Bernoulli **18**(1), 2012, 322–359
DOI: 10.3150/10-BEJ334

Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices

ZONGMING MA

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: zongming@wharton.upenn.edu

The distributions of the largest and the smallest eigenvalues of a p-variate sample covariance matrix S are of great importance in statistics. Focusing on the null case where nS follows the standard Wishart distribution $W_p(I, n)$, we study the accuracy of their scaling limits under the setting: $n/p \to \gamma \in (0, \infty)$ as $n \to \infty$. The limits here are the orthogonal Tracy–Widom law and its reflection about the origin.

With carefully chosen rescaling constants, the approximation to the rescaled largest eigenvalue distribution by the limit attains accuracy of order $\mathrm{O}(\min(n, p)^{-2/3})$. If $\gamma > 1$, the same order of accuracy is obtained for the smallest eigenvalue after incorporating an additional log transform. Numerical results show that the relative error of approximation at conventional significance levels is reduced by over 50% in rectangular and over 75% in 'thin' data matrix settings, even with $\min(n, p)$ as small as 2.

Keywords: eigenvalues of random matrices; Laguerre orthogonal ensemble; principal component analysis; rate of convergence; Tracy–Widom distribution; Wishart distribution

1. Introduction

Understanding the behavior of the extreme eigenvalues of a sample covariance matrix S is important in a large number of multivariate statistical problems. As an example, consider one of the most common inference problems: testing the null hypothesis that the population covariance is identity. Roy's union intersection principle [29] suggests that we reject the null hypothesis for large values of the largest eigenvalue of S (or for small values of the smallest eigenvalue). Naturally, the next question is: How should the p-value be calculated?

To address this issue, and many others, it is necessary to examine the null distributions of the extreme sample eigenvalues. In this paper, we restrict ourselves to the Gaussian framework. In particular, let X be an $n \times p$ data matrix whose row vectors are i.i.d. samples from the $N_p(0, I)$ distribution. The $p \times p$ matrix $A = X'X$ then follows a standard Wishart distribution: $A \sim W_p(I, n)$, and is called a (real) white Wishart matrix. The ordered eigenvalues of A are denoted by $\lambda_1 \geq \cdots \geq \lambda_p$. Our interest lies in λ_1 and λ_p, as $A = nS$.

The exact evaluation of the marginal distributions of these eigenvalues is difficult, even in the null case considered here. See, for example, Muirhead [24], Section 9.7. An alternative approach is to approximate them by their asymptotic limits. For the problem we are concerned with, Anderson [2], Chapter 13, summarized the classical results under the conventional asymptotic regime: p holds fixed and n tends to infinity.

However, for a wide range of modern data sets (microarray data, stock prices, weather forecasting, etc.), the number of features p is very large while the number of observations n is much smaller than or just comparable to p. For these situations, the classical asymptotics is not always appropriate and different asymptotic theories are needed. Borrowing tools from random matrix theory, especially those established by Tracy and Widom [32–34], Johnstone [15] showed that under the asymptotic regime

$$p \to \infty, \qquad n = n(p) \to \infty \quad \text{and} \quad n/p \to \gamma \in (0, \infty), \tag{1}$$

the largest eigenvalue λ_1 in A has the weak limit

$$\frac{\lambda_1 - \mu_p}{\sigma_p} \xrightarrow{\mathcal{D}} F_1, \tag{2}$$

where the centering and scaling constants are defined as

$$\mu_p = \left(\sqrt{n-1} + \sqrt{p}\right)^2, \qquad \sigma_p = \left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}. \tag{3}$$

Here F_1 denotes the orthogonal-Tracy–Widom law [33], the scaling limit of the largest eigenvalue in real Gaussian Wigner matrices. Slightly prior to [15], as a by-product of his analysis on the random growth model, Johansson [14] proved that the scaling limit for the largest eigenvalue in the complex white Wishart matrix is the unitary Tracy–Widom law F_2. Recently, El Karoui [9] extended the asymptotic regime (1) to include the cases where $n/p \to 0$ or ∞. For the smallest eigenvalue, when $\gamma > 1$, Baker *et al.* [3] showed that the reflection of F_2 about the origin is the scaling limit for complex Wishart matrices, and Paul [28] gave the Tracy–Widom limits in the case where $n/p \to \infty$ for both complex and real Wishart matrices.

Although this type of asymptotic result has emerged only recently in the statistics literature, it has already found its relevance to applications with modern data. For instance, based on (2), Patterson *et al.* [27] developed a formal procedure for testing the presence of population heterogeneity with SNP (single nucleotide polymorphism) data.

From a statistical point of view, to inform the use of any asymptotic result in practice, we need to understand how closely the asymptotic limit approximates the finite sample distributions. In the motivating example, this dictates the accuracy of the nominal p-value.

In this paper, we first establish a rate of convergence result for the Tracy–Widom approximation to the distribution of the rescaled largest eigenvalue, but with more carefully chosen constants than (3). Set $a \wedge b = \min(a, b)$ and $m_\pm = m \pm \frac{1}{2}$. We show that modifying the centering and scaling constants to

$$\mu_{n,p} = \left(\sqrt{n_-} + \sqrt{p_-}\right)^2, \qquad \sigma_{n,p} = \left(\sqrt{n_-} + \sqrt{p_-}\right)\left(\frac{1}{\sqrt{n_-}} + \frac{1}{\sqrt{p_-}}\right)^{1/3} \tag{4}$$

results in better approximation. The difference between the distribution of $(\lambda_1 - \mu_{n,p})/\sigma_{n,p}$ and F_1 reduces to the 'second order', being $\mathrm{O}((n \wedge p)^{-2/3})$ rather than $\mathrm{O}((n \wedge p)^{-1/3})$, that would apply by using (3). See Theorem 1. Numerical work in Section 2.2.1 suggests that the improvement is substantial.

Further assuming $\gamma > 1$ in (1), we find that, with a log transform, the scaling limit of $\log \lambda_p$ is the reflected Tracy–Widom law G_1 (defined by $G_1(s) = 1 - F_1(-s)$) [28]. Moreover, with appropriate rescaling constants, the accuracy of the limit also reaches the second order: $\mathrm{O}(p^{-2/3})$. See Theorem 2 and Section 2.2.2.

In the literature, El Karoui [10] established a parallel result for Johansson's theorem for the largest eigenvalue on the complex domain and Choup [6] studied the same problem via an Edgeworth expansion approach. Recently, Johnstone [16] obtained both scaling limit and convergence rates for the extreme eigenvalues of an F-matrix, on both complex and real domains. As is usual in the Random Matrix Theory literature, results on the real domain are founded in part on those for complex data but require significant additional constructs and arguments. This is explained for our setting in Sections 3 and 4.

The rest of the paper is organized as follows. In Section 2, we present theorems for both the largest and the smallest eigenvalues, together with supporting numerical results, related statistical settings, a real data example and a brief discussion. Section 3 proves the theorem on the largest eigenvalue and Section 4 sketches the proof of the one on the smallest eigenvalue. Finally, Section 5 establishes necessary Laguerre polynomial asymptotics, which is first used without proof in Section 3. Technical details are collected in the Appendix.

2. Main results and their applications

In this section, we first state two main theorems of this paper, which are concerned with the convergence rates of the largest and the smallest eigenvalues in finite Wishart matrices to their Tracy–Widom limits. The theorems are then complemented and further justified by a series of numerical experiments, in which the Tracy–Widom approximation is reasonably good even when n and/or p are as small as 2. After that, we review several related statistical settings and consider a real data example. Finally, we end the section with a brief discussion.

2.1. Main theorems

We begin with the largest eigenvalue, for which we have the following rate of convergence result.

Theorem 1. *Let $A \sim W_p(I, n)$ with $n \neq p$ and λ_1 its largest eigenvalue. Define $(\mu_{n,p}, \sigma_{n,p})$ as in* (4). *Under condition* (1), *for any given s_0, there exists an integer $N_0(s_0, \gamma)$, such that when $n \wedge p \geq N_0(s_0, \gamma)$ and is even, for all $s \geq s_0$,*

$$|P\{\lambda_1 \leq \mu_{n,p} + \sigma_{n,p} s\} - F_1(s)| \leq C(s_0)(n \wedge p)^{-2/3} \exp(-s/2),$$

where $C(\cdot)$ is continuous and non-increasing.

We also obtain an analogous result for the smallest eigenvalue. Refine condition (1) to

$$p \to \infty, \qquad p + 1 \leq n = n(p) \to \infty \quad \text{and} \quad n/p \to \gamma \in (1, \infty). \tag{5}$$

Define $\mu_{n,p}^{-} = (\sqrt{n_-} - \sqrt{p_-})^2$, $\sigma_{n,p}^{-} = (\sqrt{n_-} - \sqrt{p_-})(1/\sqrt{p_-} - 1/\sqrt{n_-})^{1/3}$, and let

$$\tau_{n,p}^{-} = \frac{\sigma_{n,p}^{-}}{\mu_{n,p}^{-}}, \qquad \nu_{n,p}^{-} = \log(\mu_{n,p}^{-}) + \frac{1}{8}(\tau_{n,p}^{-})^2. \tag{6}$$

Then we have the following theorem.

Theorem 2. *Let $A \sim W_p(I, n)$ with $n - 1 \geq p$ and λ_p as its smallest eigenvalue. Define $(\nu_{n,p}^{-}, \tau_{n,p}^{-})$ as in (6). Under condition (5), we have*

$$\frac{\log \lambda_p - \nu_{n,p}^{-}}{\tau_{n,p}^{-}} \xrightarrow{\mathcal{D}} G_1$$

with $G_1(s) = 1 - F_1(-s)$, the reflected Tracy–Widom law.

In addition, for any given s_0, there exists an integer $N_0(s_0, \gamma)$, such that when $p \geq N_0(s_0, \gamma)$ and is even, for all $s \geq s_0$,

$$|P\{\log \lambda_p \leq \nu_{n,p}^{-} - \tau_{n,p}^{-} s\} - G_1(-s)| \leq C(s_0) p^{-2/3} \exp(-s/2),$$

where $C(\cdot)$ is continuous and non-increasing.

While we only prove rigorous bounds for even p, numerical experiments show that the approximation works just as well in the odd case, and for the largest eigenvalue, also in the square case. See Tables 1 and 2.

2.2. Numerical performance

An important motivation for the current study is to promote practical use of the Tracy–Widom approximation. To this end, we conduct here a set of experiments to investigate its numerical quality.

2.2.1. *The largest eigenvalue*

Distributional approximation. We first computed the empirical cumulative probabilities of λ_1 (after rescaling), at a collection of F_1 percentiles, using $R = 40\,000$ replications. This is done for three different categories of (n, p) combinations: (1) the square case, where $n = p = 2$, 5, 25 and 100; (2) the rectangular case, where $p = 2$, 5, 25 and 100 and n/p is fixed at 4:1; (3) the 'thin' case, where $p = 5$ and 10 but n/p could be as high as 100:1 and 1000:1. In some sense, this category could also be thought of as in the situation where $n/p \to \infty$ as discussed in [9]. For comparison purpose, we rescaled λ_1 using both the new constants (4) and the old ones (3). The results are summarized in Table 1.

Numerical accuracy with the new constants could be viewed from two aspects. First, for the conventional significance levels of 10%, 5% and 1% that correspond to right tails of the distributions, the approximation looks good even when p is as small as 2! In addition, it improves as

Table 1. Simulations for finite $n \times p$ vs. Tracy–Widom limit: the largest eigenvalue. For each (n, p) combination, we show in the first row empirical cumulative probabilities for λ_1, rescaled by (4), and the second row, with parentheses, rescaled by (3), both computed from $R = 40\,000$ repeated draws from $W_p(n, I)$ using the method in [7]. Conventional significance levels are highlighted in bold font and the last row gives approximate standard errors based on binomial sampling. F_1 was computed by the method in [8] with percentiles obtained via inverse interpolation

Percentiles	−3.8954	−3.1804	−2.7824	−1.9104	−1.2686	−0.5923	0.4501	0.9793	2.0234
TW	0.01	0.05	0.10	0.30	0.50	0.70	**0.90**	**0.95**	**0.99**
2×2	0.000	0.000	0.000	0.034	0.379	0.690	**0.908**	**0.953**	**0.988**
	(0.000)	(0.000)	(0.000)	(0.015)	(0.345)	(0.669)	(0.902)	(0.950)	(0.987)
5×5	0.000	0.002	0.021	0.218	0.465	0.702	**0.908**	**0.954**	**0.989**
	(0.000)	(0.002)	(0.020)	(0.213)	(0.460)	(0.698)	(0.907)	(0.953)	(0.989)
25×25	0.003	0.031	0.075	0.280	0.492	0.700	**0.902**	**0.951**	**0.990**
	(0.003)	(0.030)	(0.075)	(0.280)	(0.491)	(0.699)	(0.902)	(0.951)	(0.990)
100×100	0.007	0.041	0.091	0.294	0.501	0.704	**0.902**	**0.951**	**0.990**
	(0.007)	(0.041)	(0.091)	(0.294)	(0.501)	(0.704)	(0.902)	(0.951)	(0.990)
8×2	0.000	0.001	0.012	0.196	0.456	0.702	**0.909**	**0.955**	**0.990**
	(0.000)	(0.004)	(0.031)	(0.270)	(0.532)	(0.754)	(0.928)	(0.964)	(0.992)
20×5	0.001	0.018	0.054	0.259	0.483	0.704	**0.906**	**0.954**	**0.990**
	(0.002)	(0.028)	(0.073)	(0.303)	(0.531)	(0.737)	(0.921)	(0.962)	(0.992)
100×25	0.006	0.040	0.088	0.292	0.498	0.701	**0.901**	**0.950**	**0.989**
	(0.008)	(0.047)	(0.100)	(0.314)	(0.523)	(0.721)	(0.910)	(0.955)	(0.991)
400×100	0.009	0.048	0.096	0.299	0.502	0.702	**0.902**	**0.951**	**0.990**
	(0.010)	(0.053)	(0.104)	(0.312)	(0.516)	(0.714)	(0.908)	(0.954)	(0.991)
500×5	0.010	0.049	0.098	0.296	0.502	0.705	**0.906**	**0.955**	**0.990**
	(0.020)	(0.083)	(0.150)	(0.385)	(0.589)	(0.772)	(0.933)	(0.969)	(0.994)
1000×10	0.010	0.051	0.101	0.300	0.504	0.707	**0.902**	**0.952**	**0.991**
	(0.017)	(0.077)	(0.138)	(0.366)	(0.571)	(0.757)	(0.923)	(0.963)	(0.994)
5000×5	0.012	0.056	0.107	0.307	0.509	0.707	**0.905**	**0.953**	**0.992**
	(0.027)	(0.097)	(0.169)	(0.402)	(0.602)	(0.779)	(0.933)	(0.969)	(0.994)
$10\,000 \times 10$	0.012	0.055	0.108	0.308	0.504	0.706	**0.905**	**0.954**	**0.991**
	(0.021)	(0.084)	(0.150)	(0.378)	(0.580)	(0.763)	(0.929)	(0.967)	(0.994)
$2 \times$ SE	0.001	0.002	0.003	0.005	0.005	0.005	0.003	0.002	0.001

p becomes larger and starts to match the finite distributions almost exactly when p is no greater than 25. See the last three columns of Table 1. Second, when p is large, for instance, in the 100×100 and 400×100 cases, F_1 provides reasonable approximation over the whole range of interest.

As regards the comparison between different rescaling constants, neither choice seems superior to the other in the square cases (see the first block of Table 1). However, when the ratio n/p is changed to 4:1 or higher (see the second and the third blocks), the improvement by using new constants (4) is self-evident.

Table 2. Simulations for finite $n \times p$ vs. Tracy–Widom limit: the smallest eigenvalue. For each (n, p) combination, empirical cumulative probabilities are computed for $(\log \lambda_p - \nu_{n,p}^{-})/\tau_{n,p}^{-}$ using $R = 40\,000$ draws from $W_p(I, n)$. Methods for sampling, computing F_1 and obtaining percentiles are the same as in Table 1. Conventional significance levels are highlighted in bold font and the last line gives approximate standard errors based on binomial sampling

Percentiles	3.8954	3.1804	2.7824	1.9104	1.2686	0.5923	−0.4501	−0.9793	−2.0234
RTW	0.99	0.95	0.90	0.70	0.50	0.30	**0.10**	**0.05**	**0.01**
4×2	1.000	1.000	0.998	0.893	0.625	0.326	**0.087**	**0.041**	**0.009**
10×5	0.999	0.995	0.976	0.798	0.555	0.310	**0.095**	**0.047**	**0.011**
50×25	0.997	0.973	0.931	0.728	0.515	0.302	**0.097**	**0.048**	**0.010**
200×100	0.993	0.960	0.913	0.713	0.509	0.306	**0.103**	**0.050**	**0.010**
8×2	1.000	0.992	0.969	0.792	0.554	0.314	**0.095**	**0.046**	**0.010**
20×5	0.999	0.977	0.939	0.740	0.522	0.301	**0.096**	**0.047**	**0.009**
100×25	0.993	0.960	0.915	0.713	0.505	0.298	**0.098**	**0.048**	**0.009**
400×100	0.992	0.954	0.904	0.701	0.500	0.298	**0.100**	**0.049**	**0.010**
$2 \times$ SE	0.001	0.002	0.003	0.005	0.005	0.005	0.003	0.002	0.001

As a remark, better performance on right tails and improvement by using the new constants, as reflected in this simulation study, agree well with the mathematical statement in Theorem 1.

Approximate percentiles. We can also use F_1 to calculate approximate percentiles for the finite distributions, whose accuracy can be measured by the relative error $r_\alpha = \theta_\alpha^{TW}/\theta_\alpha - 1$. Here, θ_α is the exact 100αth percentile of the rescaled largest eigenvalue in the finite $n \times p$ model and θ_α^{TW} is its counterpart from F_1.

In Figure 1, we plot r_α for $\alpha = 0.95$ and 0.99, with p ranging from 2 to 5 and n from 2 to 50. Although $n \wedge p$ is no greater than 5, the approximation is reasonably satisfactory. For the 95th percentile, $|r_{0.95}|$ ranges from 5% to 10% for most cases and slightly exceeds 10% only when $p = 2$ and the n/p ratio is high. The approximation works even better for the 99th percentile, with $|r_{0.99}| \leq 5\%$ for most cases. Due to computational limitation [20], we could not obtain exact percentiles when n and p are large. We expect the approximate percentiles to become more accurate as a consequence of better distributional approximation.

2.2.2. *The smallest eigenvalue*

For the smallest eigenvalue, we perform a simulation study to investigate the distributional approximation. We chose two n/p ratios: 2:1 and 4:1, both with $p = 2$, 5, 25 and 100. For each (n, p) combination, we used $R = 40\,000$ replications. The simulation results shown in Table 2 demonstrate similar performance as in the case of the largest eigenvalue and agree well with Theorem 2.

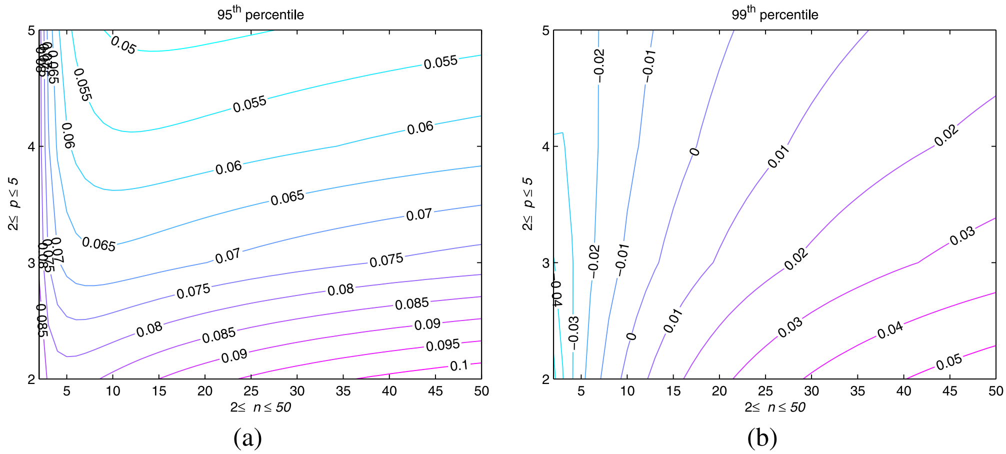

Figure 1. Plots of relative errors r_α for approximate percentiles using F_1: (a) 95th percentile; (b) 99th percentile. Exact finite $n \times p$ distributions are computed in MATLAB using Koev's implementation [20] and F_1 is computed using the method in [8]. The percentiles are obtained from inverse interpolation.

2.3. Related statistical settings

Here, we review several settings in multivariate statistics to which our results are applicable. Throughout the subsection, we only use the largest eigenvalue to illustrate.

Principal component analysis

Suppose that $X = [X_1, \dots, X_n]'$ is a Gaussian data matrix. Write the sample covariance matrix $S = (n-1)^{-1}X'HX$, where $H = I - n^{-1}\mathbf{1}\mathbf{1}'$ is the centering matrix and principal component analysis (PCA) looks for a sequence of standardized vectors $a_1, \dots, a_p$ in $\mathbb{R}^p$, such that a_i successively solves the following optimization problem:

$$\max\{a'Sa : a'a_j = 0,\, j \le i\},$$

where a_0 is the zero vector. Then, successive sample eigenvalues $\hat{\ell}_1 \ge \cdots \ge \hat{\ell}_p$ satisfy $\hat{\ell}_i = a_i' S a_i$.

One basic question in PCA application is testing the hypothesis of isotropic variation, that is, the population covariance matrix $\Sigma = \tau^2 I$. For simplicity, assume that $\tau^2 = 1$ (otherwise we divide S by τ^2). Then $(n-1)S \sim W_p(I, n-1)$. The largest eigenvalue $\hat{\ell}_1$ of S is a natural test statistic under the union intersection principle. Our result applies to $(n-1)\hat{\ell}_1$. If τ^2 is unknown, we could estimate it by $\operatorname{tr} S/p$. See [25].

Testing that a covariance matrix equals a specified matrix

Suppose that $X = [X_1, \dots, X_n]'$ has as its row vectors i.i.d. samples from the $N_p(\mu, \Sigma)$ distribution. We want to test the hypothesis H_0: $\Sigma = \Sigma_0$, where Σ_0 is a specified positive definite matrix.

Suppose μ is unknown, and let $S = (n-1)^{-1}X'HX$ be the sample covariance matrix. The union intersection test uses the largest eigenvalue of $\Sigma_0^{-1}S$, denoted by $\hat{\ell}_1(\Sigma_0^{-1}S)$, as the test statistic [23], page 130. Observe that $\hat{\ell}_1(\Sigma_0^{-1}S) = \hat{\ell}_1(\Sigma_0^{-1/2}S\Sigma_0^{-1/2})$. Under H_0, $(n-1)\Sigma_0^{-1/2}S\Sigma_0^{-1/2} \sim W_p(I, n-1)$. So, our result is available for $(n-1)\hat{\ell}_1(\Sigma_0^{-1}S)$.

Singular value decomposition

For X a real $n \times p$ matrix, there exist orthogonal matrices $U(n \times n)$ and $V(p \times p)$, such that

$$X = UDV^T,$$

where $D = \mathrm{diag}(d_1, \ldots, d_{n\wedge p}) \in \mathbb{R}^{n\times p}$ and $d_1 \geq \cdots \geq d_{n\wedge p} \geq 0$. This representation is called the singular value decomposition of X [13], Theorem 7.3.5, with d_i the ith singular value of X. Theorem 1 then provides an accurate distributional approximation for d_1^2 when the entries of X are independent standard normal random variables.

2.4. The score data example

We consider now the score data example extracted from [23]. The data set consists of the scores of 88 students on 5 subjects (mechanics, vectors, algebra, analysis and statistics). Taking account of centering, we have $n = 87$ and $p = 5$.

One might expect that there are several common factors that determine the students' performance on the tests. Moreover, one might assume that the joint effects of the common factors are observed in isotropic noises, in which case the covariance structure of the scores (after proper diagonalization) follows a spiked model $\Sigma = \tau^2\Sigma_m$, where $\tau^2 > 0$, $\Sigma_m = \mathrm{diag}(\ell_1, \ldots, \ell_m, 1, \ldots, 1)$ and $0 \leq m \leq 4$. (Note that the model $\Sigma = \tau^2\Sigma_4$ is the saturated model and is indistinguishable from $\Sigma = \tau^2\Sigma_5$.) To determine m, we are led to test a nested sequence of hypotheses H_k: $\Sigma = \tau^2\Sigma_m$ with some $m \leq k$, for $0 \leq k \leq 3$.

To compute the p-value of testing H_k, we could (i) estimate τ^2 by $\hat{\tau}_{p-k}^2$ as the mean of the $p-k$ smallest sample eigenvalues; (ii) construct the test statistic as $T_k = (n\hat{\ell}_{k+1}/\hat{\sigma}_{p-k}^2 - \mu_{n,p-k})/\sigma_{n,p-k}$; (iii) report $F_1(T_k)$ as the approximate conservative p-value. Step (iii) is justified as follows. Let $\mathcal{L}(\lambda_j|n, p, \Sigma)$ denote the law of the jth largest sample eigenvalue of a $W_p(n, \Sigma)$ matrix. By the interlacing properties of the eigenvalues [13], Theorem 7.3.9 (see also [15], Proposition 1.2), $\mathcal{L}(\lambda_1|n, p-m, I_{p-m})$ could be used to compute the conservative p-value for the null distribution $\mathcal{L}(\lambda_{k+1}|n, p, \Sigma_m)$ for all $k \geq m$, which is further approximated by F_1. We summarize the values of T_k and the corresponding p-values in Table 3.

From Table 3, we could see a noticeable difference between the values of T_k and the corresponding p-values by using different rescaling constants. The p-values obtained from the new constants are typically smaller than those from the old constants. Noting that the p-values are already conservative, the new constants (4) prevent further unnecessary conservativeness that would otherwise be caused by the old constants in this example.

Table 3. The test statistics T_k and the corresponding p-values $F_1(T_k)$ calculated using new centering and scaling constants (4) and old constants (3) for the score data

	H_0	H_1	H_2	H_3
T_k (new)	14.5934	4.3162	0.4535	1.4949
p-value (new)	$< 10^{-6}$	1.1×10^{-4}	0.0996	0.0235
T_k (old)	14.4740	4.1155	0.1803	1.1897
p-value (old)	$< 10^{-6}$	1.7×10^{-4}	0.1376	0.0371

2.5. Discussion

We discuss below two issues related to our results.

Log transform

One notable difference between Theorems 1 and 2 is the logarithmic transformation of the smallest eigenvalue before scaling.

Indeed, for the largest eigenvalue, a similar $\mathrm{O}(N^{-2/3})$ convergence rate can be obtained for the distribution of $(\log \lambda_1 - \nu_{n,p})/\tau_{n,p}$, with $\nu_{n,p} = \log(\mu_{n,p})$ and $\tau_{n,p} = \sigma_{n,p}/\mu_{n,p}$. However, when n or p is small, its numerical results are not as good as those obtained from direct scaling. In comparison, for the smallest eigenvalue, the transform yields substantial numerical improvement. Therefore, we recommend the log transform for the smallest eigenvalue.

As no theoretical analysis justifying the choice of the transform is currently available, we attempt some heuristics in the following. First, observe that sample covariance matrices are positive semidefinite. So, for λ_p, the hard lower bound at 0 truncates the left tail of its density function on any linear scale, and hence obstructs the asymptotic approximation by G_1 that is supported on the whole real line. However, by a map $x \mapsto \log x$, we map the support to the whole real line and avoid the 'hard edge' effect. The largest eigenvalue does not necessarily benefit from this transform, for it is on the 'soft edge', that is, the right edge of the covariance matrix spectrum, which does not have a deterministic upper bound. Such heuristics are supported by related studies on Gaussian Wigner matrices [17] and F-matrices [16].

Software

There have been works on the numerical evaluation of the Tracy–Widom distributions [4,5,8] and the exact finite $n \times p$ distributions of the extreme eigenvalues [19,20]. In addition, the author and colleagues have developed an `R` package `RMTstat` [18] that is intended to provide an interface for using the Tracy–Widom approximation in multivariate statistical analysis.

3. The largest eigenvalue

This section is devoted to the proof of Theorem 1. We use the operator norm convergence framework developed in [35], for the joint eigenvalue distribution of white Wishart matrices is essentially the same as the Laguerre orthogonal ensemble in random matrix theory (RMT).

In the proof, we first give the determinantal representations for the finite and limiting distribution functions and work out explicit formulas for related kernels, in which Widom's formula (12) plays the central role. Then, a Lipschitz-type inequality shows that the difference in determinants is bounded by the difference in kernels. The representation of the finite sample kernel involves weighted generalized Laguerre polynomials, while that of the limiting kernel uses Airy function. A decomposition of the kernel difference then enables us to transfer bounds on the convergence of Laguerre polynomials to Airy function to bounds on the kernel difference and eventually to bounds on the difference of the probabilities.

3.1. Determinantal laws

Following RMT notational convention, we replace the dimension parameter p of a white Wishart matrix A by N, and use x_i instead of λ_i to denote its eigenvalues. Henceforth, we assume that N is even, $n = n(N) \geq N + 1$ and $n/N \to \gamma \in [1, \infty)$ as $N \to \infty$. The cases $\gamma \in (0, 1]$ are easily obtained by interchanging n and N.

In the RMT literature, for an integer $N \geq 2$ and any $\alpha > -1$, the Laguerre orthogonal ensemble with parameters N and α, denoted by LOE(N, α), refer to joint eigenvalue density

$$\tilde{p}_N(x_1, \ldots, x_N) = \frac{1}{Z_{N,\alpha}} \prod_{1 \leq j < k \leq N} (x_j - x_k) \prod_{j=1}^{N} x_j^{\alpha} \mathrm{e}^{-x_j/2}, \tag{7}$$

where $x_1 \geq \cdots \geq x_N \geq 0$. If further α is a non-negative integer, (7) matches the density function of ordered eigenvalues $x_1 \geq \cdots \geq x_N \geq 0$ from a white Wishart matrix $A \sim W_N(I, n)$, with

$$\alpha = n - N - 1. \tag{8}$$

Henceforth, we identify the LOE(N, α) model with eigenvalues of $A \sim W_N(I, n)$ by (8). Thinking of α and n as functions of N, in what follows we sometimes drop explicit dependence of certain quantities on them.

For LOE(N, α), [34], Section 9, features the following determinantal formula

$$\tilde{F}_{N,1}(x') = P\{x_1 \leq x'\} = \sqrt{\det(I - K_N \chi)}. \tag{9}$$

Here $\chi = \mathbf{1}_{x > x'}$ and K_N is an operator with 2×2 matrix kernel

$$K_N(x, y) = (LS_{N,1})(x, y) + K^{\varepsilon}(x, y), \tag{10}$$

where

$$L = \begin{pmatrix} I & -\partial_2 \\ \varepsilon_1 & T \end{pmatrix}, \qquad K^{\varepsilon} = \begin{pmatrix} 0 & 0 \\ -\varepsilon(x - y) & 0 \end{pmatrix}.$$

In L, ∂_2 is the differential operator with respect to the second argument, ε_1 is the convolution operator acting on the first argument with the kernel $\varepsilon(x - y) = \frac{1}{2} \operatorname{sgn}(x - y)$ and $TK(x, y) = K(y, x)$ for any kernel K.

To give an explicit formula for $S_{N,1}$, introduce the generalized Laguerre polynomials $\{L_k^\alpha\}_{k=0}^\infty$ ([31], Chapter V), which are orthogonal on $[0,\infty)$ with weight function $x^\alpha \mathrm{e}^{-x}$. The normalized and weighted versions of them become

$$\phi_k(x;\alpha) = h_k^{-1/2} x^{\alpha/2} \mathrm{e}^{-x/2} L_k^\alpha(x), \qquad k = 0, \ldots, \tag{11}$$

with $h_k = \int_0^\infty L_k^\alpha(x)^2 x^\alpha \mathrm{e}^{-x}\,\mathrm{d}x = (k+\alpha)!/k!$. Widom [36] derived a formula for $S_{N,1}$, which can be rewritten in a form more convenient to us [1], equation (4.3), as

$$\begin{aligned} S_{N,1}(x,y) = S_{N,2}(x,y) &+ \frac{N!}{4\Gamma(N+\alpha)} x^{\alpha/2} \mathrm{e}^{-x/2} \left[\frac{\mathrm{d}}{\mathrm{d}x} L_N^\alpha(x) \right] \\ &\times \int_0^\infty \operatorname{sgn}(y-z) z^{\alpha/2-1} \mathrm{e}^{-z/2} [L_N^\alpha(z) - L_{N-1}^\alpha(z)]\,\mathrm{d}z, \end{aligned} \tag{12}$$

where $S_{N,2}$ is the unitary correlation kernel

$$S_{N,2}(x,y) = \sum_{k=0}^{N-1} \phi_k(x;\alpha)\phi_k(y;\alpha).$$

Let $a_N = \sqrt{N(N+\alpha)}$, and define as in [10], Section 2, functions

$$\begin{aligned} \phi(x;\alpha) &= (-1)^N \sqrt{\frac{a_N}{2}} \phi_N(x;\alpha-1) x^{-1/2} \mathbf{1}_{x\ge 0}, \\ \psi(x;\alpha) &= (-1)^{N-1} \sqrt{\frac{a_N}{2}} \phi_{N-1}(x;\alpha+1) x^{-1/2} \mathbf{1}_{x\ge 0}. \end{aligned} \tag{13}$$

Write $a \diamond b$ for the operator with kernel $(a\diamond b)(x,y) = \int_0^\infty a(x+z) b(y+z)\,\mathrm{d}z$. Then $S_{N,2}$ has the integral representation [10,15]

$$S_{N,2}(x,y) = \int_0^\infty \phi(x+z)\psi(y+z) + \psi(x+z)\phi(y+z)\,\mathrm{d}z = (\phi\diamond\psi + \psi\diamond\phi)(x,y). \tag{14}$$

By [31], equations (5.1.13) and (5.1.14), the second term on the right-hand side of (12) equals

$$-\frac{N!}{4\Gamma(N+\tilde{\alpha})} x^{\alpha/2} \mathrm{e}^{-x/2} L_{N-1}^{\alpha+1}(x) \int_0^\infty \operatorname{sgn}(y-z) z^{\alpha/2-1} \mathrm{e}^{-z/2} L_N^{\alpha-1}(z)\,\mathrm{d}z = \psi(x)(\varepsilon\phi)(y).$$

Hence, we obtain

$$S_{N,1}(x,y) = S_{N,2}(x,y) + \psi(x)(\varepsilon\phi)(y) \tag{15}$$

with $S_{N,2}(x,y)$ given in (14). Together with (9) and (10), this gives the determinantal representation of the finite sample distribution on the original scale.

The Tracy–Widom limit has a corresponding determinantal representation [35]

$$F_1(s') = \sqrt{\det(I - K_{\mathrm{GOE}} f)}, \tag{16}$$

where $f = \mathbf{1}_{s>s'}$ and the operator K_{GOE} has the matrix kernel

$$K_{\mathrm{GOE}}(s,t) = \begin{pmatrix} S(s,t) & SD(s,t) \\ IS(s,t) & S(t,s) \end{pmatrix} + K^{\varepsilon}(s,t).$$

Introduce the right tail integration operator $\tilde{\varepsilon}$ as in [16], where $(\tilde{\varepsilon}g)(s) = \int_s^\infty g(u)\,\mathrm{d}u$ and for kernel $K(s,t)$, $(\tilde{\varepsilon}_1 K)(s,t) = \int_s^\infty K(u,t)\,\mathrm{d}u$. Also write $a \otimes b$ for the rank one operator with kernel $(a\otimes b)(s,t) = a(s)b(t)$. Then the entries of K_{GOE} are

$$\begin{aligned} S(s,t) &= \bigl(S_A - \tfrac{1}{2}\mathrm{Ai}\otimes\tilde{\varepsilon}\mathrm{Ai}\bigr)(s,t) + \tfrac{1}{2}\mathrm{Ai}(s), \\ SD(s,t) &= -\partial_2\bigl(S_A(s,t) - \tfrac{1}{2}\mathrm{Ai}\otimes\tilde{\varepsilon}\mathrm{Ai}\bigr)(s,t), \\ IS(s,t) &= -\tilde{\varepsilon}_1\bigl(S_A - \tfrac{1}{2}\mathrm{Ai}\otimes\tilde{\varepsilon}\mathrm{Ai}\bigr)(s,t) - \tfrac{1}{2}(\tilde{\varepsilon}\mathrm{Ai})(s) + \tfrac{1}{2}(\tilde{\varepsilon}\mathrm{Ai})(t). \end{aligned} \tag{17}$$

Here $S_A(s,t) = (\mathrm{Ai}\diamond\mathrm{Ai})(s,t)$ is the Airy kernel, and $\mathrm{Ai}(\cdot)$ is the Airy function ([26], page 53, equation (8.01)).

Let $G = \frac{1}{\sqrt{2}}\mathrm{Ai}$, and define matrix operators

$$\tilde{L} = \begin{pmatrix} I & -\partial_2 \\ -\tilde{\varepsilon}_1 & T \end{pmatrix}, \qquad L_1 = \begin{pmatrix} I & 0 \\ -\tilde{\varepsilon}_1 & 0 \end{pmatrix}, \qquad L_2 = \begin{pmatrix} 0 & 0 \\ \tilde{\varepsilon}_2 & I \end{pmatrix}.$$

We can write K_{GOE} in a compact form as

$$K_{\mathrm{GOE}} = \tilde{L}(S_A - G\otimes\tilde{\varepsilon}G) + L_1\bigl(G\otimes\tfrac{1}{\sqrt{2}}\bigr) + L_2\bigl(\tfrac{1}{\sqrt{2}}\otimes G\bigr) + K^{\varepsilon}. \tag{18}$$

3.2. Rescaling the finite sample kernel

Under the current RMT notation, the rescaling constants (4) are translated to

$$\mu_{n,N} = \bigl(\sqrt{n_-} + \sqrt{N_-}\bigr)^2, \qquad \sigma_{n,N} = \bigl(\sqrt{n_-} + \sqrt{N_-}\bigr)\left(\frac{1}{\sqrt{n_-}} + \frac{1}{\sqrt{N_-}}\right)^{1/3}. \tag{19}$$

Introduce the linear transformation $\tau(s) = \mu_{n,N} + s\sigma_{n,N}$ and let $F_{N,1}(\cdot) = \tilde{F}_{N,1}(\tau(\cdot))$ be the distribution function of $\tau^{-1}(x_1)$, that is, the largest eigenvalue of $A \sim W_N(I,n)$, rescaled by (19).

Define the rescaled kernel $\bar{K}_\tau$ as

$$\bar{K}_\tau(s,t) = \sqrt{\tau'(s)\tau'(t)}K_N(\tau(s),\tau(t)) = \sigma_{n,N}K_N(\tau(s),\tau(t)). \tag{20}$$

Since K_N and $\bar{K}_\tau$ share the spectrum, $F_{N,1}(s') = \sqrt{\det(I - \bar{K}_\tau f)}$.

To work out a representation for $\bar{K}_\tau$, apply the τ-scaling to ϕ, ψ and $S_{N,2}$ to define

$$\phi_\tau(s) = \sigma_{n,N}\phi(\mu_{n,N} + s\sigma_{n,N}), \qquad \psi_\tau(s) = \sigma_{n,N}\psi(\mu_{n,N} + s\sigma_{n,N}) \tag{21}$$

and

$$S_\tau(s,t)=\sigma_{n,N}S_{N,2}(\mu_{n,N}+s\sigma_{n,N},\mu_{n,N}+t\sigma_{n,N})=(\phi_\tau\diamond\psi_\tau+\psi_\tau\diamond\phi_\tau)(s,t). \tag{22}$$

Then we obtain from (15) that

$$S_\tau^R(s,t)=\sqrt{\tau'(s)\tau'(t)}S_{N,1}(\tau(s),\tau(t))=S_\tau(s,t)+\psi_\tau(s)(\varepsilon\phi_\tau)(t). \tag{23}$$

This, together with (10) and (20), leads to

$$\bar{K}_\tau(s,t)=\begin{pmatrix} I & -\sigma_{n,N}^{-1}\cdot\partial_2 \\ \sigma_{n,N}\cdot\varepsilon_1 & T\end{pmatrix}S_\tau^R(s,t)+\sigma_{n,N}K^\varepsilon(s,t).$$

Observe that $\det(I-\bar{K}_\tau f)$ remains unchanged if we divide the lower left entry by $\sigma_{n,N}$ and multiply the upper right entry by $\sigma_{n,N}$. Thus, we obtain

$$F_{N,1}(s')=\sqrt{\det(I-K_\tau f)} \tag{24}$$

with

$$K_\tau(s,t)=(LS_\tau^R)(s,t)+K^\varepsilon(s,t). \tag{25}$$

To match the representation (18) of K_{GOE}, and to facilitate later arguments, it is helpful to rewrite LS_τ^R, and hence K_τ, using $\tilde{\varepsilon}$. To this end, observe that $\int\psi_\tau=0$ and let

$$\beta_N=\frac{1}{2}\int_{-\infty}^{\infty}\phi_\tau(s)\,\mathrm{d}s. \tag{26}$$

By the identity $(\varepsilon g)(s)=\frac{1}{2}\int g-(\tilde{\varepsilon}g)(s)$, we obtain $\varepsilon\phi_\tau=\beta_N-\tilde{\varepsilon}\phi_\tau$ and $\varepsilon\psi_\tau=-\tilde{\varepsilon}\psi_\tau$, and so

$$LS_\tau^R=L(S_\tau-\psi_\tau\otimes\tilde{\varepsilon}\phi_\tau)+\beta_N L(\psi_\tau\otimes 1).$$

Now $L=\tilde{L}+E$ with $E=\begin{pmatrix}0 & 0\\ \varepsilon_1+\tilde{\varepsilon}_1 & 0\end{pmatrix}$. Since $2(\varepsilon_1+\tilde{\varepsilon}_1)$ equals integration over $\mathbb{R}$ in the first argument and $\int\psi_\tau=0$, we obtain

$$\begin{aligned}LS_\tau^R&=\tilde{L}(S_\tau-\psi_\tau\otimes\tilde{\varepsilon}\phi_\tau)+ES_\tau+\beta_N\tilde{L}(\psi_\tau\otimes 1)\\ &=\tilde{L}(S_\tau-\psi_\tau\otimes\tilde{\varepsilon}\phi_\tau)+\beta_N L_1(\psi_\tau\otimes 1)+\beta_N L_2(1\otimes\psi_\tau).\end{aligned}$$

The second equality holds, for $(ES_\tau)_{21}=\frac{1}{2}\int_{-\infty}^{\infty}S_\tau(u,t)\,\mathrm{d}t=\beta_N\int_0^\infty\psi_\tau(t+z)\,\mathrm{d}z=\beta_N(\tilde{\varepsilon}\psi_\tau)(t)$. Finally, this gives K_τ a similar decomposition to that of K_{GOE}

$$K_\tau=\tilde{L}(S_\tau-\psi_\tau\otimes\tilde{\varepsilon}\phi_\tau)+L_1(\psi_\tau\otimes\beta_N)+L_2(\beta_N\otimes\psi_\tau)+K^\varepsilon. \tag{27}$$

3.3. Generalized Fredholm determinants

For any fixed $s_0 \in \mathbb{R}$, we are interested in the convergence rate of $F_{N,1}(s')$ to $F_1(s')$ for all $s' \geq s_0$. In what follows, we show that this relies on the operator convergence of K_τ to K_{GOE}.

First, we note that the determinants in (9), (16) and (24) are not the usual Fredholm determinants (see, e.g., [21] for an introduction to the Fredholm determinant), as the ε term on the lower-left position of the matrix kernels is not of trace class. Tracy and Widom [35] first observed the problem and proposed a solution by introducing weighted Hilbert spaces and regularized 2-determinants, which we adopt here.

Consider the determinant in (9). Let $\tilde{\rho}$ be a weight function such that (1) its reciprocal $\tilde{\rho}^{-1} \in L^1[0,\infty)$; and (2) $S_{N,1} \in L^2((x',\infty);\tilde{\rho}) \cap L^2((x',\infty);\tilde{\rho}^{-1})$. Then ε: $L^2((x',\infty);\tilde{\rho}) \to L^2((x',\infty);\tilde{\rho}^{-1})$ is Hilbert–Schmidt and K_N can be regarded as a 2×2 matrix kernel on the space $L^2((x',\infty);\tilde{\rho}) \oplus L^2((x',\infty);\tilde{\rho}^{-1})$. In addition, by the second condition on $\tilde{\rho}$, the diagonal elements of K_N are trace class on $L^2((x',\infty);\tilde{\rho})$ and $L^2((x',\infty);\tilde{\rho}^{-1})$, respectively.

For a Hilbert–Schmidt operator T with eigenvalues μ_k, its regularized 2-determinant [12] is defined as $\det_2(I - T) \equiv \prod_k (1-\mu_k)\mathrm{e}^{\mu_k}$. If the diagonal elements of T are trace class, then we define the generalized Fredholm determinant for T as

$$\det(I - T) = \det_2(I - T)\exp(-\operatorname{tr} T). \tag{28}$$

As remarked in [35], the definition (28) is independent of the choice of $\tilde{\rho}$ and allows the derivation in [34] that yields (9), (10) and eventually (15).

Change the domain to (s',∞) with $s' = \tau^{-1}(x')$ and the weight function to $\rho = \tilde{\rho} \circ \tau$, and abbreviate $L^2((s',\infty);\varrho)$ as $L^2(\varrho)$ for any suitable ϱ. Then, K_τ and K_{GOE} are members of the operator class $\mathcal{A}$ of 2×2 Hilbert–Schmidt operator matrices on $L^2(\rho) \oplus L^2(\rho^{-1})$ with trace class diagonal entries. Definition (28) and previous derivations in Section 3.2 remain valid.

In order to make the latter argument more explicit, it is convenient to make a specific choice of the weight function ρ. In particular, on the s-scale, we choose

$$\rho(s) = 1 + \exp(|s|). \tag{29}$$

This implies that on the x-scale, we specify the weight function $\tilde{\rho} = \rho \circ (\tau^{-1})$ as

$$\tilde{\rho}(x) = 1 + \exp(|x - \mu_{n,N}|/\sigma_{n,N}).$$

It is straightforward to verify that the required conditions are all satisfied.

With rigorous definition of the determinants, we now relate the convergence of $F_{N,1}$ to F_1 to that of K_τ to K_{GOE}. First of all, simple manipulation leads to

$$|F_{N,1}(s') - F_1(s')| \leq \frac{|F_{N,1}^2(s') - F_1^2(s')|}{F_1(s_0)} = \frac{1}{F_1(s_0)} |\det(I - K_\tau) - \det(I - K_{\mathrm{GOE}})|. \tag{30}$$

To bound the difference between the determinants, we have the following Lipschitz-type inequality. Here and after, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the trace class norm and Hilbert–Schmidt norm, respectively.

Proposition 1. *Let* $A, B \in \mathcal{A}$, *and* $\det(I - A)$, $\det(I - B)$ *defined as in* (28). *If* $\sum_{i=1}^{2} \|A_{ii} - B_{ii}\|_1 + \sum_{i\neq j} \|A_{ij} - B_{ij}\|_2 \le 1/2$, *then*

$$|\det(I - A) - \det(I - B)| \le M(B)\left(\sum_{i=1}^{2} \|A_{ii} - B_{ii}\|_1 + \sum_{i\neq j} \|A_{ij} - B_{ij}\|_2\right), \tag{31}$$

where $M(B) = 2|\det(I - B)| + 2\exp[2(1 + \|B\|_2)^2 + \sum_i \|B_{ii}\|_1]$.

Proof. [16], Proposition 3, established a similar bound to (31), but with $M(B)$ replaced by

$$C(A, B) = |e^{-\operatorname{tr} A}| \exp\left[\frac{1}{2}(1 + \|A\|_2 + \|B\|_2)^2\right] + |\det{}_2(I - B)| \frac{|e^{-\operatorname{tr} A} - e^{-\operatorname{tr} B}|}{|\operatorname{tr} A - \operatorname{tr} B|}.$$

We now bound $C(A, B)$ by the above claimed constant $M(B)$.

Observe that for $|x| \le 1/2$, $|e^x - 1| \le 2|x|$. Therefore, when $\sum_{i=1}^{2} \|A_{ii} - B_{ii}\|_1 + \sum_{i\neq j} \|A_{ij} - B_{ij}\|_2 \le 1/2$, we have $|\operatorname{tr} A - \operatorname{tr} B| \le \sum_{i=1}^{2} \|A_{ii} - B_{ii}\|_1 \le 1/2$, which in turn implies $|e^{-\operatorname{tr} A} - e^{-\operatorname{tr} B}| \le 2|\operatorname{tr} A - \operatorname{tr} B||e^{-\operatorname{tr} B}|$. Hence, for the terms in $C(A, B)$, we have

$$\begin{aligned}
|e^{-\operatorname{tr} A}| &\le |e^{-\operatorname{tr} B} - e^{-\operatorname{tr} A}| + |e^{-\operatorname{tr} B}| \le |e^{-\operatorname{tr} B}|(2|\operatorname{tr} A - \operatorname{tr} B| + 1) \\
&\le |e^{-\operatorname{tr} B}|\left(2\sum_i \|A_{ii} - B_{ii}\|_1 + 1\right) \le 2\exp(\|B_{11}\|_1 + \|B_{22}\|_1)
\end{aligned}$$

and

$$|\det{}_2(I - B)| \frac{|e^{-\operatorname{tr} A} - e^{-\operatorname{tr} B}|}{|\operatorname{tr} A - \operatorname{tr} B|} \le 2|\det{}_2(I - B)||e^{-\operatorname{tr} B}| = 2|\det(I - B)|.$$

Moreover, we observe that

$$\begin{aligned}
1 + \|A\|_2 + \|B\|_2 &\le 1 + 2\|B\|_2 + \|A - B\|_2 \\
&\le 1 + 2\|B\|_2 + \sum_{i=1}^{2} \|A_{ii} - B_{ii}\|_1 + \sum_{i\neq j} \|A_{ij} - B_{ij}\|_2 \\
&\le 2 + 2\|B\|_2.
\end{aligned}$$

Plugging all these bounds into $C(A, B)$, we obtain the claimed form of $M(B)$. □

Remark 1. Proposition 1 refines [16], Proposition 3, by having the leading constant $M(B)$ of the bound depend only on B, which is important for deriving properties of the $C(s_0)$ function later.

3.4. Decomposition of $K_\tau - K_{\mathrm{GOE}}$

By Proposition 1, to prove Theorem 1 is essentially to control the entrywise convergence rate of K_τ to K_{GOE}. To this end, we construct a telescopic decomposition of $K_\tau - K_{\mathrm{GOE}}$ into sums of simpler matrix kernels whose entries are more tractable.

To explain the intuition behind the decomposition, we introduce constants $\tilde{\mu}_{n,N}$ and $\tilde{\sigma}_{n,N}$ as

$$\tilde{\mu}_{n,N} = \left(\sqrt{n_+} + \sqrt{N_+}\right)^2, \qquad \tilde{\sigma}_{n,N} = \left(\sqrt{n_+} + \sqrt{N_+}\right)\left(\frac{1}{\sqrt{n_+}} + \frac{1}{\sqrt{N_+}}\right)^{1/3}. \tag{32}$$

In [10], it was shown that $(\mu_{n,N}, \sigma_{n,N}) = (\tilde{\mu}_{n-1,N-1}, \tilde{\sigma}_{n-1,N-1})$ is 'optimal' for ψ_τ in the sense that $|\psi_\tau - G| = \mathrm{O}(N^{-2/3})$, but suboptimal for ϕ_τ as $|\phi_\tau - G| = \mathrm{O}(N^{-1/3})$. However, later in Proposition 2, we will show that $|\phi_\tau - G - \Delta_N G'| = \mathrm{O}(N^{-2/3})$ for

$$\Delta_N = \frac{\tilde{\mu}_{n-1,N-1} - \tilde{\mu}_{n-2,N}}{\tilde{\sigma}_{n-2,N}} = \mathrm{O}(N^{-1/3}). \tag{33}$$

(For a proof, see Section A.5.) These bounds suggest that, in the decomposition, we align ψ_τ with G, and ϕ_τ with $G + \Delta_N G'$.

Let $G_N = G + \Delta_N G'$ and $S_{A_N} = G \diamond G_N + G_N \diamond G$. We obtain

$$S_{A_N} - G \otimes \tilde{\varepsilon} G_N = S_A - G \otimes \tilde{\varepsilon} G$$

for

$$\int_0^\infty G(s+z)G'(t+z) + G'(s+z)G(t+z)\,\mathrm{d}z = \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}z}[G(s+z)G(t+z)]\,\mathrm{d}z = -G(s)G(t).$$

This, together with (18) and (27), leads to the decomposition

$$\begin{aligned} K_\tau - K_{\mathrm{GOE}} &= \tilde{L}(S_\tau - S_{A_N}) + \tilde{L}(G \otimes \tilde{\varepsilon} G_N - \psi_\tau \otimes \tilde{\varepsilon}\phi_\tau) \\ &\quad + L_1\left(\psi_\tau \otimes \beta_N - G \otimes \tfrac{1}{\sqrt{2}}\right) + L_2\left(\beta_N \otimes \psi_\tau - \tfrac{1}{\sqrt{2}} \otimes G\right). \end{aligned} \tag{34}$$

3.5. Laguerre asymptotics and operator bounds

Here we collect a set of intermediate results to be used repeatedly in the proof of Theorem 1.

To start with, we consider the asymptotics of ϕ_τ and ψ_τ and their derivatives. Recalling that $G = \frac{1}{\sqrt{2}}\mathrm{Ai}$ and $G_N = G + \Delta_N G'$, we have the following.

Proposition 2. *Let ϕ_τ, ψ_τ and Δ_N be defined as in (21) and (33). Assume that (8) holds, and that as $N \to \infty$, $n = n(N) \to \infty$ with $n/N \to \gamma \in [1, \infty)$. Then, for any given s_0, there exists an integer $N_0(s_0, \gamma)$ such that when $N \ge N_0(s_0, \gamma)$, for all $s \ge s_0$,*

$$|\psi_\tau(s)|, |\psi_\tau'(s)| \le C(s_0)\exp(-s), \tag{35}$$

$$|\phi_\tau(s)|, |\phi_\tau'(s)| \le C(s_0)\exp(-s), \tag{36}$$

$$|\psi_\tau(s) - G(s)|, |\psi_\tau'(s) - G'(s)| \le C(s_0)N^{-2/3}\exp(-s), \tag{37}$$

$$|\phi_\tau(s) - G_N(s)|, |\phi_\tau'(s) - G_N'(s)| \le C(s_0)N^{-2/3}\exp(-s), \tag{38}$$

where $C(\cdot)$ is continuous and non-increasing.

Integrating these bounds over $[s, \infty)$, we know that they remain valid if we replace ψ_τ, ϕ_τ, G and G_N with $\tilde{\varepsilon}\psi_\tau, \tilde{\varepsilon}\phi_\tau, \tilde{\varepsilon}G$ and $\tilde{\varepsilon}G_N$ on the left-hand sides. The proof of Proposition 2 involves careful Liouville–Green analysis on the solution of certain differential equations and will be discussed in detail later in Section 5.

On the other hand, for G and G_N, we have the following bounds from [26], page 394. Note that the bounds for G_N and G'_N do not depend on N, for Δ_N is uniformly bounded.

Lemma 1. *Fix $\beta > 0$ and $k \geq 0$. Then, for all $s \geq s_0$,*

$$|s^k G(s)|, |s^k G_N(s)|, |s^k G'(s)|, |s^k G'_N(s)| \leq C(s_0) \exp(-\beta s),$$

where $C(s_0)$ is continuous and non-increasing.

For a proof of the lemma, see [22]. Integrating the bounds for $|G|$ and $|G_N|$ over $[s, \infty)$, we obtain that $|\tilde{\varepsilon}G|$ and $|\tilde{\varepsilon}G_N|$ are also bounded by $C(s_0)\mathrm{e}^{-\beta s}$.

For a later operator convergence argument, we will need simple bounds for certain norms of operator $D : L^2(\rho_1) \to L^2(\rho_2)$ with kernel $D(u, v) = \alpha(u)\beta(v)(a \diamond b)(u, v)$, where $\{\rho_1, \rho_2\} \subset \{\rho, \rho^{-1}\}$ with ρ given in (29). In particular, we have

Lemma 2 ([16]). *Let $D : L^2(\rho_1) \to L^2(\rho_2)$ have kernel $D(u, v) = \alpha(u)\beta(v)(a \diamond b)(u, v)$. Suppose that $\{\rho_1, \rho_2\} \subset \{\rho, \rho^{-1}\}$ and that, for $u \geq s'$,*

$$|\alpha(u)| \leq \alpha_0 \mathrm{e}^{\alpha_1 u}, \qquad |\beta(u)| \leq \beta_0 \mathrm{e}^{\beta_1 u}, \qquad |a(u)| \leq a_0 \mathrm{e}^{-a_1 u}, \qquad |b(u)| \leq b_0 \mathrm{e}^{-b_1 u}, \tag{39}$$

with $a_1 - \alpha_1, b_1 - \beta_1 \geq 1$. Then the Hilbert–Schmidt norm satisfies

$$\|D\|_2 \leq C \frac{\alpha_0 \beta_0 a_0 b_0}{a_1 + b_1} \exp[-(a_1 + b_1 - \alpha_1 - \beta_1)s' + |s'|], \tag{40}$$

where $C = C(a_1, \alpha_1, b_1, \beta_1)$. If $\rho_1 = \rho_2$, the trace norm $\|D\|_1$ satisfies the same bound.

3.6. Operator convergence: Proof of Theorem 1

Abbreviate the terms in the decomposition (34) as

$$K_\tau - K_{\mathrm{GOE}} = \delta^R + \delta_0^F + \delta_1^F + \delta_2^F.$$

We work out below entrywise bounds for each of these δ terms and then apply Proposition 1 to complete the proof of Theorem 1. In what follows, we use the abbreviation $D^{(k)} f$, $k = -1, 0, 1$ to denote $\tilde{\varepsilon} f$, f and f', respectively. Moreover, the unspecified norm $\|\cdot\|$ denotes the Hilbert–Schmidt norm $\|\cdot\|_2$ for off-diagonal entries and trace class norm $\|\cdot\|_1$ for diagonal ones.

δ^R term

Recall that $\delta^R = \tilde{L}(S_\tau - S_{A_N})$ with $S_\tau = \phi_\tau \diamond \psi_\tau + \psi_\tau \diamond \phi_\tau$ and $S_{A_N} = G_N \diamond G + G \diamond G_N$. Regardless of the signs, we have the following unified expression for the entries of δ^R:

$$\begin{aligned}(\delta^R)_{ij} &= D^{(k)}(\phi_\tau - G_N) \diamond D^{(l)}\psi_\tau + D^{(k)}G_N \diamond D^{(l)}(\psi_\tau - G) \\ &\quad + D^{(k)}(\psi_\tau - G) \diamond D^{(l)}\phi_\tau + D^{(k)}G \diamond D^{(l)}(\phi_\tau - G_N),\end{aligned} \tag{41}$$

for $i, j \in \{1, 2\}$, $k \in \{-1, 0\}$ and $l \in \{0, 1\}$. By Proposition 2 and Lemma 1, we find that for any of the four terms in (41), condition (39) is satisfied with $\alpha_0 = \beta_0 = 1$, $\alpha_1 = \beta_1 = 0$, $a_1 = b_1 = 1$ and $\{a_0, b_0\} = \{C(s_0), C(s_0)N^{-2/3}\}$. So Lemma 2 implies

$$\|(\delta^R)_{ij}\| \le C(s_0)N^{-2/3}\exp(-2s' + |s'|). \tag{42}$$

By a simple triangle inequality, we can choose $C(s_0)$ in the last display as the sum of products of continuous and non-increasing functions, which can be seen from the term $(\alpha_0\beta_0 a_0 b_0)/(a_1 + b_1)$ in (40). Moreover, the term C in (40) is a universal constant for fixed a_1, α_1, b_1 and β_1 here. Hence, the final $C(s_0)$ function remains continuous and non-increasing.

Finite rank terms

For a rank one operator $a \otimes b : L^2(\rho_1) \to L^2(\rho_2)$ with kernel $a(s)b(t)$, its norm is

$$\|a \otimes b\| = \|a\|_{2,\rho_2}\|b\|_{2,\rho_1^{-1}}.$$

Here, the norm can be either trace class or Hilbert–Schmidt, since the two agree for rank one operators. In addition, for any ϱ, $\|a\|_{2,\varrho}^2 = \int_{s'}^\infty |a(s)|^2\varrho(s)\,\mathrm{d}s$. Now consider matrices of rank one operators on $L^2(\rho) \otimes L^2(\rho^{-1})$. Write $\|\cdot\|_+$ and $\|\cdot\|_-$ for $\|\cdot\|_{2,\rho}$ and $\|\cdot\|_{2,\rho^{-1}}$, respectively. [16], equation (213) gives the following bound

$$\begin{pmatrix} \|a_{11} \otimes b_{11}\|_1 & \|a_{12} \otimes b_{12}\|_2 \\ \|a_{21} \otimes b_{21}\|_2 & \|a_{22} \otimes b_{22}\|_1 \end{pmatrix} \le \begin{pmatrix} \|a_{11}\|_+\|b_{11}\|_- & \|a_{12}\|_+\|b_{12}\|_+ \\ \|a_{21}\|_-\|b_{21}\|_- & \|a_{22}\|_-\|b_{22}\|_+ \end{pmatrix}. \tag{43}$$

First consider δ_0^F. We reorganize it as

$$\begin{aligned}\delta_0^F &= -\tilde{L}(\psi_\tau \otimes \tilde{\varepsilon}\phi_\tau - G \otimes \tilde{\varepsilon}G_N) \\ &= -\tilde{L}[\psi_\tau \otimes \tilde{\varepsilon}(\phi_\tau - G_N) + (\psi_\tau - G) \otimes \tilde{\varepsilon}G_N] = \delta_0^{F,1} + \delta_0^{F,2}.\end{aligned}$$

The entries of $\delta_0^{F,i}$, $i = 1, 2$, are all of the form $a \otimes b$, with a and b chosen from $D^{(k)}\psi_\tau$, $D^{(k)}(\phi_\tau - G_N)$, $D^{(k)}(\psi_\tau - G)$ and $D^{(k)}G_N$, for $k \in \{-1, 0, 1\}$.

Observe that for $\eta \ge 2$ we have

$$\int_{s'}^\infty \exp(-\eta s)\rho^{\pm 1}(s)\,\mathrm{d}s \le \frac{4}{\eta - 1}\exp(-\eta s' \pm |s'|) \le \frac{8}{\eta}\exp(-\eta s' + |s'|). \tag{44}$$

Together with Proposition 2 and Lemma 1, this implies

$$\left\|D^{(k)}\psi_\tau\right\|_\pm^2, \left\|D^{(k)}G_N\right\|_\pm^2 \le C(s_0)\exp(-2s'+|s'|),$$
$$\left\|D^{(k)}(\psi_\tau - G)\right\|_\pm^2, \left\|D^{(k)}(\phi_\tau - G_N)\right\|_\pm^2 \le C(s_0)N^{-4/3}\exp(-2s'+|s'|).$$

These bounds, together with the triangle inequality and (43), yield

$$\begin{aligned}\|(\delta_0^F)_{11}\|_1 &\le \|\psi_\tau \otimes \tilde{\varepsilon}(\phi_\tau - G_N)\|_1 + \|(\psi_\tau - G)\otimes \tilde{\varepsilon}G_N\|_1\\ &\le \|\psi_\tau\|_+\|\tilde{\varepsilon}(\phi_\tau - G_N)\|_- + \|\psi_\tau - G\|_+\|\tilde{\varepsilon}G_N\|_-\\ &\le C(s_0)N^{-2/3}\exp(-2s'+|s'|).\end{aligned}$$

Similarly, we obtain the bounds for the other entries. In summary, we have

$$\|(\delta_0^F)_{ij}\| \le C(s_0)N^{-2/3}\exp(-2s'+|s'|). \tag{45}$$

Switch to δ_1^F and δ_2^F. Recall that $\delta_1^F = L_1(\psi_\tau \otimes \beta_N - G\otimes \frac{1}{\sqrt{2}})$ and $\delta_2^F = L_2(\beta_N \otimes \psi_\tau - \frac{1}{\sqrt{2}}\otimes G)$. Due to their similarity, we take δ_1^F as an example and the same analysis applies to δ_2^F with obvious modification. We further decompose δ_1^F as

$$\delta_1^F = L_1\big[(\psi_\tau - G)\otimes\beta_N + G\otimes\big(\beta_N - \tfrac{1}{\sqrt{2}}\big)\big].$$

By (43), the essential elements we need to bound are $\|D^{(k)}(\psi_\tau - G)\|_\pm$, $\|D^{(k)}G\|_\pm$ and $\|1\|_-$ for $k=-1$ and 0. The bounds related to $D^{(k)}(\psi_\tau - G)$ have already been obtained. For the other two terms, (44) and Lemma 1 give

$$\left\|D^{(k)}G\right\|_\pm^2 \le C(s_0)\exp(-2s'+|s'|)$$

and

$$\|1\|_-^2 = \int_{s'}^\infty [1+\exp(|s|)]^{-1}\,\mathrm{d}s \le \int_{-\infty}^\infty \exp(-|s|)\,\mathrm{d}s \le 2.$$

Since $\beta_N - \frac{1}{\sqrt{2}} = \mathrm{O}(N^{-1})$ (for a proof, see Section A.5), we have

$$\begin{aligned}\|(\delta_1^F)_{11}\|_1 &\le \|(\psi_\tau - G)\otimes\beta_N\|_1 + \left\|G\otimes\big(\beta_N - 1/\sqrt{2}\big)\right\|_1\\ &\le \|(\psi_\tau - G)\|_+\|\beta_N\|_- + \|G\|_+\left\|\beta_N - 1/\sqrt{2}\right\|_-\\ &\le C(s_0)N^{-2/3}\exp(-s'+|s'|/2) + C(s_0)N^{-1}\exp(-s'+|s'|/2)\\ &\le C(s_0)N^{-2/3}\exp(-s'/2).\end{aligned}$$

In a similar vein, the same bound can be obtained for $\|(\delta_1^F)_{12}\|_2$ and entries of δ_2^F. Therefore, we conclude that

$$\|(\delta_1^F)_{ij}\|, \|(\delta_2^F)_{ij}\| \le C(s_0)N^{-2/3}\exp(-s'/2). \tag{46}$$

Now we prove Theorem 1.

Proof of Theorem 1. By the decomposition (34) and bounds (42), (45) and (46), the triangle inequality gives the following bound for the norm of each entry in $K_\tau - K_{\mathrm{GOE}}$:

$$\|(K_\tau - K_{\mathrm{GOE}})_{ij}\| \le C(s_0)N^{-2/3}\exp(-s'/2).$$

We then apply Proposition 1 with $A = K_\tau$ and $B = K_{\mathrm{GOE}}$ to get

$$|\det(I - K_\tau) - \det(I - K_{\mathrm{GOE}})| \le M(K_{\mathrm{GOE}})C(s_0)N^{-2/3}\exp(-s'/2), \tag{47}$$

where $M(K_{\mathrm{GOE}}) = 2\det(I - K_{\mathrm{GOE}}) + 2\exp\{2(1 + \|K_{\mathrm{GOE}}\|_2)^2 + \sum_i \|K_{\mathrm{GOE},ii}\|_1\}$.

For the first term in $M(K_{\mathrm{GOE}})$, we have $\det(I - K_{\mathrm{GOE}}) = F_1^2(s') \le 1$. On the other hand, we have

$$\|K_{\mathrm{GOE}}\|_2 \le \sum_{i,j}\|(K_{\mathrm{GOE}})_{ij}\|_2 \le \sum_i \|(K_{\mathrm{GOE}})_{ii}\|_1 + \sum_{i\ne j}\|(K_{\mathrm{GOE}})_{ij}\|_2.$$

In principle, one can show that, for each (i, j), $\|(K_{\mathrm{GOE}})_{ij}\| \le C(s_0)$, with $C(s_0)$ continuous and non-increasing. Take $\|(K_{\mathrm{GOE}})_{11}\|_1$ as an example. Let H_τ and G_τ be Hilbert–Schmidt operators with kernels $\phi_\tau(x + y)$ and $\psi_\tau(x + y)$, respectively, then as an operator

$$(K_{\mathrm{GOE}})_{11} = H_\tau G_\tau + G_\tau H_\tau + G \otimes \frac{1}{\sqrt{2}} - G \otimes \tilde{\varepsilon}G.$$

Since $\|AB\|_1 \le \|A\|_2\|B\|_2$, we have

$$\|(K_{\mathrm{GOE}})_{11}\|_1 \le 2\|H_\tau\|_2\|G_\tau\|_2 + \frac{1}{\sqrt{2}}\|G\|_{2,\rho}\|1\|_{2,\rho^{-1}} + \|G\|_{2,\rho}\|\tilde{\varepsilon}G\|_{2,\rho^{-1}}.$$

Each norm on the right-hand side of the last inequality is the square root of an integral of a positive function on (s', ∞) or $(s', \infty)^2$ that is bounded by the corresponding integral over (s_0, ∞) or $(s_0, \infty)^2$, which in turn is continuous and non-increasing in s_0. Hence, $\|(K_{\mathrm{GOE}})_{11}\|_1 \le C(s_0)$. A similar argument applies to other entries. So, we can control $M(K_{\mathrm{GOE}})$ by a continuous and non-increasing $C(s_0)$. Finally, we complete the proof by noting (30) and the fact that $1/F_1(s_0)$ is continuous and non-increasing. □

4. The smallest eigenvalue

This section is dedicated to the proof of Theorem 2.

Recall that two key components in the proof of Theorem 1 were: (1) determinantal representations for both the finite and the limiting distributions; (2) a closed-form formula for the finite sample kernel that yields a convenient decomposition of its difference from the limiting kernel.

In what follows, we first establish the rate of convergence for matrices with even dimensions. This is achieved by working out the above two components in the case of the smallest eigenvalue. Then, we prove weak convergence for matrices with odd dimensions using an interlacing property of the singular values.

4.1. Determinantal formula

As before, we follow RMT notation to replace p with N, and identify LOE(N,α) with eigenvalues of $A \sim W_N(I,n)$ by (8).

Assume that N is even. For the smallest eigenvalue x_N, for any $x' \geq 0$, [34] gives

$$1 - \tilde{F}_{N,N}(x') = P\{x_N > x'\} = \sqrt{\det(I - K_N\chi)}, \tag{48}$$

where $\chi = \mathbf{1}_{0 \leq x \leq x'}$ and K_N is given in (10).

Due to a nonlinear transformation to be introduced, the formula (12) that we previously used to represent $S_{N,1}$, the key component in K_N, is not most appropriate here. Instead, we find an alternative (yet equivalent) formula given in [1], Proposition 4.2, more convenient. Indeed, let

$$\bar{\phi}_k(x;\alpha) = (-1)^k \sqrt{\frac{a_N}{2}} \phi_k(x;\alpha) x^{-1/2} \mathbf{1}_{x \geq 0}, \tag{49}$$

with $a_N = \sqrt{N(N+\alpha)}$. Then [1], Proposition 4.2, asserts that

$$S_{N,1}(x,y;\alpha) = \sqrt{\frac{y}{x}} S_{N-1,2}(x,y;\alpha+1) + \sqrt{\frac{N-1}{N}} \bar{\phi}_{N-1}(x;\alpha+1)(\varepsilon\bar{\phi}_{N-2})(y;\alpha+1). \tag{50}$$

We write out the explicit dependence of these kernels on the parameter α as they are different on the two sides of the equation. As a comparison, the previous representation (15) could be rewritten as

$$S_{N,1}(x,y;\alpha) = S_{N,2}(x,y;\alpha) + \bar{\phi}_{N-1}(x;\alpha+1)(\varepsilon\bar{\phi}_N)(y;\alpha-1).$$

Its equivalence to (50) is given in the Appendix of [1].

Now, introduce the nonlinear transformation

$$\pi(s) = \exp(\nu_{n,N}^- - s\tau_{n,N}^-), \tag{51}$$

where $\nu_{n,N}^-$ and $\tau_{n,N}^-$ are the rescaling constants in (6), with p replaced by N. Incorporating the transformation into K_N, we define

$$\bar{K}_\pi(s,t) = \sqrt{\pi'(s)\pi'(t)} K_N(\pi(s),\pi(t)). \tag{52}$$

Let $F_{N,N}$ be the distribution of $(\log x_N - \nu_{n,N}^-)/\tau_{n,N}^-$. Fix s_0, for any $s' = \pi^{-1}(x') \geq s_0$ and $f = \mathbf{1}_{s \geq s'}$, since $\det(I - K_N\chi) = \det(I - \bar{K}_\pi f)$, we obtain $1 - F_{N,N}(-s') = \sqrt{\det(I - \bar{K}_\pi f)}$. Thinking of K_π as a Hilbert–Schmidt operator with trace class diagonal entries on $L^2([s',\infty);\rho) \oplus L^2([s',\infty);\rho^{-1})$ for proper weight function ρ, we can drop f.

Now consider the representation of $\bar{K}_\pi$. For $b_N = \sqrt{(N-1)/N}$, let

$$\phi_\pi(s) = -\sqrt{b_N}\pi'(s)\bar{\phi}_{N-2}\big(\pi(s);\alpha+1\big), \qquad \psi_\pi(s) = \sqrt{b_N}\pi'(s)\bar{\phi}_{N-1}\big(\pi(s);\alpha+1\big). \tag{53}$$

Using [11], Proposition 5.4.2, we obtain

$$S_{N-1,2}\big(\pi(s),\pi(t);\alpha+1\big) = (\pi'(s)\pi'(t))^{-1/2}(\phi_\pi \diamond \psi_\pi + \psi_\pi \diamond \phi_\pi)(s,t).$$

On the other hand, simple manipulation yields that the second term in (50), with $x=\pi(s)$ and $y=\pi(t)$, equals $(-\pi'(s))^{-1}\psi_\pi(s)(\varepsilon\phi_\pi)(t)$. Thus, $S_{N,1}(\pi(s),\pi(t)) = (-\pi'(s))^{-1}S_\pi^R(s,t)$ with

$$S_\pi^R(s,t) = (\phi_\pi \diamond \psi_\pi + \psi_\pi \diamond \phi_\pi)(s,t) + (\psi_\pi \otimes \varepsilon\phi_\pi)(s,t). \tag{54}$$

In addition, we have

$$\begin{aligned}
(-\partial_2 S_{N,1})(\pi(s),\pi(t)) &= \frac{-\partial_t S_{N,1}(\pi(s),\pi(t))}{\partial_t \pi(t)} = \frac{-1}{\pi'(s)\pi'(t)} \cdot [-\partial_2 S_\pi^R(s,t)],\\
(\varepsilon_1 S_{N,1})(\pi(s),\pi(t)) &= \int_0^\infty \varepsilon(\pi(s)-z) S_{N,1}(z,\pi(t))\,\mathrm{d}z\\
&= \int_{-\infty}^\infty \varepsilon(s-u) S_{N,1}(\pi(u),\pi(t))\pi'(u)\,\mathrm{d}u = -(\varepsilon_1 S_\pi^R)(s,t).
\end{aligned}$$

Supplying these equations to (10), we obtain that

$$\bar{K}_\pi(\pi(s),\pi(t)) = U(s)(LS_\pi^R + K^\varepsilon)(s,t)U^{-1}(t)$$

with $U(s) = \operatorname{diag}(1/\sqrt{-\pi'(s)}, -\sqrt{-\pi'(s)})$. Observe that $\det(I-\bar{K}_\pi)$ remains unchanged if we premultiply $\bar{K}_\pi$ with $U^{-1}(s_0)$ and postmultiply it with $U(s_0)$. Denoting the resulting kernel by K_π, we obtain that

$$K_\pi(s,t) = Q_N(s)(LS_\pi^R + K^\varepsilon)(s,t)Q_N^{-1}(t) \tag{55}$$

with $Q_N(s) = U^{-1}(s_0)U(s) = \operatorname{diag}(\sqrt{\pi'(s_0)/\pi'(s)}, \sqrt{\pi'(s)/\pi'(s_0)})$ and that $1-F_{N,N}(-s') = \sqrt{\det(I-K_\pi)}$.

Recall that $G_1(-s') = 1-F_1(s')$. So, $F_{N,N}(-s') - G_1(-s') = F_1(-s') - [1-F_{N,N}(-s')]$. Similar to (30), we obtain

$$|F_{N,N}(-s') - G_1(-s')| \le \frac{1}{F_1(s_0)} |\det(I-K_\pi) - \det(I-K_{\mathrm{GOE}})|.$$

Thus, as in the case of the largest eigenvalue, by Proposition 1, to prove Theorem 2 is to control the entrywise norm of $K_\pi - K_{\mathrm{GOE}}$. For this purpose, a convenient decomposition of $K_\pi - K_{\mathrm{GOE}}$ is crucial, to which we now turn.

4.2. Kernel difference decomposition

We derive below a decomposition of $K_\pi - K_{\mathrm{GOE}}$. Despite the differences in actual formulas, the general guideline of the decomposition is the same as that in Section 3.4.

To start with, we rewrite (55) using the right tail integration operator $\tilde{\varepsilon}$. To this end, observe that $\int \psi_\pi = 0$ and that

$$\tilde{\beta}_N = \frac{1}{2}\int_{-\infty}^{\infty} \phi_\pi(s)\,\mathrm{d}s = \frac{(N-1)^{1/4}(n-1)^{1/4}}{2^{(n-N)/2}(N-1)} \frac{\Gamma((N+1)/2)}{\Gamma(n/2)} \left[\frac{\Gamma(n-1)}{\Gamma(N-1)}\right]^{1/2} = \frac{1}{\sqrt{2}} + \mathrm{O}(N^{-1}).$$

By the same argument that leads to (27), we obtain

$$K_\pi(s,t) = Q_N(s)(K_\pi^R + K_{\pi,1}^F + K_{\pi,2}^F + K^\varepsilon)(s,t)Q_N^{-1}(t),$$

with the unspecified components given by

$$K_\pi^R = \tilde{L}(S_\pi - \psi_\pi \otimes \tilde{\varepsilon}\phi_\pi), \qquad K_{\pi,1}^F = L_1(\psi_\pi \otimes \tilde{\beta}_N), \qquad K_{\pi,2}^F = L_2(\tilde{\beta}_N \otimes \psi_\pi).$$

Define $\tilde{\Delta}_N = (\nu_{n,N}^- - \nu_{n-1,N-1}^-)/\tau_{n-1,N-1}^- = \mathrm{O}(N^{-1/3})$ and $\tilde{G}_N = G + \tilde{\Delta}_N G'$. For $\tilde{S}_{A_N} = G \diamond \tilde{G}_N + \tilde{G}_N \diamond G$, we have $\tilde{S}_{A_N} - G \otimes \tilde{\varepsilon}\tilde{G}_N = S_A - G \otimes \tilde{\varepsilon} G$. Abbreviate the terms in (18) as

$$K_{\mathrm{GOE}} = K^R + K_1^F + K_2^F + K^\varepsilon.$$

Then,

$$\begin{aligned} K_\pi^R - K^R &= \tilde{L}(S_\pi - S_A - \psi_\pi \otimes \tilde{\varepsilon}\phi_\pi + G \otimes \tilde{\varepsilon} G) \\ &= \tilde{L}(S_\pi - \tilde{S}_{A_N}) - \tilde{L}(\psi_\pi \otimes \tilde{\varepsilon}\phi_\pi - G \otimes \tilde{\varepsilon}\tilde{G}_N) = \delta^{R,I} + \delta_0^F. \end{aligned}$$

Further define

$$\begin{aligned} \delta^{R,D}(s,t) &= Q_N(s)K_\pi^R(s,t)Q_N^{-1}(t) - K_\pi^R(s,t), \\ \delta_i^F(s,t) &= Q_N(s)K_{\pi,i}^F(s,t)Q_N^{-1}(t) - K_i^F(s,t), \qquad i = 1, 2, \\ \delta^\varepsilon(s,t) &= Q_N(s)K^\varepsilon(s,t)Q_N^{-1}(t) - K^\varepsilon(s,t). \end{aligned}$$

Our final decomposition of $K_\pi - K_{\mathrm{GOE}}$ is

$$K_\pi - K_{\mathrm{GOE}} = \delta^{R,D} + \delta^{R,I} + \delta_0^F + \delta_1^F + \delta_2^F + \delta^\varepsilon. \tag{56}$$

We remark that Proposition 2 remains valid if we replace ϕ_τ and ψ_τ with ϕ_π and ψ_π, respectively. The proof is similar to that to be presented in Section 5 for Proposition 2. With these estimates, for each term in (56), we apply Lemma 2 to bound their entrywise norms as in Section 3.6. This completes the proof of the rate of convergence part in Theorem 2.

4.3. Weak convergence in the odd N case

We now establish weak convergence to the reflected Tracy–Widom law in the odd N case. This is achieved by employing an interlacing property of the singular values. The strategy follows from [30], Remark 5.

Assume that N is odd and $n-1 \geq N$. Let X_{N+1} be an $(n+1) \times (N+1)$ matrix with i.i.d. $N(0,1)$ entries and X_N the $n \times N$ matrix obtained by deleting the last row and the last column of X_{N+1}. Denote the smallest singular values of X_{N+1} and X_N by ι_{N+1} and ι_N, respectively. We apply [13], Theorem 7.3.9, twice to obtain that $\iota_N \leq \iota_{N+1}$. Repeat the deletion operation on X_N to obtain the $(n-1) \times (N-1)$ matrix X_{N-1} and denote its smallest singular value by ι_{N-1}. Then we obtain the 'sandwich' relation: $\iota_{N-1} \leq \iota_N \leq \iota_{N+1}$.

Observe that for $k = N-1$, N and $N+1$, $X_k' X_k$ are white Wishart matrices with the smallest eigenvalues $x_k = \iota_k^2$. In addition, as $N \to \infty$ and $n/N \to \gamma > 1$,

$$(\nu_{n,N}^{-} - \nu_{n-1,N-1}^{-})/\tau_{n-1,N-1}^{-} = \mathrm{O}(N^{-1/3}) \quad \text{and} \quad \tau_{n,N}^{-}/\tau_{n-1,N-1}^{-} = 1 + \mathrm{O}(N^{-1}).$$

They together imply that the weak limits for the odd N and the even N sequences must be the same. This completes the proof of Theorem 2.

5. Laguerre polynomial asymptotics

In this section, we complete the proof of Proposition 2. The proof has the following components. First, we take the Liouville–Green approach to analyze an intermediate function that is connected to both ϕ_τ and ψ_τ. After recollecting some previous results in [10,15] for ψ_τ, we give a detailed analysis of ψ_τ', $\psi_\tau' - G'$ and also strengthen a previous bound on $\psi_\tau - G$. Finally, we transfer the bounds on quantities related to ψ_τ to those related to ϕ_τ by a change of variable argument.

5.1. Liouville–Green approach

Recall $(\tilde{\mu}_{n,N}, \tilde{\sigma}_{n,N})$ in (32) and α in (8). We introduce the intermediate function

$$F_{n,N}(x) = (-1)^N \tilde{\sigma}_{n,N}^{-1/2} \sqrt{N!/n!}\, x^{\alpha/2+1} \mathrm{e}^{-x/2} L_N^{\alpha+1}(x) \tag{57}$$

as in [15], equation (5.1), and [10], Section 2.2.2. (Note: $\alpha = \alpha_N - 1$ for the constant α_N used in [15] and [10].) Then ϕ_τ is related to $F_{n,N}$ as

$$\psi_\tau(s) = \frac{1}{\sqrt{2}} \left(\frac{N^{1/4}(n-1)^{1/4} \tilde{\sigma}_{n-1,N-1}^{1/2} \sigma_{n,N}}{\tilde{\mu}_{n-1,N-1}} \right) F_{n-1,N-1}(\mu_{n,N} + s\sigma_{n,N}) \left(\frac{\tilde{\mu}_{n-1,N-1}}{\mu_{n,N} + s\sigma_{n,N}} \right).$$

Replacing the subscripts $(n-1, N-1)$ by $(n-2, N)$ in $\tilde{\mu}_{n-1,N-1}, \tilde{\sigma}_{n-1,N-1}$ and $F_{n-1,N-1}$ on the right-hand side, we also obtain the expression for $\phi_\tau(s)$.

Due to the close connection of ψ_τ and ϕ_τ to $F_{n,N}$, the key element in the proof of Proposition 2 becomes asymptotic analysis of $F_{n,N}$ and its derivative. To this end, the Liouville–Green (LG) theory set out in Olver [26], Chapter 11, is useful, for it comes with ready-made bounds on the difference between $F_{n,N}$ and the Airy function, and also on the difference between their derivatives.

To start with, we observe that $F_{n,N}$ satisfies a second-order differential equation,

$$F''_{n,N}(x) = \left\{\frac{1}{4} - \frac{\kappa_N}{x} + \frac{\lambda_N^2 - 1/4}{x^2}\right\} F_{n,N}(x), \tag{58}$$

with $\kappa_N = \frac{1}{2}(n+N+1)$ and $\lambda_N = \frac{1}{2}(n-N)$. By rescaling $x = \kappa_N \xi$, setting $w_N(\xi) = F_{n,N}(x)$, the equation becomes

$$w''_N(\xi) = \{\kappa_N^2 f(\xi) + g(\xi)\} w_N(\xi),$$

where

$$f(\xi) = \frac{(\xi - \xi_-)(\xi - \xi_+)}{4\xi^2}, \qquad g(\xi) = \frac{1}{4\xi^2}.$$

The zeros of f are given by $\xi_\pm = 2 \pm \sqrt{4 - \omega_N^2}$ for $\omega_N = 2\lambda_N/\kappa_N$. They are called the turning points of the differential equation, for each separates an interval in which the solutions are oscillating from one in which they are of exponential type. The LG approach introduces a new independent variable, ζ, and dependent variable, W, as

$$\zeta \left(\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\right)^2 = f(\xi), \qquad W = \left(\frac{\mathrm{d}\zeta}{\mathrm{d}\xi}\right)^{1/2} w_N.$$

Then the differential equation takes the form $W''(\zeta) = \{\kappa_N^2 \zeta + v(\omega_N, \zeta)\} W(\zeta)$. Without the perturbation term $v(\omega_N, \zeta)$, this is the Airy equation having linearly independent solutions in terms of Airy functions $\mathrm{Ai}(\kappa_N^{2/3}\zeta)$ and $\mathrm{Bi}(\kappa_N^{2/3}\zeta)$. We focus on approximating the recessive solution $\mathrm{Ai}(\kappa_N^{2/3}\zeta)$.

Let $\hat{f} = f/\zeta$. [26], Theorem 11.3.1, gives that

$$w_N(\xi) \propto \hat{f}^{-1/4}(\xi)\{\mathrm{Ai}(\kappa_N^{2/3}\zeta) + \varepsilon_2(\kappa_N, \xi)\},$$

where, uniformly for $\xi \in [2, \infty)$, the error term ε_2 satisfies

$$|\varepsilon_2(\kappa_N, \xi)| \le (\mathcal{M}/\mathcal{E})(\kappa_N^{2/3}\zeta)\left[\exp\left\{\frac{\lambda_0}{\kappa_N} F(\omega_N)\right\} - 1\right], \tag{59}$$

$$|\partial_\xi \varepsilon_2(\kappa_N, \xi)| \le \kappa_N^{2/3} \hat{f}^{1/2}(\xi)(\mathcal{N}/\mathcal{E})(\kappa_N^{2/3}\zeta)\left[\exp\left\{\frac{\lambda_0}{\kappa_N} F(\omega_N)\right\} - 1\right]. \tag{60}$$

In the bounds, $\mathcal{M}, \mathcal{E}$ are the modulus and weight functions for the Airy function and $\mathcal{N}$ the phase function for its derivative ([26], pages 394–396). On the real line, $\mathcal{E} \ge 1$ and is increasing, $0 \le \mathcal{M} \le 1$ and $\mathcal{N} \ge 0$. Moreover, for all x,

$$|\mathrm{Ai}(x)| \le (\mathcal{M}/\mathcal{E})(x), \qquad |\mathrm{Ai}'(x)| \le (\mathcal{N}/\mathcal{E})(x). \tag{61}$$

As $x \to \infty$, their asymptotics are given by

$$\mathcal{E}(x) \sim \sqrt{2}\mathrm{e}^{(2/3)x^{3/2}}, \qquad \mathcal{M}(x) \sim \pi^{-1/2} x^{-1/4}, \qquad \mathcal{N}(x) \sim \pi^{-1/2} x^{1/4}. \tag{62}$$

In addition, in the bounds (59) and (60), $\lambda_0 \doteq 1.04$ and the analysis in [10], A.3, shows that, uniformly for $\xi \in [2, \infty)$, for large enough N,

$$\exp\left\{\frac{\lambda_0}{\kappa_N} F(\omega_N)\right\} - 1 \le N^{-2/3}. \tag{63}$$

Come back to $F_{n,N}$. The alignment in [10], equation (5) and A.1, shows that

$$F_{n,N}(x) = r_N \kappa_N^{1/6} \tilde{\sigma}_{n,N}^{1/2} \hat{f}^{-1/4}(\xi)\{\mathrm{Ai}(\kappa_N^{2/3}\zeta) + \varepsilon_2(\kappa_N, \xi)\},$$

with $r_N = 1 + \mathrm{O}(N^{-1})$. Let $R_N(\xi) = (\zeta'(\xi)/\zeta'_N)^{-1/2}$ with $\zeta'_N = \zeta'(\xi_+)$. As $(\zeta'_N)^{-1} = \kappa_N^{1/3}\tilde{\sigma}_{n,N}$ and $\hat{f}(\xi) = \zeta'(\xi)^2$, we can rewrite $F_{n,N}$ as

$$F_{n,N}(x) = r_N R_N(\xi)\{\mathrm{Ai}(\kappa_N^{2/3}\zeta) + \varepsilon_2(\kappa_N, \xi)\}. \tag{64}$$

This representation serves as the starting point for all the subsequent asymptotic analysis on ϕ_τ, ψ_τ and their derivatives.

From now on, without notice, all the inequalities are understood to hold uniformly for $N \ge N_0(s_0, \gamma)$.

5.2. Summary of previous analysis: Bound for $|\psi_\tau(s)|$

Here, we summarize the previous analysis of $F_{n,N}$ in [10,15], which gives the desired bound for $|\psi_\tau(s)|$ in (35) and a crude estimate for $|\psi_\tau - G|$.

Let $x_{n,N}(s) = \tilde{\mu}_{n,N} + s\tilde{\sigma}_{n,N}$ and define

$$\theta_{n,N}(x_{n,N}(s)) = F_{n,N}(x_{n,N}(s))\left(\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)}\right). \tag{65}$$

As $\tilde{\sigma}_{n,N}^{-1/2} N^{1/6} < 1$, we obtain that, for all $s \ge 0$,

$$|F_{n,N}(x_{n,N}(s))| \le |F_{n,N}(x_{n,N}(s))\tilde{\sigma}_{n,N}^{1/2} N^{-1/6}| \le C\exp(-s),$$

where the latter inequality was obtained in [15], A.8. If $s_0 < 0$, then $\xi = x_{n,N}(s)/\kappa_N \ge 2$ uniformly for all $s \ge s_0$. In addition, Lemma 3 later shows that $|R_N(\xi)| \le 1 + CN^{-2/3}|s|$ for $s \in [s_0, 0]$. Therefore, we apply (59), (63) and (64) to obtain that

$$|F_{n,N}(x_{n,N}(s))| \le 2r_N|R_N(\xi)|(\mathcal{M}/\mathcal{E})(\kappa_N^{2/3}\zeta) \le 4,$$

uniformly for $s \in [s_0, 0]$. Hence, $|F_{n,N}(x_{n,N}(s))| \le C\exp(-s)$ for all $s \ge s_0$. Moreover, we note that $\tilde{\sigma}_{n,N}/\tilde{\mu}_{n,N} = \mathrm{O}(N^{-2/3})$. So, when $N \ge N_0(s_0)$, for all $s \ge s_0$,

$$\tilde{\mu}_{n,N}/x_{n,N}(s) \le (1 + s_0\tilde{\sigma}_{n,N}/\tilde{\mu}_{n,N})^{-1} \le 2.$$

Hence, uniformly for $s \geq s_0$,

$$|\theta_{n,N}(x_{n,N}(s))| \leq C(s_0)\exp(-s). \tag{66}$$

Finally, for any $\varrho_N = 1 + \mathrm{O}(N^{-1})$, El Karoui [10], Section 3.2, showed that, for all $s \geq s_0$,

$$|\varrho_N \theta_{n,N}(x_{n,N}(s)) - \mathrm{Ai}(s)| \leq C(s_0) N^{-2/3}\exp(-s/2).$$

For $\psi_\tau(s)$, observe that $(\mu_{n,N}, \sigma_{n,N}) = (\tilde{\mu}_{n-1,N-1}, \tilde{\sigma}_{n-1,N-1})$. Using Sterling's formula, we obtain that $\psi_\tau(s) = \frac{1}{\sqrt{2}}\rho_N \theta_{n-1,N-1}(x_{n-1,N-1}(s))$ for some $\rho_N = 1 + \mathrm{O}(N^{-1})$. Then, we apply the last two displays to obtain

$$|\psi_\tau(s)| \leq C(s_0)\exp(-s), \qquad |\psi_\tau(s) - G(s)| \leq C(s_0)N^{-2/3}\exp(-s/2), \tag{67}$$

uniformly for $s \geq s_0$.

Here, the first inequality gives the bound for $|\psi_\tau|$, while the bound on $|\psi_\tau(s) - G(s)|$ could be further improved; see (75). Note that we cannot apply these results directly to ϕ_τ since the 'optimal' rescaling constants $(\tilde{\mu}_{n-2,N}, \tilde{\sigma}_{n-2,N})$ for $F_{n-2,N}$ do not agree with the global constants $(\mu_{n,N}, \sigma_{n,N})$.

5.3. Asymptotics of $|\psi'_\tau(s)|$, $|\psi'_\tau(s) - G'(s)|$ and $|\psi_\tau(s) - G(s)|$

Here, we derive bounds on $|\psi'_\tau|$ and $|\psi'_\tau - G'|$ and refine the bound on $|\psi_\tau(s) - G(s)|$.

5.3.1. *Bound for* $|\psi'_\tau(s)|$

To obtain bounds for $|\psi'_\tau|$, we study $|\partial_s \theta_{n,N}(x_{n,N}(s))|$. By the triangle inequality,

$$\begin{aligned} |\partial_s \theta_{n,N}(x_{n,N}(s))| &\leq \left|\tilde{\sigma}_{n,N} F'_{n,N}(x_{n,N}(s)) \frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)}\right| + \left|\tilde{\sigma}_{n,N} F_{n,N}(x_{n,N}(s)) \frac{\tilde{\mu}_{n,N}}{x^2_{n,N}(s)}\right| \\ &= T_{N,1}(s) + T_{N,2}(s). \end{aligned} \tag{68}$$

In what follows, we deal with the two terms in order.

The $T_{N,1}$ *term.* Recall that $\tilde{\mu}_{n,N}/x_{n,N}(s) \leq 2$ for large N. So, we focus on $\tilde{\sigma}_{n,N} F'_{n,N}$, which can be decomposed as $\tilde{\sigma}_{n,N} F'_{n,N} = \sum_{i=1}^4 D^i_{n,N}$, with

$$D^1_{n,N} = r_N \tilde{\sigma}_{n,N} \kappa_N^{-1} R'_N(\xi)\{\mathrm{Ai}(\kappa_N^{2/3}\xi) + \varepsilon_2(\kappa_N, \xi)\}, \qquad D^2_{n,N} = r_N[R_N^{-1}(\xi) - 1]\mathrm{Ai}'(\kappa_N^{2/3}\zeta),$$
$$D^3_{n,N} = r_N \mathrm{Ai}'(\kappa_N^{2/3}\zeta), \qquad D^4_{n,N} = r_N \tilde{\sigma}_{n,N}\kappa_N^{-1} R_N(\xi)\partial_\xi \varepsilon_2(\kappa_N, \xi).$$

Due to different strategies used for the asymptotics on the s-scale, we divide $[s_0, \infty)$ into $I_{1,N} \cup I_{2,N}$, with $I_{1,N} = [s_0, s_1 N^{1/6})$ and $I_{2,N} = [s_1 N^{1/6}, \infty)$. The choice of s_1 is worked out in Section A.6. Here, we note that $s_1 \geq 1$ and that, for $s \geq s_1$,

$$\mathcal{E}^{-1}(\kappa_N^{2/3}\zeta) \leq C\exp(-3s/2) \leq C\exp(-s). \tag{69}$$

In addition, we will repeatedly use the following facts.

Lemma 3. *Under the conditions of Proposition 2, when $N \geq N_0(s_0,\gamma)$, for all $s \in I_{1,N}$,*

$$|R_N'(\xi)| \leq C\gamma^{-1/2}(1+\gamma), \qquad |R_N(\xi)-1| \leq CN^{-2/3}|s|,$$
$$|\kappa_N^{2/3}\zeta - s| \leq (CN^{-2/3}s^2) \wedge \tfrac{1}{2}|s| \wedge 1.$$

Proof of Lemma 3 is given in [22].

Case $s \in I_{1,N}$. Consider $D_{n,N}^1$ first. Recall that $r_N = 1 + \mathrm{O}(N^{-1})$. Together with Lemma 3, this implies

$$|r_N\tilde{\sigma}_{n,N}\kappa_N^{-1}R_N'(\xi)| \leq CN^{-2/3}. \tag{70}$$

On the other hand, as $0 \leq \mathcal{M} \leq 1$, (59), (61) and (63) together imply

$$|\mathrm{Ai}(\kappa_N^{2/3}\zeta) + \varepsilon_2(\kappa_N,\xi)| \leq C(\mathcal{M}/\mathcal{E})(\kappa_N^{2/3}\zeta) \leq C\mathcal{E}^{-1}(\kappa_N^{2/3}\zeta).$$

For $s \geq 0$, Lemma 3 implies $\kappa_N^{2/3}\zeta \geq s/2$. Since $\mathcal{E}$ is monotone increasing, by (62),

$$|\mathrm{Ai}(\kappa_N^{2/3}\zeta) + \varepsilon_2(\kappa_N,\xi)| \leq C\mathcal{E}^{-1}(s/2) \leq C\mathrm{e}^{-(1/(3\sqrt{2}))s^{3/2}} \leq C\exp(-s).$$

If $s_0 \leq 0$, we can replace the C on the rightmost side with $C(s_0) = \max\{C, \max_{s\in[3s_0/2,0]}\mathcal{E}^{-1}(s)\}$, which is continuous and non-increasing in s_0. Together with (70), we obtain that

$$|D_{n,N}^1| \leq C(s_0)N^{-2/3}\exp(-s).$$

(Here and after, we derive more stringent bounds with the $N^{-2/3}$ term whenever possible. Although they are not necessary for bounding $|\psi_\tau'|$, they are useful in the later study of $|\psi_\tau'(s) - G'(s)|$.)

For $D_{n,N}^2$, we first have $|r_N R_N^{-1}(\xi) - 1| \leq r_N|R_N^{-1}(\xi) - 1| + |r_N - 1|$. Lemma 3 implies that $|R_N^{-1}(\xi) - 1| \leq CN^{-2/3}|s|$. Observing that $|r_N - 1| = \mathrm{O}(N^{-1})$, we obtain

$$|r_N R_N^{-1}(\xi) - 1| \leq CN^{-2/3}|s|.$$

For $|\mathrm{Ai}'(\kappa_N^{2/3}\zeta)|$, when $s \geq 0$, Lemma 3 gives $\kappa_N^{2/3}\zeta \in [s/2, 3s/2]$. This, together with Lemma 1, implies that

$$|\mathrm{Ai}'(\kappa_N^{2/3}\zeta)| \leq C\exp(-3s/2). \tag{71}$$

If $s_0 < 0$, we can replace the C on the right-hand side with $C(s_0) = \max\{C, \max_{[3s_0/2,0]}|\mathrm{Ai}'(s)|\}$, which is continuous and non-increasing. Then the last two displays give

$$|D_{n,N}^2| \leq C(s_0)N^{-2/3}|s|\exp(-3s/2) \leq C(s_0)N^{-2/3}\exp(-s).$$

For $D_{n,N}^3$, we recall that $r_N = 1 + \mathrm{O}(N^{-1})$. Together with (71), this implies that

$$|D_{n,N}^3| \leq C(s_0)\exp(-s).$$

For $D^4_{n,N}$, since $r_N = 1 + \mathrm{O}(N^{-1})$, $\zeta'(\xi) = \hat{f}^{1/2}(\xi)$ and $\zeta'_N = \kappa_N^{1/3}/\tilde{\sigma}_{n,N}$, (60) and (63) imply

$$\begin{aligned} |D^4_{n,N}| &= |r_N \tilde{\sigma}_{n,N} \kappa_N^{-1} R_N(\xi) \partial_\xi \varepsilon_2(\kappa_N, \xi)| \\ &\le C N^{-2/3} \tilde{\sigma}_{n,N} \kappa_N^{-1/3} R_N(\xi) (\mathcal{N}/\mathcal{E})(\kappa_N^{2/3}\zeta) \\ &= C N^{-2/3} R_N^{-1}(\xi) (\mathcal{N}/\mathcal{E})(\kappa_N^{2/3}\zeta). \end{aligned}$$

Lemma 3 implies that $R_N^{-1}(\xi) \le C$ and $\kappa_N^{2/3}\zeta \in [s/2, 3s/2]$, uniformly on $I_{1,N}$. So, (62) gives

$$(\mathcal{N}/\mathcal{E})(\kappa_N^{2/3}\zeta) \le C s^{1/4} \mathrm{e}^{-(1/(3\sqrt{2}))s^{3/2}} \le C \exp(-s)$$

for all $s \ge 0$. And if $s_0 < 0$, we can replace the C on the rightmost side with $C(s_0) = \max\{C, \max_{s \in [3s_0/2, 0]} (\mathcal{N}/\mathcal{E})(s)\}$, which is continuous and non-increasing in s_0. All these elements together lead to

$$|D^4_{n,N}| \le C(s_0) N^{-2/3} \exp(-s).$$

Combining all the bounds on the $D^i_{n,N}$ terms, we obtain that $T_{N,1} \le C(s_0)\exp(-s)$ on $I_{1,N}$.

Case $s \in I_{2,N}$. In this case, we define $\tilde{D}^1_{n,N} = D^1_{n,N}$ and $\tilde{D}^2_{n,N} = D^2_{n,N} + D^3_{n,N} + D^4_{n,N}$. Consider $\tilde{D}^1_{n,N}$ first. By (59), (61) and (63), we obtain that for $N \ge N_0(s_0, \gamma)$,

$$|\tilde{D}^1_{n,N}| \le C \tilde{\sigma}_{n,N} \kappa_N^{-1} |R'_N/R_N|(\xi) R_N(\xi) (\mathcal{M}/\mathcal{E})(\kappa_N^{2/3}\zeta).$$

Observe that, uniformly on $I_{2,N}$,

$$\tilde{\sigma}_{n,N} \kappa_N^{-1} |R'_N/R_N|(\xi) \le C, \qquad R_N(\xi) \mathcal{M}(\kappa_N^{2/3}\zeta) \le Cs. \tag{72}$$

For a proof of (72), see [22]. On the other hand, (69) holds on $I_{2,N}$. Thus,

$$|\tilde{D}^1_{n,N}| \le Cs \exp(-3s/2) \le Cs^4 \exp(-s) \le C N^{-2/3} \exp(-s).$$

For $\tilde{D}^2_{n,N}$, we can write it as $\tilde{D}^2_{n,N} = r_N R_N(\xi)[\mathrm{Ai}'(\kappa_N^{2/3}\zeta) R_N^{-2}(\xi) + \tilde{\sigma}_{n,N}\kappa_N^{-1} \partial_\xi \varepsilon_2(\kappa_N, \xi)]$. By (60), (61) and (63) and the identity $R_N^{-1} = \tilde{\sigma}_{n,N}^{-1/2} \kappa_N^{1/6} \hat{f}^{1/4}$, we get the bound

$$|\tilde{D}^2_{n,N}| \le C R_N^{-1}(\xi) (\mathcal{N}/\mathcal{E})(\kappa_N^{2/3}\zeta).$$

(62) suggests that $R_N^{-1}(\xi)\mathcal{N}(\kappa_N^{2/3}\zeta) \le C R_N^{-1}(\xi)\kappa_N^{1/6}\zeta^{1/4} = C f^{1/4}(\xi)\tilde{\sigma}_{n,N}^{1/2} \le C\tilde{\sigma}_{n,N}^{1/2}$. The last inequality holds as $f \le 4$ for $s \in I_{2,N}$. On the other hand, $\tilde{\sigma}_{n,N} \le C(\gamma) N^{1/3} \le Cs^4$ for large N. Assembling all the pieces, we obtain $R_N^{-1}(\xi)\mathcal{N}(\kappa_N^{2/3}\zeta) \le Cs^2$. Together with (69), this implies

$$|\tilde{D}^2_{n,N}| \le Cs^2 \exp(-3s/2) \le Cs^{-4} \exp(-s) \le C N^{-2/3} \exp(-s).$$

Therefore, $T_{N,1} \le C N^{-2/3} \exp(-s)$ on $I_{2,N}$.

The $T_{N,2}$ term. This term is relatively easy to bound. Note that $\tilde{\sigma}_{n,N}/\tilde{\mu}_{n,N} = \mathrm{O}(N^{-2/3})$ and that $T_{N,2}(s) = |\theta_{n,N}(x_{n,N}(s))\tilde{\sigma}_{n,N}/x_{n,N}(s)|$. So, for all $s \geq s_0$, $N \geq N_0(s_0)$,

$$|\tilde{\sigma}_{n,N}/x_{n,N}(s)| = |s + \tilde{\mu}_{n,N}/\tilde{\sigma}_{n,N}|^{-1} \leq C(s_0)N^{-2/3}.$$

Together with (66), this implies that for all $s \geq s_0$, $T_{N,2}(s) \leq C(s_0)N^{-2/3}\exp(-s)$.

Summing up. By (68), the bounds on $T_{N,1}$ and $T_{N,2}$ transfer to

$$|\partial_s \theta_{n,N}(x_{n,N}(s))| \leq C(s_0)\exp(-s) \tag{73}$$

uniformly for $s \geq s_0$. On the other hand, we note that

$$\psi_\tau'(s) = \tfrac{1}{\sqrt{2}}\rho_N \partial_s \theta_{n-1,N-1}(x_{n-1,N-1}(s)),$$

with $\rho_N = 1 + \mathrm{O}(N^{-1})$. Thus, (73) implies the desired bound on $|\psi_\tau'|$ in (35).

5.3.2. *Bound for* $|\psi_\tau'(s) - G'(s)|$

By the triangle inequality, we bound $|\psi_\tau'(s) - G'(s)|$ as

$$\begin{aligned} |\psi_\tau'(s) - G'(s)| \leq{}& \tfrac{1}{\sqrt{2}}|\rho_N - 1||\partial_s\theta_{n-1,N-1}(x_{n-1,N-1}(s))| \\ &+ \tfrac{1}{\sqrt{2}}|\partial_s\theta_{n-1,N-1}(x_{n-1,N-1}(s)) - \mathrm{Ai}'(s)|. \end{aligned} \tag{74}$$

As $\rho_N = 1 + \mathrm{O}(N^{-1})$, by (73), we bound the first term by $C(s_0)N^{-1}\exp(-s)$. In what follows, to bound the second term in (74), we focus on $|\partial_s\theta_{n,N}(x_{n,N}(s)) - \mathrm{Ai}'(s)|$, which can first be split into two parts as:

$$\begin{aligned} &|\partial_s\theta_{n,N}(x_{n,N}(s)) - \mathrm{Ai}'(s)| \\ &\quad\leq \left|\tilde{\sigma}_{n,N}F_{n,N}'(x_{n,N}(s))\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)} - \mathrm{Ai}'(s)\right| + \left|\tilde{\sigma}_{n,N}F_{n,N}(x_{n,N}(s))\frac{\tilde{\mu}_{n,N}}{x_{n,N}^2(s)}\right| \\ &\quad= \mathcal{T}_{N,1}(s) + \mathcal{T}_{N,2}(s). \end{aligned}$$

The $\mathcal{T}_{N,1}(s)$ term. For this term, we separate the arguments on $I_{1,N} = [s_0, s_1N^{1/6})$ and $I_{2,N} = [s_1N^{1/6}, \infty)$.

Case $s \in I_{1,N}$. On $I_{1,N}$, we decompose $\mathcal{T}_{N,1}(s)$ as $\mathcal{T}_{N,1}(s) = \sum_{i=1}^5 \mathcal{D}_{n,N}^i$, with $\mathcal{D}_{n,N}^i = D_{n,N}^i\tilde{\mu}_{n,N}/x_{n,N}(s)$ for $i = 1, 2$ and 4, and

$$\mathcal{D}_{n,N}^3 = r_N\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)}[\mathrm{Ai}'(\kappa_N^{2/3}\zeta) - \mathrm{Ai}'(s)], \qquad \mathcal{D}_{n,N}^5 = \left[r_N\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)} - 1\right]\mathrm{Ai}'(s).$$

Observe that $|\tilde{\mu}_{n,N}/x_{n,N}(s)| \leq 2$ on $I_{1,N}$. Thus, by previous bounds on $D_{n,N}^i$, we obtain that, for $i = 1, 2$ and 4, $|\mathcal{D}_{n,N}^i| \leq C(s_0)N^{-2/3}\exp(-s)$.

Consider $\mathcal{D}^3_{n,N}$. By the Taylor expansion, for some s^* between $\kappa_N^{2/3}\zeta$ and s,

$$|\mathrm{Ai}'(\kappa_N^{2/3}\zeta) - \mathrm{Ai}'(s)| \leq |\mathrm{Ai}''(s^*)||\kappa_N^{2/3}\zeta - s| = |s^*\mathrm{Ai}(s^*)||\kappa_N^{2/3}\zeta - s|,$$

where the equality comes from the identity $\mathrm{Ai}''(s) = s\mathrm{Ai}(s)$. By Lemma 3, we have that $|\kappa_N^{2/3}\zeta - s| \leq CN^{-2/3}s^2$, and that s^* lies between $\frac{1}{2}s$ and $\frac{3}{2}s$. The latter, together with Lemma 1, implies that, for $s \geq 0$,

$$|s^*\mathrm{Ai}(s^*)| \leq C\exp(-3s/2).$$

If $s_0 \leq 0$, we then have $s^* \in [\frac{3}{2}s, 0]$, and hence we can replace C on the right-hand side with $C(s_0) = \max\{C, \max_{s\in[3s_0/2,0]} |s\mathrm{Ai}(s)|\}$. Observe that $r_N = 1 + \mathrm{O}(N^{-1})$ and that $|\tilde{\mu}_{n,N}/x_{n,N}(s)| \leq 2$. We thus conclude that

$$|\mathcal{D}^3_{n,N}| \leq C(s_0)N^{-2/3}s^2\exp(-3s/2) \leq C(s_0)N^{-2/3}\exp(-s).$$

Switch to $\mathcal{D}^5_{n,N}$. We first note that

$$\begin{aligned}\left|r_N\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)} - 1\right| &\leq r_N\left|\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)} - 1\right| + |r_N - 1| \\ &= r_N|s|\left|s + \frac{\tilde{\mu}_{n,N}}{\tilde{\sigma}_{n,N}}\right|^{-1} + |r_N - 1| \leq CN^{-2/3}|s| + CN^{-1}.\end{aligned}$$

The last inequality holds as $\tilde{\sigma}_{n,N}/\tilde{\mu}_{n,N} = \mathrm{O}(N^{-2/3})$, $r_N = 1 + \mathrm{O}(N^{-1})$, and for large N, $|s + \tilde{\mu}_{n,N}/\tilde{\sigma}_{n,N}| \geq \frac{1}{2}\tilde{\mu}_{n,N}/\tilde{\sigma}_{n,N}$ uniformly for $s \in I_{1,N}$. On the other hand, Lemma 1 implies that $|\mathrm{Ai}'(s)| \leq C(s_0)\exp(-3s/2)$. Putting the two parts together, we obtain

$$|\mathcal{D}^5_{n,N}| \leq C(s_0)N^{-2/3}(|s| + CN^{-1/3})\exp(-3s/2) \leq C(s_0)N^{-2/3}\exp(-s).$$

Assembling all the bounds on the $\mathcal{D}^i_{n,N}$'s, we obtain that, on $I_{1,N}$,

$$\mathcal{T}_{N,1}(s) \leq C(s_0)N^{-2/3}\exp(-s).$$

Case $s \in I_{2,N}$. In this case, we could act more heavy-handedly. In particular, by the asymptotics of $T_{N,1}(s)$ on $I_{2,N}$ and Lemma 1, we have

$$\begin{aligned}\mathcal{T}_{N,1}(s) &\leq \left|\tilde{\sigma}_{n,N}F'_{n,N}(x_{n,N}(s))\frac{\tilde{\mu}_{n,N}}{x_{n,N}(s)}\right| + |\mathrm{Ai}'(s)| \leq CN^{-2/3}\exp(-s) + C\exp(-3s/2) \\ &\leq CN^{-2/3}\exp(-s) + CN^{-2/3}s^4\exp(-3s/2) \leq CN^{-2/3}\exp(-s).\end{aligned}$$

The $\mathcal{T}_{N,2}(s)$ term. The $\mathcal{T}_{N,2}(s)$ term is the same as $T_{N,2}(s)$ defined previously in the study of $\partial_s\theta_{n,N}(x_{n,N}(s))$ and hence we quote the bound derived there directly as

$$\mathcal{T}_{N,2}(s) \leq C(s_0)N^{-2/3}\exp(-s) \qquad \text{for all } s \geq s_0.$$

Summing up. Combining the bounds on $\mathcal{T}_{N,1}$ and $\mathcal{T}_{N,2}$, we have, uniformly for $s \geq s_0$,

$$|\partial_s \theta_{n,N}(x_{n,N}(s)) - \mathrm{Ai}'(s)| \leq C(s_0) N^{-2/3} \exp(-s).$$

By the discussion following (74), we obtain the desired bound on $|\psi'_\tau(s) - G'(s)|$ in (37).

5.3.3. *Improved bound for* $|\psi_\tau - G|$

The bound on $|\psi'_\tau(s) - G'(s)|$, together with (67), can lead to a tighter bound for $|\psi_\tau(s) - G(s)|$ as the following:

$$\begin{aligned} |\psi_\tau(s) - G(s)| &= \left| \int_s^{2s} [\psi'_\tau(t) - G'(t)]\,\mathrm{d}t - [\psi_\tau(2s) - G(2s)] \right| \\ &\leq \int_s^{2s} |\psi'_\tau(t) - G'(t)|\,\mathrm{d}t + |\psi_\tau(2s) - G(2s)| \\ &\leq \int_s^{2s} C(s_0) N^{-2/3} \mathrm{e}^{-t}\,\mathrm{d}t + C(s_0) N^{-2/3} \exp(-s) \leq C(s_0) N^{-2/3} \exp(-s). \end{aligned} \tag{75}$$

This is exactly what we claimed in Proposition 2.

5.4. Asymptotics for quantities related to $\phi_\tau(s)$

In this part, we employ a trick in [15] to transfer the bounds on the quantities related to ψ_τ to those related to ϕ_τ.

Recall that, for $\tilde{\rho}_N = 1 + \mathrm{O}(N^{-1})$ (see Section A.5 for its proof),

$$\phi_\tau(s) = \frac{1}{\sqrt{2}} \tilde{\rho}_N F_{n-2,N}(x_{n-1,N-1}(s)) \frac{\tilde{\mu}_{n-2,N}}{x_{n-1,N-1}(s)}.$$

If the $x_{n-1,N-1}(s)$ term on the right-hand side were $x_{n-2,N}(s)$, then all the bounds we have proved for ψ_τ would also be valid for ϕ_τ. As this is not the case, we introduce a new independent variable s' as:

$$x_{n-1,N-1}(s) = x_{n-2,N}(s'), \tag{76}$$

that is, $s' = (\tilde{\mu}_{n-1,N-1} - \tilde{\mu}_{n-2,N})/\tilde{\sigma}_{n-2,N} + s\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N}$. (The readers are expected not to confuse it with the s' that previously appeared in Section 3.1.) Then ϕ_τ can be rewritten as

$$\phi_\tau(s) = \frac{1}{\sqrt{2}} \tilde{\rho}_N F_{n-2,N}(x_{n-2,N}(s')) \frac{\tilde{\mu}_{n-2,N}}{x_{n-2,N}(s')} = \frac{1}{\sqrt{2}} \tilde{\rho}_N \theta_{n-2,N}(x_{n-2,N}(s')).$$

Recalling the definition of Δ_N in (33), we have $s' - s = \Delta_N + [\tilde{\sigma}_{n-1,N-1}\tilde{\sigma}_{n-2,N}^{-1}]s$, with

$$\Delta_N = \mathrm{O}(N^{-1/3}), \qquad 1 \leq \tilde{\sigma}_{n-1,N-1}\tilde{\sigma}_{n-2,N}^{-1} = 1 + \mathrm{O}(N^{-1}). \tag{77}$$

Bounds for $|\phi_\tau(s)|$ and $|\phi_\tau'(s)|$

Recall previous bounds on $|\theta_{n,N}(x_{n,N}(s))|$ and $|\partial_s\theta_{n,N}(x_{n,N}(s))|$. Together with (77), they imply that, for all $s \geq s_0$,

$$|\phi_\tau(s)| \leq C(s_0)\exp(-s') \leq C(s_0)\exp(-s)$$

and

$$\begin{aligned}|\phi_\tau'(s)| &= \frac{1}{\sqrt{2}}\tilde{\rho}_N|\partial_s\theta_{n-2,N}(x_{n-2,N}(s'))| \\ &= \frac{1}{\sqrt{2}}\tilde{\rho}_N|\partial_{s'}\theta_{n-2,N}(x_{n-2,N}(s'))|\frac{\mathrm{d}s'}{\mathrm{d}s} \\ &\leq C(s_0)\exp(-s')\frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}} \leq C(s_0)\exp(-s).\end{aligned}$$

Bounds for $|\phi_\tau(s) - G_N(s)|$ and $|\phi_\tau'(s) - G_N'(s)|$

We consider $|\phi_\tau(s) - G_N(s)|$ in detail and the derivation for the bound on $|\phi_\tau'(s) - G_N'(s)|$ is essentially the same.

By the definition of s' and the identity $\mathrm{Ai}''(s) = s\mathrm{Ai}(s)$, we obtain the Taylor expansion

$$\begin{aligned}G(s') &= G(s) + (s'-s)G'(s) + \frac{1}{2}(s'-s)^2G''(s^*) \\ &= G_N(s) + \frac{1}{\sqrt{2}}\left[\frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}} - 1\right]s\mathrm{Ai}'(s) + \frac{1}{2\sqrt{2}}(s'-s)^2s^*\mathrm{Ai}(s^*),\end{aligned}$$

with s^* lying in between s and s'. By the previous discussion on $|\psi_\tau(s) - G(s)|$, this leads to

$$\begin{aligned}|\phi_\tau(s) - G_N(s)| &\leq C(s_0)N^{-2/3}\exp(-s') + CN^{-1}|s\mathrm{Ai}'(s)| + C(s'-s)^2|s^*\mathrm{Ai}(s^*)| \\ &\leq C(s_0)N^{-2/3}\exp(-s) + C(s'-s)^2|s^*\mathrm{Ai}(s^*)|.\end{aligned} \tag{78}$$

To further bound the last term, we split $[s_0, \infty)$ into $I_{1,N} \cup I_{2,N}$. For $s \in I_{1,N}$,

$$(s-s')^2 = \left[\Delta_N + \left(\frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}} - 1\right)s\right]^2 \leq [CN^{-1/3} + CN^{-1}s]^2 \leq (CN^{-2/3}) \wedge 1.$$

So $|s^*| \leq |s| + 1$, and Lemma 1 implies that

$$C(s-s')^2|s^*\mathrm{Ai}(s^*)| \leq C(s_0)N^{-2/3}\exp(-s).$$

On $I_{2,N}$, (77) implies that $s' \geq s/2$, and hence $s^* \geq s/2$. Together with Lemma 1, this implies

$$C(s'-s)^2|s^*\mathrm{Ai}(s^*)| \leq Cs^{-4}\cdot|(s^*)^7\mathrm{Ai}(s^*)| \leq CN^{-2/3}\exp(-s).$$

Therefore, we have shown that, for all $s \geq s_0$, the last term in (78) is further controlled by

$C(s_0)N^{-2/3}\exp(-s)$, which in turn gives the desired bound for $|\phi_\tau - G_N|$. It is not hard to check that all the $C(s_0)$ functions in the above analysis could be continuous and non-increasing.

Appendix

In the Appendix, we collect technical details that led to some of the claims previously made in the main text. Section A.5 gives proofs to properties of a number of constants. Section A.6 works out the details on the choice of s_1, which was used to decompose the interval $[s_0, \infty)$ in Section 5.

A.5. Properties of β_N, ρ_N, $\tilde{\rho}_N$, Δ_N and $\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N}$

Property of β_N

We are to show that $\beta_N = \frac{1}{\sqrt{2}} + \mathrm{O}(N^{-1})$. By definition, we know

$$\begin{aligned}
\beta_N &= \frac{1}{2}\int_{-\infty}^{\infty}\phi_\tau(s)\,\mathrm{d}s = \frac{1}{2}\int_0^{\infty}\phi(x;\alpha)\,\mathrm{d}x \\
&= \frac{N^{1/4}(n-1)^{1/4}\Gamma^{1/2}(N+1)}{2\sqrt{2}\Gamma^{1/2}(n)} \times \int_0^{\infty} x^{(\alpha-1)/2}\mathrm{e}^{-x/2}L_N^{\alpha}(x)\,\mathrm{d}x \\
&= \frac{2^{-\alpha/2}N^{1/4}(n-1)^{1/4}\Gamma^{1/2}(n)\Gamma((1/2)(N+3))}{(N+1)\Gamma^{1/2}(N+1)\Gamma((1/2)(n+1))}.
\end{aligned}$$

Applying Sterling's formula $\Gamma(z) = (2\pi/z)^{1/2}(z/\mathrm{e})^z(1+\mathrm{O}(z^{-1}))$, we obtain that

$$\begin{aligned}
\beta_N &= \frac{(2\pi/n)^{1/4}(n/\mathrm{e})^{n/2}[4\pi/(N+3)]^{1/2}[(N+3)/(2\mathrm{e})]^{(N+3)/2}}{[2\pi/(N+1)]^{1/4}[(N+1)/\mathrm{e}]^{(N+1)/2}[4\pi/(n+1)]^{1/2}[(n+1)/(2\mathrm{e})]^{(n+1)/2}} \\
&\quad \times \frac{2^{-\alpha/2}N^{1/4}(n-1)^{1/4}}{N+1}\big(1+\mathrm{O}(N^{-1})\big) \\
&= \frac{1}{\sqrt{2\mathrm{e}}}\left(1-\frac{1}{n+1}\right)^{n/2}\left(1+\frac{2}{N+1}\right)^{(N+1)/2+3/4}\big(1+\mathrm{O}(N^{-1})\big) \\
&= \frac{1}{\sqrt{2}} + \mathrm{O}(N^{-1}).
\end{aligned}$$

Properties ρ_N and $\tilde{\rho}_N$

We want to show that $\rho_N, \tilde{\rho}_N = 1 + \mathrm{O}(N^{-1})$. Consider ρ_N first. By definition, we have

$$\rho_N = \frac{N^{1/4}(n-1)^{1/4}\tilde{\sigma}_{n-1,N-1}^{1/2}\sigma_{n,N}}{\mu_{n,N}} = \frac{N^{1/4}(n-1)^{1/4}\tilde{\sigma}_{n-1,N-1}^{3/2}}{\tilde{\mu}_{n-1,N-1}}.$$

Plugging in the definition of $\tilde{\sigma}_{n-1,N-1}$ and $\tilde{\mu}_{n-1,N-1}$, we obtain that

$$\begin{aligned}\rho_N &= N^{1/4}(n-1)^{1/4}\left(\sqrt{N-\frac{1}{2}}+\sqrt{n-\frac{1}{2}}\right)^{-1/2}\left(\frac{1}{\sqrt{N-1/2}}+\frac{1}{\sqrt{n-1/2}}\right)^{1/2}\\ &= \left(\frac{N}{N-1/2}\right)^{1/4}\left(\frac{n-1}{n-1/2}\right)^{1/4} = 1+\mathrm{O}(N^{-1}).\end{aligned}$$

For $\tilde{\rho}_N$, we have

$$\begin{aligned}\tilde{\rho}_N &= \frac{N^{1/4}(n-1)^{1/4}\tilde{\sigma}_{n-2,N}^{1/2}\sigma_{n,N}}{\tilde{\mu}_{n-2,N}} = \frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}}\frac{N^{1/4}(n-1)^{1/4}\sigma_{n-2,N}^{3/2}}{\mu_{n-2,N}}\\ &= \frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}}N^{1/4}(n-1)^{1/4}\left(\sqrt{N+\frac{1}{2}}+\sqrt{n-\frac{3}{2}}\right)^{-1/2}\left(\frac{1}{\sqrt{N+1/2}}+\frac{1}{\sqrt{n-3/2}}\right)^{1/2}\\ &= \frac{\sigma_{n-1,N-1}}{\sigma_{n-2,N}}\left(\frac{N}{N+1/2}\right)^{1/4}\left(\frac{n-1}{n-3/2}\right)^{1/4} = 1+\mathrm{O}(N^{-1}).\end{aligned}$$

The last equality holds since $\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N} = 1+\mathrm{O}(N^{-1})$ as claimed in (33), which is to be shown below.

Property of Δ_N

Recall the definition $\Delta_N = (\tilde{\mu}_{n-1,N-1} - \tilde{\mu}_{n-2,N})/\tilde{\sigma}_{n-2,N}$. By [10], A.1.2, the numerator $\tilde{\mu}_{n-1,N-1} - \tilde{\mu}_{n-2,N} = \mathrm{O}(1)$. For the denominator, let $\gamma_{n,N} = (n-\frac{3}{2})/(N+\frac{1}{2})$. We then have

$$\begin{aligned}\frac{1}{\tilde{\sigma}_{n-2,N}} &= \left(\sqrt{N+\frac{1}{2}}+\sqrt{n-\frac{3}{2}}\right)^{-1}\left(\frac{1}{\sqrt{N+1/2}}+\frac{1}{\sqrt{n-3/2}}\right)^{-1/3}\\ &= \frac{1}{1+\gamma_{n,N}^{1/2}}(1+\gamma_{n,N}^{-1/2})\left(N+\frac{1}{2}\right)^{-1/3} = \mathrm{O}(N^{-1/3}).\end{aligned}$$

The last equality holds since $\gamma_{n,N}$ is bounded below for all $n > N$. Combining the two parts, we establish that $\Delta_N = \mathrm{O}(N^{-1/3})$.

Property of $\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N}$

We now switch to prove that

$$1 \le \tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N} = 1+\mathrm{O}(N^{-1}).$$

[10], A.1.3, showed that $\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N} = 1+\mathrm{O}(N^{-1})$. On the other hand, we have from the second-to-last display of [10], A.1.3, that

$$\left(\frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}}\right)^3 = \left[1+\frac{\sqrt{n/N}-\sqrt{N/n}}{n+N}+\mathrm{O}(n^{-2})\right]\left[1+\frac{1}{2}\left(\frac{1}{n}+\frac{1}{N}\right)+\mathrm{O}(n^{-2})\right].$$

Both terms become greater than 1 when $N \geq N_0(\gamma)$, and hence $\tilde{\sigma}_{n-1,N-1}/\tilde{\sigma}_{n-2,N} \geq 1$ for large N. Actually, the inequality holds for any $n > N \geq 2$. However, what we have proved here is sufficient for our argument in Section 5.4.

A.6. Choice of s_1 and its consequences

The key point in our choice of s_1 is to ensure that when $s \geq s_1$, we have

$$\tfrac{2}{3}\kappa_N \zeta^{3/2} \geq \tfrac{3}{2}s. \tag{79}$$

To this end, recall that in [15], A.8, one could choose $\tilde{s}_1(\gamma) = C(\gamma)(1+\delta)$ with some $\delta > 0$, such that when $s \geq \tilde{s}_1(\gamma)$, we have $\sqrt{f(\xi)} \geq 2/\tilde{\sigma}_{n,N}$ and hence if $s \geq 4\tilde{s}_1(\gamma)$,

$$\frac{2}{3}\kappa_N \zeta^{3/2} = \kappa_N \int_{\xi_+}^{\xi} \sqrt{f(z)}\,\mathrm{d}z \geq \kappa_N \frac{2}{\tilde{\sigma}_{n,N}}\bigl(s - \tilde{s}_1(\gamma)\bigr)\frac{\tilde{\sigma}_{n,N}}{\kappa_N} = 2\bigl(s - \tilde{s}_1(\gamma)\bigr) \geq \frac{3}{2}s.$$

Moreover, by the analysis in [10], A.6.4, $\tilde{s}_1(\gamma)$ could be chosen independently of γ and hence we could define our s_1 to be

$$s_1 = 4\tilde{s}_1,$$

which is independent of γ and such that (79) holds. Moreover, we also require that $s_1 \geq 1$.

After specifying our choice of s_1, we spell out two of its consequences. The first of them is that when $s \geq s_1 \geq 1$,

$$\mathcal{E}^{-1}(\kappa_N^{2/3}\zeta) \leq C\exp(-3s/2) \leq C\exp(-s). \tag{80}$$

This is from the observation that $\mathcal{E}(x) \geq C\exp(2x^{3/2}/3)$ and hence

$$\mathcal{E}^{-1}(\kappa_N^{2/3}\zeta) \leq C\exp\biggl(-\frac{2}{3}\kappa_N\zeta^{3/2}\biggr) \leq C\exp(-3s/2).$$

The other consequence is about the behavior of s' defined in (76) when $s \geq s_1$. Remembering that $s_1 \geq 1$, we then have that when $s \geq s_1$ and $N \geq N_0(\gamma)$,

$$s' - \frac{s}{2} = \Delta_N + \biggl(\frac{\tilde{\sigma}_{n-1,N-1}}{\tilde{\sigma}_{n-2,N}} - \frac{1}{2}\biggr)s \geq \Delta_N + \frac{s_1}{2} \geq \Delta_N + \frac{1}{2} \geq 0. \tag{81}$$

The last inequality holds when $N \geq N_0(\gamma)$, for $\Delta_N = \mathrm{O}(N^{-1/3})$.

Acknowledgements

I am most grateful to Professor Iain Johnstone for numerous discussions. Thanks also go to Professor Debashis Paul for kindly sharing an unpublished manuscript. I am grateful to the editor, an associate editor and an anonymous referee for their helpful comments that led to improvement on the presentation of the paper. This work is supported in part by Grants NSF DMS-05-05303 and NIH EB R01 EB001988.

References

[1] Adler, M., Forrester, P.J., Nagao, T. and van Moerbeke, P. (2000). Classical skew orthogonal polynomials and random matrices. *J. Statist. Phys.* **99** 141–170. MR1762659

[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Hoboken, NJ: Wiley. MR1990662

[3] Baker, T.H., Forrester, P.J. and Pearce, P.A. (1998). Random matrix ensembles with an effective extensive external charge. *J. Phys. A* **31** 6087–6101. MR1637735

[4] Bejan, A.I. (2005). Largest eigenvalues and sample covariance matrices. Tracy–Widom and Painlevé II: computational aspects and realization in S-plus with applications. Preprint.

[5] Bornemann, F. (2011). On the numerical evaluation of distributions in random matrix theory: A review. *Markov Process. Related Fields* **16** 803–866.

[6] Choup, L.N. (2006). Edgeworth expansion of the largest eigenvalue distribution function of GUE and LUE. *Int. Math. Res. Not.* Art. ID 61049 1–33. MR2233711

[7] Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. MR1936554

[8] Edelman, A. and Persson, P.O. (2002). Numerical methods for eigenvalue distributions of random matrices. Technical report, Massachusetts Institute of Technology.

[9] El Karoui, N. (2006). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and $p/n \to \infty$. Available at arXiv:math/0309355v1.

[10] El Karoui, N. (2006). A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.* **34** 2077–2117. MR2294977

[11] Forrester, P.J. (2010). *Log-Gases and Random Matrices. London Mathematical Society Monographs Series* **34**. Princeton, NJ: Princeton Univ. Press. MR2641363

[12] Gohberg, I., Goldberg, S. and Krupnik, N. (2000). *Traces and Determinants of Linear Operators. Operator Theory: Advances and Applications* **116**. Basel: Birkhäuser. MR1744872

[13] Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge: Cambridge Univ. Press. MR0832183

[14] Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991

[15] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[16] Johnstone, I.M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

[17] Johnstone, I.M. and Ma, Z. (2010). Fast approach to the Tracy–Widom law at the edge of GOE and GUE. Preprint.

[18] Johnstone, I.M., Ma, Z., Perry, P.O. and Shahram, M. (2009). RMTstat: Distributions, statistics and tests derived from random matrix theory. R package version 0.2.

[19] Koev, P. (2011). Random matrix statistics. Unpublished manuscript.

[20] Koev, P. and Edelman, A. (2006). The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comp.* **75** 833–846. MR2196994

[21] Lax, P.D. (2002). *Functional Analysis*. New York: Wiley. MR1892228

[22] Ma, Z. (2010). Supplement to "Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices". Available at http://www-stat.wharton.upenn.edu/~zongming/research.html.

[23] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press. MR0560319

[24] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley. MR0652932

[25] Nadler, B. (2010). On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. Available at http://www.wisdom.weizmann.ac.il/~nadler/Publications/publications.html.
[26] Olver, F.W.J. (1974). *Asymptotics and Special Functions*. London: Academic Press. MR0435697
[27] Patterson, N., Price, A.L. and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet*. **2** e190.
[28] Paul, D. (2006). Distribution of the smallest eigenvalue of Wishart(N, n) when $N/n \to 0$. Unpublished manuscript.
[29] Roy, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann*. *Math*. *Statist*. **24** 220–238. MR0057519
[30] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J*. *Statist*. *Phys*. **108** 1033–1056. MR1933444
[31] Szegö, G. (1975). *Orthogonal Polynomials*, 4th ed. Providence, RI: Amer. Math. Soc.
[32] Tracy, C.A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm*. *Math*. *Phys*. **159** 151–174. MR1257246
[33] Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm*. *Math*. *Phys*. **177** 727–754. MR1385083
[34] Tracy, C.A. and Widom, H. (1998). Correlation functions, cluster functions, and spacing distributions for random matrices. *J*. *Statist*. *Phys*. **92** 809–835. MR1657844
[35] Tracy, C.A. and Widom, H. (2005). Matrix kernels for the Gaussian orthogonal and symplectic ensembles. *Ann*. *Inst*. *Fourier* (*Grenoble*) **55** 2197–2207. MR2187952
[36] Widom, H. (1999). On the relation between orthogonal, symplectic and unitary matrix ensembles. *J*. *Statist*. *Phys*. **94** 347–363. MR1675356

Received April 2009 and revised August 2010