*Journal of Statistical Physics, Vol. 108, Nos. 5/6, September 2002 (© 2002)*

# A Note on Universality of the Distribution of the Largest Eigenvalues in Certain Sample Covariance Matrices

**Alexander Soshnikov**[1]

*Received April 20, 2001; revised June 7, 2001*

Recently Johansson[(21)] and Johnstone[(16)] proved that the distribution of the (properly rescaled) largest principal component of the complex (real) Wishart matrix $X^*X(X^tX)$ converges to the Tracy–Widom law as $n$, $p$ (the dimensions of $X$) tend to $\infty$ in some ratio $n/p \to \gamma > 0$. We extend these results in two directions. First of all, we prove that the joint distribution of the first, second, third, etc. eigenvalues of a Wishart matrix converges (after a proper rescaling) to the Tracy–Widom distribution. Second of all, we explain how the combinatorial machinery developed for Wigner random matrices in refs. 27, 38, and 39 allows to extend the results by Johansson and Johnstone to the case of $X$ with non-Gaussian entries, provided $n - p = O(p^{1/3})$. We also prove that $\lambda_{\max} \leqslant (n^{1/2} + p^{1/2})^2 + O(p^{1/2}\log(p))$ (a.e.) for general $\gamma > 0$.

**KEY WORDS:** Sample covariance matrices; principal component; Tracy–Widom distribution.

## 1. INTRODUCTION

Sample covariance matrices were introduced by statisticians about seventy years ago (refs. 1 and 2). There is a large literature on the subject (see, e.g., refs. 3–19). We start with the real case.

### 1.1. Real Sample Covariance Matrices

The ensemble consists of $p$-dimensional random matrices $A_p = X^tX$ ($X^t$ denotes a transpose matrix), where $X$ is an $n \times p$ matrix with independent real random entries $x_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$ such that

Dedicated to David Ruelle and Yakov Sinai on the occasion of their 65th birthdays.

[1] University of California, Department of Mathematics, One Shields Avenue, Davis, California 95616; e-mail: soshniko@math.ucdavis.edu

(i)

$$\mathbf{E}x_{ij} = 0, \tag{1.1}$$

$$\mathbf{E}(x_{ij})^2 = 1, \tag{1.2}$$

$$1 \leqslant i \leqslant n, \qquad 1 \leqslant j \leqslant p.$$

To prove the results of Theorems 2 and 3 later we will need some additional assumptions:

(ii) The random variables $x_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$, have symmetric laws of distribution.

(iii) All moments of these random variables are finite; in particular (ii) implies that all odd moments vanish.

(iv) The distributions of $x_{ij}$, decay at infinity at least as fast as a Gaussian distribution, namely

$$\mathbf{E}(x_{ij})^{2m} \leqslant (\text{const}\, m)^m. \tag{1.3}$$

Here and below we denote by const various positive real numbers that do not depend on $n, p, i, j$.

Complex sample covariance matrices are defined in a similar way.

## 1.2. Complex Sample Covariance Matrices

The ensemble consists of $p$-dimensional random matrices $A_p = X^*X$ ($X^*$ denotes a complex conjugate matrix), where $X$ is an $n \times p$ matrix with independent complex random entries $x_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$, such that

(i′)

$$\mathbf{E}x_{ij} = 0, \tag{1.4}$$

$$\mathbf{E}(x_{ij})^2 = 0, \tag{1.5}$$

$$\mathbf{E}\,|x_{ij}|^2 = 1, \tag{1.6}$$

$$1 \leqslant i \leqslant n, \qquad 1 \leqslant j \leqslant p.$$

The additional assumptions in the complex case mirror those from the real case:

(ii′) The random variables $\operatorname{Re} x_{ij}$, $\operatorname{Im} x_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$, have symmetric laws of distribution.

(iii′) All moments of these random variables are finite; in particular (ii′) implies that all odd moments vanish.

(iv′) The distributions of $\mathrm{Re}\, x_{ij}$, $\mathrm{Im}\, x_{ij}$ decay at infinity at least as fast as a Gaussian distribution, namely

$$\mathbf{E}\, |x_{ij}|^{2m} \leqslant (\text{const}\, m)^m. \tag{1.7}$$

**Remark 1.** The archetypical examples of sample covariance matrices is a $p$ variate Wishart distribution on $n$ degrees of freedom with identity covariance. It corresponds to

$$x_{ij} \sim N(0, 1), \qquad 1 \leqslant i \leqslant n, \qquad 1 \leqslant j \leqslant p, \tag{1.8}$$

in the real case, and

$$\mathrm{Re}\, x_{ij}, \mathrm{Im}\, x_{ij} \sim N(0, 1), \qquad 1 \leqslant i \leqslant n, \qquad 1 \leqslant j \leqslant p, \tag{1.9}$$

in the complex case.

It was proved in refs. 14, 17, and 19 that if (i) ((i′) in the real case) is satisfied,

$$n/p \to \gamma \geqslant 1, \qquad \text{as} \quad p \to \infty, \qquad \text{and} \qquad \mathbf{E}\, |x_{ij}|^{2+\delta} < \text{const} \tag{1.10}$$

then the empirical distribution function of the eigenvalues of $A_p/n$ converges to a non-random limit

$$G_{A_p/n}(x) = \frac{1}{p}\, \#\{\lambda_k^{(p)} \leqslant x, k = 1,..., n\} \to G(x) \quad \text{(a.s.)}. \tag{1.11}$$

where

$$\lambda_1^p \geqslant \lambda_2^p \geqslant \cdots \geqslant \lambda_p^p$$

are the eigenvalues (all real) of $A_p/n$, and $G(x)$ is defined by its density $g(x)$:

$$g(x) = \begin{cases} \frac{\gamma}{2\pi x} \sqrt{(b-x)(x-a)}, & a \leqslant x \leqslant b, \\ 0, & \text{otherwise}, \end{cases}$$

$$a = (1-\gamma^{-1/2})^2, \qquad b = (1+\gamma^{-1/2})^2.$$

Since the spectrum of $XX^*$ differs from the spectrum of $X^*X$ only by $(n-p)$ null eigenvalues, the limiting spectral distribution in the case

$0<\gamma<1$ remains the same, except for an atom of mass $1-\gamma$ at the origin. From now on we will always assume that $p \leqslant n$, however our results remain valid for $p>n$ as well.

The distribution of the largest eigenvalues attracts a special attention (see, e.g., ref. 16, Section 1.2). It was shown by Geman[12] in the i.i.d. case that if $\mathbf{E}\,|x_{ij}|^{6+\delta}<\infty$ the largest eigenvalue of $A_p/n$ converges to $(1+\gamma^{-1/2})^2$ almost surely. A few years later Yin, Bai, Krishnaiah and Silverstein[4, 18] showed (in the i.i.d. case) that the finiteness of the fourth moment is a necessary and sufficient condition for the almost sure convergence (see also ref. 20). These results state that $\lambda_{\max}(A_p)=(n^{1/2}+p^{1/2})^2+o(n+p)$. However no results were known about the rate of the convergence until recently Johansson[21] and Johnstone[16] proved the following theorem in the Gaussian (real and complex) cases.

**Theorem.** Suppose that a matrix $A_p=X^tX$ $(A_p=X^*X)$ has a real (complex) Wishart distribution (defined in Remark 1 above) and $n/p \to \gamma>0$. Then

$$\frac{\lambda_{\max}(A_p)-\mu_{n,p}}{\sigma_{n,p}}$$

where

$$\mu_{n,p}=(n^{1/2}+p^{1/2})^2, \tag{1.12}$$

$$\sigma_{n,p}=(n^{1/2}+p^{1/2})(n^{-1/2}+p^{-1/2})^{1/3} \tag{1.13}$$

converges in distribution to the Tracy–Widom law ($F_1$ in the real case, $F_2$ in the complex case).

**Remark 2.** Tracy–Widom distribution was discovered by Tracy and Widom in refs. 22 and 23. They found that the limiting distribution of the (properly rescaled) largest eigenvalue of a Gaussian symmetric (Gaussian Hermitian) matrix is given by $F_1(F_2)$, where

$$F_1(x)=\exp\left\{-\tfrac{1}{2}\int_x^{\infty} q(t)+(x-t)\,q^2(t)\,dt\right\}, \tag{1.14}$$

$$F_2(x)=\exp\left\{-\int_x^{\infty}(x-t)\,q^2(t)\,dt\right\}, \tag{1.15}$$

and $q(x)$ is such that it solves the Painlevé II differential equation

$$d^2q(x)/dx^2 = xq(x) + 2q^3(x) \tag{1.16}$$

$$q(x) \sim \mathrm{Ai}(x) \qquad \text{as} \quad x \to +\infty \tag{1.17}$$

where $\mathrm{Ai}(x)$ is the Airy function. Tracy and Widom also derived the expressions for the limiting distribution of the second largest, third largest, etc. eigenvalues as well. Since their discovery the field has exploded with a number of fascinating papers with applications to combinatorics, representation theory, probability, statistics, mathematical physics, in which Tracy–Widom law appears as a limiting distribution (for recent surveys we refer the reader to ref. 24 and 25).

**Remark 3.** It should be noted that Johansson studied the complex case and Johnstone did the real case. Johnstone also gave an alternative proof in the complex case. We also note that Johnstone has $n-1$ instead of $n$ in the center and scaling constants $\mu_{n,p}, \sigma_{n,p}$ in the real case. While this change clearly does not affect the limiting distribution of the largest eigenvalues, the choice of $n-1$ is more natural if one uses in the proof the asymptotics of Laguerre polynomials.

**Remark 4.** On a physical level of rigor the results similar to those from the Johansson–Johnstone Theorem (in the complex case) were derived by Forrester in ref. 10.

While it was not specifically pointed there, the results obtained in ref. 16 imply that the joint distribution of the first, second, third,..., $k$th, $k = 1, 2,\ldots$, largest eigenvalues converges (after the rescaling (1.12), (1.13)) to the limiting distribution derived by Tracy–Widom in refs. 22 and 23. In the complex case one can think about the limiting distribution as the distribution of the first $k$ (from the right) particles in the determinantal random point field with the correlation kernel given by the Airy kernel (2.8). We remind the reader that a random point field is called determinantal with a correlation kernel $S(x, y)$ if its correlation functions are given by

$$\rho_k(x_1,\ldots, x_k) = \det_{1 \leqslant i, j \leqslant k} S(x_i, x_j), \qquad k = 1, 2,\ldots \tag{1.18}$$

(for more information on determinantal random point field we refer the reader to ref. 26). In the real case the situation is slightly more complicated (correlation functions are given by the square roots of determinants, see Section 2, Lemma 1 and Remark 6). We claim the following result to be true:

**Theorem 1.** The joint distribution of the first, second, third, etc. largest eigevalues (rescaled as in (1.12), (1.13)) of a real (complex) Wishart matrix converges to the distribution given by the Tracy–Widom law (i.e., the limiting distribution of the first, second, etc. rescaled eigenvalues for GOE ($\beta=1$, real case) or GUE ($\beta=2$, complex case) correspondingly).

Theorem 1 is proved in Section 2. Our next result generalizes Theorem 1 to the non-Gaussian case, provided $n-p=O(n^{1/3})$.

**Theorem 2.** Let a real (complex) sample covariance matrix satisfy the conditions (i−iv)((i′−iv′)) and $n-p=O(p^{1/3})$. Then the joint distribution of the first, second, third, etc. largest eigenvalues (rescaled as in (1.12), (1.13)) converge to the Tracy–Widom law with $\beta=1(2)$.

Similar result for Wigner random matrices was proven in ref. 27. For other results on universality in random matrices we refer the reader to refs. 28–33.

While we expect the result of Theorem 2 to be true whenever $n/p\to\gamma>0$, we do not know at this moment how to extend our technique to the case of general $\gamma$. In this paper we settle for a weaker result.

**Theorem 3.** Let a real (complex) sample covariance matrix satisfy (i)–(iv) ((i′)–(iv′)) and $n/p\to\gamma>0$. Then

$$\text{(a)}\quad \mathbf{E}\,\mathrm{Trace}\,A_p^m=\frac{(\sqrt{\gamma}+1)\,\gamma^{1/4}}{2\sqrt{\pi}}\,\frac{p\mu_{n,p}^m}{m^{3/2}}\,(1+o(1))\qquad\text{if}\quad m=o(\sqrt{p}).$$

$$\text{(b)}\quad \mathbf{E}\,\mathrm{Trace}\,A_p^m=O\left(\frac{p\mu_{n,p}^m}{m^{3/2}}\right)\qquad\text{if}\quad m=O(\sqrt{p}).$$

As a corollary of Theorem 3 we have

**Corollary 1.**

$$\lambda_{\max}(A_p)\leqslant\mu_{n,p}+O(p^{1/2}\log(p))\quad\text{(a.e.)}.$$

We prove Theorem 1 in Section 2, Theorem 2 in Section 3 and Theorem 3 and Corollary 1 in Section 4.

The author would like to thank Craig Tracy for useful conversations.

## 2. WISHART DISTRIBUTION

The analysis in the Gaussian cases is simplified a great deal by the exact formulas for the joint distribution of the eigenvalues and the $k$-point correlation functions, $k = 1, 2, \ldots$ . In the complex case the density of the joint distribution of the eigenvalues is given by:[15]

$$P_p(x_1, \ldots, x_p) = c_{n,p}^{-1} \prod_{1 \leqslant i < j \leqslant p} (x_i - x_j)^2 \prod_{j=1}^{p} x_j^{\alpha_p} \exp(-x_j), \qquad \alpha_p = n - p, \tag{2.1}$$

where $c_{n,p}$ is a normalization constant. Using a standard argument from Random Matrix Theory[34] one can rewrite $P_p(x_1, \ldots, x_p)$ as

$$\frac{1}{p!} \det_{1 \leqslant i, j \leqslant p} S_p(x_i, x_j) \tag{2.2}$$

where

$$S_p(x, y) = \sum_{j=0}^{p-1} \varphi_j^{(\alpha_p)}(x)\, \varphi_j^{(\alpha_p)}(y) \tag{2.3}$$

is the reproducing (Christoffel–Darboux) kernel of the Laguerre orthonormalized system

$$\varphi_j^{(\alpha_p)}(x) = \sqrt{\frac{j!}{(j+\alpha_p)!}}\, x^{\alpha_p/2} \exp(-x/2)\, L_j^{\alpha_p}(x), \tag{2.4}$$

and $L_j^{\alpha_p}$ are the Laguerre polynomials.[35] This allows one to write the $k$-point correlation functions as

$$\rho_k^{(p)}(x_1, \ldots, x_k) = \det_{1 \leqslant i, j \leqslant k} S_p(x_i, x_j), \qquad k = 1, 2, \ldots, p \tag{2.5}$$

(for more information on correlation functions we refer the reader to refs. 18, 34, and 36). As a by-product of the results in ref. 16 Johnstone showed that after the rescaling

$$x = \mu_{n,p} + \sigma_{n,p} s \tag{2.7}$$

the (rescaled) kernel

$$\sigma_{n,p} S_p(\mu_{n,p} + \sigma_{n,p} s_1, \mu_{n,p} + \sigma_{n,p} s_2) \tag{2.7}$$

converges to the Airy kernel

$$S(s_1, s_2) = \frac{A(s_1) \cdot A'(s_2) - A'(s_1) \cdot A(s_2)}{s_1 - s_2} = \int_0^{+\infty} \mathrm{Ai}(s_1 + t)\, \mathrm{Ai}(s_2 + t)\, dt. \tag{2.8}$$

The convergence is pointwise and also in the trace norm on any $(t, \infty)$, $t \in R^1$.

In the real Wishart case the formula for the joint distribution of the eigenvalues was independently discovered by several groups of statisticians at the end of thirties (see refs. 1 and 2):

$$P_p(x_1,..., x_p) = \text{const}_{n,p}^{-1} \prod_{1 \leqslant i<j \leqslant p} |x_i - x_j| \prod_{j=1}^{p} x_j^{\alpha_p/2} \exp(-x_j/2), \quad \alpha_p = n-1-p. \tag{2.9}$$

(note that in the real case $\alpha_p = n-1-p$, while in the complex case it was $n-p$.) The $k$-point correlation function has a form similar to (2.2), (2.3) however it is now equal to a square root of the determinant, and $K_p(x, y)$ is a $2 \times 2$ matrix kernel (see, e.g., refs. 16 and 37):

$$\rho_k^{(p)}(x_1,..., x_k) = (\det_{1 \leqslant i, j \leqslant k} K_p(x_i, x_j))^{1/2}, \qquad k = 1,..., p, \tag{2.10}$$

where (in the even $p$ case)

$$K_p^{(1,1)}(x, y) = S_p(x, y) + \psi(x)(\epsilon\phi)(y) \tag{2.11}$$

$$K_p^{(1,2)}(x, y) = (S_p D)(x, y) - \psi(x)\, \phi(y) \tag{2.12}$$

$$K_p^{(2,1)}(x, y) = (\epsilon S_p)(x, y) - \epsilon(x-y) + (\epsilon\psi)(x)(\epsilon\phi)(y) \tag{2.13}$$

$$K_p^{(2,2)}(x, y) = K_p^{(1,1)}(y, x), \tag{2.14}$$

operator $\epsilon$ denotes convolution with the kernel

$$\epsilon(x-y) = \frac{1}{2}\,\text{sign}(x-y), \qquad (SD)(x, y) = -\frac{\partial S(x, y)}{\partial y},$$

and $\psi(x)$, $\phi(x)$ are defined as follows

$$\psi(x) = (-1)^p \frac{(p(p+\alpha_p))^{1/4}}{2^{1/2}} (\sqrt{p+\alpha_p}\, \xi_p(x) - \sqrt{p}\, \xi_{p-1}(x)) \tag{2.15}$$

$$\phi(x) = (-1)^p \frac{(p(p+\alpha_p))^{1/4}}{2^{1/2}} (\sqrt{p}\, \xi_p(x) - \sqrt{p+\alpha_p}\, \xi_{p-1}(x)) \tag{2.16}$$

$$\xi_k(x) = \varphi_k^{(\alpha_p)}(x)/x. \tag{2.17}$$

**Remark 5.** The formulas for $K_p(x, y)$ in the odd $p$ case are slightly different. However since we are interested in the asymptotic behavior of the largest eigenvalues it is enough to consider only even $p$ case. Indeed, one can carry very similar calculations in the odd $p$ case and obtain the same limiting kernel $K(x, y)$ as we got in Lemma 1. Or one may observe that the limiting distribution of the largest (rescaled) eigenvalues must be the same in the even $p$ and odd $p$ cases as implied by the following argument. Consider an $(n+p)\times(n+p)$ real symmetric (self-adjoint) matrix $B=(b_{ij})$, $1\leqslant i, j\leqslant n+p$,

$$b_{ij}=\begin{cases} x_{i,\,j-n}, & \text{if} \quad 1\leqslant i\leqslant n, \quad n+1\leqslant j\leqslant n+p \\ \bar{x}_{j,\,i-p}, & \text{if} \quad p+1\leqslant i\leqslant n+p, \quad 1\leqslant j\leqslant p \\ 0, & \text{otherwise,} \end{cases}$$

Then the non-zero eigenvalues of $B^2$ and $X^*X$ coincide. If we now consider a matrix $\tilde{X}$ obtained by deleting the first row and the last column of $X$ and construct the corresponding matrix $\tilde{B}$, then by the mini-max principle we have $\lambda_k(B)\geqslant\lambda_k(\tilde{B})$, $k=1, 2,\dots$. Repeating this procedure once more we see that the $k$th eigenvalue of $X^*X$ for odd $p$ is sandwiched between the $k$th eigenvalues for $p+1$ and $p-1$.

The machinery developed in ref. 16 allows us to obtain the following result about the pointwise convergence of the entries of $K_p(x, y)$.

**Lemma 1.**

(a) $$\sigma_{n,p}K_p^{(1,1)}(\mu_{n,p}+\sigma_{n,p}s_1,\, \mu_{n,p}+\sigma_{n,p}s_2)$$
$$\to S(s_1, s_2)+\frac{1}{2}\operatorname{Ai}(s_1)\int_{-\infty}^{s_2}\operatorname{Ai}(t)\,dt, \tag{2.18}$$

$$\sigma_{n,p}K_p^{(2,2)}(\mu_{n,p}+\sigma_{n,p}s_1,\, \mu_{n,p}+\sigma_{n,p}s_2)$$
$$\to S(s_2, s_1)+\frac{1}{2}\operatorname{Ai}(s_2)\int_{-\infty}^{s_1}\operatorname{Ai}(t)\,dt, \tag{2.19}$$

$$\sigma_{n,p}^2K_p^{(1,2)}(\mu_{n,p}+\sigma_{n,p}s_1,\, \mu_{n,p}+\sigma_{n,p}s_2)$$
$$\to -\frac{1}{2}\operatorname{Ai}(s_1)\operatorname{Ai}(s_2)-\frac{\partial}{\partial s_2}S(s_1, s_2), \tag{2.20}$$

$$K_p^{(2,1)}(\mu_{n,p}+\sigma_{n,p}s_1,\, \mu_{n,p}+\sigma_{n,p}s_2)$$
$$\to -\int_0^{+\infty}du\left(\int_{s_1+u}^{+\infty}\operatorname{Ai}(v)\,dv\right)\operatorname{Ai}(s_2+u)$$
$$-\epsilon(x-y)+\frac{1}{2}\int_{s_2}^{s_1}\operatorname{Ai}(u)\,du+\frac{1}{2}\int_{s_1}^{+\infty}\operatorname{Ai}(u)\,du\int_{-\infty}^{s_2}\operatorname{Ai}(v)\,dv. \tag{2.21}$$

(b) Convergence in (2.18)–(2.21) is uniform on $[\tilde{s}_1, +\infty) \times [\tilde{s}_2, +\infty)$ as $p \to \infty$ for any $\tilde{s}_1 > -\infty$, $\tilde{s}_2 > -\infty$. It is also true that the error terms are $O(e^{-\text{const}(s_1+s_2)})$ uniformly in $p$ with some const $> 0$.

**Remark 6.** Lemma 1 implies the convergence of the rescaled $k$-point correlation functions $\sigma_{n,p}^k \rho_k^{(p)}(x_1,..., x_k)$, $x_i = \mu_{n,p} + \sigma_{n,p} s_i$, $i = 1,..., k$, $k = 1, 2,...$, to

$$\rho_k(s_1,..., s_k) = (\det_{1 \leqslant i,j \leqslant k} K(s_i, s_j))^{1/2},$$

where the entries of $K(s, t) = (K_{ij}(s, t))_{i,j=1,2}$ are given by the r.h.s. of (2.18)–(2.21). The limiting correlation functions coincide with the limiting correlation functions at the edge of the spectrum in the Gaussian Orthogonal Ensemble (see, e.g., ref. 11) (it also should be noted that the formulas (1.15) and (1.16) we gave in ref. 27 for $K(s, t)$ must be replaced by (2.18)–(2.21)).

*Proof of Lemma 1.* The proof is a consequence of (2.11)–(2.14), (1.12) and (1.13) and the asymptotic formulas for the Laguerre polynomials $L_j^{\alpha_p}(x)$, $\alpha_p \to \infty$, $j \sim \alpha_p$, near the turning point derived in ref. 16. Later we prove (2.18) and (2.21). (2.19) immediately follows from (2.14) and (2.18). (2.20) is established in a similar way to (2.18), (2.21). To prove (2.18) we employ a very useful integral representation for $S_p(x, y)$:[37]

$$S_p(x, y) = \int_0^{+\infty} \phi(x+z)\, \psi(y+z) + \psi(x+z)\, \phi(y+z)\, dz, \tag{2.22}$$

where $\phi(x)$, $\psi(x)$ are defined in (2.15)–(2.17).

The asymptotic behavior of $\phi(x)$, $\psi(x)$ was studied by Johnstone[16] who proved

$$\sigma_{n,p}\phi(\mu_{n,p} + \sigma_{n,p}s),\ \sigma_{n,p}\psi(\mu_{n,p} + \sigma_{n,p}s) \to \frac{1}{\sqrt{2}} \mathrm{Ai}(s) \tag{2.23}$$

and that the l.h.s. at (2.24) is exponentially small for large $s_1, s_2$ (uniformly in $p$.) While Johnstone stated only pointwise convergence in (2.22) his results (see (3.7), (5.1), (5.19), (5.18), (5.22)–(5.24) and (6.11) from ref. 16) actually imply that the convergence is uniform on any $[s, +\infty)$. This together with (2.22) gives us

$$\sigma_{n,p} S_p(\mu_{n,p} + \sigma_{n,p}s_1,\ \mu_{n,p} + \sigma_{n,p}s_2) \to S(s_1, s_2), \tag{2.24}$$

where the convergence is uniform on any $[\tilde{s}_1, \infty) \times [\tilde{s}_2, \infty)$. To deal with the second term at the r.h.s. of (2.11),

$$\psi(x)(\epsilon\phi)(y) = \psi(x)\left(\tfrac{1}{2}\int_0^\infty \phi(u)\,du - \int_y^\infty \phi(u)\,du\right), \tag{2.25}$$

we use

$$\frac{1}{2}\int_0^\infty \phi(u)\,du \to \frac{1}{\sqrt{2}} \tag{2.26}$$

(see ref. 16, Appendix A7). (2.23), (2.25) and (2.26) imply

$$\sigma_{n,p}\psi(\mu_{n,p}+\sigma_{n,p}s_1)(\epsilon\phi)(\mu_{n,p}+\sigma_{n,p}s_2) \to \tfrac{1}{2}\,\mathrm{Ai}(s_1)\int_{-\infty}^{s_2} \mathrm{Ai}(t)\,dt. \tag{2.27}$$

This proves (2.18).

To establish (2.21) we consider separately $(\epsilon S_p)(x, y)$ and $(\epsilon\psi)(x)$ $(\epsilon\phi)(y)$. We have

$$\epsilon S_p(x, y) = \left(\tfrac{1}{2}\int_0^{+\infty} du - \int_x^{+\infty} du\right)\int_0^{+\infty} \phi(u+z)\,\psi(y+z)$$
$$+\psi(u+z)\,\phi(y+z)\,dz \tag{2.28}$$
$$= \tfrac{1}{2}\int_0^{+\infty}\left(\int_z^{+\infty} \phi(u)\,du\,\psi(y+z)\right)dz \tag{2.29}$$
$$-\int_0^{+\infty}\left(\int_{x+z}^{+\infty} \phi(u)\,du\,\psi(y+z)\right)dz \tag{2.30}$$
$$+\tfrac{1}{2}\int_0^{+\infty}\left(\int_z^{+\infty} \psi(u)\,du\,\phi(y+z)\right)dz \tag{2.31}$$
$$-\int_0^{+\infty}\left(\int_{x+z}^{+\infty} \psi(u)\,du\,\phi(y+z)\right)dz \tag{2.32}$$

Let us fix $s_1, s_2$ and consider

$$x = \mu_{n,p}+\sigma_{n,p}s_1, \qquad y = \mu_{n,p}+\sigma_{n,p}s_2. \tag{2.33}$$

It follows from (2.23) that the integrals (2.30) and (2.32) converge to

$$-\tfrac{1}{2}\int_0^{+\infty} du\left(\int_{s_1+u}^{+\infty} \mathrm{Ai}(v)\,dv\right)\mathrm{Ai}(s_2+u).$$

Let us now write (2.29) as

$$\frac{1}{2}\int_0^{+\infty}\left(\int_0^{+\infty}\phi(u)\,du\,\psi(y+z)\right)dz \tag{2.34}$$

$$-\frac{1}{2}\int_0^{+\infty}\left(\int_0^{z}\phi(u)\,du\,\psi(y+z)\right)dz \tag{2.35}$$

$$=\frac{1}{\sqrt{2}}\int_0^{+\infty}\psi(y+z)\,dz \tag{2.36}$$

$$-\frac{1}{2}\int_0^{+\infty}\left(\int_0^{z}\phi(u)\,du\,\psi(y+z)\right)dz \tag{2.37}$$

Using (2.23) one can see that (2.36) converges to $\frac{1}{2}\int_{s_2}^{\infty}\mathrm{Ai}(u)\,du$.

The integral (2.37) tends to zero as $p\to\infty$. Indeed, suppose that $n-p\to+\infty$ [the case $n-p=O(1)$ can be treated by using the classical asymptotic formulas for Laguerre polynomials for fixed $\alpha$ (see, e.g., ref. 35)]. Let us write $\int_0^{+\infty}(\int_0^z\phi(u)\,du\,\psi(y+z))\,dz$ as

$$\int_0^{\sqrt{p}}\left(\int_0^{z}\phi(u)\,du\,\psi(y+z)\right)dz+\int_{\sqrt{p}}^{\infty}\left(\int_0^{z}\phi(u)\,du\,\psi(y+z)\right)dz. \tag{2.38}$$

Similar calculations to the ones from Appendix 7 of ref. 16 show that for $z<\sqrt{p}$

$$\int_0^z\phi(u)\,du=O((\mathrm{const}\;p)^{-(n-p)/4}),\qquad \text{where}\quad \mathrm{const}>0.$$

This estimate coupled with the following (rather rough) bounds

$$\begin{aligned}\int_0^{\sqrt{p}}|\psi(y+z)|\,dz &\leqslant p^{1/4}\left(\int_y^{\infty}\psi(z)^2\,dz\right)^{1/2}\\ &\leqslant \mathrm{const}\;p^{1/4}\left(\left(\int_y^{\infty}\varphi_p^{\alpha_p}(z)^2\,dz\right)^{1/2}+\left(\int_y^{\infty}\varphi_{p-1}^{\alpha_p}(z)^2\,dz\right)^{1/2}\right)\\ &=O(p^{1/4})\end{aligned}$$

take care of the first term in (2.38). If $z\geqslant\sqrt{p}$ one has

$$\begin{aligned}|\psi(y+z)|&=|\psi(\mu_{n,p}+\sigma_{n,p}(s_2+z/\sigma_{n,p}))|\\ &<\exp(-\mathrm{const}(s_2+z/p^{1/3})),\quad \mathrm{const}>0,\end{aligned}$$

where we have used the exponential decay of $\psi(\mu_{n,p}+\sigma_{n,p}s)$ for large $s$ (see (2.23), (2.15)–(2.17) and ref. 16, formula (5.1)). Since

$$\left|\int_0^z \phi(u)\,du\right| \leqslant \sqrt{z}\left(\int_0^z \phi(u)^2\,du\right)^{1/2} \leqslant \text{const}\; p\sqrt{z},$$

we conclude that (2.37) is o(1). Using $\int_0^\infty \psi(u)\,du=0$ one can prove in a similar fashion that (2.31) is also o(1). To establish (2.21) we are left with estimating

$$(\epsilon\psi)(x)(\epsilon\phi)(y)=\left(\frac{1}{2}\int_0^\infty \psi(u)\,du-\int_x^\infty \psi(u)\,du\right)\left(\frac{1}{2}\int_0^\infty \phi(u)\,du-\int_y^\infty \phi(v)\,dv\right)$$

$$=\left(-\int_x^\infty \psi(u)\,du\right)\left(\frac{1}{\sqrt{2}}+o(1)-\int_y^\infty \phi(v)\,dv\right).$$

Using (2.23) and (2.33) we derive that the last expression converges to $-\frac{1}{2}\int_{s_1}^\infty \mathrm{Ai}(u)\,du+\frac{1}{2}\int_{s_1}^{+\infty} \mathrm{Ai}(u)\,du\int_{-\infty}^{s_2} \mathrm{Ai}(v)\,dv$. This finishes the proof of (2.21). To obtain (2.20) we use (2.23) and

$$\sigma_{n,p}^2\phi'(\mu_{n,p}+\sigma_{n,p}s),\qquad \sigma_{n,p}^2\psi'(\mu_{n,p}+\sigma_{n,p}s)\to\frac{1}{\sqrt{2}}\mathrm{Ai}'(s). \tag{2.39}$$

which follows from the machinery developed in ref. 16. Lemma 1 is proven.

Theorem 1 now follows from

**Lemma 2.** Suppose that we are given random point fields $\mathbf{F}, \mathbf{F_n}$, $n=1,2,\dots$ with the $k$-point correlation functions $\rho_k(x_1,\dots,x_k)$, $\rho_k^{(n)}(x_1,\dots,x_k)$ $k=1,2,\dots$ such that the number of particles in $(a,\infty)$ (denoted by $\#(a,\infty)$) is finite $\mathbf{F}-$, a.e., for any $a>-\infty$, $\mathbf{F}$ is uniquely determined by its correlation functions and the distribution of the numbers of particles in the finite intervals (w.r.t. $F$) is uniquely determined by the moments. Then the following diagram holds:

$$(d)\Rightarrow(c)\Rightarrow(b)\Leftrightarrow(a),$$

where

(a) The joint distribution of the first, second,..., $k$th rightmost particles in $\mathbf{F_n}$ converges to the joint distribution of the first, second,..., $k$th rightmost particles in $\mathbf{F}$ for any $k\geqslant 1$.

(b) The joint distribution of $\#(a_1,b_1),\dots,\#(a_l,b_l)$, $l\geqslant 1$ in $\mathbf{F_n}$ converges to the corresponding distribution in $\mathbf{F}$ for any collection of disjoint intervals $(a_1,b_1),\dots,(a_l,b_l)$, $a_j>-\infty$, $b_j\leqslant+\infty$, $j=1,\dots,l$, $l=1,\dots$ .

(c) $\rho_k(x_1,\ldots, x_k)$ is integrable on $[t, \infty)^k$ for any $t \in R^1$, $k = 1, 2,\ldots$ and

$$\int_{(a_1, b_1)^{k_1} \times \cdots \times (a_l, b_l)^{k_l}} \rho_k^{(n)}(x_1,\ldots, x_k)\, dx_1 \cdots dx_k \tag{2.40}$$

$$\to \int_{(a_1, b_1)^{k_1} \times \cdots \times (a_l, b_l)^{k_l}} \rho_k(x_1,\ldots, x_k)\, dx_1 \cdots dx_k \tag{2.41}$$

for any disjoint intervals $(a_1, b_1),\ldots, (a_l, b_l)$, $a_j > -\infty$, $b_j \leqslant +\infty$, $j = 1,\ldots, l$, $l = 1,\ldots, k$, $k_1 + \cdots + k_l = k$, $k = 1, 2,\ldots$.

(d) For any $k \geqslant 1$ the Laplace transform

$$\int \exp\left(\sum_{j = 1,\ldots, k} t_j x_j\right) \rho_k(x_1,\ldots, x_k)\, dx_1 \cdots dx_k$$

is finite for $t_1 \in [c_1^{(k)}, d_1^{(k)}],\ldots, t_k \in [c_k^{(k)}, d_k^{(k)}]$, where $c_j^{(k)} < d_j^{(k)}$, $d_j^{(k)} > 0$, $j = 1,\ldots, k$, and

$$\int \exp\left(\sum_{j = 1,\ldots, k} t_j x_j\right) \rho_k^{(n)}(x_1,\ldots, x_k)\, dx_1 \cdots dx_k \tag{2.42}$$

$$\to \int \exp\left(\sum_{j = 1,\ldots, k} t_j x_j\right) \rho_k(x_1,\ldots, x_k)\, dx_1 \cdots dx_k \tag{2.43}$$

for such $t_1,\ldots, t_k$ as $n \to \infty$.

*Proof of Lemma 2.* (d) $\Rightarrow$ (c)

Suppose that (d) holds. Fix some positive $\tilde{t}_1 \in (c_1^{(k)}, d_1^{(k)}),\ldots, \tilde{t}_k \in (c_k^{(k)}, d_k^{(k)})$. Denote by $H_n(dx_1,\ldots, dx_k)$, $H(dx_1,\ldots, dx_k)$, the probability measures on $R^k$ with the densities

$$h_n(x_1,\ldots, x_k) = Z_n^{-1} \exp\left(\sum_{j = 1,\ldots, k} \tilde{t}_j x_j\right) \rho_k^{(n)}(x_1,\ldots, x_k),$$

$$h(x_1,\ldots, x_k) = Z^{-1} \exp\left(\sum_{j = 1,\ldots, k} \tilde{t}_j x_j\right) \rho_k(x_1,\ldots, x_k),$$

where $Z_n$, $Z$ are the normalization constants (it is easy to see that $Z_n \to Z$.). The constructed sequence of probability measures is tight (by Helly theorem), moreover their distributions decay (at least) exponentially for large (positive and negative) $x_1,\ldots, x_k$ uniformly in $n$. It follows from the tightness of $\{H_n\}$ that all we have to show is that any limiting point of $H_n$

coincides with $H$. Suppose that a subsequence of $H_n$ weakly converges to $\bar{H}$. Then $\bar{H}$ must have a finite Laplace transform for $c_1^{(k)}-\tilde{t}_1 \leqslant \operatorname{Re} t_1 \leqslant d_1^{(k)}-\tilde{t}_1,\dots, c_k^{(k)}-\tilde{t}_k \leqslant \operatorname{Re} t_k \leqslant d_k^{(k)}-\tilde{t}_k$ and the Laplace transforms of $H_n$ must converge to the Laplace transforms of $\bar{H}$ in this strip. Since the Laplace transforms of $\bar{H}, H$ are analytic in the strip and coincide for $t_1 \in [c_1^{(k)}, d_1^{(k)}],\dots, t_k \in [c_k^{(k)}, d_k^{(k)}]$ they must coincide in the whole strip. Applying the inverse Laplace transform we obtain that $\bar{H}$ coincides with $H$. It follows then that

$$\int_{(a_1, b_1)\times \cdots \times (a_k, b_k)} \rho_k^{(n)}(x_1,\dots, x_k)\, dx_1 \cdots dx_k$$

$$\to \int_{(a_1, b_1)\times \cdots \times (a_k, b_k)} \rho_k(x_1,\dots, x_k)\, dx_1 \cdots dx_k$$

for any finite $a_j < b_j$, $j=1,\dots, k$, and the exponential decay of $\rho_k^{(n)}(x_1,\dots, x_k)$, $\rho_k(x_1,\dots, x_k)$, for large positive $x_1,\dots, x_k$, implies that this still holds for $b_j = +\infty$, $j=1,\dots, k$.

(c) $\Rightarrow$ (b) We remind the reader that the integral in (2.40) is equal to the $(k_1,\dots, k_l)$th factorial moment

$$\mathbf{E} \prod_{j=1,\dots,l} \frac{(\#(a_j, b_j))!}{(\#(a_j, b_j)-k_j)!}$$

of the numbers of particles in the disjoint intervals $(a_1, b_1),\dots, (a_l, b_l)$. Since the joint distribution of the numbers of particles in the boxes is uniquely determined by the moments, the convergence of moments implies the convergence of the distributions of $\#(a_1, b_1),\dots, \#(a_l, b_l)$.

(b) $\Leftrightarrow$ (a) Trivial. Observe that

$$P(\lambda_1 \leqslant s_1, \lambda_2 \leqslant s_2,\dots, \lambda_k \leqslant s_k)$$
$$= P(\#(s_1, +\infty)=0, \quad \#(s_2, +\infty) \leqslant 1,\dots, \quad \#(s_k, +\infty) \leqslant k-1).$$

Lemma 2 is proven.

*Proof of Theorem 1.* It is worth noting that the limiting random point fields defined in (1.18), (2.8) (complex case) and in Remark 6 (real case) are uniquely determined by their correlation functions (see, e.g., ref. 26). Indeed, $\sum_{k=0}^{\infty} \left(\frac{1}{k!}\int_{A_k} \rho_k(x_1,\dots, x_k)\, dx_1 \cdots dx_k\right)^{1/k} = \infty$ for any finite interval $A$. Similarly, the distribution of the number of particles in any finite interval is uniquely determined by the moments. To prove Theorem 1 in the complex case we use a general fact for the ensembles with determinantal correlation functions that the generating function of the numbers of particles in the boxes is given by the Fredholm determinant

$$\mathbf{E} \prod_{j=1,\dots,l} z_j^{\#(a_j,b_j)} = \det\left(\mathrm{Id} + \sum_{j=1,\dots,k} (z_j - 1)\, S_p \chi_{(a_j,b_j)}\right) \tag{2.44}$$

(see, e.g., refs. 26 and 36), where $\chi_{(a,b)}$ is the operator of the multiplication by the indicator of $(a,b)$. Trace class convergence of $S_p$ to $K$ on any $(a,\infty)$, $a > -\infty$, implies the convergence of the Fredholm determinants, which together with Lemma 2 proves Theorem 1 in the complex case. To prove Theorem 1 in the real case we observe that Lemma 1 implies that after rescaling $x_i = \mu_{n,p} + \sigma_{n,p} s_i$, $i = 1, 2,\dots$ condition (2.40) and (2.41) of Lemma 2, part (c) is satisfied. Theorem 1 is proven.

## 3. PROOF OF THEOREM 2

The proof of Theorem 2 heavily relies on the results obtained in refs. 27, 38, and 39. We start with

**Lemma 3.** Let $A_p$ be either a real sample covariance matrix (i)–(iv) or complex sample covariance matrix ((i′)–(iv′)) and $n-p = O(p^{1/3})$ as $p \to \infty$. Then there exists some const $> 0$ such that for any $t_1, t_2,\dots,t_k > 0$ and

$$m_p^{(1)} = [t_1 \cdot p^{\frac{2}{3}}],\dots, m_p^{(k)} = [t_k \cdot p^{\frac{2}{3}}],$$

the following estimate holds:

(a)

$$\mathbf{E} \prod_{i=1}^{k} \mathrm{Trace}\, A_p^{m_p^{(i)}} \leqslant \mathrm{const}^k \prod_{i=1}^{k} \frac{\mu_{n,p}^{m_p^{(i)}}}{t_i^{3k/2}} \exp\left(\mathrm{const} \sum_{i=1}^{k} t_i^3\right) \tag{3.1}$$

(b) If $A_p, \tilde{A}_p$ belong to two different ensembles of random real (complex) sample covariance matrices satisfying (i)–(iv) ((i′)–(iv′)), and $n-p = O(p^{1/3})$, then

$$\mathbf{E} \prod_{i=1}^{k} \mathrm{Trace}\, A_p^{m_p^{(i)}} - \mathbf{E} \prod_{i=1}^{k} \mathrm{Trace}\, \tilde{A}_p^{m_p^{(i)}} \tag{3.2}$$

tends to zero as $p \to \infty$.

*Proof of Lemma 3.* Lemma 3 is the analogue of Theorem 3 in ref. 27 and is proved in the same way. Since the real and the complex cases are very similar, we will consider here only the real case. As we explained earlier, we can assume without loss of generality that $p \leqslant n$. Our arguments will be the most transparent when $k=1$ and the matrix entries $\{x_{ij}\}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$ are identically distributed. Construct a $n \times n$ random real symmetric Wigner matrix $M_n = (y_{ij})$, $1 \leqslant i, j \leqslant n$ such that $y_{ij} = y_{ji}$, $i \leqslant j$ are independent identically distributed random variables with the same distribution as $x_{11}$. Then

$$\mathbf{E} \text{ Trace } A_p^{m_p} \leqslant \mathbf{E} \text{ Trace } M_n^{2m_p}. \tag{3.3}$$

To see this we consider separately the left and the right hand sides of the inequality. We start by calculating the mathematical expectation of Trace $A_p^{m_p}$. Clearly,

$$\mathbf{E} \text{ Trace } A_p^{m_p} = \sum_{\mathscr{P}} \mathbf{E} x_{i_1, i_0} x_{i_1, i_2} x_{i_3, i_2} x_{i_3, i_4} \cdots x_{i_{2m_p-1}, i_{2m_p-2}} x_{i_{2m_p-1}, i_0}. \tag{3.4}$$

The sum in (3.4) is taken over all closed paths $\mathscr{P} = \{i_0, i_1, \ldots, i_{2m_p-1}, i_0\}$, with a distinguished origin, in the set $\{1, 2, \ldots n\}$ with the condition

**C1.** $i_t \in \{1, 2, \ldots, p\}$ for odd $t$

satisfied. We consider the set of vertices $\{1, 2, \ldots n\}$ as a nonoriented graph in which any two vertices are joined by an unordered edge. Since the distributions of the random variables $x_{ij}$ are symmetric, we conclude that if a path $\mathscr{P}$ gives a nonzero contribution to (3.4) then the following condition C2 also must hold:

**C2.** The number of occurrences of each edge is even.

Indeed, due to the independence of $\{x_{ij}\}$, the mathematical expectation of the product factorizes as a product of mathematical expectations of random variables corresponding to different edges of the path. Therefore if some edge appears in $\mathscr{P}$ odd number of times at least one factor in the product will be zero. Condition C2 is a necessary but not sufficient condition on $\mathscr{P}$ to give a non-zero contribution in (3.4). To obtain a necessary and sufficient condition let us note that an edge $i_k = j$, $i_{k+1} = g$, $k = 0, \ldots, 2m_p - 1$, contributes $x_{jg}$ for odd $k$ and $x_{gj}$ for even $k$. Clearly the number of apperances in each non-zero term of (3.4) must be even both for $x_{jg}$ and $x_{gj}$. This leads to

**C3.** For any edge $\{j, g\}$, $j, g \in \{1, 2, \ldots, n\}$, the number of times we pass $\{j, g\}$ in the direction $j \to g$ at odd moments of time $2k+1$,

$k = 0, 1,\dots, m_p$, plus the number of times we pass $\{j, g\}$ in the direction $g \to j$ at even moments of time $2k$, $k = 0, 1,\dots$ must be even.

Let us now consider the r.h.s. of (3.3). We can write

$$\mathbf{E}\,\mathrm{Trace}\, M_n^{2m_p} = \sum_{\mathscr{P}} \mathbf{E} y_{i_0, i_1} y_{i_1, i_2} y_{i_2, i_3} y_{i_3, i_4} \cdots y_{i_{2m_p-2}, i_{2m_p-1}} y_{i_{2m_p-1}, i_0}, \tag{3.5}$$

where the sum again is over all closed paths $\mathscr{P} = \{i_0, i_1,\dots, i_{2m_p-1, i_0}\}$, with a distinguished origin, in the set $\{1, 2,\dots, n\}$. Since $M_n$ is a square $n \times n$ real symmetric matrix conditions C1 and C3 are no longer needed. In particular the necessary and sufficient condition on a path $\mathscr{P}$ to give a non-zero contribution to (3.5) is C2. It does not matter in which direction we pass an edge $\{jg\}$, because both steps $j \to g$ and $g \to j$ give us $y_{jg} = y_{gj}$. Using the inequalities $\mathbf{E}x_{jg}^{2r}\mathbf{E}x_{gj}^{2q} \leqslant \mathbf{E}y_{jg}^{2r+2q}$ we show that each term in (3.4) is not greater than the corresponding term in (3.5) and, therefore, obtain (3.3). (3.1) (in the case $k = 1$) then immediately follows from Theorem 3 of ref. 27 (the matrix $A_n$ considered there differs from $M_n$ by a factor $\frac{1}{2\sqrt{n}}$). In the general case the proof of (3.1) and (3.2) is essentially identical to the one given in ref. 27. In particular, part b) of Lemma 3 follows from the fact that the l.h.s. at (3.2) is given by a subsum over paths that, in addition to C1–C3 have at least one edge appeared four times or more. As we showed in ref. 27 the contribution of such paths tends to zero as $n \to \infty$. Lemma 3 is proven.

**Remark 7.** If the condition $n - p = O(p^{1/3})$ in Lemma 3 and Theorem 2 is not satisfied the machinery from refs. 27, 38, and 39 does not work, essentially for the following reason: when we decide which vertex to choose during the moment of self-intersection (as explained in Section 4 of ref. 39) the number of choices for odd moments of time is smaller because of the constrain C1. If we now use the same bound as for the even moments of time (the one similar to the bound at the bottom of p. 725 of ref. 39) the estimate becomes rough when $n - p$ is much greater then $p^{1/3}$. Therefore new combinatorial ideas are needed.

As corollaries of Lemma 3 we obtain

**Corollary 2.** There exist const $> 0$ such that for any $s = o(p^{1/3})$

$$\mathbf{P}(\lambda_1(A_p) > \mu_{n,p} + \sigma_{n,p} s) < \mathrm{const}\, \exp(-\mathrm{const}\, s)$$

**Corollary 3.**

$$\int_{(-\infty,\, \mu_{n,p}+\sigma_{n,p}p^{1/6}]^k} \exp\left(\sum_{j=1,\dots,k} t_j s_j\right) \bar{\rho}_k^{(p)}(s_1,\dots,s_k)\, ds_1 \cdots ds_k \tag{3.6}$$

$$\to \int_{R^k} \exp\left(\sum_{j=1,\dots,k} t_j s_j\right) \rho_k(s_1,\dots,s_k)\, ds_1 \cdots ds_k \tag{3.7}$$

for any $t_1 > 0,\dots, t_k > 0$ as $n \to \infty$, where

$$\bar{\rho}_k^{(p)}(s_1,\dots,s_k) = (\sigma_{n,p})^k\, \rho_k^{(p)}(\mu_{n,p}+\sigma_{n,p}s_1,\dots, \mu_{n,p}+\sigma_{n,p}s_k)$$

is the rescaled $k$-point correlation function and $\rho_k(s_1,\dots,s_k)$ is defined in Section 2, Remark 6 by the r.h.s. of (2.18)–(2.21).

To prove Corollary 2 we use the Chebyshev inequality

$$\mathbf{P}(\lambda_1(A_p) > \mu_{n,p}+\sigma_{n,p}s) \leqslant \frac{\mathbf{E}\lambda_1(A_p)^{\sigma_{n,p}}}{(\mu_{n,p}+\sigma_{n,p}s)^{p^{2/3}}} \leqslant \frac{\mathbf{E}\ \mathrm{Trace}\ A_p^{\sigma_{n,p}}}{(\mu_{n,p}+\sigma_{n,p}s)^{p^{2/3}}}$$

and Lemma 3. As a result of Corollary 2 we obtain that with probability $O(\exp(-\mathrm{const}\, p^{1/6}))$ the largest eigenvalue is not greater than $\mu_{n,p}+\sigma_{n,p}p^{1/6}$. Therefore, it is enough to study only the eigenvalues in $(-\infty, \mu_{n,p}+\sigma_{n,p}p^{1/6}]$ (with very high probability there are no eigenvalues outside). To prove Corollary 3 we first note that Lemma 3 implies

$$\int_{(-\infty,\, \mu_{n,p}+\sigma_{n,p}p^{1/6}]^k} \exp\left(\sum_{j=1,\dots,k} t_j s_j\right) \bar{\rho}_k^{(p)}(s_1,\dots,s_k)\, ds_1 \cdots ds_k \tag{3.8}$$

$$\leqslant \frac{\mathrm{const}^k}{\prod_{i=1}^k t_i^{3k/2}} \exp\left(\mathrm{const} \cdot \sum_{i=1}^k t_i^3\right), \tag{3.9}$$

with some const $> 0$. To see this we write

$$\begin{aligned}
&\mathbf{E} \sum^{*} \prod_{j=1}^{j=k} \exp(t_j(\lambda_{i_j} - \mu_{n,p})/\sigma_{n,p}) \\
&\quad \leqslant \mathbf{E} \sum^{*} \prod_{j=1}^{j=k} (\lambda_{i_j}/\mu_{n,p})^{[t_j \mu_{n,p}/\sigma_{n,p}]}(1+o(1)) \\
&\quad \leqslant \mathbf{E} \prod_{1}^{k} \mathrm{Trace}\ A_p^{[t_j \mu_{n,p}/\sigma_{n,p}]} \mu_{n,p}^{-(\sum_1^k t_j)\, \mu_{n,p}/\sigma_{n,p}}(1+o(1))
\end{aligned}$$

where the sum $\sum^*$ is over all $k$-tuples of non-coinciding indices $(i_1, i_2,\dots, i_k)$,

$1 \leqslant i_j \leqslant p$, $j = 1,..., k$, such that $\lambda_{i_j} < \mu_{n,p} + \sigma_{n,p} p^{1/6}$, $j = 1,..., k$, and apply Lemma 3, (a). Part (b) of Lemma 3 implies that the differences between left hand sides of (3.8) for different ensembles of random matrices (i)–(iv) ((i′)–(iv′)) tend to 0. Finally we note that in the Gaussian case the l.h.s. of (3.8) converges, which in turn implies the convergence for arbitrary ensemble of sample covariance matrices. For the details we refer the reader to the analogous arguments in ref. 27. Corollary 3 is proven.

Theorem 2 now follows from Lemma 2, part (d) and Corollary 3.

## 4. PROOF OF THEOREM 3

In order to estimate the r.h.s. of (3.4) we assume some familiarity of the reader with the combinatorial machinery developed in refs. 27, 38, and 39. In particular we refer the reader to [28] (Section 2, Definitions 1 and 2) or ref. 39 (Section 4, Definitions 1–4) how we defined (a) marked and unmarked instants, (b) a partition of all verices into the classes $\mathcal{N}_0, \mathcal{N}_1,..., \mathcal{N}_m$ and c) paths of the type $(n_0, n_1,..., n_m)$, where $\sum_0^m n_k = n$, $\sum_0^m k n_k = m$(for simplicity we omit a subindex $p$ in $m_p$ throughout this section). Let us first estimate a subsum of (3.4) over the paths of some fixed type $(n_0, n_1,..., n_m)$. Essentially repeating the arguments from refs. 38 and 39 we can bound it from above by

$$p^{n_1+1} \frac{(n-n_1)!}{n_0!\, n_1!,..., n_m!} \frac{m!}{\prod_{k=2}^m (k!)^{n_k}} \prod_{k=2}^m (\text{const } k)^{k \cdot n_k} \sum_{X \in \Omega_m} (n/p)^{\#(X)}, \tag{4.1}$$

where the sum $\sum_{X \in \Omega_m}$ is over all possible trajectories

$$X = \{x(t) \geqslant 0,\, x(t+1) - x(t) = -1, +1,\, t = 0,..., 2m-1,\, x(o) = x(2m) = 0\}$$

and $\#(X) = \#(t : x(t+1) - x(t) = +1,\, t = 2k,\, k = 0,..., m-1)$.

The only differences between the estimates (4.1) in this paper and (4.4) and (4.27) in ref. 39 are

(a) the number of ways we can choose the vertices from $\mathcal{N}_1$ is estimated from above by $p^{n_1}(n/p)^{\#(X)}/n_1!$ not by $n(n-1)\cdots(n-n_1+1)/n_1!$, because of the restriction C1 from the last section ,

(b) we have in (4.1) the factor $(\text{const}2)^{2n_2}$ instead of $3^r$ in (4.27) of ref. 39, which is perfectly fine since $r \leqslant n_2$ (by $r$ we denoted in ref. 39 the number of so-called "non-closed" verices from $\mathcal{N}_2$), and

(c) there is no factor $\frac{1}{n^m}$ in (4.1) because of the different normalization. Let us denote by $g_m(y) = \sum_{X \in \Omega_m} y^{\#(X)}$(observe that $g_m(1) = |\Omega_m| = \frac{2m!}{m!\,(m+1)!}$ are just Catalan numbers).

Consider the generating function $G(z, y) = \sum_{m=0}^{\infty} g_m(y)\, z^m$, $g_0(y) = 1$. It is not difficult to see (by representing $g_m(y)$ as a sum over the first instants of the return of the trajectory to the origin) that

$$\begin{aligned} G(z, y) &= 1 + yzG'(z, y)\, G(z, y) \\ G'(z, y) &= 1 + zG'(z, y)\, G(z, y), \end{aligned} \tag{4.2}$$

where

$$G'(z, y) = \sum_{m=0}^{\infty} g'_m(y)\, z^m, \quad g'_m(y) = \sum_{X \in \Omega_m} y^{\#'(X)} \qquad \text{and}$$

$$\#'(X) = \#(t : x(t+1) - x(t) = +1,\, t = 2k+1,\, k = 0,..., m-1). \tag{4.3}$$

Solving (4.2) we obtain

$$\begin{aligned} G(z, y) &= \frac{-yz + z + 1 - \sqrt{((y-1)\, z - 1)^2 - 4z}}{2z} \\ &= \frac{-yz + z + 1 - (y-1)\sqrt{(z - z_1)(z - z_2)}}{2z}, \end{aligned} \tag{4.4}$$

where $z_1 = 1/(\sqrt{y} + 1)^2$, $z_2 = 1/(\sqrt{y} - 1)^2$, and we take the branch of $\sqrt{(z - z_1)(z - z_2)}$, analytic everywhere outside $[z_1, z_2]$, such that that $\sqrt{(0 - z_1)(0 - z_2)} = 1/(y-1)$. Therefore

$$g_m(y) = -\frac{y-1}{4\pi i} \oint_{|z| = z_1 - \epsilon} \frac{\sqrt{(z - z_1)(z - z_2)}}{z^{m+2}}, \qquad m \geqslant 1, \tag{4.5}$$

where the integration is counter-clockwise. An exercise in complex analysis gives us

$$\begin{aligned} &\oint_{|z| = z_1 - \epsilon} \frac{\sqrt{(z - z_1)(z - z_2)}}{z^{m+2}} \\ &\quad = -2i \frac{\sqrt{z_2 - z_1}}{z_1^{m+1/2} m^{3/2}} \int_0^{\infty} \sqrt{t} \exp(-t)\, dt (1 + o(1)) \\ &\quad = \frac{2i\sqrt{\pi}\, y^{1/4}(\sqrt{y} + 1)}{(y-1)} \frac{(\sqrt{y} + 1)^m}{m^{3/2}} (1 + o(1)). \end{aligned} \tag{4.6}$$

Therefore

$$g_m(y)=\frac{y^{1/4}(\sqrt{y}+1)}{2\sqrt{\pi}}\frac{(\sqrt{y}+1)^m}{m^{3/2}}(1+o(1)),\tag{4.7}$$

and the subsum of (3.4) over the paths of the type $(n_0, n_1,..., n_m)$ is bounded from above by

$$\begin{aligned}&\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}p^{n_1+1}\frac{(n-n_1)!}{n_0!\,n_1!\cdots n_m!}\frac{m!}{\prod_{k=2}^m (k!)^{n_k}}\\&\times\prod_{k=2}^m(\text{const}\,k)^{k\cdot n_k}\frac{(\sqrt{n/p}+1)^m}{m^{3/2}}(1+o(1))\\&\leqslant\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}\frac{p\mu_{n,p}^m}{m^{3/2}}\frac{1}{p^{m-n_1}}\frac{(n-n_1)!}{n_0!\,n_1!\cdots n_m!}\\&\times\frac{m!}{\prod_{k=2}^m (k!)^{n_k}}\prod_{k=2}^m(\text{const}\,k)^{k\cdot n_k}(1+o(1))\end{aligned}\tag{4.8}$$

(the constant const may have changed). Using the inequality $m!<n_1!\times m^{m-n_1}$ and $\sum_{k=1}^m kn_k=m$, $\sum_{k=1}^m \;=n-n_0$ we obtain

$$\begin{aligned}(4.8)&\leqslant\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}\frac{p\mu_{n,p}^m}{m^{3/2}}n^{-\Sigma_2^k kn_k}n^{\Sigma_2^m n_k}m^{\Sigma_2^m kn_k}\prod_{k=2}^m\frac{(\text{const}^k)^{n_k}}{n_k!}\\&\leqslant\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}\frac{p\mu_{n,p}^m}{m^{3/2}}\left(\prod_2^m\frac{1}{n_k!}\left(\frac{(\text{const}\,m)^k}{n^{k-1}}\right)^{n_k}\right)\end{aligned}\tag{4.9}$$

Now we can estimate the sum of (4.9) over $(n_0, n_1,..., n_m)$, $0<\sum_{k=2}^m kn_k\leqslant m$ as

$$\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}\frac{p\mu_{n,p}^m}{m^{3/2}}\left(\exp\left(\sum_{k=2}^m\frac{(\text{const}\,m)^k}{n^{k-1}}\right)-1\right)\tag{4.10}$$

Since for $m=o(p^{1/2})$

$$\sum_{k=2}^m\frac{(\text{const}\,m)^k}{n^{k-1}}=O(m^2/n)=o(1)\tag{4.11}$$

we see that the subsum of (3.4) over $\mathscr{P}$ with $\sum_{k=2}^m n_k>0$ is $o(\frac{p\mu_{n,p}^m}{m^{3/2}})$. Finally we note that the subsum over the paths of the type $(n-m, m, 0, 0,..., 0)$ is

$$\frac{(n/p)^{1/4}(\sqrt{n/p}+1)}{2\sqrt{\pi}}\frac{p\mu_{n,p}^m}{m^{3/2}}(1+o(1)),\tag{4.12}$$

because for such paths we can choose the vertices from $\mathcal{N}_1$ exactly in $p^m(n/p)^{\#(X)}(1+o(1))$ different ways (if $m=o(p^{1/2})$), and the first point of a path in $p$ different ways. Combining (4.11) and (4.12) we prove the first part of Theorem 3. To prove part b) we observe that if $m=O(p^{1/2})$, the l.h.s. of (4.11) is still $O(m^2/n)$, which together with (4.10) and (4.12) finishes the proof. Theorem 3 is proven.

To derive Corollary 1 from Theorem 3 we apply Chebyshev's inequality (similarly to the proof of Corollary 2 in Section 3) and Borel–Cantelli lemma.

## REFERENCES


1. R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
2. S. Wilks, *Mathematical Statistics* (Wiley, New York, 1967).
3. Z. D. Bai and Y. Q. Yin, Convergence of the semicircle law, *Ann. Probab.* **16**:863–875 (1988).
4. Z. D. Bai, J. W. Silverstein, and Y. Q. Yin, A note on the largest eigenvalue of a large dimensional sample covariance matrix, *J. Mult. Anal.* **26**:166–168 (1988).
5. Z. D. Bai, Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices, *Ann. Probab.* **21**:649–672 (1993).
6. Z. D. Bai and J. W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, *Ann. Probab.* **26**:316–345 (1998).
7. Z. D. Bai, Methodologies in spectral analysis of large dimensional random matrices. A review, *Statist. Sinica* **9**:611–677 (1999).
8. B. V. Bronk, Exponential ensemble for random matrices, *J. Math. Phys.* **6**:228–237 (1965).
9. A. Edelman, The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type, *Linear Algebra Appl.* **159**:55–80 (1991).
10. P. J. Forrester, The spectral edge of random matrix ensembles, *Nucl. Phys. B.* **402**:709–728 (1994).
11. P. J. Forrester, T. Nagao, and G. Honner, Correlations for the orthogonal-unitary and symplectic-unitary transitions at the hard and soft edges, *Nucl. Phys. B.* **553**:601–643 (1999).
12. S. Geman, A limit theorem for the norm of random matrices, *Ann. Probab.* **8**:252–261 (1980).
13. A. Guionnet, Large deviation upper bounds and central limit theorems for band matrices and non-commutative functionals of Gaussian large random matrices, preprint (2000).
14. U. Grenader and J. Silverstein, Spectral analysis of networks with random topologies, *SIAM J. Appl. Math.* **32**:499–519 (1977).
15. A. T. James, Distribution of matrices variates and latent roots derived from normal samples, *Ann. Math. Statist.* **35**:475–501–237 (1965).
16. I.M. Johnstone, On the distribution of the largest principal component, *Ann. Statist.* **29**, (2001).
17. V. Marchenko and L. Pastur, The eigenvalue distribution in some ensembles of random matrices, *Math. USSR Sbornik* **1**:457–483 (1967).
18. Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, On the limit of the largest eigenvalue of a large dimensional sample covariance matrix, *Probab. Theory Related Fields* **78**:509–521 (1988).

19. K. W. Wachter, The strong limits of random matrix spectra for sample matrices of independent elements, *Ann. Probab.* **6**:1–18 (1978).
20. J. Silverstein, On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix, *J. Multivariate Anal.* **30**:307–311 (1989).
21. K. Johansson, Shape fluctuations and random matrices, *Comm. Math. Phys.* **209**: 437–476, (2000).
22. C. A. Tracy and H. Widom, Level-spacing distribution and Airy kernel, *Comm. Math. Phys.* **159**:151–174 (1994).
23. C. A. Tracy and H. Widom, On orthogonal and symplectic matrix ensembles, *Comm. Math. Phys.* **177**:727–754 (1996).
24. D. Aldous and P. Diaconis, Longest increasing subsequences: From patience sorting to the Baik–Deift–Johansson theorem, *Bull. Amer. Math. Soc. (N.S.)* **36**:413–432 (1999).
25. P. Deift, Integrable Systems and Combinatorial Theory, *Notices Amer. Math. Soc.* **47**: 631–640 (2000).
26. A. Soshnikov, Determinantal random point fields, *Russian Math. Surveys* **55**:923–975 (2000).
27. A. Soshnikov, Universality at the edge of the spectrum in Wigner random matrices, *Comm. Math. Phys.* **207**:697–733 (1999).
28. L. A. Pastur and M. Shcherbina, Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles, *J. Statist. Phys.* **86**:109–147 (1997).
29. P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou, Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory, *Comm. Pure Appl. Math.* **52**:1335–1425, (1999).
30. P. Bleher and A. Its, Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model, *Ann. of Math.* **150**:185–266 (2000).
31. K. Johansson, Universality of local spacing distributions in certain Hermitian Wigner matrices, *Comm. Math. Phys.* **215**:683–705 (2001).
32. E. Brézin and A. Zee, Universality of the correlations between eigenvalues of large random matrices, *Nuclear Phys. B.* **402**:613–627 (1993).
33. E. Kanzieper, V. Frelikher, Spectra of large matrices: a method of study, in *Diffusive Waves in Complex Media* (Kluwer, Dodrecht, 1999), pp. 165–211.
34. M. L. Mehta, *Random Matrices*, 2nd ed. (Academic Press, New York, 1991).
35. G. Szego, *Orthogonal Polynomials*, 3rd ed. (American Mathematical Society, 1967).
36. C. A. Tracy and H. Widom, Correlation functions, cluster functions, and spacing distribution for random matrices, *J. Statist. Phys.* **92**:809–835 (1998).
37. H. Widom, On the relation between orthogonal, symplectic and unitary matrix ensembles, *J. Statist. Phys.* **94**:347–363 (1999).
38. Ya. Sinai, A. Soshnikov, Central limit theorem for traces of large random symmetric matrices, *Bol. Soc. Brasil. Mat. (N.S.)* **29**:1–24 (1998).
39. Ya. Sinai and A. Soshnikov, A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices, *Funct. Anal. Appl.* **32**:114–131 (1998).