt{n}U}(x)$ is as in the proof of Lemma 2.7. These two facts imply $E|K_n L_n - 1| \to 0$ as $n \to \infty$. □

Proof of Corollary 1.1 We first prove the unitary case. Fix $n \ge 1$. By a perturbation theorem for singular values, see, e.g., Corollary 7.3.8 from [37],

$$\begin{aligned} \max_{1 \le i \le q} |\sqrt{\lambda_i(A^*A)} - \sqrt{\lambda_i(B^*B)}| &\le \max_{1 \le i \le q} \sqrt{\lambda_i((A-B)^*(A-B))} \\ &\le \sqrt{tr((A-B)^*(A-B))} \end{aligned} \tag{2.24}$$

for any $p \times q$ matrices A and B. This says that $(\lambda_1(A^*A), \ldots, \lambda_q(A^*A))$ is a continuous vector of A. Then there exists a Borel set $H \subset \mathbb{R}^{pq}$ such that

$$\begin{aligned} &\left\{f_n(\lambda_1(X_n^*X_n), \ldots, \lambda_1(X_n^*X_n)) \in F\right\} = \{X_n \in H\} \ \text{ and} \\ &\left\{f_n(\lambda_1(U_n^*U_n), \ldots, \lambda_1(U_n^*U_n)) \in F\right\} = \{U_n \in H\}. \end{aligned}$$

Theorem 1 says that $P(X_n \in H) - P(U_n \in H) \to 0$. The conclusion follows.

The same argument also applies to orthogonal case because of Theorem 10. □

3 Proofs of Theorems 3–7

Let F_1 and F_2 be two probability cumulative distribution functions (c.d.f.). Recall the Levy distance (see, e.g., p. 278 from [15])

$$L(F_1, F_2) = \inf\{\epsilon > 0;\ F_1(x - \epsilon) - \epsilon \le F_2(x) \le F_1(x + \epsilon) + \epsilon \text{ for all } x \in \mathbb{R}\}.$$

This distance characterizes the weak convergence of probability distributions, i.e., for a sequence of c.d.f.'s $\{F, F_n;\ n \geq 1\}$, the assertion $\int_{\mathbb{R}} g(x)\, dF_n(x) \to \int_{\mathbb{R}} g(x)\, dF(x)$ for every bounded continuous function g, is equivalent to that $L(F_n, F) \to 0$, which again is equivalent to that $F_n(x) \to F(x)$ for every continuous point x of F. See, e.g., Theorem 3 on p.278 from [15].

Lemma 3.1 *Let $F_n;\ n \geq 1$, be a sequence of c.d.f.'s, and F be a continuous and deterministic c.d.f. such that $L(F_n, F) \to 0$ in probability as $n \to \infty$. Then $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$ in probability.*

Proof If $L(F_n, F) < a$ for some $a > 0$, then $F(x-a) - a \leq F_n(x) \leq F(x+a) + a$ for all $x \in \mathbb{R}$. Since $L(F_n, F) \to 0$, for any $\epsilon > 0$ there exists $N \geq 1$ such that

$$P\left(F(x-\epsilon) - \epsilon \leq F_n(x) \leq F(x+\epsilon) + \epsilon \ \text{ for all } x \in \mathbb{R}\right) > 1 - \epsilon$$

for $n \geq N$. If the event in the above parenthesis holds, then $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq \sup_{|x-y|\leq\epsilon} |F(x) - F(y)| + \epsilon$. Since $F(x)$ is continuous, and $F(+\infty) = 1$ and $F(-\infty) = 0$, we know that $F(x)$ is uniformly continuous. Thus $\sup_{|x-y|\leq\delta} |F(x) - F(y)| < \epsilon$ as $\delta \in (0, \delta_0)$ for some $\delta_0 > 0$. Therefore, for any $\epsilon \in (0, \delta_0)$,

$$P\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| < 2\epsilon\right) > 1 - \epsilon$$

as $n \geq N$. This proves the lemma. □

Lemma 3.2 *Let X be an $n \times n$ matrix with complex entries, and $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Let μ_X be the empirical law of $\lambda_i,\ 1 \leq i \leq n$. Let μ be a probability measure. Then $\rho(\mu_X, \mu)$ is a continuous function in the entries of X, where ρ is as in* (1.7).

Proof For convenience of discussion, denote by $\lambda_i(X),\ 1 \leq i \leq n$, the eigenvalues of X. First, note that $\int_{\mathbb{R}} f(x)\, \mu_X(dx) = (1/n) \sum_{i=1}^n f(\lambda_i(X))$. Then by the triangle inequality, for any permutation π of $1, 2, \ldots, n$,

$$\begin{aligned}
|\rho(\mu_X, \mu) - \rho(\mu_Y, \mu)| &\leq \frac{1}{n} \sup_{\|f\|_L \leq 1} \left| \sum_{i=1}^n f(\lambda_{\pi(i)}(X)) - f(\lambda_i(Y)) \right| \\
&\leq \max_{1\leq i\leq n} \sup_{\|f\|_L \leq 1} |f(\lambda_{\pi(i)}(X)) - f(\lambda_i(Y))| \\
&\leq \max_{1\leq i\leq n} |\lambda_{\pi(i)}(X) - \lambda_i(Y)|,
\end{aligned}$$

where in the last step we use the Lipschitz property of $f:\ |f(x) - f(y)| \leq |x - y|$ for any x and y. Since the above inequality is true for any permutation π, we have that

$$|\rho(\mu_X, \mu) - \rho(\mu_Y, \mu)| \leq \min_{\pi} \max_{1\leq i\leq n} |\lambda_{\pi(i)}(X) - \lambda_i(Y)|.$$

By Theorem 2 from [9], which is a generalization of [57], we have that

$$\min_{\pi} \max_{1\le i\le n} |\lambda_{\pi(i)}(X) - \lambda_i(Y)| \le 2^{2-1/n}\|X - Y\|_2^{1/n}(\|X\|_2 + \|Y\|_2)^{1-1/n},$$

where $\|X\|_2$ is the operator norm of X. Let $X = (x_{ij})$ and $Y = (y_{ij})$. We know that $\|X\|_2 \le \sum_{1\le i,j\le n} |x_{ij}|^2$. Therefore

$$\begin{aligned}&|\rho(\mu_X, \mu) - \rho(\mu_Y, \mu)|\\ &\le 2(\|X\|_2 + \|Y\|_2)^{1-1/n} \cdot \left(\sum_{1\le i,j\le n} |x_{ij} - y_{ij}|^2\right)^{1/(2n)}.\end{aligned}$$

□

We will use the following results in later proofs.

Theorem 8 (Theorem 5 in [43]) *For each $n \ge 2$, there exists matrices $\Gamma_n = (\gamma_{ij})_{1\le i,j\le n}$ and $\mathbf{Y}_n = (y_{ij})_{1\le i,j\le n}$ whose $2n^2$ elements are random variables defined on the same probability space such that*

(i) *the law of Γ_n is the normalized Haar measure on the orthogonal group O_n;*
(ii) *$\{y_{ij}; 1 \le i, j \le n\}$ are i.i.d. random variables with the standard normal distribution;*
(iii) *set $\epsilon_n(m) = \max_{1\le i\le n, 1\le j\le m} |\sqrt{n}\gamma_{ij} - y_{ij}|$ for $m = 1, 2, \ldots, n$. Then*

$$\begin{aligned}P(\epsilon_n(m) \ge rs + 2t) &\le 4me^{-nr^2/16} + 3mn\\ &\times\left(\frac{1}{s}e^{-s^2/2} + \frac{1}{t}\left(1 + \frac{t^2}{3(m+\sqrt{n})}\right)^{-n/2}\right)\end{aligned}$$

for any $r \in (0, 1/4)$, $s > 0$, $t > 0$, and $m \le (r/2)n$.

Theorem 9 (Theorem 6 in [43]) *For each $n \ge 2$, there exists two $n \times n$ matrices $\Gamma_n = (\gamma_{jk})$ and $Y_n = ((x_{jk} + iy_{jk})/\sqrt{2})$ such that γ_{jk}'s, x_{jk}'s and y_{jk}'s are random variables defined on the same probability space, and*

(i) *the law of Γ_n is the normalized Haar measure on the unitary group $U(n)$;*
(ii) *the $2n^2$ random variables $\{x_{jk}, y_{jk}; 1 \le j, k \le n\}$ are independent standard normals;*
(iii) *set $\epsilon_n(m) = \max_{1\le j\le n, 1\le k\le m} |\sqrt{n}\gamma_{jk} - (x_{jk} + iy_{jk})/\sqrt{2}|$ for $m = 1, 2, \ldots, n$. Then*

$$P(\epsilon_n(m) \ge rs + 2t) \le 4me^{-nr^2/8} + mne^{-s^2} + \frac{6mn}{t}\left(1 + \frac{t^2}{12(m + t\sqrt{n})}\right)^{-n}$$

for any $r \in (0, 1/4)$, $s > 0$, $t > 0$, and $m \le (r/2)n$.

Theorem 10 (Theorem 1 in [42]) *Let* $\{p_n;\ n \geq 1\}$ *and* $\{q_n;\ n \geq 1\}$ *be two sequences of positive integers such that* $p_n \to +\infty$, $p_n = o(\sqrt{n})$ *and* $p_n/q_n \to c$ *for some constant* $c \in (0, \infty)$ *as* $n \to \infty$. *Let* U_n *be the upper-left* $p_n \times q_n$ *block of an* $n \times n$ *Haar invariant random orthogonal matrix. Let* X_n *be a* $p_n \times q_n$ *matrix whose* $p_n q_n$ *entries are i.i.d. standard real normal random variables. Then* $\lim_{n\to\infty} \|\mathcal{L}(\sqrt{n}U_n) - \mathcal{L}(X_n)\| = 0$.

The order that $p = o(\sqrt{n})$ is also proved to be the best in Jiang [42]. However, we do not need it in later applications.

Proof of Theorem 3 We only prove the unitary case, the orthogonal case is similar.

Let X_n be as in Theorem 1. For a $q \times q$ square matrix A, write $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_q(A)$ for all the eigenvalues of A. By Theorem 2.16 from [3] (see also [7] and [71]) that

$$\frac{\lambda_{\max}(X_n^* X_n)}{p} \to (1+\sqrt{c})^2 \quad \text{and} \quad \frac{\lambda_{\min}(X_n^* X_n)}{p} \to (1-\sqrt{c})^2 \tag{3.1}$$

in probability by the given condition that $q \to +\infty$ and $q/p \to c \in (0, +\infty)$. Thus

$$P\left(\left|\frac{\lambda_{\max}(X_n^* X_n)}{p} - (1+\sqrt{c})^2\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$ for any $\epsilon > 0$. Then the first assertion in (1.4) follows from Corollary 1.1. The second holds by the same argument since $f_n(\lambda_1(X_n^* X_n), \ldots, \lambda_q(X_n^* X_n)) = \lambda_{\min}(X_n^* X_n)$ as in (1.3) is a continuous function in the entries of $X_n^* X_n$ by (2.24). □

Proof of Theorem 4 We will only prove the unitary case, the orthogonal case is similar.

By Lemma 3.1, we only need to show that the Levy distance between $F_{p,q}$ and F goes to zero in probability, that is, $L(F_{p,q}, F) \to 0$ in probability as $n \to +\infty$.

Let Γ_n and Y_n be as in Theorem 9, and $Y_{p,q}$ be the upper-left sub-matrix of Y_n. Let $\tilde{F}$ be the empirical distribution of eigenvalues of $Y_{p,q}^* Y_{p,q}/p$. Apply the difference inequality Lemma 2.7 from Bai [3] for $A = \sqrt{n}U_{p,q}^*/\sqrt{p}$ and $B = Y_{p,q}^*/\sqrt{p}$ to obtain that

$$\begin{aligned} L^4(F_{p,q}, \tilde{F}_{p,q}) \leq \frac{2}{p^4} tr\left\{(\sqrt{n}U - Y_{p,q})^*(\sqrt{n}U - Y_{p,q})\right\} \\ \times \left\{tr(nU^*U) + tr(Y_{p,q}^* Y_{p,q})\right\}. \end{aligned}$$

Recalling the definition of $\epsilon_n(m)$ in Theorem 9, we have that

$$tr\left\{(\sqrt{n}U - Y_{p,q})^*(\sqrt{n}U - Y_{p,q})\right\} \leq pq\epsilon_n(q).$$

Also, $\|A\| := \sqrt{tr(A^*A)}$ is a norm, we obtain from the triangle inequality that

$$\begin{aligned} tr(nU^*U) = \|\sqrt{n}U\|^2 &\le (\|\sqrt{n}U - Y_{p,q}\| + \|Y_{p,q}\|)^2 \\ &\le 2pq\epsilon_n(q) + 2tr(Y_{p,q}^* Y_{p,q}). \end{aligned}$$

The above inequalities lead to

$$L^4(F_{p,q}, \tilde{F}_{p,q}) \le \frac{2p\epsilon_n(q)}{p} \cdot \left(\frac{2q\epsilon_n(q)}{p} + \frac{3}{p^2} \sum_{1\le i\le p, 1\le j\le q} (x_{ij}^2 + y_{ij}^2) \right).$$

By the Law of Large Numbers, the second term in the parenthesis above goes to $3c$ in probability as $n \to +\infty$. We have to show that

$$\epsilon_n(q) \to 0 \tag{3.2}$$

in probability as $q = o(n/\log n)$ and $n \to \infty$. First, this is true for the orthogonal case by Theorem 3 from [42]. Second, applying the same argument in the beginning of the proof of Theorem 3 from [42] to Theorem 9, (3.2) also holds for the unitary case. Therefore, $L(F_{p,q}, \tilde{F}_{p,q}) \to 0$ in probability. By the Machenko–Pastur law, see, e.g., Theorem 2.5 from [3] or [52], $L(\tilde{F}_{p,q}, F) \to 0$ in probability. Therefore, by the triangle inequality, $L(F_{p,q}, F) \to 0$ in probability as $n \to +\infty$. □

Proof of Theorem 5 We only prove the unitary case, the orthogonal case is exactly the same. Let X_p be a $p \times p$ matrix with entries being i.i.d. standard complex normals. By the circular law, see, e.g., Bai [4], $\rho(\nu_n, \mu) \to 0$ in probability, that is, $P(\rho(\nu_n, \mu) \ge \epsilon) \to 0$ for any $\epsilon > 0$, where ν_n is the empirical distribution of eigenvalues of $(1/\sqrt{p})X_p$. Now, by Theorem 1 and Lemma 3.2, $P(\rho(\mu_n, \mu) \ge \epsilon) - P(\rho(\nu_n, \mu) \ge \epsilon) \to 0$ for any $\epsilon > 0$. The conclusion follows. □

Proof of Theorem 6 (i) Let $f_n(x_1, \ldots, x_q) = \left(\sum_{i=1}^q f_1(x_i), \ldots, \sum_{i=1}^q f_k(x_i)\right) - c_n$. Note that p and q here correspond to N and n, respectively, in Theorem 1.1 from [6]. By Theorem 1.1 from [6], $\left(\sum_{i=1}^q f_1(\lambda_i), \ldots, \sum_{i=1}^q f_k(\lambda_i)\right) - c_n$ converges weakly to a k-dimensional normal distribution $N_k(\mu, \Sigma)$, for certain μ, Σ and c_n, where λ_i's are eigenvalues of a $q \times q$ matrix with i.i.d. entries of the standard complex normal distribution. Then Corollary 1.1 yields the result.

(ii) It is similar to (i). We omit it. □

Proof of Theorem 7 Johansson [46] showed that $(\lambda_1(X_n^* X_n) - \mu_{pq})/\sigma_{pq}$ converges to F_2 when the entries of X_n are i.i.d. standard complex Gaussian random variables. Johnstone [47] proved that the result is still true if F_2 is replaced by F_1 when the entries of X_n are i.i.d. standard real normals. Then the results follow from Corollary 1.1 by taking

$$f_n(x_1, \ldots, x_q) = \frac{\max_{1\le i\le q} x_i - \mu_{np}}{\sigma_{np}}, \quad x_1 \ge 0, \ldots, x_q \ge 0,$$

and $F = (-\infty, x]$ for any $x \in \mathbb{R}$. □

Acknowledgements The author thanks Professors Zhidong Bai and Peter Forrester very much for inspiring discussions and emails. He thanks Professor Amir Dembo for some constructive suggestions on this paper. He appreciates very much that the referees made many helpful suggestions.

References

1. Absil, P.A., Edelman, A., Koev, P.: On the largest principal angle between random subspaces. Preprint
2. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 2nd edn. Wiley, New York (1984)
3. Bai, Z.D.: Methodologies in spectral analysis of large dimensional random matrices, a review. Statist. Sin. **9**, 611–677 (1999)
4. Bai, Z.D.: Circular law. Ann. Probab. **25**, 494–529 (1997)
5. Bai, Z.D., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices. Science Press, Marricville (2006)
6. Bai, Z.D., Silverstein, J.W.: CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**, 553–605 (2004)
7. Bai, Z.D., Yin, Y.Q.: Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Ann. Probab. **21**, 1275–1294 (1993)
8. Bai, Z.D., Yin, Y.Q.: Convergence to the semicircle law. Ann. Probab. **16**, 863–875 (1988)
9. Bhatia, R., Elsner, L., Krause, G.: Bounds for variation of the roots of a polynomial and the eigenvalues of a matrix. Linear Algebra Appl. **142**, 195–209 (1990)
10. Baranger, H.U., Mello, P.A.: Mesoscopic transport through chaotic cavities: a random S-matrix theory approach. Phys. Rev. Lett. **73**, 142–145 (1994)
11. Beenakker, C.W.J.: Random-matrix theory of quantum transport. Rev. Modern Phys. **69**, 731–808 (1997)
12. Bohigas, O.: Random matrix theories and chaotic dynamics. In: Giannoni, M., Voros, A., Zinn-Justin, J. (eds.) Chaos and Quantum Physics, pp. 89–199. Elsevier, Amsterdam (1991)
13. Bryc, W., Dembo, A., Jiang, T.: Spectral measure of large random Hankel, Markov and Toeplitz matrices. Ann. Probab. **34**, 1–38 (2006)
14. Capitaine, M., Casalis, M.: Asymptotic Freeness by generalized moments for Gaussian and Wishart matrices. Application to Beta random matrices. Indiana Univ. Math. J. **53**, 397–431 (2004)
15. Chow, Y.S., Teicher, H.: Probability Theory, Independence, Interchangeability, Martingales, 3rd edn. Springer Texts in Statistics, Heidelberg (1997)
16. Collins, B.: Intégrales Matricielles et Probabilitiés Non-commutatives. Thèse de Doctorat of Université Paris 6 (2003)
17. Collins, B.: Product of random projections, Jacobi ensembles and universality problems arising from free probability. Probab. Theory Relat. Fields **133**, 315–344 (2005)
18. Conrey, J.B., Farmer, D.W., Mezzadri, F., Snaith, N.C.: Ranks of Elliptic Curves and Random Matrix Theory (London Mathematical Society Lecture Notes Series). Cambridge University Press, Cambridge (2007)
19. Constantine, A.G.: Some non-central distribution problems in multivariate analysis. Ann. Math. Stat. **34**, 1270–1285 (1963)
20. D'Aristotle, A., Diaconis, P., Newman, C.: Brownian Motion and the Classical Groups, Probability, Statisitca and their applications. In: Athreya, K. et al. (eds.) Papers in Honor of Rabii Bhattacharaya, pp. 97–116. Institute of Mathematical Statistics, Beechwood, OH (2003)
21. Diaconis, P.: Patterns in eigenvalues: the 70th Josiah Willard Gibbs Lecture. Bull. Am. Math. Soc. **40**, 155–178 (2003)
22. Diaconis, P., Eaton, M., Lauritzen, L.: Finite deFinetti theorem in linear models and multivariate analysis. Scand. J. Statist. **19**(4), 289–315 (1992)
23. Deift, P.: Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach (Courant Lecture Notes), American Mathematical Society, Providence (2000)
24. Dudley, R.M.: Real Analysis and Probability, 2nd edn. Cambridge University Press, Cambridge (2002)
25. Dumitriu, I., Edelman, A.: Matrix models for beta ensembles. J. Math. Phys. **43**, 5830–5847 (2002)
26. Dyson, F.J.: The three fold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics. J. Math. Phys. **3**, 1199–1215 (1962)

27. Eaton, M.: Multivariate Statistics: A Vector Space Approach (Wiley Series in Probability and Statistics). Wiley, New York (1983)
28. Edelman, A.: The circular law and the probability that a random matrix has k real eigenvalues. J. Multivar. Anal. **60**, 188–202 (1997)
29. Forrester, P.: Quantum conductance problems and the Jacobi ensemble. J. Phys. A: Math. Gen. **39**, 6861–6870 (2004)
30. Forrester, P.: Log-gases and Random matrices, Prepint. http://www.ms.unimelb.edu.au/~matpjf/matpjf.html (2007)
31. Geman, S.: The spectral radius of large random matrices. Ann. Probab. **14**, 1318–1328 (1986)
32. Girko, V.L.: Circle law. Theory Probab. Appl. **4**, 694–706 (1984a)
33. Girko, V.L.: On the circle law. Theory Probab. Math. Statist. **28**, 15–23 (1984b)
34. Grenander, U., Silverstein, J.W.: Spectral analysis of networks with random topologies. SIAM J. Appl. Math. **32**, 499–519 (1977)
35. Guhr, T., Müller-Groelling, A., Weidenmüller, H.: Random matrix theories in quantum physics: common concepts. Phys. Rep. **199**, 189–425 (1998)
36. Hiai, F., Petz, D.: Large deviations for functions of two random projection matrices. To appear “Acta Szeged”
37. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1990)
38. Hsu, P.L.: On the distribution of the roots of certain determinantal equations. Ann. Eugene **9**, 250–258 (1939)
39. Hwang, C.R.: A brief survey on the spectral radius and the spectral distribution of large dimensional random matrices with i.i.d. entries. Random Matrices and Their Applications, Contemporary Mathematics, vol. 50, pp. 145–152. AMS, Providence (1986)
40. James, A.T.: Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. **35**, 475–501 (1964)
41. Jiang, T.: A variance formula for quantum conductance. Preprint
42. Jiang, T.: How many entries of a typical orthogonal matrix can be approximated by independent normals?. Ann. Probab. **34**(4), 1497–1529 (2006)
43. Jiang, T.: Maxima of entries of Haar distributed matrices. Probab. Theory Relat. Fields **131**, 121–144 (2005)
44. Jiang, T.: The asymptotic distributions of the largest entries of sample correlation matrices. Ann. Appl. Probab. **14**(2), 865–880 (2004)
45. Jonsson, D.: Some limit theorem for eigenvalues of a sample covariance matrices. J. Multivar. Anal. **12**, 1–38 (1982)
46. Johansson, K.: Shape fluctation and random matrices. Comm. Math. Phys. **209**, 437–476 (2000)
47. Johnstone, I.: On the distribution of the largest eigenvalue in principal component analysis. Ann. Stat. **29**(2), 295–327 (2001)
48. Katz, N., Sarnak, P.: Random Matrices, Frobenius Eigenvalues, and Monodromy. AMS, Providence (1999)
49. Kilip, R.: Gaussian fluctuations for beta Ensembles. Preprint. http://xxx.lanl.gov/abs/math/0703140 (2007)
50. Koev, P., Dumitriu, I.: Distribution of the extreme eigenvalues of the complex Jacobi random matrix ensemble. To appear in Siam J. Matrix Anal. Appl. (2005)
51. Ledoux, M.: Differential operators and spectral distributions of invariant ensembles from the classical orthogonal polynomials. The continuous case. J. Theor. Probab. **9**, 177–208 (2004)
52. Marchenko, V.A., Pastur, L.A.: Distribution of some sets of random matrices. Math. USSR-Sb. **1**, 457–483 (1967)
53. Mezzadri, F., Snaith, N.C.: Recent Perspectives in Random Matrix Theory and Number Theory (London Mathematical Society Lecture Note Series). Cambridge University Press, Cambridge (2005)
54. Mehta, M.L.: Random Matrices, 2nd edn. Academic Press, Boston (1991)
55. Muirhead, R.J.: Aspects of Multivariate Statistical Theory (Wiley Series in Probability and Statistics), 2nd edn. Wiley-Interscience, New York (2005)
56. Petz, D., Réffy, J.: Large deviations for the empirical eigenvalue density of truncated Haar unitary matrices. Probab. Theory Relat. Fields **133**, 175–189 (2005)
57. Phillips, D.: Improving spectral-variation bound with Chebyshev polynomials. Linear Algebra Appl. **133**, 165–173 (1990)
58. Pichard, J.L., Jalabert, R.A., Beenakker, C.W.J.: Universal signature of chaos in ballistic transport. Europhys. Lett. **27**, 255–260 (1994)

59. Silverstein, J.W.: The smallest eigenvalue of a large dimensional Wishart matrix. Ann. Probab. **13**, 1364–1368 (1985)
60. Silverstein, J.W.: Comments on a result of Yin, Bai and Krishnaiah for large dimensional multivariate F matrices. J. Multivar. Anal. **15**, 408–409 (1984a)
61. Silverstein, J.W.: Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. J. Multivar. Anal. **15**, 295–324 (1984b)
62. Sinai, Y., Soshnikov, A.: Central limit theorems for traces of large symmetric matrices with independent matrix elements. Bol. Soc. Brasil Mat. (N.S.) **29**, 1–24 (1998)
63. Tracy, C.A., Widom, H.: Level-spacing distributions and Airy kernal. Comm. Math. Phys. **159**, 151–174 (1994)
64. Tracy, C.A., Widom, H.: On the orthogonal and symplectic matrix ensembles. Comm. Math. Phys. **177**, 727–754 (1996)
65. Tracy, C.A., Widom, H.: The distribution of largest eigenvalue in the Gaussian ensembles. In: van Diejen, J., Vinet, L. (eds.) Calogero-Moser-Sutherland Models, vol. 4 of CRM, Series in Mathematical Physics, pp. 461–472. Springer, Berlin (2000)
66. Tulino, A.M., Verdu, S.: Random Matrix Theory And Wireless Communications. Now Publishers, Hanover (2004)
67. Wachter, K.W.: The strong limits of random matrix spectra for sample matrices of independent elements. Ann. Probab. **6**, 1–18 (1978)
68. Wigner, E.P.: On the distribution of roots of certain symmetric matrices. Ann. Math. **67**, 325–328 (1958)
69. Yin, Y.Q.: LSD for a class of random matrices. J. Multivar. Anal. **20**, 50–68 (1986)
70. Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R.: Limiting behavior of the eigenvalues of multivariate F matrix. J. Multivar. Anal. **13**, 508–516 (1983)
71. Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R.: On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. Probab. Theory Relat. Fields **78**, 509–521 (1988)
72. Życzkowski, K., Sommers, H.J.: Truncation of random unitary matrices. J. Phys. A: Math. Gen. **33**, 2045–2057 (2000)