FAST APPROACH TO THE TRACY–WIDOM LAW AT THE EDGE OF GOE AND GUE[1]

BY IAIN M. JOHNSTONE AND ZONGMING MA

Stanford University and University of Pennsylvania

We study the rate of convergence for the largest eigenvalue distributions in the Gaussian unitary and orthogonal ensembles to their Tracy–Widom limits.

We show that one can achieve an $O(N^{-2/3})$ rate with particular choices of the centering and scaling constants. The arguments here also shed light on more complicated cases of Laguerre and Jacobi ensembles, in both unitary and orthogonal versions.

Numerical work shows that the suggested constants yield reasonable approximations, even for surprisingly small values of N.

1. Introduction. The celebrated papers of Tracy and Widom (1994, 1996) described the limiting distributions of the largest eigenvalues of the Gaussian unitary and orthogonal ensembles (GUE and GOE), respectively. The purpose of this article is to show that an appropriate choice of centering and scaling allows us to establish a rate of convergence in these results, and further that this rate can be understood as "second order," being $O(N^{-2/3})$ rather than the $O(N^{-1/3})$ that would otherwise apply.

The Gaussian ensembles refer, as is usual, to eigenvalue densities of $x = (x_1, \ldots, x_N)$ given by

$$f(x) = c_{N\beta} \prod_{i=1}^{N} e^{-\beta x_i^2/2} \prod_{i<j} |x_i - x_j|^\beta,$$

with $\beta = 1$ corresponding to GOE_N and $\beta = 2$ to GUE_N, the subscript being shown only when clarity dictates. The corresponding matrix models specify that f is the density of the eigenvalues x of a symmetric or Hermitian random matrix M with independent entries on and above the diagonal, whose density function is given by

$$g(M) = c'_{N\beta} \exp\{-(\beta/2) \operatorname{tr} M^2\}.$$

Our principal rate of convergence results follow. The Tracy–Widom distributions are denoted $F_\beta(s)$ for $\beta = 1, 2$.

THEOREM 1. *Let $x_{(1)}$ denote the largest eigenvalue of a sample from* GUE_N, *and*

$$\mu_N = \sqrt{2N}, \qquad \tau_N = 2^{-1/2}N^{-1/6}.$$

Given s_0, there exists $C = C(s_0)$ such that for $s \geq s_0$,

$$\left| P\{(x_{(1)} - \mu_N)/\tau_N \leq s\} - F_2(s) \right| \leq CN^{-2/3}e^{-s}.$$

THEOREM 2. *Let $x_{(1)}$ denote the largest eigenvalue of a sample from* GOE_{N+1}, *with $N + 1$ even, and*

(1) $$\mu_N = \sqrt{2N+1}, \qquad \tau_N = 2^{-1/2}N^{-1/6}.$$

Given s_0, there exists $C = C(s_0)$ such that for $s \geq s_0$,

(2) $$\left| P\{(x_{(1)} - \mu_N)/\tau_N \leq s\} - F_1(s) \right| \leq CN^{-2/3}e^{-s/2}.$$

We use index $N + 1$ (rather than N) because of a key formula relating the Gaussian orthogonal ensemble GOE_{N+1} to the Gaussian unitary ensemble GUE_N, (36) below. The centering and scaling constants carry subscripts N rather than $N + 1$ for this reason.

Our interest in these results is threefold. First, they provide the simplest case of a class of such $O(N^{-2/3})$ convergence results for the classical orthogonal polynomial ensembles—the other two being the Laguerre and Jacobi ensembles—in both orthogonal and unitary versions. These results are of interest in statistics because they show that the Tracy–Widom approximation is accurate enough to replace exact evaluation of the finite LOE and JOE probabilities for many applied purposes where highly accurate values are not necessary [Johnstone (2009)]. The results of this paper focus on the corresponding phenomenon for the simplest case of GUE and GOE. Since the LOE and JOE proofs are lengthy analyses with Laguerre and Jacobi polynomial asymptotics, respectively, this paper outlines the approach in the simplest case.

Second, our interest was stimulated by Choup (2009), which provided the leading terms in an Edgeworth expansion of the largest eigenvalue distribution of GOE, and remarked that the $N^{-1/3}$ correction term does not vanish in GOE. As our earlier results on $O(N^{-2/3})$ convergence for LOE and JOE would suggest, a similar $O(N^{-2/3})$ property for GOE with a suitable specific centering, it seemed, therefore, of interest to verify the conjecture in this setting. Although we subsequently learned of an error in the argument of Choup (2009) (private communication), it was an important stimulus for this work.

Third, we find it of interest that adjustment of μ_N and τ_N to secure $O(N^{-2/3})$ convergence yields an approximation, that is, adequate—for some purposes—for surprisingly small values of N.

TABLE 1

GUE approximation. For each percentile α shown in the top row, let $s_{2\alpha}$ be the quantile
$F_2(s_{2\alpha}) = \alpha$, and $s_{N\alpha} = \mu_N + \tau_N s_{2\alpha}$ for μ_N and τ_N specified in Theorem 1. Table entries are
$P\{x_{(1)} \leq s_{N\alpha}\}$ computed using Bornemann's code for $E_2^{(n)}(0; [s_{N\alpha}, \infty))$

N	μ_N	0.01	0.05	0.1	0.3	0.5	0.7	0.9	0.95	0.99
2	2.000	0.026	0.087	0.149	0.359	0.549	0.732	0.913	0.958	0.992
5	3.162	0.017	0.068	0.125	0.331	0.526	0.717	0.907	0.954	0.991
10	4.472	0.014	0.061	0.115	0.319	0.516	0.711	0.905	0.953	0.991
25	7.071	0.012	0.056	0.108	0.310	0.509	0.706	0.902	0.951	0.991
50	10.000	0.011	0.054	0.105	0.307	0.506	0.704	0.902	0.951	0.990
75	12.247	0.011	0.053	0.104	0.305	0.504	0.703	0.901	0.951	0.990
100	14.142	0.011	0.052	0.103	0.304	0.504	0.702	0.901	0.951	0.990
200	20.000	0.011	0.051	0.102	0.303	0.502	0.701	0.901	0.950	0.990
500	31.623	0.010	0.051	0.101	0.301	0.501	0.701	0.900	0.950	0.990

To illustrate, first in GUE, Table 1 shows the exact probabilities $P\{x_{(1)} \leq \mu_N + \tau_N s_{2\alpha}\}$ for quantiles $s_{2\alpha}$ of the limiting F_2 distribution, computed using the finite GUE function provided in the MATLAB toolbox RMTFredholm [Bornemann (2010)].

In fact, our proof suggests a slightly different centering value, $\mu_N = (\sqrt{2N-1} + \sqrt{2N+1})/2$, which differs from $\sqrt{2N}$ in relative terms by only $O(N^{-4})$. However, Table 2 shows an observable improvement at very small values of N.

Our interest is primarily with GOE, for which software for exact computation appears to be as yet unavailable. Table 3 shows Monte Carlo simulations of $P\{x_{(1)} \leq \mu_N + \tau_N s_{1\alpha}\}$ for quantiles $s_{1\alpha}$ of the F_1 limit, based on $R = 10^6$ replications. The corresponding 95% confidence intervals have half-width $2\sqrt{p_\alpha(1 - p_\alpha)} \times 10^{-3}$ which decreases from 0.001 at $p_\alpha = 0.5$ to 0.0002 at $p_\alpha = 0.01$ and 0.99. Thus the tabulated values should be correct to within ±0.001.

Two features of the numerical results deserve note. First the approximations are somewhat better in the near right tail than in the left. This is presumably because

TABLE 2

GUE approximation: as for Table 1, but with $\mu_N = (\sqrt{2N-1} + \sqrt{2N+1})/2$

N	μ_N	0.01	0.05	0.1	0.3	0.5	0.7	0.9	0.95	0.99
2	1.984	0.025	0.083	0.143	0.349	0.538	0.723	0.909	0.955	0.992
5	3.158	0.017	0.067	0.123	0.328	0.523	0.715	0.906	0.954	0.991
10	4.471	0.014	0.061	0.115	0.318	0.515	0.710	0.904	0.952	0.991
25	7.071	0.012	0.056	0.108	0.310	0.508	0.706	0.902	0.951	0.990
50	10.000	0.011	0.054	0.105	0.306	0.505	0.704	0.901	0.951	0.990

TABLE 3

GOE approximation. Let quantiles $s_{1\alpha}$ be defined by $F_1(s_{1\alpha}) = \alpha$ for values of α shown in the top row, and μ_N and τ_N be given by (1). *Based on $R = 10^6$ replications drawn from GOE, table entries are the fraction of replications of $s_{(1)} = (x_{(1)} - \mu_N)/\tau_N$ satisfying $s_{(1)} \leq s_{1\alpha}$*

$N+1$	0.01	0.05	0.1	0.3	0.5	0.7	0.9	0.95	0.99
2	0.010	0.045	0.090	0.279	0.483	0.698	0.914	0.963	0.995
5	0.012	0.053	0.103	0.300	0.500	0.704	0.907	0.956	0.993
10	0.011	0.053	0.103	0.302	0.502	0.703	0.904	0.954	0.992
25	0.011	0.052	0.103	0.302	0.502	0.702	0.902	0.952	0.991
50	0.011	0.052	0.102	0.301	0.501	0.701	0.901	0.951	0.991
75	0.011	0.051	0.102	0.302	0.501	0.701	0.901	0.951	0.990
100	0.010	0.051	0.101	0.301	0.501	0.701	0.901	0.951	0.990
200	0.011	0.051	0.101	0.301	0.501	0.701	0.901	0.951	0.990
500	0.010	0.050	0.100	0.300	0.500	0.700	0.901	0.951	0.990

the underlying approximation of Hermite polynomials by the Airy function is anchored at the turning point 0 of the Airy equation $A''(s) = s A(s)$, which lies in the right tail at about the 83rd percentile of F_1 and the 97th percentile of F_2.

Second, the errors in Tables 1–3 all have the same sign, suggesting that a further shift in the approximating distribution might improve accuracy. We experimented in GOE with small changes of the form, setting $N_+ = N + 1/2$,

$$(3) \qquad \mu_N(\gamma) = (2N_+ - \gamma N_+^{-1/3})^{1/2}, \qquad \tau_N(c) = 2^{-1/2}(N + c)^{-1/6},$$

and obtained good results, Table 4 and Figure 1, for $\gamma = 1/5$ and $c = 1$.

TABLE 4

GOE approximation. As for Table 3, but using $s'_{(1)}$ defined using $\mu_N(\gamma)$ and $\tau_N(c)$ given by (3), *with $\gamma = 1/5$ and $c = 1$*

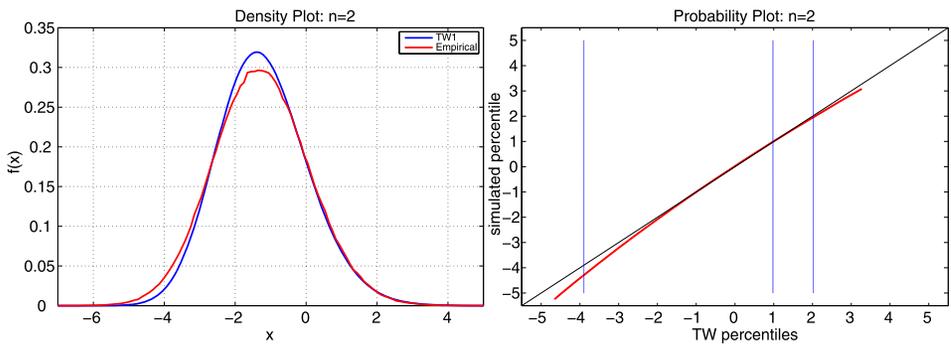| $N+1$	0.01	0.05	0.1	0.3	0.5	0.7	0.9	0.95	0.99
2	0.022	0.073	0.127	0.319	0.505	0.696	0.897	0.950	0.991
3	0.018	0.067	0.120	0.315	0.505	0.699	0.899	0.951	0.991
4	0.017	0.063	0.116	0.311	0.505	0.699	0.900	0.951	0.991
5	0.015	0.061	0.114	0.310	0.504	0.700	0.901	0.951	0.991
10	0.013	0.056	0.107	0.305	0.502	0.700	0.901	0.951	0.991
25	0.011	0.053	0.104	0.302	0.501	0.700	0.901	0.951	0.990
50	0.011	0.052	0.102	0.301	0.500	0.699	0.900	0.950	0.990
75	0.011	0.052	0.102	0.302	0.500	0.700	0.900	0.950	0.990
100	0.010	0.051	0.101	0.301	0.500	0.700	0.900	0.951	0.990
200	0.011	0.051	0.101	0.301	0.500	0.700	0.900	0.950	0.990
500	0.010	0.050	0.100	0.300	0.500	0.700	0.901	0.951	0.990

FIG. 1.   *Left panel*: *estimated density function for* $s'_{(1)} = (x_{(1)} - \mu'_N(1/5))/\tau_N(1)$, *compare* (3), *based on* $R = 10^6$ *samples from* GOE_2, *compared to the Tracy–Widom density function* f_1. *Right panel*: *probability plot of percentiles of the* F_1 *distribution on horizontal axis versus ordered values of* $s_{(1)}$ *on the vertical axis, based on the same samples from* GOE_2. *Vertical lines mark the* 1*st*, 95*th and* 99*th percentiles of* F_1.

These values differ from μ_N and τ_N of Theorem 2 by by relative errors of $O(N^{-4/3})$ and $O(N^{-1})$, respectively, and so have no effect on the validity of Theorem 2. However, they provide a substantial numerical improvement, especially in the right tail for values of N below 10. Indeed, for some purposes, the approximation in the right tail would be adequate, even for $N = 2$.

Outline of proof. We use the operator norm convergence framework developed in Tracy and Widom (2005); our focus, of course, is on achieving the second order convergence rate results. We use the Fredholm determinant representations for the finite and limiting distribution functions in terms of the two-point correlation kernels. A bound of Seiler–Simon, along with its orthogonal case analog, bounds the difference in Fredholm determinants in terms of the kernels. In turn, the kernels have integral representations in terms of weighted Hermite polynomials, and so we transfer bounds on convergence of Hermite polynomials to the Airy function to bounds on the kernels and hence to bounds on the probabilities.

Convenient uniform bounds on the convergence of weighted Hermite polynomials to the Airy function come from Liouville–Green theory, which analyzes convergence of the solutions of the second-order differential equation satisfied by the Hermite polynomials to those of the equation for the Airy function.

The correlation kernels for finite N involve polynomials of both degrees N and $N - 1$, each with its own Liouville–Green centering u_N and u_{N-1}. The overall centering μ_N for the kernel and distribution function must combine u_N and u_{N-1} appropriately to ensure that the generic $O(N^{-1/3})$ error terms cancel to uncover $O(N^{-2/3})$ convergence. In the unitary case, simple averaging suffices: $\mu_N = (u_N + u_{N-1})/2$. For the orthogonal setting, we use a formula expressing the GOE_{N+1} kernel in terms of the GUE_N kernel plus a rank one kernel, and obtain cancellation of $O(N^{-1/3})$ errors from these two components.

The Hermite polynomial approximation results are summarized in Section 2. The unitary proof, in Section 3, is a necessary preparation for the orthogonal case in Section 4.

Reproducible code: MATLAB files to produce the figures and tables are available at the second author's website.

Related work. Convergence rate results at $O(N^{-2/3})$ were obtained by El Karoui (2006) for LUE, Johnstone (2008) for JUE and JOE, and Ma (2012) for LOE. The study of Edgeworth-type expansions for GUE and LUE was initiated by Choup (2006, 2008), who noted that $N^{-1/3}$ terms in these expansions can be removed by specific choices of the centering constant.

The Tracy–Widom limit laws for the largest eigenvalue hold much more generally—such universality results are an active subject of research. For Hermitian Wigner matrices, see Tao and Vu (2010) and references therein, and for covariance matrices Soshnikov (2002) and Péché (2009).

2. Hermite polynomial asymptotics. The Hermite polynomials, $H_k(x)$ in notation of Szegő [(1967), Chapter 4], are orthogonal with respect to the weight function $w(x) = e^{-x^2}$ on $(-\infty, \infty)$. The "oscillator wave functions" are normalized, weighted versions

$$\phi_k(x) = h_k^{-1/2} e^{-x^2/2} H_k(x),$$

with $h_k = \int H_k^2(x) e^{-x^2} \, dx = \sqrt{\pi} 2^k k!$

Classical Plancherel–Rotach asymptotics for $H_N(x)$ near the largest zero, Szegő [(1967), page 201] and Anderson, Guionnet and Zeitouni [(2010), Section 3.7.2], establish that, for $m_N = \sqrt{2N}$ and $\tau_N = 2^{-1/2} N^{-1/6}$,

$$(4) \qquad (2N)^{1/4} \tau_N \phi_N(m_N + s\tau_N) \to A(s),$$

where throughout we use A to denote the Airy function Ai.

We will need to explicitly bound the error in the convergence in (4). There is now a substantial literature on asymptotic approximations to Hermite polynomials, using, for example, the steepest descent method for integrals [e.g., Shi (2008)], the nonlinear steepest descent method for Riemann–Hilbert problems [e.g., Wong and Zhang (2007)] and recurrence relations [e.g., Wang and Wong (2011)]. Much of this recent attention has focused on expansions for $H_N(\sqrt{2N+1}\xi)$ and $\phi_N(\sqrt{2N+1}\xi)$ that are valid uniformly for large regions of ξ.

For this work, however, we need more detailed information for $\xi = 1 + \sigma_N s$ near 1, and specifically uniform bounds for the error of Airy approximation for both ϕ_N *and its derivative* that have exponential decay in the variable s and rate $N^{-2/3}$; cf. Proposition 1 below. We have not found this extra detail explicitly in the literature, and since the Liouville–Green discussion of Olver [(1974), Chapter 11] comes with ready-made bounds for approximation error for both ϕ_N and ϕ_N', we use this as a starting point for extracting, in the Appendix, the specific bounds we

need. In this section, we explain just enough of the approach to describe the bounds
we need.

The Liouville–Green (LG) approach relies on the fact that Hermite polynomials,
and hence ϕ_N, satisfy a second order differential equation,

$$(5) \qquad \phi_N''(x) = \{x^2 - (2N+1)\}\phi_N(x).$$

Rescaling the x axis via $x = \sqrt{2N+1}\,\xi$, and setting $w_N(\xi) = \phi_N(x)$, the equation
takes the form

$$(6) \qquad w_N''(\xi) = \kappa_N^2 f(\xi) w_N(\xi),$$

with

$$\kappa_N = 2N+1, \qquad f(\xi) = \xi^2 - 1.$$

The *turning points* of the differential equation are the zeros of f, namely
$\xi_\pm = \pm 1$, so named because each separates an interval in which the solution is of
exponential type from one in which the solution oscillates. The LG transformation
introduces new independent and dependent variables ζ and W via the equations

$$(7) \qquad \zeta\left(\frac{d\zeta}{d\xi}\right)^2 = f(\xi), \qquad W = \left(\frac{d\zeta}{d\xi}\right)^{1/2} w_N.$$

More precisely, we take

$$(8) \qquad (2/3)\zeta^{3/2}(\xi) = \int_1^\xi \sqrt{f(\xi')}\, d\xi'.$$

The transform W approximately satisfies the Airy equation $W''(\zeta) = \kappa_N^2 \zeta W(\zeta)$,
which has linearly independent solutions in terms of Airy functions, traditionally
denoted by $\mathrm{Ai}(\kappa^{2/3}\zeta)$ and $\mathrm{Bi}(\kappa^{2/3}\zeta)$. Our interest lies in approximating the *reces-
sive* solution $\mathrm{Ai}(\kappa^{2/3}\zeta)$.

As described in more detail in the Appendix, the error in the Liouville–Green
approximation can be bounded, and one arrives at

$$(9) \qquad \bar{\phi}_N(x) = (2N)^{1/4}\tau_N\phi_N(x) = \bar{e}_N r(\xi)\{A(\kappa_N^{2/3}\zeta) + O(N^{-1})\}.$$

Here $r(\xi) = [\dot\zeta(\xi)/\dot\zeta(1)]^{-1/2}$ is approximately 1 for ξ near 1, and $\bar{e}_N = 1 +
O(N^{-1})$. This is, then, a version of (4) with an error term of order $O(N^{-1})$, but
with Airy function argument $\kappa_N^{2/3}\zeta$ rather than s.

We focus on x near $u_N = \sqrt{2N+1}$, that is, on ξ near the upper turning point
$\xi_+ = 1$. Introduce the rescaled variable s through $\xi = 1 + \sigma_N s$. To more closely
match the result (4), we want σ_N to be chosen so that the Airy function argument

$$(10) \qquad \kappa_N^{2/3}\zeta(1 + \sigma_N s) \approx s$$

for s in a suitably large range. A Taylor expansion of the left-hand side yields

$$(11) \qquad \kappa_N^{2/3}\big(\zeta(1) + \sigma_N \dot\zeta(1)s + \tfrac{1}{2}\sigma_N^2 s^2 \ddot\zeta^*\big).$$
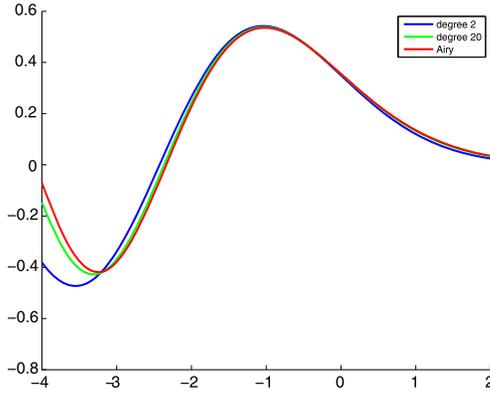
FIG. 2. *Plots of $\bar{\phi}_N(u_N + \tau_N s)$ for $N = 2, 20$ compared with the Airy function $A(s)$.*

Since $\zeta(1) = 0$ and $\dot{\zeta}(1) = 2^{1/3}$, as follows from (7) and (8), we obtain (10) by any choice of the form

$$\sigma_N = \tfrac{1}{2} N^{-2/3}(1 + o(1)).$$

For such a choice, (76) shows that

$$\kappa_N^{2/3} \zeta(1 + \sigma_N s) = s + O(s^2 N^{-2/3}).$$

Thus, to replace $A(\kappa_N^{2/3} \zeta)$ in (9) by $A(s)$ entails, in general, accepting an error term of $O(N^{-2/3})$ instead of $O(N^{-1})$, and so we use this error scale henceforth.

With the specific choice $\sigma_N = \tau_N/u_N$, we will show that for $s \geq s_L$,

$$|\bar{\phi}_N(u_N + s\tau_N) - A(s)| \leq C N^{-2/3} e^{-s/2}.$$

Figure 2 shows that, for values of s corresponding to the bulk of the support of F_1, the approximation is tolerably good even for $N = 2$.

In fact, since the two-point correlation functions depend on both ϕ_N and ϕ_{N-1}, we need such approximations both for $\bar{\phi}_N$ and for $\bar{\phi}_{N-1}$. For $\bar{\phi}_{N-1} = (2N)^{1/4} \tau_N \phi_{N-1}$, the corresponding turning point is at $u_{N-1} = \sqrt{2N - 1}$, though we still use the same scale factor τ_N. We use the notation $\bar{\phi}_{Nj}$, with $Nj = N$ or $N - 1$, respectively, to refer to both cases. In addition, for the GOE case, bounds on the convergence of the derivative is also required. In the Appendix, we establish

PROPOSITION 1. *Let $s_L \in \mathbb{R}$. For $s \geq s_L$, we have*

$$(12) \qquad |\bar{\phi}_{Nj}(u_{Nj} + s\tau_N)| \leq C e^{-s},$$

$$(13) \qquad |\bar{\phi}_{Nj}(u_{Nj} + s\tau_N) - A(s)| \leq C N^{-2/3} e^{-s/2},$$

where the error bounds are uniform in $s \geq s_L$ and $N \geq N_0(s_L)$. The same bounds hold, with modified constants C, when $\bar{\phi}_{Nj}$ and A are replaced by $\tau_N \bar{\phi}'_{Nj}$ and A'.

We record here also some corresponding exponential decay bounds for the Airy function and its derivatives $A^{(i)}$. Indeed, given s_L, there exist constants $C_i(s_L)$ such that

$$(14) \qquad \left| s^k A^{(i)}(s) \right| \leq C_i(s_L) e^{-s}, \qquad s \geq s_L, i = 0, 1, 2.$$

Proposition 1 provides good Airy approximations for both $\bar{\phi}_N$ and $\bar{\phi}_{N-1}$, but with differing centering values, u_N and u_{N-1}, respectively. To obtain scaling limits for the correlation kernels in GUE and GOE, we need to combine these centerings in a manner appropriate to each case.

It is convenient to express these centering shifts in the rescaled variable s. Thus, set

$$(15) \qquad \begin{aligned} \phi_\tau(s; k) &= \bar{\phi}_N\big(u_N + \tau_N(s + k\Delta_N)\big), \\ \psi_\tau(s; l) &= \bar{\phi}_{N-1}\big(u_{N-1} + \tau_N(s + l\Delta_N)\big), \end{aligned}$$

where

$$\Delta_N = (u_N - u_{N-1})/\tau_N = N^{-1/3}\big(1 + 2^{-5}N^{-2} + O(N^{-4})\big)$$

—indeed $2^{1/3}\Delta_2 = 1.0080!$ We obtain extensions of Proposition 1: indeed from (12),

$$\left| \phi_\tau(s; k) \right| \leq C e^{-(s + k\Delta_N)} \leq C e^{-s},$$

and a similar bound holds for $\left| \psi_\tau(s; k) \right|$. From (15) and Proposition 1,

$$(16) \qquad \begin{aligned} \phi_\tau(s; k) &= A(s + k\Delta_N) + O(N^{-2/3} e^{-s/2}) \\ &= A(s) + k\Delta_N A'(s) + O(N^{-2/3} e^{-s/2}), \end{aligned}$$

since $\frac{1}{2}(k\Delta_N)^2 |A''(s^*)| \leq C N^{-2/3} e^{-s}$ using (14) and $|s^* - s| \leq k\Delta_N$.

More generally, but by identical arguments, for $r = 0, 1$ we have

$$(17) \qquad \phi_\tau^{(r)}(s; k) = A^{(r)}(s) + k\Delta_N A^{(r+1)}(s) + O(N^{-2/3} e^{-s/2}),$$

and correspondingly for $\psi_\tau^{(r)}(s; k)$. As a byproduct of a steepest descent analysis for Laguerre polynomials, Choup (2006, 2008) derived a three-term asymptotic expansion for ϕ_N and ϕ_{N-1} whose first two terms agree with (16), though without the uniform error bounds in N and s of Proposition 1.

3. Unitary case.

PROOF OF THEOREM 1. The argument has three components: first we recall determinantal representations of the eigenvalue probabilities $F_{N,2}(x_0)$ and limiting value $F_2(s_0)$, along with integral representations of the associated correlation kernels. Second we set up the rescaling that connects x_0 and s_0, and finally establish the convergence bounds.

The two-point correlation kernel for GUE_N

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

has a useful integral representation [Tracy and Widom (1996), equation (57)]. Set[2]

(18) $$\phi(x) = (2N)^{1/4}\phi_N(x), \qquad \psi(x) = (2N)^{1/4}\phi_{N-1}(x),$$

then

(19) $$S_{N,2}(x, y) = \frac{1}{2}\int_0^\infty [\phi(x + z)\psi(y + z) + \psi(x + z)\phi(y + z)]\,dz.$$

The distribution of $x_{(1)}$ may be expressed as a Fredholm determinant

(20) $$F_{N,2}(x_0) = P\left\{\max_{1 \le k \le N} x_k \le x_0\right\} = \det(I - S_{N,2}\chi_0),$$

where $\chi_0(x) = I_{(x_0,\infty)}(x)$ and the operator $S_{N,2}\chi_0$ is defined via

$$(S_{N,2}\chi_0)g(x) = \int_{x_0}^\infty S_{N,2}(x, y)g(y)\,dy.$$

Equivalently, we may speak of $S_{N,2}$ as an operator on $L_2(x_0, \infty)$ with kernel $S_{N,2}(x, y)$. On this understanding, we drop further explicit reference to χ_0.

Now change variables, setting $x = \tau(s) = \mu_N + \tau_N s$, with μ_N yet to be determined, and $x_0 = \tau(s_0)$. Set also

(21) $$S_\tau(s, t) = \tau_N S_{N,2}(\mu_N + \tau_N s, \mu_N + \tau_N t).$$

Defining

(22) $$\phi_\tau(s) = \tau_N \phi(\mu_N + \tau_N s), \qquad \psi_\tau(s) = \tau_N \psi(\mu_N + \tau_N s),$$

it is clear that (19) becomes

(23) $$S_\tau(s, t) = \frac{1}{2}\int_0^\infty [\phi_\tau(s + z)\psi_\tau(t + z) + \psi_\tau(s + z)\phi_\tau(t + z)]\,dz.$$

Since $S_{N,2}$ and S_τ have the same eigenvalues, $\det(I - S_{N,2}) = \det(I - S_\tau)$, and so

(24) $$P\{(\max x_k - \mu_N)/\tau_N \le s_0\} = \det(I - S_\tau).$$

Tracy and Widom (1994) showed that the limiting distribution F_2 also has a determinantal representation

$$F_2(s_0) = \det(I - S_A),$$

---

[2]Note: our definitions differ by a factor $\sqrt{2}$ from those of Tracy and Widom.

where S_A denotes the Airy operator on $L^2(s_0, \infty)$ with the kernel having the form

$$(25) \qquad S_A(s, t) = \int_0^\infty A(s + z)A(t + z)\,dz.$$

To derive bounds on the convergence of $F_{N,2}(x_0)$ to $F_2(s_0)$, we use a bound due to Seiler and Simon (1975),

$$(26) \quad |\det(I - S_\tau) - \det(I - S_A)| \leq \|S_\tau - S_A\|_1 \exp(\|S_\tau\|_1 + \|S_A\|_1 + 1).$$

Here $\|\cdot\|_1$ denotes trace class norm on operators on $L^2(s_0, \infty)$. This bound reduces the convergence question to study of convergence of the kernel $S_\tau(s, t)$ to $S_A(s, t)$.

Given functions a and b, denote by $a \diamond b$ the operator having kernel

$$(a \diamond b)(s, t) = \int_0^\infty a(s + z)b(t + z)\,dz.$$

In this notation, the kernel difference becomes

$$S_\tau - S_A = \tfrac{1}{2}(\phi_\tau \diamond \psi_\tau + \psi_\tau \diamond \phi_\tau) - A \diamond A.$$

To facilitate convergence arguments, we rewrite this as

$$8(S_\tau - S_A) = (\phi_\tau + \psi_\tau + 2A) \diamond (\phi_\tau + \psi_\tau - 2A)$$
$$+ (\phi_\tau + \psi_\tau - 2A) \diamond (\phi_\tau + \psi_\tau + 2A) - (\phi_\tau - \psi_\tau) \diamond (\phi_\tau - \psi_\tau).$$

Recall that the centering constant μ_N was left unspecified in the definitions of ϕ_τ and ψ_τ in (22). We now choose μ_N so that each term in the preceding decomposition is $O(N^{-2/3})$. This amounts to choosing the shifts k and l in (15) to satisfy two constraints. First, the centerings $\mu_N = u_N + k\tau_N \Delta_N$ and $\mu_N = u_{N-1} + l\tau_N \Delta_N$ must agree, so that necessarily $l = k + 1$. Second, the $N^{-1/3}$ term must drop out in the expansion for $\phi_\tau + \psi_\tau$ given by (17), so that $l = -k$. We therefore must have, for the present unitary case,

$$(27) \qquad \begin{aligned} \phi_\tau(s) &= \phi_\tau(s; -\tfrac{1}{2}), \\ \psi_\tau(s) &= \psi_\tau(s; \tfrac{1}{2}) \end{aligned}$$

which entails that $\mu_N = (u_N + u_{N-1})/2$ as was used in Table 2. From Proposition 1 and the succeeding discussion we obtain

COROLLARY 1 (Complex Case). *Let ϕ_τ and ψ_τ be defined by (27) and (15). Given $s_L \in \mathbb{R}$, there exists $C = C(s_L)$ such that for $N \geq N(s_L)$ and $s \geq s_L$,*

$$(28) \qquad\qquad |\phi_\tau(s)|, |\psi_\tau(s)| \leq Ce^{-s},$$

$$(29) \qquad |\phi_\tau(s) - A(s)|, |\psi_\tau(s) - A(s)| \leq CN^{-1/3}e^{-s/2},$$

$$(30) \qquad\qquad |\phi_\tau(s) + \psi_\tau(s) - 2A(s)| \leq CN^{-2/3}e^{-s/2}.$$

We will need some simple bounds for certain norms of $a \diamond b$. In the unitary case, we need the trace norm of $a \diamond b$ as an operator on $L^2(s_0, \infty)$. In the orthogonal case, we need the weighted L^2-spaces $L^2((s_0, \infty), \rho(s) \, ds)$ and $L^2((s_0, \infty), \rho^{-1}(s) \, ds)$ for a weight function ρ such that the reciprocal $\rho^{-1} \in L^1(\mathbb{R})$. Further details are given in Section 4. For some $\gamma \geq 0$, let

$$(31) \qquad \rho(s) = e^{\gamma |s|}.$$

In this section $\gamma = 0$, while values of $\gamma > 0$ will be specified later for GOE.

PROPOSITION 2. *Let weight functions ρ_1, ρ_2 be chosen from $\{\rho, 1/\rho\}$, where ρ is given by* (31), *and consider the Hilbert–Schmidt norm of operator $a \diamond b : L^2(\rho_2) \to L^2(\rho_1)$. Assume that, for $s \geq s_0$,*

$$|a(s)| \leq a_N e^{-a_1 s}, \qquad |b(s)| \leq b_N e^{-b_1 s}.$$

If $0 \leq \gamma < 2 \min(a_1, b_1)$, then

$$(32) \qquad \|a \diamond b\|_{HS} \leq C \frac{a_N b_N}{a_1 + b_1} e^{-(a_1 + b_1)s_0 \pm \gamma |s_0|},$$

where $C = C(a_1, b_1, \gamma) = [(a_1 - \gamma/2)(b_1 - \gamma/2)]^{-1/2}$. If $\rho_1 = \rho_2$, then the trace norm satisfies the same bound.

This is a special case of Johnstone [(2008), Lemma 7]. In the present unitary case, we apply Proposition 2, with $\gamma = 0$, to bound the trace norm of each term on the right-hand side, using (28)–(30). For each of the three terms, we find that $a_1 + b_1 \geq 1$ and $a_N b_N \leq C N^{-2/3}$, so that

$$\|S_\tau - S_A\|_1 \leq C N^{-2/3} e^{-s_0}.$$

We may similarly conclude that

$$\|S_\tau\|_1 \leq C e^{-2s_0}, \qquad \|S_A\|_1 \leq C e^{-2s_0}.$$

Indeed, bounds (28) and (14) show that in each case, a_N and $b_N \leq C(s_0)$, and that $a_1 = b_1 = 1$.

Combining the two previous displays with the Seiler–Simon bound (26), we obtain Theorem 1. $\square$

4. Orthogonal case.

To establish Theorem 2, we again follow the outline of proof given in Section 1.

$1°$. Assume that $N + 1$ is even. Tracy and Widom (1998) gave a derivation[3] of the determinant representation

$$(33) \qquad P\left\{ \max_{1 \leq k \leq N+1} x_k \leq x_0 \right\} = \sqrt{\det(I - K_{N+1}\chi_0)}.$$

---

[3] Sinclair (2009) extended Tracy and Widom's derivation to cover $N + 1$ odd, but we do not pursue this here. See also Forrester and Mays (2009).

Here K_{N+1} is a 2×2-matrix valued operator

(34) $K_{N+1}(x, y) = (LS_{N+1,1})(x, y) + K^\varepsilon(x, y),$

where

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

Here ∂_2 denotes the operator of partial differentiation with respect to the second variable, and ε_1 the operator of convolution in the first variable with the function $\varepsilon(x) = \frac{1}{2} \operatorname{sgn}(x)$. Thus $(\varepsilon_1 S)(x, y) = \int \varepsilon(x - u) S(u, y) \, du$. Finally T denotes transposition of variables $TS(x, y) = S(y, x)$. The scalar kernel

$$S_{N+1,1}(x, y) = \sum_{n=0}^{N} \phi_n(x)\phi_n(y) + \sqrt{\frac{N+1}{2}}\phi_N(x)\varepsilon\phi_{N+1}(y),$$

and Adler et al. (2000) observe that it may be rewritten as

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

where ϕ and ψ are as defined at (18). The orthogonal kernel is thus expressed in terms of the unitary kernel and a rank one remainder term. The formula allows convergence results from the unitary case to be reused, with relatively minor modification.

2°. The limiting distribution has a corresponding determinantal representation

$$F_1(s_0) = \sqrt{\det(I - K_{\mathrm{GOE}})}.$$

To state the Tracy and Widom (2005) form for K_{GOE}, and for the convergence argument to follow, it is helpful to rewrite expressions involving ε in terms of the right-tail integration operator $(\tilde{\varepsilon} g)(s) = \int_s^\infty g(u) \, du$ and for kernels $A(s, t)$ in the form $(\tilde{\varepsilon}_1 A)(s, t) = \int_s^\infty A(u, t) \, du$. This is due to the oscillatory behavior of the Airy function in the *left* tail. We write $A \otimes B$ for the operator whose kernel is $A(s)B(t)$. The Tracy–Widom expression states that

(37) $K_{\mathrm{GOE}}(s, t) = \begin{pmatrix} S(s, t) & SD(s, t) \\ IS(s, t) - \varepsilon(s - t) & S(t, s) \end{pmatrix},$

and the entries of K_{GOE} are given by

$$S(s, t) = \left(S_A - \tfrac{1}{2}A \otimes \tilde{\varepsilon} A\right)(s, t) + \tfrac{1}{2}A(s),$$

(38) $SD(s, t) = -\partial_2\left(S_A - \tfrac{1}{2}A \otimes \tilde{\varepsilon} A\right)(s, t),$

$$IS(s, t) = -\tilde{\varepsilon}_1\left(S_A - \tfrac{1}{2}A \otimes \tilde{\varepsilon} A\right)(s, t) - \tfrac{1}{2}(\tilde{\varepsilon} A)(s) + \tfrac{1}{2}(\tilde{\varepsilon} A)(t),$$

where S_A is the Airy kernel defined at (25).

Defining operator matrices

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

we may rewrite (37) in the form

$$(39) \qquad K_{\text{GOE}} = \tilde{L}\big(S_A - \tfrac{1}{2} A \otimes \tilde{\varepsilon} A\big) + \tfrac{1}{2} L_1 A(s) + \tfrac{1}{2} L_2 A(t) + K^{\varepsilon}.$$

3°. We turn to a linear rescaling of formulas (33) and (34). We again set $x = \tau(s) = \mu_N + \tau_N s$ and $y = \tau(t) = \mu_N + \tau_N t$, but now with $\mu_N = \mu_N^R$ to be determined anew in this orthogonal case; see 4° below. Define ϕ_τ and ψ_τ as before by (22); we occasionally write ϕ_τ^R and ψ_τ^R to emphasize the different centering. We have

$$
\begin{aligned}
(40) \qquad S_\tau^R(s,t) &:= \tau_N S_{N+1,1}(\tau(s), \tau(t)) \\
&= \tau_N S_{N,2}(\tau(s), \tau(t)) + \tfrac{1}{2} \tau_N \phi(\tau(s))(\varepsilon\psi)(\tau(t)) \\
&= S_\tau(s,t) + \tfrac{1}{2}\phi_\tau(s)(\varepsilon\psi_\tau)(t),
\end{aligned}
$$

where we have used $(\varepsilon\psi)(y) = (\varepsilon\psi_\tau)(t)$ for a linear rescaling.

Now $\det(I - K_{N+1}\chi_0) = \det(I - \bar{K}_\tau)$, where

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

where K^ε was defined at (35). Since $\det(I - \bar{K}_\tau)$ is unchanged if the lower left entry is divided by τ_N and the upper right entry multiplied by τ_N,

$$(41) \qquad\qquad \det(I - K_{N+1}\chi_0) = \det(I - K_\tau),$$

where K_τ is an operator with matrix kernel

$$(42) \qquad\qquad K_\tau(s,t) = (LS_\tau^R)(s,t) + K^\varepsilon(s,t).$$

Now we rewrite LS_τ^R using $\tilde{\varepsilon}$ and $\tilde{\varepsilon}_1$. First, define

$$(43) \qquad\qquad \beta_{N-1} = \frac{1}{2}\int_{-\infty}^\infty \psi_\tau = \frac{1}{2}(2N)^{1/4}\int_{-\infty}^\infty \phi_{N-1},$$

and observe that $\varepsilon\psi_\tau = \beta_{N+1} - \tilde{\varepsilon}\psi_\tau$. Thus

$$LS_\tau^R = L\big(S_\tau - \tfrac{1}{2}\phi_\tau \otimes \tilde{\varepsilon}\psi_\tau\big) + \tfrac{1}{2}\beta_{N-1} L(\phi_\tau \otimes 1).$$

Now $L = \tilde{L} + \begin{pmatrix} 0 & 0 \\ \varepsilon_1 + \tilde{\varepsilon}_1 & 0 \end{pmatrix}$ and $2(\varepsilon_1 + \tilde{\varepsilon}_1)$ amounts to integration over $\mathbb{R}$ in the first slot. From (23), after interchanging orders of integration and using $\int \phi_\tau = 0$, we obtain

$$\int_{-\infty}^\infty S_\tau(s,t)\,ds = \int_0^\infty \beta_{N-1}\phi_\tau(t+z)\,dz = \beta_{N-1}\tilde{\varepsilon}\phi_\tau(t),$$

and then

$$[(L - \tilde{L})(S_\tau - \phi_\tau \otimes \tilde{\varepsilon}\psi_\tau)]_{2,1} = \tfrac{1}{2}\beta_{N-1} \otimes \tilde{\varepsilon}\phi_\tau$$

as the only nonzero entry of the matrix on the left-hand side. Combining the last two displays with (42), we get

$$(44) \qquad K_\tau = \tilde{L}\big(S_\tau - \tfrac{1}{2}\phi_\tau \otimes \tilde{\varepsilon}\psi_\tau\big) + \tfrac{1}{2}\beta_{N-1}[L_1\phi_\tau(s) + L_2\phi_\tau(t)] + K^\varepsilon.$$

4°. We now look at the $(1, 1)$ terms in (39) and (44) in order to see, somewhat informally, how the choice $\mu_N^R = u_N$ leads to $O(N^{-2/3})$ convergence. Thus, we examine the difference

$$(45) \qquad \big[S_\tau - \tfrac{1}{2}\phi_\tau \otimes \tilde{\varepsilon}_\tau\big] - \big[S_A - \tfrac{1}{2}A \otimes \tilde{\varepsilon}A\big] + \tfrac{1}{2}[\beta_{N-1}\phi_\tau - A].$$

From definitions (15) and expansions (17), this choice of μ_N^R corresponds to

$$(46) \qquad \begin{aligned} \phi_\tau^R(s) &= \phi_\tau(s; 0) = A(s) + O(N^{-2/3}), \\ \psi_\tau^R(s) &= \psi_\tau(s; 1) = A(s) + \Delta_N A'(s) + O(N^{-2/3}). \end{aligned}$$

We write $A_N = A + \Delta_N A'$ and define

$$S_{A_N} = \tfrac{1}{2}(A \diamond A_N + A_N \diamond A).$$

From representation (23) and (46), $S_\tau = S_{A_N} + O(N^{-2/3})$, while the identity

$$S_{A_N} = S_A - \tfrac{1}{2}\Delta_N A \otimes A$$

follows from

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

Thus

$$S_\tau = S_A - \tfrac{1}{2}\Delta_N A \otimes A + O(N^{-2/3}).$$

Since $\tilde{\varepsilon}A' = -A$, we have $\tilde{\varepsilon}A_N = \tilde{\varepsilon}A - \Delta_N A$, and so

$$\tfrac{1}{2}\phi_\tau \otimes \tilde{\varepsilon}\psi_\tau = \tfrac{1}{2}A \otimes \tilde{\varepsilon}A - \tfrac{1}{2}\Delta_N A \otimes A + O(N^{-2/3}).$$

Forming the difference of the last two displays, we see an important cancellation of the $O(N^{-1/3})$ terms involving Δ_N, and hence that the first two terms of (45) together are $O(N^{-2/3})$.

A computation with the recursion relation for Hermite polynomials and then Stirling's formula [with its $O(N^{-1})$ error term] shows that, as $N \to \infty$,

$$\beta_{N-1} = \left(\frac{\pi N}{2}\right)^{1/4} \frac{\sqrt{(N-1)!}}{2^{(N-1)/2}((N-1)/2)!} = 1 + O(N^{-1}).$$

From this and (46), it follows that the final term of (45) is also $O(N^{-2/3})$.

5°. To prepare for the convergence argument for the 2×2 matrix kernels, we combine (39) and (44). Noting also from our considerations above that

$$S_A - \tfrac{1}{2} A \otimes \tilde{\varepsilon} A = S_{A_N} - \tfrac{1}{2} A \otimes \tilde{\varepsilon} A_N,$$

we obtain the basic difference representation

(47)
$$\begin{aligned}
K_\tau - K_{\mathrm{GOE}} &= \tilde{L}(S_\tau - S_{A_N}) - \tfrac{1}{2}\tilde{L}(\phi_\tau \otimes \tilde{\varepsilon}\psi_\tau - A \otimes \tilde{\varepsilon} A_N) \\
&\quad + \tfrac{1}{2}L_1[\beta_{N-1}\phi_\tau - A](s) + \tfrac{1}{2}L_2[\beta_{N-1}\phi_\tau - A](t),
\end{aligned}$$

from which we may expect to show $O(N^{-2/3})$ convergence, in view of the fact that $\phi_\tau, \psi_\tau, S_\tau$ and β_{N-1} merge, respectively, with A, A_N, S_{A_N} and 1 at rates of at least $O(N^{-2/3})$.

6°. We now turn to study the convergence of

(48)
$$F_{N+1,1}(s_0) = P\{(x_{(1)} - \mu_N)/\tau_N \le s_0\} = \sqrt{\det(I - K_\tau)}$$

to $F_1(s_0) = \sqrt{\det(I - K_{\mathrm{GOE}})}$. Tracy and Widom (2005) describe with some care the nature of the operator convergence of K_{N+1} to K_{GOE} for the Gaussian finite N ensemble. We adopt their framework of weighted L^2 spaces and regularized 2-determinants. Thus, let ρ be a weight function such that $\rho^{-1} \in L^1(\mathbb{R})$ and all $\phi_N \in L^2(\rho)$. Write $L^2(\rho)$ and $L^2(\rho^{-1})$ for the spaces $L^2((s_0, \infty), \rho(s)\,ds)$ and $L^2((s_0, \infty), \rho^{-1}(s)\,ds)$, respectively.

We consider K_τ and K_{GOE} as members of the collection $\mathcal{B}$ of 2×2 Hilbert–Schmidt operator matrices $B = (B_{ij}, i, j = 1, 2)$ on $L^2(\rho) \oplus L^2(\rho^{-1})$ whose diagonal entries are trace class. Note that $\varepsilon : L^2(\rho) \to L^2(\rho^{-1})$ as a consequence of the assumption that $\rho^{-1} \in L^1$. The specific ρ that we use is defined in (31) with $\gamma > 0$.

To analyze the convergence of $p_{N+1} = F_{N+1,1}(s_0)$ to $p_\infty = F_1(s_0)$, we note that their difference is bounded by $|p_{N+1}^2 - p_\infty^2|/p_\infty$, so that we are led to the difference of determinants

(49)
$$|F_{N+1,1}(s_0) - F(s_0)| \le C(s_0)|\det(I - K_\tau) - \det(I - K_{\mathrm{GOE}})|.$$

A Seiler–Simon-type bound on the matrix operator determinant for operators in $\mathcal{B}$ is established in Johnstone (2008).

PROPOSITION 3. *For $B, B' \in \mathcal{B}$, we have*

$$|\det(I - B) - \det(I - B')| \le C(B, B')\Delta(B - B'),$$

where

$$\Delta(B) = \sum_{i=1}^{2} \|B_{ii}\|_1 + \sum_{i \ne j} \|B_{ij}\|_2.$$

The coefficient has the form $C(B, B') = \sum_{j=1}^{2} c_{1j}(\operatorname{tr} B, \operatorname{tr} B')c_{2j}(B, B')$, where c_{1j} and c_{2j} are continuous functions, the latter with respect to the strong (Hilbert–Schmidt norm) topology.

Insert the conclusion of Proposition 3 into (49) to obtain

$$(50) \qquad |F_{N+1,1}(s_0) - F_1(s_0)| \leq C(s_0, K_\tau, K_{\text{GOE}})\Delta(K_\tau - K_{\text{GOE}}).$$

We exploit decomposition (47), which we write in the form

$$K_\tau - K_{\text{GOE}} = \delta^I + \delta_0^F + \delta_1^F + \delta_2^F$$

to distinguish a term involving integral kernels, $\delta^I = \tilde{L}(S_\tau - S_{A_N})$ from terms involving finite rank operators. We establish trace norm bounds for the diagonal elements and Hilbert–Schmidt bounds for the off-diagonal entries. The distinction between the two norms is moot for the finite rank terms δ_i^F, so the trace bounds are actually also needed only for the δ^I term.

For each term, we show $\|\delta_{ij}\| \leq CN^{-2/3}$, so that $\Delta(K_\tau - K_{\text{GOE}})$ is bounded above by $CN^{-2/3}$. We have both $\|K_\tau - K_{\text{GOE}}\|_2$ and $\text{tr}\, K_\tau - \text{tr}\, K_{\text{GOE}}$ converging to 0 at $O(N^{-2/3})$ rate, so that $C(K_\tau, K_{\text{GOE}})$ remains bounded as $N \to \infty$.

$7°$. To bound each term δ_{ij}, we need orthogonal case analogs of the uniform bounds of Corollary 1, but now for ϕ_τ^R, ψ_τ^R and their integrals and derivatives. From Proposition 1 and the succeeding discussion, we obtain

COROLLARY 2 (Real case).   *Let ϕ_τ and ψ_τ be defined by* (46) *and* (15). *Given $s_L \in \mathbb{R}$, there exists $C = C(s_L)$ such that for $N \geq N(s_L)$ and $s \geq s_L$,*

$$(51) \qquad\qquad\qquad |\phi_\tau(s)| \leq Ce^{-s},$$

$$(52) \qquad\qquad\qquad |\psi_\tau(s)| \leq Ce^{-s},$$

$$(53) \qquad\qquad\qquad |\phi_\tau(s) - A(s)| \leq CN^{-2/3}e^{-s/2},$$

$$(54) \qquad\qquad |\psi_\tau(s) - A(s) - \Delta_N A'(s)| \leq CN^{-2/3}e^{-s/2}.$$

The same bounds hold, with modified constants C, when ϕ_τ, ψ_τ, A and A' are replaced, respectively, by $\phi_\tau', \psi_\tau', A'$ and A'', or when ψ_τ, A and A' are replaced by $\tilde{\varepsilon}\psi_\tau, \tilde{\varepsilon}A$ and $\tilde{\varepsilon}A'$.

δ^I *term.* For $\delta^I = \tilde{L}[S_\tau - S_{A_N}]$, we use Proposition 2 to establish the needed Hilbert–Schmidt and trace norm bounds for each entry in the 2×2 matrix. We write

$$S_\tau - S_{A_N} = (\phi_\tau - A) \diamond \psi_\tau + A \diamond (\psi_\tau - A_N) + (\psi_\tau - A_N) \diamond \phi_\tau + A_N \diamond (\phi_\tau - A).$$

In turn, for $\partial_2(\bar{S}_\tau - S_{A_N})$ we replace the second slot arguments $\psi_\tau, (\psi_\tau - A_N)$, etc., by their derivatives, and for $\tilde{\varepsilon}(\bar{S}_\tau - S_{A_N})$, we replace the first slot arguments $(\phi_\tau - A)$, etc., by their right tail integrals.

Consider, for example, the first term $(\phi_\tau - A) \diamond \psi_\tau$. We apply Proposition 2 using (51) and (53) to set

$$a_N = CN^{-2/3}, \qquad b_N = C, \qquad a_1 = \tfrac{1}{2}, b_1 = 1.$$

The argument is entirely parallel when ∂_2 and $\tilde{\varepsilon}_1$ is applied to $(\phi_\tau - A) \diamond \psi_\tau$, and also for each of the second through fourth terms. Thus, if D_{ij} denotes any matrix entry in any component of δ^I, we obtain

$$(55) \qquad \|D_{ij}\| \leq CN^{-2/3} e^{-3s_0/2 + \gamma|s_0|}.$$

Finite rank terms. As Tracy and Widom (2005) note, the norm of a rank-one kernel $u(x)v(y)$, when regarded as an operator $u \otimes v$ taking $L^2(\rho_1)$ to $L^2(\rho_2)$ is given by

$$(56) \qquad \|u \otimes v\| = \|u\|_{2,\rho_2} \|v\|_{2,\rho_1^{-1}}.$$

Here the norm can be trace, Hilbert–Schmidt or operator norm, since all agree for a rank-one operator.

The finite rank terms include ones of the form $\tilde{L}(a \otimes \tilde{\varepsilon}b)$. We use (56) to establish entrywise bounds

$$(57) \qquad \begin{pmatrix} \|a \otimes \tilde{\varepsilon}b\| & \|a \otimes b\| \\ \|\tilde{\varepsilon}a \otimes \tilde{\varepsilon}b\| & \|\tilde{\varepsilon}b \otimes a\| \end{pmatrix} \leq \begin{pmatrix} A_+ B_- & A_+ B_+ \\ A_- B_- & A_+ B_- \end{pmatrix},$$

where

$$A_+ = \|a\|_+, \qquad B_+ = \|b\|_+,$$
$$A_- = \|\tilde{\varepsilon}a\|_-, \qquad B_- = \|\tilde{\varepsilon}b\|_-.$$

Indeed, for the (i,j)th entry, apply (56) to $a_{ij} \otimes b_{ij} : L^2(\rho_j) \to L^2(\rho_i)$, where $\rho_1 = \rho$ and $\rho_2 = \rho^{-1}$. On the right, and henceforth, we abbreviate the L^2 norms on $L^2(\rho)$ and $L^2(\rho^{-1})$ by $\|\cdot\|_+$ and $\|\cdot\|_-$, respectively.

Let us indicate how this applies to

$$-2\delta_0^F = \tilde{L}[\phi_\tau \otimes \tilde{\varepsilon}(\psi_\tau - A_N) + (\phi_\tau - A) \otimes \tilde{\varepsilon}A_N].$$

Consider the first term on the right-hand side—the second term is similar—and apply (57) with $a = \phi_\tau, b = \psi_\tau - A_N$. From Corollary 2 we have

$$A_+^2 = \|\phi_\tau\|_+^2 = \int_{s_0}^\infty \phi_\tau^2 \rho \leq C(\gamma) e^{-2s_0 + \gamma|s_0|},$$

$$B_-^2 = \|\tilde{\varepsilon}(\psi_\tau - A_N)\|_-^2 \leq C(\gamma) N^{-4/3} e^{-s_0 + \gamma|s_0|}$$

and with similar bounds, respectively, for A_-^2 and B_+^2. Hence

$$(58) \qquad A_\pm B_\pm \leq C(\gamma) N^{-2/3} e^{-3s_0/2 + \gamma|s_0|}.$$

Turning to the the δ_1^F, δ_2^F terms, we have

$$2\delta_1^F = \begin{pmatrix} (u_{N1} - A) \otimes 1 & 0 \\ -(u_{N2} - \tilde{\varepsilon}A) \otimes 1 & 0 \end{pmatrix}, \qquad 2(\delta_2^F)^t = \begin{pmatrix} 0 & 1 \otimes (u_{N2} - \tilde{\varepsilon}A) \\ 0 & 1 \otimes (u_{N1} - A) \end{pmatrix}$$

with $u_{N1} = \beta_{N-1}\phi_\tau$ and $u_{N2} = \beta_{N-1}\tilde{\varepsilon}\phi_\tau$. Using (57), we find that the norms of the terms in the first column of δ_1^F are bounded by $\|u_{N1} - A\|_+ \|1\|_-$ and $\|u_{N2} -$

$\tilde{\varepsilon} A \|_- \|1\|_-$ while the norms of the second column of $(\delta_2^F)^t$ are bounded by the same quantities interchanged.

From the definitions, and with $s_0 \geq 0$, we have $\|1\|_-^2 \leq \gamma^{-1} e^{-\gamma s_0}$ and

$$\|u_{N1} - A\|_+ \leq |\beta_{N-1} - 1| \|\phi_\tau\|_+ + \|\phi_\tau - A\|_+.$$

Note that $|\beta_{N-1} - 1| = O(N^{-1})$. Using also the bounds of Corollary 2,

$$\|u_{N1} - A\|_+ \leq (CN^{-1} e^{-s_0} + CN^{-2/3} e^{-s_0/2}) e^{\gamma s_0/2}$$

and $\|u_{N1} - A\|_+ \|1\|_- \leq CN^{-2/3} e^{-s_0/2}$. The term $\|u_{N2} - \tilde{\varepsilon} A\|_-$ is bounded analogously.

We finally assemble the bounds obtained from (55), (58) and the analysis of δ_i^F and only track the tail dependence on s_0 for $s_0 > 0$. Thus (50) is bounded by

$$CN^{-2/3}(e^{-3s_0/2 + \gamma s_0} + e^{-s_0/2}),$$

where the second term results from δ_1^F and δ_2^F. It is clear that $\gamma = 1$ yields a bound $CN^{-2/3} e^{-s_0/2}$.

APPENDIX A: HERMITE POLYNOMIAL ASYMPTOTICS NEAR LARGEST ZERO

Define new independent and dependent variables ζ and W via the equations (7), which put (6) into the form

$$(59) \qquad \frac{d^2 W}{d\zeta^2} = \{\kappa^2 \zeta + \psi(\zeta)\} W,$$

where the perturbation term $\psi(\xi) = \dot{\zeta}^{-1/2} (d^2/d\zeta^2)(\dot{\zeta}^{1/2})$. If the perturbation term $\psi(\zeta)$ were absent, the equation $d^2 W/d\zeta^2 = \kappa^2 \zeta W$ would have linearly independent solutions in terms of the Airy functions $\text{Ai}(\kappa^{2/3} \zeta)$ and $\text{Bi}(\kappa^{2/3} \zeta)$. Our interest is in approximating the *recessive* solution $\text{Ai}(\kappa^{2/3} \zeta)$, so write the relevant solution of (59) as $W_2(\zeta) = \text{Ai}(\kappa^{2/3} \zeta) + \eta(\zeta)$. In terms of the original independent and dependent variables w and ξ, the solution W_2 becomes

$$(60) \qquad w_2(\xi, \kappa) = \dot{\zeta}^{-1/2}(\xi) \{A(\kappa^{2/3} \zeta) + \varepsilon_2(\xi, \kappa)\}.$$

Olver (1974)—hereafter abbreviated as [O]—provides, in his Theorem 11.3.1, an explicit bound for $\eta(\zeta)$ and hence ε_2 and its derivative. To describe these error bounds even in the oscillatory region of $A(x)$, [O] introduces a positive weight function $\mathsf{E}(x) \geq 1$ and positive moduli functions $\mathsf{M}(x) \leq 1$ and $\mathsf{N}(x)$ such that for all x,

$$(61) \qquad |A(x)| \leq \mathsf{M}(x) \mathsf{E}^{-1}(x), \qquad |A'(x)| \leq \mathsf{N}(x) \mathsf{E}^{-1}(x).$$

[Here, $\mathsf{E}^{-1}(x)$ denotes $1/\mathsf{E}(x)$.] In addition,

$$(62) \qquad A(x) = 2^{-1/2} \mathsf{M}(x) \mathsf{E}^{-1}(x), \qquad x \geq c \doteq -0.37,$$

and the asymptotics as $x \to \infty$ are given by

$$(63) \quad \mathsf{E}(x) \sim \sqrt{2}e^{(2/3)x^{3/2}}, \qquad \mathsf{M}(x) \sim \pi^{-1/2}x^{-1/4} \quad \text{and} \quad \mathsf{N}(x) \sim \pi^{-1/2}x^{1/4}.$$

The key bounds of [O, Theorem 11.3.1] then state, for $\xi > 0$ and $\hat{f}(\xi) = f(\xi)/\xi$,

$$(64) \qquad |\varepsilon_2(\xi, \kappa)| \leq (\mathsf{M}/\mathsf{E})(\kappa^{2/3}\zeta)\left[\exp\left\{\frac{\lambda_0}{\kappa}\mathcal{V}(\zeta)\right\} - 1\right],$$

$$(65) \qquad |\partial_\xi \varepsilon_2(\xi, \kappa)| \leq \kappa^{2/3}N^{-1}\hat{f}^{1/2}(\xi)(\mathsf{N}/\mathsf{E})(\kappa^{2/3}\zeta),$$

where $\lambda_0 \doteq 1.04$. For $\kappa^{2/3}\zeta \geq c$, (62) shows that the coefficient in (64) is just $\sqrt{2}A(\kappa^{2/3}\zeta)$. Here $\mathcal{V}(\zeta) = \mathcal{V}_{[\xi,\infty]}(H)$ is the total variation on $[\xi, \infty]$ of the *error control* function $H(\xi) = -\int_0^{\zeta(\xi)} |v|^{-1/2}\psi(v)\,dv$. From [O, page 403] we have $\lambda_0\mathcal{V}_{[\xi,\infty)}(H) \leq 0.28$ and hence

$$(66) \qquad \exp\left\{\frac{\lambda_0}{\kappa}\mathcal{V}(\zeta)\right\} - 1 \leq 1/N.$$

Application to Hermite polynomials. In the case of Hermite polynomials, transformed as in (6), the points $\pm\infty$ are irregular singularities, and the points $\xi_\pm = \pm 1$ are turning points. We are interested in behavior near the upper turning point ξ_+, which is located near the largest (scaled) zero of H_N. Using (8), the independent variable $\zeta(\xi)$ is given in terms of $f(\xi)$ by

$$(67) \qquad (2/3)\zeta^{3/2}(\xi) = \tfrac{1}{2}\xi(\xi^2 - 1)^{1/2} - \tfrac{1}{2}\log(\xi + (\xi^2 - 1)^{1/2})$$

for $\xi \geq 1$, and by

$$(2/3)(-\zeta)^{3/2}(\xi) = \tfrac{1}{2}[\cos^{-1}\xi - \xi(1 - \xi^2)^{1/2}],$$

for $\xi \leq 1$. The function $\zeta(\xi)$ is increasing and C^2 on $(0, \infty)$ (e.g., [O, page 399]), with $\ddot{\zeta}(\xi)$ nonnegative and bounded. It is easily seen that $\zeta \to \infty$ as $\xi \to \infty$, and more precisely, from (67), that

$$(68) \qquad (2/3)\zeta^{3/2}(\xi) = \tfrac{1}{2}(\xi^2 - \log\xi - \tfrac{1}{2} - \log 2) + O(\xi^{-2}),$$

from which it follows that

$$(69) \qquad \dot{\zeta}(\xi) \sim (4\xi/3)^{1/3} \qquad \text{as } \xi \to \infty.$$

We remark that $\dot{\zeta} = \dot{\zeta}(1)$ is easily evaluated using L'Hôpital's rule. From (67), as $\xi \to 1$, we have $\dot{\zeta}^2(\xi) = (\xi^2 - 1)/\zeta(\xi) \to 2/\dot{\zeta}$, so that $\dot{\zeta} = 2^{1/3}$. In addition, we shall need the function

$$r(\xi) = [\dot{\zeta}(\xi)/\dot{\zeta}]^{-1/2},$$

which is positive on $(0, \infty)$ since $\zeta(\xi)$ is strictly increasing. Both $r(\xi)$ and $\dot{r}(\xi)$ are continuous on $[0, \infty)$, and as $\xi \to \infty$ we have $r(\xi) \sim (2\xi/3)^{-1/6}$ and $\dot{r}(\xi) \sim c_1\xi^{-7/6}$, so that $r(\xi)$ and $r'(\xi)$ are both bounded on $[0, \infty)$.

Bound (64) has a double asymptotic property in ξ and κ which will be useful. First, suppose that N, and hence κ, are held fixed. As $\xi \to \infty$, $\mathcal{V}(\zeta) \to 0$ and so from (64) and its following remarks $\varepsilon_2(\xi, \kappa) = o(A(\kappa^{2/3}\zeta))$. Consequently, as $\xi \to \infty$ we have $w_2(\xi, \kappa) \sim \dot{\zeta}^{-1/2}(\xi)A(\kappa^{2/3}\zeta)$. If the weighted polynomial $w_N(\xi)$ is a recessive solution of (6), then it must be proportional to w_2, so that $w_N(\xi) = c_N w_2(\xi, \kappa)$. Now c_N may be identified by comparing the growth of $w_N(\xi)$ as $\xi \to \infty$ with that of $w_2(\xi, \kappa)$ (Appendix B),

$$(70) \qquad c_N = e^{\theta''/N} \kappa_N^{1/6}(2/N)^{1/4},$$

where $\theta'' = O(1)$. Now we can use (60) to write $\phi_N(x) = w_N(\xi)$ in terms of the Airy approximation. Below, we write $\bar{e}_N$ for any term, that is, uniformly $1 + O(N^{-1})$. Hence

$$(2N)^{1/4}\phi_N(x) = \bar{e}\, 2^{1/2}\kappa_N^{1/6} w_2(\xi, \kappa).$$

Set $N_+ = N + 1/2$ and $\bar{\tau}_N = 2^{-1/2}N_+^{-1/6}$. Since $2^{1/2}\kappa_N^{1/6}\dot{\zeta}(\xi)^{-1/2} = \bar{\tau}_N^{-1}r(\xi)$, and using the Airy approximation (60) to $w_2(\xi, \kappa)$, we finally have

$$(71) \qquad (2N)^{1/4}\bar{\tau}_N\phi_N(x) = \bar{e}_N r(\xi)\{A(\kappa_N^{2/3}\zeta) + \varepsilon_2(\xi, \kappa_N)\}.$$

Approximations at degree N and $N - 1$. The kernel $S_{N,2}(x, y)$ is expressed in terms of the two functions $\phi_{N-1}(x)$ and $\phi_N(x)$, which need separate Liouville–Green asymptotic approximations. Thus, for example, in comparing the two cases, we have $\kappa_N = 2N + 1$ and $\kappa_{N-1} = 2N - 1$. The turning point $\xi_+ = 1$ and the transformation $\zeta(\xi)$ of (67) are the same in both cases, hence so is $r(\xi)$. The analog of (71) then states

$$(2N - 2)^{1/4}\bar{\tau}_{N-1}\phi_{N-1}(x) = \bar{e}_{N-1}r(\xi)\{A(\kappa_{N-1}^{2/3}\zeta) + \varepsilon_2(\xi, \kappa_{N-1})\}.$$

Rather than $\bar{\tau}_{Nj} = 2^{-1/2}N_\pm^{-1/6}$, we will use the single factor $\tau_N = 2^{-1/2}N^{-1/6}$ in the work below. Clearly, we may replace both $(2N)^{1/4}\bar{\tau}_N$ in (71) and $(2N - 2)^{1/4}\bar{\tau}_{N-1}$ in the preceding display by $(2N)^{1/4}\tau_N$ at cost of multiplicative error terms $e_{Nj} = 1 + O(N^{-1})$.

To summarize then, with the convention that quantities with subscript Nj differ for $Nj = N, N - 1$, while those with subscript N do not, we have

$$(72) \qquad (2N)^{1/4}\tau_N\phi_{Nj}(x) = e_{Nj}r(\xi)\{A(\kappa_{Nj}^{2/3}\zeta) + \varepsilon_2(\xi, \kappa_{Nj})\}.$$

Denote the left-hand side of (72) by $\bar{\phi}_{Nj}$. We seek a uniform bound on the Airy approximation. If we write $x = \sqrt{\kappa_{Nj}}\xi$ in the form $u_{Nj} + s\tau_N$, then we have in particular $u_N = \sqrt{2N + 1}$ and $u_{N-1} = \sqrt{2N - 1}$. In turn,

$$\xi = 1 + s\tau_N/\sqrt{\kappa_{Nj}} = 1 + s\sigma_{Nj},$$

where we define

$$(73) \quad \sigma_{Nj} = \tau_N/u_{Nj} = 2^{-1/2}N^{-1/6}(2N \pm 1)^{-1/2} = 2^{-1}N^{-2/3}(1 + O(N^{-1})).$$

We turn now to the proof of Proposition 1. We first record some properties of the map $s \to \kappa_{Nj}^{2/3} \zeta(1 + \sigma_{Nj}s)$, which we sometimes abbreviate as $\kappa^{2/3}\zeta$.

LEMMA 1. *Given $s_L \in \mathbb{R}$,*

$$(74) \qquad |\kappa^{2/3}\zeta - s| \leq |s|/4 \qquad for\ s_L \leq s \leq N^{1/6}, N \geq N_0,$$

$$(75) \qquad |\kappa^{2/3}\zeta| \leq C|s|/4 \qquad for\ s_L \leq s \leq CN^{2/3},\ all\ N.$$

PROOF. Expand $\zeta(\xi)$ about the turning point $\xi_+ = 1$:

$$(76) \qquad \kappa_{Nj}^{2/3}\zeta(1 + \sigma_{Nj}s) = \kappa_{Nj}^{2/3}\sigma_{Nj}\dot{\zeta}s + \tfrac{1}{2}\kappa_{Nj}^{2/3}\sigma_{Nj}^2 s^2 \ddot{\zeta}(\xi^*).$$

We note from the definitions that

$$(77) \qquad \kappa_{Nj}^{2/3}\sigma_{Nj}\dot{\zeta} = (1 \pm 1/(2N))^{1/6} = 1 + \delta_N,$$

with $|\delta_N| \leq N^{-1}$ for all $N \geq 1$. Since $0 \leq \ddot{\zeta}$ is bounded, we find that

$$(78) \qquad |\kappa^{2/3}\zeta - s| \leq \left(\frac{1}{N} + \frac{M|s|}{N^{2/3}}\right)|s|,$$

again for all $N \geq 1$. If $s < N^{1/6}$, then the right-hand side is bounded by $|s|/4$ for $N \geq N_0(M, s_L)$. If $|s| < N^{2/3}$, then we have (75) for $C = C(s_L, M)$. $\square$

We consider some global bounds, valid for $s \geq s_L$, or equivalently for $\xi \geq 1 + s_L\sigma_{Nj}$.

LEMMA 2. *Let $s_L < 0$. Let $\xi = 1 + \sigma_{Nj}s$ with σ_{Nj} satisfying (73). There exists $C = C(s_L)$ such that for $s \geq s_L$,*

$$\mathsf{E}^{-1}(\kappa_{Nj}^{2/3}\zeta) \leq Ce^{-2s},$$

$$\mathsf{N}(\kappa^{2/3}\zeta) \leq C(1 + |s|^{1/3}).$$

Some immediate consequences: using (61) and $\mathsf{M} \leq 1$, for $s \geq s_L$,

$$(79) \qquad |A(\kappa_{Nj}^{2/3}\zeta(\xi))| \leq |\mathsf{M}/\mathsf{E}|(\kappa^{2/3}\zeta) \leq Ce^{-2s},$$

$$(80) \qquad |\varepsilon_2(\xi, \kappa)| \leq N^{-1}|\mathsf{N}/\mathsf{E}|(\kappa^{2/3}\zeta) \leq CN^{-1}e^{-2s},$$

$$(81) \qquad |A'(\kappa_{Nj}^{2/3}\zeta(\xi))| \leq |\mathsf{N}/\mathsf{E}|(\kappa^{2/3}\zeta) \leq C(1 + |s|^{1/3})e^{-2s}.$$

PROOF. First, since $f(\xi) = (\xi + 1)(\xi - 1) \geq 2\sigma_{Nj}s$, we use (77) to observe that for $s \geq r^2$,

$$\kappa_{Nj}\sigma_{Nj}\sqrt{f} \geq \sqrt{2}\kappa_{Nj}\sigma_{Nj}^{3/2}\sqrt{s} \geq e_{Nj}r.$$

Hence, from (8), again for $s \geq r^2$,

$$\frac{2}{3}\kappa_{Nj}\zeta^{3/2} = \kappa_{Nj}\int_1^\xi \sqrt{f} = e_{Nj}r(s - r^2).$$

Now choose r large enough so that for $N > N_0$ and $j = N, N - 1$, we have $e_{Nj}r \geq 1$. From (63) we have $\mathsf{E}^{-1}(s) \leq C\exp(-\frac{2}{3}s^{3/2})$ for $s \geq 0$, and so in particular for $s \geq r^2$,

$$\mathsf{E}^{-1}(\kappa_{Nj}^{2/3}\zeta) \leq C(r)e^{-s}.$$

For $s \in [s_L, r^2]$, we simply use the bound $\mathsf{E} \geq 1$.

For the second statement, we will use the bound $\mathsf{N}(s) \leq 1 + |s|^{1/4}$ [O, pages 396–397]. First, for $s \leq N^{2/3}$, using the bound on N and (75), we obtain $\mathsf{N}(\kappa^{2/3}\zeta) \leq C(1 + |s|^{1/4})$. When $s \geq N^{2/3}$, we use (67) to bound

$$(2/3)\zeta^{3/2}(\xi) \leq \int_1^\xi t\,dt \leq \xi^2/2 \leq 1 + \sigma_{Nj}^2 s^2 \leq c_0\sigma_{Nj}^2 s^2.$$

From (73) we have $\kappa_{Nj}\sigma_{Nj}^2 = 2N^{-1/3}$ and so $\kappa_{Nj}\zeta^{3/2} \leq c_1 s^2$ for all N and hence $\mathsf{N}(\kappa^{2/3}\zeta) \leq 1 + c_1^{1/6}s^{1/3}$ as required. $\quad\square$

PROOF OF PROPOSITION 1. We begin from the formula

$$(82) \qquad \bar{\phi}_{Nj}(u_{Nj} + s\tau_N) = e_{Nj}r(\xi)\{A(\kappa_{Nj}^{2/3}\zeta) + \varepsilon_2(\xi, \kappa_{Nj})\}.$$

The bound (12) then follows from (80), (79) and boundedness of $r(\xi)$. To ease notation, we will, as needed, drop subscripts from $e_{Nj}, \sigma_{Nj}, \kappa_{Nj}$, and τ_N, writing $\bar{e}$ for a term, that is, generically $1 + O(N^{-1})$.

For the next bound, we differentiate (82), obtaining

$$(83) \qquad \tau_N\bar{\phi}'_{Nj}(u_{Nj} + s\tau_N) = D_1 + D_2 + D_3,$$

with the component terms given by

$$D_1 = \bar{e}\sigma\dot{r}(\xi)[A(\kappa^{2/3}\zeta) + \varepsilon_2(\xi, \kappa)],$$
$$D_2 = \bar{e}r(\xi)A'(\kappa^{2/3}\zeta)\sigma\kappa^{2/3}\dot{\zeta}(\xi),$$
$$D_3 = \bar{e}r(\xi)\sigma\,\partial_\xi\varepsilon_2(\xi, \kappa).$$

Since $\dot{r}(\xi)$ is bounded, we again use (80) and (79) to conclude that $|D_1| \leq C\sigma_N e^{-2s}$ for all $s \geq s_L$. From (77) and (69), we observe that

$$(84) \qquad \sigma_N\kappa_N^{2/3}\dot{\zeta}(\xi) \leq C|\xi|^{1/3} \leq C(1 + \sigma_N^{1/3}|s|^{1/3}).$$

Turning to the second term, we have from (84) and (81) that

$$|D_2| \leq C|A'(\kappa^{2/3}\zeta)\sigma\kappa^{2/3}\dot{\zeta}(\xi)| \leq C(1 + \sigma_N^{1/3}|s|^{1/3})(1 + |s|^{1/3})e^{-2s} \leq Ce^{-s}.$$

Using (65) and (66), we can rewrite D_3 as

$$|D_3| \leq CN^{-1} \cdot r(\xi)\sigma_N \kappa_N^{2/3} \dot{\zeta}(\xi) \cdot (\mathsf{N/E})(\kappa^{2/3}\zeta).$$

Using also boundedness of $r(\xi)$, (84) and (81), we conclude for all $s \geq s_L$,

$$|D_3| \leq CN^{-1} \cdot \sqrt{2} \cdot (1 + \sigma_N^{1/3}|s|^{1/3})(1 + |s|^{1/3})e^{-2s} \leq CN^{-2/3}e^{-s}.$$

This completes the proof of bound (12) for $\tau_N \bar{\phi}_{Nj}$. $\quad\square$

Turning to the error bound (13) and its analog for $\tau_N \bar{\phi}_{Nj}$, we first note that we may confine attention to $s \in [s_L, N^{1/6}]$, since for $s \geq N^{1/6}$, the bounds follow trivially from (12) and its analog and (14).

We use the decomposition suggested by (82),

$$\begin{aligned}
\bar{\phi}_{Nj}(x) - A(s) &= [e_{Nj}r(\xi) - 1]A(\kappa_{Nj}^{2/3}\zeta) + [A(\kappa_{Nj}^{2/3}\zeta) - A(s)] \\
&\quad + e_{Nj}r(\xi)\varepsilon_2(\xi, \kappa_{Nj}) \\
&= E_{N1} + E_{N2} + E_{N3}.
\end{aligned}$$

For the E_{N1} term, first use $\xi = 1 + \sigma_{Nj}s$ to write

$$(85) \qquad |\dot{\zeta}(\xi)/\dot{\zeta} - 1| = \left| \int_1^{1 + s\sigma_{Nj}} \ddot{\zeta}(u)/\dot{\zeta} \, du \right| \leq Cs\sigma_{Nj},$$

since $\ddot{\zeta}(u)$ is bounded for $u \in [1, 1 + s\sigma_{Nj}] \subset [1, 1 + N^{-1/2}]$. Together with $r(\xi) = [\dot{\zeta}(\xi)/\dot{\zeta}]^{-1/2}$, this yields

$$(86) \qquad |e_{Nj}r(\xi) - 1| \leq C(1 + s)\sigma_{Nj}.$$

Combined with (79), we obtain $|E_{N1}| \leq C\sigma_{Nj}(1 + s)e^{-2s} \leq CN^{-2/3}e^{-s/2}$.

For the E_{N2} term, we use (74) and (78) to write

$$\begin{aligned}
|A(\kappa_{Nj}^{2/3}\zeta) - A(s)| &\leq C|s|N^{-2/3}(N^{-1/3} + |s|)\sup\{|A'(t)| : \tfrac{3}{4}s \leq t \leq \tfrac{5}{4}s\} \\
&\leq CN^{-2/3}e^{-s/2},
\end{aligned}$$

uniformly for $s \in [s_L, N^{1/6}]$, where we used (14).

Finally, for the E_{N3} term, (80) and boundedness of r imply that

$$|E_{N3}| \leq e_{Nj}r(\xi)(\mathsf{M/E})(\kappa_{Nj}^{2/3}\zeta)N^{-1} \leq CN^{-1}e^{-2s}.$$

We turn now to the proof of (13) for $\tau_N \bar{\phi}_{Nj}$ on $[s_L, N^{1/6}]$. Using (83), we may write the difference $\tau_N \bar{\phi}'_{Nj}(x) - A'(s)$ as $D_1(s) + [D_2(s) - A'(s)] + D_3(s)$. Now decompose $D_2(s) - A'(s) = G_1 + G_2 + G_3$, with

$$G_1 = [\bar{e}r(\xi) - 1][\dot{\zeta}(\xi)/\dot{\zeta}]A'(\kappa^{2/3}\zeta), \qquad G_2 = [\dot{\zeta}(\xi)/\dot{\zeta} - 1]A'(\kappa^{2/3}\zeta)$$

and $G_3 = A'(\kappa^{2/3}\zeta) - A'(s)$. Combining (86), (81) and then (85), we find

$$|G_1| \leq C(1+s)\sigma_N \cdot 2 \cdot C(1+|s|^{1/3})e^{-2s} \leq CN^{-2/3}e^{-s/2},$$
$$|G_2| \leq Cs\sigma_N \cdot C(1+|s|^{1/3})e^{-2s} \leq CN^{-2/3}e^{-s/2}.$$

G_3 is treated in exactly the same manner as the E_{N2} term above, additionally using the equation $A''(x) = xA(x)$.

APPENDIX B: IDENTIFICATION OF c_N

We first remark that as $\zeta \to \infty$ when $\xi \to \infty$, we may substitute $A(x) \sim [2\sqrt{\pi}x^{1/4}]^{-1}\exp\{-(2/3)x^{3/2}\}$ into (60), along with $\dot\zeta^{-1/2} = [\zeta/f(\xi)]^{1/4}$ from (7) to obtain

$$w_2(\xi, \kappa) \sim [2\sqrt{\pi}]^{-1}\kappa^{-1/6}f^{-1/4}(\xi)\exp\{-(2/3)\kappa\zeta^{3/2}\}.$$

Consequently we may express c_N in terms of the limit

$$c_N = \lim_{\xi \to \infty} w_N(\xi) \cdot 2\sqrt{\pi}\kappa^{1/6}f^{1/4}(\xi) \cdot \exp\{(2/3)\kappa\zeta^{3/2}\}.$$

Write N_+ for $N + 1/2$. Since $w_N(\xi) = \phi_N(x)$ with $x = \sqrt{2N_+}\xi$, and since $H_N(x) \sim 2^N x^N$, we have as $\xi \to \infty$,

$$w_N(\xi) = h_N^{-1/2}e^{-N_+\xi^2}H_N(\sqrt{2N_+}\xi) \sim h_N^{-1/2}e^{-N_+\xi^2+N\log\xi}2^N(2N_+)^{N/2},$$

and $f^{1/4}(\xi) = e^{(\log\xi)/2+O(\xi^{-2})}$, while from (68)

$$\exp\{(2/3)\kappa\zeta^{3/2}\} = e^{N_+\xi^2-N_+\log\xi-N_+/2-N_+\log 2+O(\xi^{-2})}.$$

Multiply the last three quantities: the coefficients of ξ^2 and $\log\xi$ cancel, leaving ξ-dependence of only $O(\xi^{-2})$ as $\xi \to \infty$. Hence

$$c_N = 2\sqrt{\pi}\kappa^{1/6}h_N^{-1/2}(2N_+)^{N/2}e^{-N_+/2}2^{N-N_+},$$

and noting that $(N/2)\log N_+ = (N/2)\log N + 1/4 + O(N^{-1})$, we get

$$\kappa^{-1/6}c_N\sqrt{h_N} = \sqrt{2\pi}\exp\left\{\frac{N}{2}\log 2 + \frac{N}{2}\log N - \frac{N}{2} + O\left(\frac{1}{N}\right)\right\}.$$

Applying Stirling's formula to $h_N = \sqrt{\pi}2^N N!$, and dividing into the previous display yields (70).

REFERENCES

ADLER, M., FORRESTER, P. J., NAGAO, T. and VAN MOERBEKE, P. (2000). Classical skew orthogonal polynomials and random matrices. *J. Stat. Phys.* **99** 141–170. MR1762659

ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

BORNEMANN, F. (2010). On the numerical evaluation of distributions in random matrix theory: A review. *Markov Process. Related Field* **16** 803–866.

CHOUP, L. N. (2006). Edgeworth expansion of the largest eigenvalue distribution function of GUE and LUE. *Int. Math. Res. Not.* Art. ID 61049, 32. MR2233711

CHOUP, L. N. (2008). Edgeworth expansion of the largest eigenvalue distribution function of Gaussian unitary ensemble revisited. *J. Math. Phys.* **49** 033508, 16. MR2406805

CHOUP, L. N. (2009). Edgeworth expansion of the largest eigenvalue distribution function of Gaussian orthogonal ensemble. *J. Math. Phys.* **50** 013512, 22. MR2492622

EL KAROUI, N. (2006). A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.* **34** 2077–2117. MR2294977

FORRESTER, P. J. and MAYS, A. (2009). A method to calculate correlation functions for $\beta = 1$ random matrices of odd size. *J. Stat. Phys.* **134** 443–462. MR2485724

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

JOHNSTONE, I. M. (2009). Approximate null distribution of the largest root in multivariate analysis. *Ann. Appl. Stat.* **3** 1616–1633. MR2752150

MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359.

OLVER, F. W. J. (1974). *Asymptotics and Special Functions*. Academic Press, New York–London. MR0435697

PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670

SEILER, E. and SIMON, B. (1975). An inequality among determinants. *Proc. Natl. Acad. Sci. USA* **72** 3277–3278. MR0433247

SHI, W. (2008). A globally uniform asymptotic expansion of the Hermite polynomials. *Acta Math. Sci. Ser. B Engl. Ed.* **28** 834–842. MR2462927

SINCLAIR, C. D. (2009). Correlation functions for $\beta = 1$ ensembles of matrices of odd size. *J. Stat. Phys.* **136** 17–33. MR2525224

SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444

SZEGŐ, G. (1967). *Orthogonal Polynomials*, 3rd ed. Amer. Math. Soc., Providence, RI. MR0310533

TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449

TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

TRACY, C. A. and WIDOM, H. (1998). Correlation functions, cluster functions, and spacing distributions for random matrices. *J. Stat. Phys.* **92** 809–835. MR1657844

TRACY, C. A. and WIDOM, H. (2005). Matrix kernels for the Gaussian orthogonal and symplectic ensembles. *Ann. Inst. Fourier (Grenoble)* **55** 2197–2207. MR2187952

WANG, X. S. and WONG, R. (2011). Asymptotics of orthogonal polynomials via recurrence relations. Technical report. Available at arXiv:1101.4371.

WONG, R. and ZHANG, L. (2007). Global asymptotics of Hermite polynomials via Riemann-Hilbert approach. *Discrete Contin. Dyn. Syst. Ser. B* **7** 661–682 (electronic). MR2291866

DEPARTMENT OF STATISTICS	DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY	THE WHARTON SCHOOL
SEQUOIA HALL	UNIVERSITY OF PENNSYLVANIA
STANFORD, CALIFORNIA 94305	PHILADELPHIA, PENNSYLVANIA 19104
USA	USA
E-MAIL: imj@stanford.edu	E-MAIL: zongming@wharton.upenn.edu