AN EDGE CLT FOR THE LOG DETERMINANT OF WIGNER ENSEMBLES

BY IAIN M. JOHNSTONE¹, YEGOR KLOCHKOV²,†, ALEXEI ONATSKI²,‡ AND DAMIAN PAVLYSHYN¹,*

¹Department of Statistics, Stanford University, imj@stanford.edu; *damianp@stanford.edu
²Faculty of Economics, University of Cambridge, †yk376@cam.ac.uk; ‡ao319@cam.ac.uk

We derive a Central Limit Theorem (CLT) for \( \log |\det (W_N - E_N)| \), where \( W_N \) is a Wigner matrix, and \( E_N \) is local to the edge of the semi-circle law. Precisely, \( E_N = 2 + N^{-2/3} \sigma_N \) with \( \sigma_N \) being either a constant (possibly negative), or a sequence of positive real numbers, slowly diverging to infinity so that \( \sigma_N \ll \log^2 N \). We also extend our CLT to cover spiked Wigner matrices. Our interest in the CLT is motivated by its applications to statistical testing in critically spiked models and to the fluctuations of the free energy in the spherical Sherrington-Kirkpatrick model of statistical physics.

1. Introduction. Let \( W_N = (\xi_{ij}/\sqrt{N}) \) be an \( N \times N \) real or complex Wigner matrix; in particular \( W_N \) is Hermitian and for \( i \geq j \) the entries \( \xi_{ij} \) are independent with mean zero, the variances \( E|\xi_{ij}|^2 = 1 \) for \( i \neq j \) and are bounded for \( i = j \). Our conditions, fully specified in section 5.1, imply that the empirical distribution of the eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_N \) of \( W_N \) converges to the Wigner semi-circle law \( \rho_{sc} \) on \([-2,2]\) and that the largest eigenvalue \( \lambda_1 \) converges almost surely to the right edge 2, see, for example, [3].

The logarithmic linear statistic

\[
\mathcal{L}_N = \sum_{j=1}^N \log |\lambda_j - E|
\]

arises in several applications; we focus below in particular on statistical testing in ‘spiked’ models and on the fluctuation behavior of the free energy in the spherical Sherrington-Kirkpatrick (SSK) model of statistical physics. Suppose initially that \( E > 2 \) is fixed. In this case \( \mathcal{L}_N - N \int f \, d\rho_{sc} \) is asymptotically Gaussian with finite variance that depends on the first four moments of the entries of \( W_N \). Since \( f(z) = \log |z - E| \) is analytic in a neighborhood of the semi-circle support, this follows from general CLTs for linear statistics, e.g. [51].

This paper concerns Gaussian behavior near, at, or just inside the edge:

\[
E = E_N = 2 + \sigma_N N^{-2/3}, \quad -\gamma \leq \sigma_N \ll \log^2 N
\]

for some fixed \( \gamma > 0 \). Our particular motivations, detailed below, lie in certain transition zones in the spiked statistical and SSK models. Here \( E \) is sufficiently close to the edge that the functions \( f_N(z) = \log |z - E| \) do not appear to be covered even by recent mesoscopic CLTs (e.g. [35, 36]).

The basic identity

\[
L_N = \sum_{j=1}^N \log |\lambda_j - E| = \log |\det(W_N - E)|
\]

casts the linear statistic (now with the absolute value under the logarithm) as a log determinant, i.e. in terms of the characteristic polynomial of \( W_N \). The latter is the subject of a

MSC2020 subject classifications: Primary 60, 62, 82; secondary 60B20, 60F05, 60K35, 62H15, 82B44.

Keywords and phrases: CLT, Log determinant, Edge of the semi-circle law, Wigner matrix.
substantial literature, partly reviewed in Section 1.3. In particular, as pioneered by Tao and Vu [46] for $E = 0$, for Gaussian ensembles $W_N$ drawn from GUE or GOE, one can use the Trotter equivalence to cast the matrix in tridiagonal Jacobi form and analyze the recurrence satisfied by the principal minors. Lindeberg swapping is used to extend to Wigner matrices with four matching moments.

In this paper we carry out this program at the edge (1), to arrive at the following result.

**THEOREM 1.** Let $W_N$ be a Wigner matrix whose off-diagonal moments match GUE ($\alpha = 1$) or GOE ($\alpha = 2$) to third order. For edge values $E = E_N$ satisfying (1), we have

$$
(\log|\det(W_N - E)| - \mu_N) / \tau_N \xrightarrow{d} N(0, 1),
$$

with

$$
\mu_N = \frac{1}{2} N + \sigma_N N^{1/3} - \frac{2}{5} (\sigma_N \vee 0)^{3/2} - \frac{1}{6} (\alpha - 1) \log N, \quad \tau_N = \sqrt{\frac{\alpha}{3}} \log N.
$$

1.1. Two motivating applications. Although superficially unrelated, both applications involve the spherical integral

$$
Z_{\alpha,N}(\beta, M) = \int_{S^{N-1}_\alpha} \exp \{(\beta N/\alpha) u^* M u\} (du),
$$

where $(du)$ denotes normalized uniform measure on the unit sphere $S^{N-1}_\alpha = \{x : \|x\| = 1\}$ in $\mathbb{C}^N$ for $\alpha = 1$, or $\mathbb{R}^N$ for $\alpha = 2$, while $M$ is Hermitian resp. symmetric, and $\beta > 0$.

**Testing critical spiked models.** Principal Components Analysis (PCA) seeks low-dimensional summaries of high-dimensional data. In certain cases, such as genomics e.g. [43], it can be reasonable to approximate the covariance matrix as $\Sigma = \sigma^2 I + F$, where $F$ has small rank. A perennial applied question is to determine this rank, at least approximately. The simplest version is to test for the presence of a rank one component. Thus we assume

$$(\text{PCA}) \quad X_{N\times n} \text{ has i.i.d. columns } \sim N(0, \Sigma), \quad \Sigma = \sigma^2 I + h \nu \nu^*, \quad M = XX^*/n.$$ The largest eigenvalue $\lambda_1(M)$ is a natural test statistic, but its utility is limited by a phase transition first exhibited for complex data in (PCA) by Baik, Ben Arous and Péché [6] in the setting of proportional asymptotics $N/n \to y > 0$. Below the critical value, $h < \sqrt{y}$, the largest eigenvalue, after centering and scaling at rate $N^{-2/3}$, has a limiting Tracy-Widom distribution, and so carries no information about $h$.

Onatski, Moreira and Hallin [42] showed that testing below the critical value was still possible, using a likelihood ratio test of $H_0 : h = 0$ versus $H_A : h = \beta$. The asymptotic behavior of this test depends on a logarithmic linear statistic $L_N$ for $E = E(\beta)$ located outside the edge of the Marčenko-Pastur bulk. It is also noted in [42] that the likelihood ratio is exponentially small for supercritical alternatives $h > \sqrt{y}$, but left open the behavior for alternatives $\beta$ near the critical point.

It is commonly noted that the spectra of Wishart matrices $XX^*$ exhibit behavior analogous to that of simpler symmetric Wigner matrices. For us, the analog of (PCA) specifies that

$$(\text{SMD}) \quad M = h \nu \nu^* + Z/\sqrt{N},$$

where $Z$ is in general an $N \times N$ Wigner matrix, real symmetric or complex Hermitian. In the special case that $Z$ is drawn from GOE resp GUE, the term deformed GO(O/U)E is used. The BBP transition occurs at threshold $h = 1$ in these models [44, 37]. The likelihood ratio against $H_A : h = \beta < 1$ was studied, along with other spiked models in [29], again in terms
of a logarithmic linear statistic $L_N$, now with $E = \beta + 1/\beta > 2$. Again behavior for $\beta$ near 1 was left open.

One reason for the close parallel of results for PCA and SMD is that the joint density of the eigenvalues $\Lambda = (\lambda_i)_i$ of $M$ has the same form in both cases. The joint density is found by integrating over the orthogonal group corresponding to the eigenvectors; for a rank one spike the integral reduces to one over $S^N_{\alpha - 1}$. The argument goes back at least to [25], see also [29, Suppl p. 6] and [40, p. 104]. For SMD the result is

$$p(\Lambda, h) = c(\Lambda)d(h)Z_{\alpha,N}(h, \Lambda).$$

The main term $Z_{\alpha,N}(h, \Lambda)$ is given by (4), while $d(h) = \exp(-(N/\alpha)(h^2/2))$ and $c(\Lambda)$ though explicit is not needed as it disappears on taking ratios for distinct values of $h$.

For PCA, we have $d(h) = (1 + h)^{-n/\alpha}$ and in the $Z_{\alpha,N}$ term, $h$ is replaced by $nh/(N(1 + h))$. In view of the foregoing remarks, we will henceforth focus on SMD.

**SSK model.** In the spherical version of the Sherrington-Kirkpatrick model studied by Kosterlitz, Thouless and Jones [31], the vector of spins $\sigma \in \mathbb{R}^N$ is constrained to lie on the sphere $\|\sigma\|^2 = N$. The Hamiltonian is given by $H_N(\sigma) = \sum_{i < j} M_{ij}^{SSK} \sigma_i \sigma_j$, where the couplings $M_{ij}^{SSK}$ between distinct spins are random, and in [31] are independent $N^{-1/2}N(J/N^{1/2}, 1)$ variates. The partition function is then $Z_N = \int e^{\beta H_N(\sigma)} d\omega_N(\sigma)$ with $d\omega_N$ being normalized uniform measure on the sphere $\|\sigma\|^2 = N$. If $M^{SSK}$ is the corresponding symmetric matrix, then on rescaling to the unit sphere, we have

$$Z_N = \int \exp\{(\beta N/2) u^* M^{SSK} u\} (du) = Z_{2,N}(\beta, M^{SSK}).$$

Since the integrals depend only on the eigenvalues of $M$, this is exactly the integral occurring in the rank one spiked GOE model (SMD), with the sole difference that $M^{SSK}$ has vanishing diagonal.

Kosterlitz et. al. evaluated the first order limiting behavior of the free energy, finding a phase diagram for $(J, 1/\beta) \in \mathbb{R}^2_+$ with three regions: ferromagnetic for $J > 1, \beta J > 1$, paramagnetic for $\beta < 1, \beta J < 1$ and spin glass for $J < 1, \beta > 1$. In particular, we record that

$$F_N = \frac{1}{N} \log Z_N \to F(\beta) = \begin{cases} \frac{1}{4}\beta^2 & \beta < 1, J < 1 \\ \beta - \frac{1}{2}\log \beta - \frac{3}{4} & \beta > 1, J < 1 \end{cases}.$$  

Baik and Lee [7, 8] studied the second-order fluctuations of the free energy, making more general assumptions of Wigner type on the distributions of the couplings $M_{ij}$. When $\text{Var} M_{ij} = 0$, they refer to $H_N(\sigma)$ as the spherical SK Hamiltonian with ferromagnetic Curie-Weiss interaction. We refer to [7, 8] for fuller bibliographic discussion of work around the SSK model. The main results of [7, 8] show that $N^{-1/2}(F_N - F(\beta)) \overset{d}{\to} \xi$, where in the three phases respectively $(\gamma, \xi) = \left(\frac{1}{4}, \text{Gaussian}\right)$, $(1, \text{Gaussian})$, and $(\frac{2}{3}, \text{Tracy-Widom})$, suppressing details of the centering and scaling of $\xi$.

The transition regions between the three phases are studied in [7, 8] and [9]. Two transitions are settled but the spin glass to paramagnetic transition is left open. We emphasize that the open case is exactly the transition relevant to studying the likelihood ratio statistic for testing against near critical alternatives! By equating variances for $\beta$ above and below 1, [7] conjecture that the relevant scale has $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ for $b \in \mathbb{R}$.

In a companion paper [27], we apply our Theorem 1 to verify the scaling conjectured by Baik and Lee: on this scale, after centering and scaling $F_N - F(\beta)$ converges in law to a $b$-dependent linear combination of independent Tracy-Widom and Gaussian components. In
turn this implies conclusions for the null distribution of log-likelihood ratio tests of $H_0: h = J \in [0,1]$ versus critically spiked alternatives $H_A: h = \beta$.

The loglinear statistic $L_N$ and Theorem 1 are basic for this result. Briefly, the standard first step casts the spherical integral as a single contour integral

$$Z_{\alpha,N}(\beta, \Lambda) = C_{\alpha,N} \int_K e^{(N/\alpha)G_\beta(z)} \, dz,$$

where $G_\beta$ involves the loglinear statistic $L_N$

$$G_\beta(z) = (1 + bN^{-1/3} \sqrt{\log N}) z - N^{-1} \sum_{1}^{N} \log(z - \lambda_j).$$

The contour $K$ passes to the right of all eigenvalues $\lambda_j$, and is chosen to allow Laplace approximation of the integral. For $b < 0$, the vertical contour through $\tilde{\gamma}_b = 2 + b^2 N^{-2/3} \log N$ suffices, and the main approximating term involves $L_N(\tilde{\gamma}_b)$. For $b > 0$ a keyhole contour around $\lambda_1$ and for $b = 0$ a contour of steepest descent both yield a leading approximation term involving

$$-L_N(2) + (\beta - 1)N(\lambda_1 - 2).$$

In each case Theorem 1 along with further analysis of the derivatives of $G_\beta$ lead to the transition theorem. In addition, the tridiagonal Jacobi method and Lindeberg swapping tools developed here is adapted to show the asymptotic independence of $L_N(2)$ and $\lambda_1$.

1.2. Outline of approach. The analysis begins with $W_N$ drawn from a Gaussian ensemble: GUE or GOE. By a unitary/orthogonal transformation [49] the eigenvalues of $\sqrt{N} W_N$ are the same as those of

$$\sqrt{N} W_N = \begin{pmatrix} a_1 & b_1 & b_2 \\ b_1 & a_2 & b_2 \\ & \ddots & \ddots & \ddots \\ & & b_{N-1} & b_N \\ & & & b_{N-1} & a_N \end{pmatrix},$$

where $a_i \sim \mathcal{N}(0, \alpha)$ and $b_i^2 \sim \chi^2(2i/\alpha)(2/\alpha)$, $i = 1, \ldots, N$, are jointly independent, and $\alpha = 1$ for GUE and $\alpha = 2$ for GOE. Here, by definition, $\chi^2(d)$ has the density $c_d^{-1} x^{d/2-2} e^{-x/2} 1_{x>0}$ for $d > 0$ and $c_d = 2^{d/2} \Gamma(d/2)$.

Let $D_i$ denote the determinant of the $i$-th minor of $\hat{W}_N - E$. Using the cofactor expansion yields the recurrence

$$D_i = \left( \frac{a_i}{\sqrt{N}} - E \right) D_{i-1} - \frac{i - 1 + \sqrt{i - 1} c_{i-1}}{N} D_{i-2},$$

for $i \geq 1$ with the conventions $D_0 = 1, D_{-1} = 0$. Here $c_i = (b_i^2 - i) / \sqrt{i}$, so that $E c_i = 0$ and $\text{Var}(c_i) = \alpha$. A deterministic version of such a recursion ($a_i = c_{i-1} = 0$) has an explosive characteristic root $\rho_i^+$ with

$$2\rho_i^+ = -\left( E \pm \sqrt{E^2 - \frac{4(i - 1)}{N}} \right).$$

We therefore adopt the normalization

$$M_i = D_i / \prod_{j=1}^{i} |\rho_j^+|. $$
The characteristic roots of the dynamic equation describing $M_i$ can be approximated by the pair $\rho_i^\pm / |\rho_i^+|$. For $i \leq E^2 N/4$, these characteristic roots are real, the first one equals minus one, and the other decreases from zero to minus one as $i$ goes from 1 to $E^2 N/4$.

Qualitatively, for most $i$, $M_i$ and $M_{i-1}$ have opposite signs and similar magnitudes, so that $R_i = M_i/M_{i-1} + 1$ remains close to zero. However, for $i$ approaching $N$, $R_i$ starts to develop more excited dynamics.

It may be of interest to note that $M_i/M_{i-1}$ can be interpreted as normalized Sturm ratio sequence of matrix $\hat{W}_N$. Sturm ratios play a useful role in the analysis of large random matrices (see e.g. [1] or Section 1.9.3 in [20]).

In Section 2, we show that as long as the local parameter, $\sigma_N$, of the singularity is slowly diverging to infinity so that $\sigma_N \gg (\log \log N)^2$, the dynamics of $R_i$, $i = 1, ..., N$, can be well approximated by a linear one. Then we use this linear approximation to obtain a CLT for the sums of the logarithms of the normalized Sturm ratios. This leads to Theorem 1 with $\sigma_N \gg (\log \log N)^2$. In fact, for such $\sigma_N$, our proof remains valid for matrices $\hat{W}_N$ from general Gaussian $\beta$-ensembles (with $\beta = 1/\alpha \in (0, \infty)$).

To extend the theorem to slower growing and constant $\sigma_N$, Section 3 derives simple asymptotic formulae for the Stieltjes transform of the empirical spectral distribution of $\hat{W}_N$ and its derivative at the edge of the support $[-2, 2]$. These formulae and the Taylor expansion of the logarithm describe the asymptotic behavior of the log statistics at the edge with $\sigma_N \leq (\log \log N)^3$ in terms of that of the statistic with $\sigma_N \gg (\log \log N)^2$. Thus, we obtain Theorem 1 in its generality.

Our proof can be easily extended to cover spiked GUE and GOE matrices. See proposition 29 and the remark that follows that proposition.

**Extension to Wigner case.** Proving that a Wigner matrix $W_N'$ satisfies a certain property as long as a matrix $W_N$ from $\text{G(O/U/)}E$ satisfies this property is often based on the Lindeberg swapping process, where elements of $W_N$ are replaced by the elements of $W_N'$ one by one without losing the property. Typically, one needs to show that any individual swap does not change the expectation $\mathbb{E}Q(M)$ of some smooth function $Q(\cdot)$ of the matrix $M$ participating in the swapping process too much.

Although our initial interest is in the asymptotic normality of the log-determinant, we will eventually need to use Lindeberg swapping for several functionals which depend on the Stieltjes transform evaluated at $z = \hat{E} + i\eta$ for $\hat{E}$ near the edge and $\delta$ distant at least $N^{-2/3-\delta}$ from the real axis – here the gross $N^{-2/3}$ scale is that appropriate for working at the edge of the spectrum. We outline the swapping approach for the log-determinant example but with the general class of “Stieltjes edge functionals” in mind.

We adopt the method of [46], with modifications to work at the edge, and under weakened assumptions, as described below. We call a quantity $S(W_N)$ insensitive at rate $\delta_N$ if $S(W_N) - S(W_N') = O(\delta_N)$. Let $L_N(W_N) = \log |\det(W_N - \hat{E})|$. To extend the asymptotic normality of $L_N(W_N)$ to $L_N(W_N')$ it is sufficient, via a standard smoothing argument, to show that $\mathbb{E}G \circ L_N(W_N)$ is insensitive at rate $\delta_N$ for scalar functions for which $\|G^{(j)}\|_\infty \leq b_N$. For the log-determinant $\delta_N = b_N \asymp (\log N)^{-1/4}$ will work.

In an initial approximation step, we show that it suffices to replace $L_N(W)$ by a function of the Stieltjes transform $s_W = N^{-1} \text{tr}(W - z)^{-1}$

\begin{equation}
(8) \quad g(W) = N \int_{\gamma_N}^{N^{100}} \text{Im} s_W(E + i\eta) \, d\eta.
\end{equation}

Here $\gamma_N = N^{-2/3-\delta}$; to show that values $0 \leq \eta \leq \gamma_N$ can be neglected, we use an anti-concentration result that guarantees that with high probability, all eigenvalues are at least
$N^{-2/3-\delta}$-distant from $E$. This too is proved by Lindeberg swapping, now with a second Stieltjes functional, section 5.3.1.

The swapping argument is now applied to show that $\mathbb{E}Q(W_N)$ is insensitive for $Q$ of the form $(G \circ g)(W_N)$. To review this in outline, let $\gamma$ index an ordering of the independent components $\{\text{Re} \xi_{ij}, \text{Im} \xi_{ij}\}_{i<j}$ and $\{\xi_{ii}\}$ of $W_N$. Thus $\gamma$ runs over $N^2$ and $N(N+1)/2$ elements in the Hermitian and symmetric cases respectively. By convention in each case, the first $N$ values of $\gamma$ index the diagonal matrix entries. Thus $W^\gamma$ will refer to a matrix in which the elements prior to $\gamma$ come from $W'_N$ while those at $\gamma$ or later come from $W_N$.

At stage $\gamma$ in the swapping process, we can write $W^{(0)} = W^{\gamma}$, $W^{(1)} = W^{\gamma+1}$, and

$$W^{(0)} = W_0 + \frac{\xi^{(0)}}{\sqrt{N}} V, \quad W^{(1)} = W_0 + \frac{\xi^{(1)}}{\sqrt{N}} V,$$

and $W_0 = W_0^\gamma$ is independent of both $\xi^{(0)}$ and $\xi^{(1)}$. In the symmetric case, $V$ is one of the elementary matrices of the form $e_a e_a^* + e_b e_b^*$, for $1 \leq a < b \leq N$. In the Hermitian case, we add matrices $i e_a e_a^* - e_b e_b^*$. The variables $\xi^{(0)}$ and $\xi^{(1)}$ correspond to the $\gamma$th components of $W_N$ and $W'_N$ respectively. All matrices $W^\gamma$ are Wigner matrices.

To focus on individual swaps, write

$$\mathbb{E}Q(W) - \mathbb{E}Q(W') = \sum_{\gamma} \mathbb{E} \Delta_{\gamma},$$

with $\Delta_{\gamma} = Q(W^{(1)}) - Q(W^{(0)}) - Q(W^{(0)} - Q(W^{(1)})$.

We consider $W^{(0)}$ and $W^{(1)}$ as perturbations of $W_0$. Thus, set $W^{(0)} = W_0^{\gamma} + t N^{-1/2} V^{\gamma}$, and introduce $Q_{\gamma}(t) = Q(W^{(0)}_t)$. Note that this function is independent of $\xi^{(i)}$, and that

$$\Delta_{\gamma} = Q_{\gamma}(\xi^{(0)}) - Q_{\gamma}(\xi^{(1)}).$$

In a Taylor expansion of $Q_{\gamma}$, formal for now, this independence implies

$$\mathbb{E}[Q_{\gamma}(\xi^{(i)})] = \sum_j \frac{1}{j!} \mathbb{E}[Q^{(j)}(0)] \mathbb{E}[(\xi^{(i)})^j].$$

If moments match at order $j \leq k - 1$, that is, $\mathbb{E}[(\xi^{(0)})^j] = \mathbb{E}((\xi^{(1)})^j)$, then the $j$th order term in $\mathbb{E} \Delta_{\gamma}$ vanishes. If, as one expects, $Q^{(k)}(t)$ is of order $N^{-k/2} b_N$, bounding the remainder term appropriately leads to the required bounds on $\mathbb{E} \Delta_{\gamma}$. This is formalized in Proposition 20.

To show that such derivative bounds hold specifically for $Q = G \circ g$ when $g$ is as in (8), we need good bounds for $\partial_t^j g^\gamma(t)$ when $g^\gamma(t) = g(W^{(0)}_t)$. Introduce notation for the resolvent and Stieltjes transforms

$$(10) \quad R^\gamma_t = R^\gamma_t(z) = (W^{(0)}_t - z)^{-1}, \quad s^\gamma_t(\eta) = N^{-1} \text{tr} R^\gamma_t(E + i\eta).$$

The standard resolvent perturbation argument (equations (75)-(78)) shows that $\partial_t^j s^\gamma_t = c_j N^{-j/2 - 1} \text{tr}[R^\gamma_t V^j R^\gamma_t]$. This is bounded for $E$ near the edge and $\eta > N^{-2/3-\delta}$ using the entrywise local law (see Proposition 23(i)). Working at the edge allows, through use of the Ward identity, improvements in bounds because $\text{Im} R$ is small. What results (see the proof of Proposition 24) are bounds $\|\partial_t^j g^\gamma(t)\|_\infty \lesssim N^{-j/2} a_N$ with $a_N = 1$ in the log-determinant case. These bounds are useful both for reducing the number of matching moments required to three (for off-diagonal entries) and requiring only bounded variances (for diagonal entries). Combining with the derivative bounds on $G$, the chain rule shows that we obtain the desired insensitivity with $\delta_N = a_N b_N = b_N$. 

1.3. Related work. The interest in determinants of Hermitian matrices from GUE, GOE and other classical ensembles of Random Matrix Theory emerged in the 1960s from motivations in nuclear physics. The first published derivation of the joint distribution of the eigenvalues of GUE in [50] was spurred by the problem of approximating the value of $\log |\det (W_N - E)|$, where $E \in (-2, 2)$.

As pointed out by [22], a CLT for the GUE log statistic with singularity $E$ from a compact subset of $(-2, 2)$ can be obtained from Theorem 1 of Krasovsky [32]. That theorem derives detailed asymptotics for the Laplace transform of the log statistic using Riemann-Hilbert machinery. Tao and Wu [46] derive their CLT for $E = 0$ and for Wigner matrices with atom distributions that match the first four moments of the normal. Bourgade and Mody [13] relax the conditions to require only matching of the first two moments. Duy [15] describes a very elegant proof of such a CLT for Gaussian ensembles GUE/GOE and $E = 0$ based on a representation of the corresponding log-determinants in the form of sums of independent random variables.

For super-critical $E$ that lie outside an open set covering $[-2, 2]$, the CLT for log-determinants follow from the CLT for more general linear spectral statistics with only super-critical singularities. Such a CLT is well known for classical ensembles (e.g. Johansson [26]). For extensions to Wigner matrices we refer the reader to Bai and Yao [31].

For the critical regime with $E$ local to 2 the corresponding CLT has not been available. When this paper was close to completion, we learned about a recent work by Lambert and Paquette [33, 34] that obtains powerful asymptotic approximations to the logarithmic statistics with singularity local to the edge of the semi-circle law for Gaussian $\beta$-ensembles. Such approximations imply a CLT.

The analysis in [33, 34] starts from the recurrence for the minors of $z - \tilde{W}_N / 2$, equivalent to our (6) with $z$ interpreted as $E/2$. The deterministic version of their recurrence generates monic Hermite polynomials $\pi_i(z)$ orthogonal with respect to the weight $\exp(-2Nz^2)$,

$$
\begin{pmatrix}
\pi_i(z) \\
\pi_{i-1}(z)
\end{pmatrix} = \tilde{T}_i(z) \begin{pmatrix}
\pi_{i-1}(z) \\
\pi_{i-2}(z)
\end{pmatrix}
$$

with $\tilde{T}_i(z) = \left( \begin{array}{cc} z & \frac{i}{N} \\
1 & 0 \end{array} \right)$.

Lambert and Paquette point out three regimes of this recurrence: hyperbolic, parabolic, and elliptic, corresponding to the eigenvalues of $\tilde{T}_i(z)$ being, respectively, distinct real, coinciding or local to each other, and distinct complex. The regimes are associated with growing, Airy-type transitory, and oscillatory behavior of the Hermite polynomials, respectively. In terms of $E = 2z$, the recursion remains in hyperbolic or elliptic regimes for all $i = 1, \ldots, N$ as long as $|E| > 2 + \epsilon$ or $|E| < 2 - \epsilon$, respectively. It enters the parabolic regime for relatively large $i$ if $|E - 2| = O(N^{-2/3})$.

[33] studies the hyperbolic regime of the recurrence for the minors of $z - \tilde{W}_N / 2$. It covers the range $E > 2 + \sigma_N N^{-2/3}$, where $\sigma_N \gtrsim \log^{2/3} N$ in our notations. Instead of analysing the dynamics of the Sturm ratios $M_i/\tilde{M}_{i-1}$ as we do, [33] base their analysis on an approximation to the product $T_N(z)...T_2(z)$, where $T_i(z)$ are the stochastic analogues of the deterministic transfer matrices $\tilde{T}_i$.

[34] extends [33] to the parabolic regime by noting that scaled versions of the minors of $z - \tilde{W}_N / 2$ satisfy a finite difference equation which can be interpreted as a discretisation of the stochastic Airy equation. This yields a refined asymptotic approximation for the log-determinant.

Our CLT for the parabolic regime does not rely on the stochastic Airy equation machinery. Instead, we use asymptotics of 1-point correlation function for GUE to link hyperbolic and parabolic regimes. Although the resulting asymptotic approximations are less refined than those obtained in [34], they do deliver the CLT for the log-determinant. In contrast to [33, 34], we extend our asymptotic results to the general Wigner setting.
Another related and independently written paper is Augeri, Butez, and Zeitouni [4]. It deals with the CLT for $\beta$-ensembles when the singularity $E$ is a fixed number in the bulk $(-2,2)$. Such a location of the singularity implies that, as $i$ goes from 1 to $N$, the recurrence for the minors of $W_N - E$ goes through all the regimes, starting from the hyperbolic, transiting through the parabolic, and finishing in the elliptic regime. [4] refers to these regimes as “scalar”, “transitory”, and “oscillatory”. The analysis in the “scalar” regime is similar to ours. However, that of the “transition” regime is based on combinatoric arguments, whereas ours is using 1-point correlation and the asymptotics of the Stieltjes transform.

In contrast to [4], we do not analyze “oscillatory” regime because we focus on the edge singularity $E$. Although [4] only consider $E$ inside $(-2,2)$, we believe that their analysis can be extended to the edge with some extra work. Unlike [4], we do extend our results to Wigner matrices.

Finally, an interesting recent paper by Bourgade, Mody, and Pain [14] obtains a CLT for the real and imaginary parts of the log determinant of $\beta$-ensemble when the singularity is in the bulk. The proof is based on a new local law result, and is completely different from the proof used in our paper. We do not know whether the proof of [14] can be extended to the case of the singularity at the edge.

1.4. Organization and some notation. The rest of the paper is organized as follows. Section 2 states a CLT for the log determinant for the special case of slowly diverging $\sigma_N$, and outlines its proof. Section 3 extends this result to constant (possibly negative) $\sigma_N$, which yields Theorem 1. Section 4 implements the strategy of the proof outlined in section 2 and establishes key bounds on the Stieltjes transform that are postulated in section 3. Sections 5 and 6 analyze Wigner matrices and the spiked case, respectively. Relatively more technical proofs are compiled in the Supplementary Material.

Notations $a_N \ll b_N$, $a_N \lesssim b_N$, and $a_N \asymp b_N$ mean, respectively, that $a_N/b_N \to 0$, that $a_N \leq C b_N$ for some $C$ and $N$ large, and that $a_N \lesssim b_N$ and $b_N \lesssim a_N$.

2. G/BE: The CLT slightly away from the edge. In this section we establish the following analogue of Theorem 1 for general Gaussian $\beta$-ensembles in cases where $(\log \log N)^2 \ll \sigma_N \ll \log^2 N$. Hence, the location of singularity $E = E_N$ is slightly away from the edge in the sense that $(E - 2) N^{2/3}$ slowly diverges to infinity.

**Theorem 2.** Consider matrix $\tilde{W}_N$ from a (scaled) general Gaussian $\beta$-ensemble (5) with $\beta = 2/\alpha$. Let $D_N = \det(\tilde{W}_N - 2\theta_N)$, where $2\theta_N \equiv E = 2 + N^{-2/3} \sigma_N$ with $(\log \log N)^2 \ll \sigma_N \ll (\log N)^2$. Then, $$(\log |D_N| - \mu_N)/\tau_N \xrightarrow{d} N(0,1),$$

where

$$\tau_N = \sqrt{\alpha \rho(\theta_N^{-2})} \quad \text{with} \quad \rho(x) = \log \frac{1}{2} [1 + (1 - x)^{-1/2}].$$

**Remark.** The modified scaling in the above CLT naturally arises from the arguments in our proof. Note that

$$\rho(\theta_N^{-2}) = \frac{1}{4} \log N - \frac{1}{2} \log \sigma_N - \log 2 + O(N^{-1/3} \sigma_N^{1/2}),$$

so that the asymptotic variance of $\log |D_N|$ is $\frac{3}{2} \log N$, as in Theorem 1. However, our Monte Carlo experiments (which we do not report here) indicate that the scaling in Theorem 2 makes the standard normal approximation better in finite samples.
The proof of Theorem 2 is based entirely on the recurrence equation (6) with application of some well-known deviation and concentration inequalities for sums of independent random variables. In this section, we briefly outline the main steps of the proof. Details can be found in section 4.1.

Define normalized versions of the characteristic roots \( \rho_j^\pm \):

\[
(11) \quad r_i = 1 + \sqrt{1 - \frac{i - 1}{N\theta_N}}, \quad m_i = 1 - \sqrt{1 - \frac{i - 1}{N\theta_N}}.
\]

In particular, \( |\rho_i^\pm| = \theta_N r_i \). Then, from (6) and the identities \( r_i + m_i = 2 \) and \( r_i m_i = (i - 1)/N\theta_N^2 \), the normalized determinants (7) follow the recurrence

\[
(12) \quad M_i = (\alpha_i - \gamma_i - 1) M_{i-1} - (\gamma_i + \beta_i - \delta_i) M_{i-2},
\]

where

\[
\alpha_i = \frac{a_i}{\sqrt{N\theta_N}}, \quad \gamma_i = \frac{m_i}{r_i}, \quad \beta_i = \frac{\sqrt{\gamma_i} c_{i-1}}{\sqrt{N\theta_N r_{i-1}}}, \quad \delta_i = \frac{m_i}{r_i} - \frac{m_i}{r_{i-1}}.
\]

With the conventions \( M_0 = 1, M_{-1} = 0 \), and declaring \( c_0 = 0 \), so that \( \beta_1 = 0 \) along with \( \gamma_1 = \delta_1 = 0 \), equation (12) holds for \( i = 1, \ldots, N \).

Dividing both sides of (12) by \( M_{i-1} \) yields

\[
(13) \quad R_i = M_i \frac{M_i}{M_{i-1}} + 1 = \alpha_i - \gamma_i + \frac{\gamma_i + \beta_i - \delta_i}{1 - R_{i-1}},
\]

which can be rewritten as a recurrence

\[
(14) \quad R_i = \xi_i + \gamma_i R_{i-1} + \varepsilon_i,
\]

for \( i = 1, \ldots, N \), with the definitions

\[
\xi_i = \alpha_i + \beta_i,
\]

\[
(15) \quad \varepsilon_i = -\delta_i + (\beta_i - \delta_i) \frac{R_{i-1}}{1 - R_{i-1}} + \gamma_i \frac{R_{i-1}^2}{1 - R_{i-1}}.
\]

By dropping the non-linear term \( \varepsilon_i \) from (14), we now define a linear process \( \{L_i\}_{i=1}^N \) satisfying the recursion

\[
(16) \quad L_i = \xi_i + \gamma_i L_{i-1}.
\]

In particular, \( L_1 = \xi_1 = \alpha_1 \). Note that \( \{\xi_i\} \) are independent random variables, while \( \{\gamma_i\} \) are deterministic.

To establish the CLT, we study the dynamics of \( L_i \) and \( R_i \). Our proof consists of the following three steps:

1. First, derive a CLT for \( \sum_{j=1}^N L_j \) with the variance of exact order \( \log N \).
2. Then, show in the regime \( \sigma_N \gg (\log \log N)^2 \) that both \( \max_i |L_i| = o_P(N^{-1/3}) \) and \( \max_i |R_i| = o_P(N^{-1/3}) \). This allows us to use Taylor’s approximation for the logarithm, so that

\[
\log |M_N| = \sum_{j=1}^N \log |1 - R_j| = \sum_{j=1}^N (-R_j - R^2_j/2) + o_P(1).
\]

3. Finally, prove that the sum \( \sum_{j=1}^N (-R_j - R^2_j/2) \) can be replaced with \( \sum_{j=1}^N -L_j \) at the cost of some \( O_P(\log \log N) \) error term and with some explicit deterministic shift.
Achieving these objectives will show that the asymptotic behavior of \(|M_N|\) is the same as that of \(-\sum_{i=1}^{N} L_i\) up to \(O_p(\log \log N)\) and an explicit deterministic shift, and hence an appropriately centered \(|M_N|\) satisfies the same CLT as \(-\sum_{i=1}^{N} L_i\). After calculating the deterministic shift between \(\log |D_N|\) and \(\log |M_N|\), we derive the CLT for the log-determinant as required.

3. G(U/O): All the way to the edge. Theorem 2 covers singularities \(2\theta_N = E = 2 + N^{-2/3} \sigma_N\) in the range \((\log \log N)^2 \ll \sigma_N \ll \log^2 N\) for all positive \(\alpha\). We seek to extend the result to singularities at a distance of exact order \(N^{-2/3}\) away from the edge, or even (just) inside the bulk. We consider now sequences \(\sigma_N\) satisfying

\[
- \gamma \leq \sigma_N \leq \sigma_N := (\log \log N)^3 \quad \text{for some } \gamma > 0.
\]

Our extension will rely on the properties of GUE and GOE, and so covers only the cases \(\alpha = 1\) and \(\alpha = 2\). Indeed, the main tool is uniform approximation of the one-point function of GUE for regions up to and containing the spectral edge, based chiefly on results of Göetze and Tikhomirov [23].

In Section 4.2, we use the one-point function approximation to obtain the following estimates on the Stieltjes transform and its derivative near the edge.

**Proposition 3.** Suppose that \(\alpha = 1\) or \(\alpha = 2\) and let \(\sigma_N = N^{2/3}(E - 2)\) satisfy condition (17). Then

\[
\sum_{i=1}^{N} (E - \lambda_i)^{-1} - N = O_p \left( (1 + |\sigma_N|^{1/2}) N^{2/3} \right)
\]

and

\[
\sum_{i=1}^{N} (E - \lambda_i)^{-2} = O_p(N^{4/3}).
\]

**Proof of Theorem 1 in the range (17).** Let

\[
S_N(\sigma_N) := \sum_{i=1}^{N} \log |2 + N^{-2/3} \sigma_N - \lambda_i| - \mu_N(\sigma_N) = \sum_{i=1}^{N} \log |E - \lambda_i| - \mu_N(\sigma_N),
\]

where \(\mu_N(\sigma_N)\) is given by (3). The strategy is to use Proposition 3 to show that

\[
S_N(\sigma_N) - S_N(\sigma_N) = O_p \left( (\log \log N)^6 \right),
\]

so that \(S_N(\sigma_N)/\sqrt{4/3 \log N}\) has the same limiting \(\mathcal{N}(0, 1)\) distribution as \(S_N(\sigma_N)/\sqrt{4/3 \log N}\), the latter being given by Theorem 2.

Abbreviate \(\ell_N = (\log \log N)^6\) and note that \(|\sigma_N|^{3/2} \leq \sigma_N^{3/2} = o(\ell_N)\). Introducing

\[
\delta_N = N^{-2/3} (\sigma_N - \sigma_N),
\]

\[
d_i = \log |E - \lambda_i + \delta_N| - \log |E - \lambda_i| - \delta_N (E - \lambda_i)^{-1},
\]

we can decompose

\[
S_N(\sigma_N) - S_N(\sigma_N) = \sum_{i=1}^{N} \left( \log |E - \lambda_i + \delta_N| - \log |E - \lambda_i| \right) - (\sigma_N - \sigma_N) N^{1/3} + o(\ell_N)
\]

\[
= \sum_{i=1}^{N} d_i + \delta_N \left[ \sum_{i=1}^{N} (E - \lambda_i)^{-1} - N \right] + o(\ell_N).
\]
We will show that for each $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$\left| \sum_{i=1}^{N} d_i \right| \leq \delta_N^2 \sum_{i=1}^{N} (E - \lambda_i)^{-2} + o(\ell_N). \quad (20)$$

The bound (18) then follows directly from Proposition 3, since $N^{2/3}\delta_N = O(\ell_N^{1/2})$. Thus it remains to establish (20). For this we use some consequences of convergence to the Tracy-Widom law formulated in the following lemma, proved in appendix A.1.

**Lemma 4.** Suppose that $W_N$ is (scaled) GUE/GOE. Let $\gamma > 0$ be fixed, and suppose that $E = 2 + \sigma_N^{-2/3}$ with $\sigma_N > -\gamma$, and that $\bar{E} = 2 + N^{-2/3}\tilde{\sigma}_N$. Then for each $\varepsilon > 0$ small, there exists $k = k(\varepsilon, \gamma)$ such that for large $N$,

$$\mathbb{P}(\lambda_1 > E - N^{-2/3}) < \varepsilon, \quad \mathbb{P}(\lambda_k > E) < \varepsilon. \quad (21)$$

Moreover, there are constants $c_1 = c_1(\varepsilon, \gamma)$ small and $C_1 = C_1(\varepsilon, \gamma)$ large, such that for large enough $N$,

$$\mathbb{P}\left( \min_{i \leq N} |\lambda_i - E| < c_1 N^{-2/3} \right) < \varepsilon, \quad \mathbb{P}\left( \max_{i \leq k} |\lambda_i - E| > (C_1 + |\sigma_N|) N^{-2/3} \right) < \varepsilon. \quad (22)$$

Turning to (20), our first goal is to establish probabilistic bounds on $|d_i|$ for $i = 1, \ldots, N$. Let $\mu_i = E - \lambda_i$. Given $\varepsilon$ and $\gamma$, Lemma 4 yields $k, c_1, C_1$ such that the event $\mathcal{E} = \mathcal{E}(k, c_1, C_1)$ given by

$$\mathcal{E} = \left\{ \lambda_1 \leq E - N^{-2/3}, \mu_k > 0, \min_{i=1, \ldots, N} |N^{2/3}\mu_i| \geq c_1, \max_{i=1, \ldots, k} |N^{2/3}\mu_i| \leq C_1 + |\sigma_N| \right\}$$

has probability at least $1 - \varepsilon$. On event $\mathcal{E}$, for $i \geq k$, the bound $|\log(1 + x) - x| \leq x^2/2$ for $x \geq 0$ implies that

$$|d_i| \leq \frac{1}{2} \delta_N^2 \mu_i^{-2}.$$ 

Further, note that

$$|d_i| = \left| \log |\sigma_N - \sigma_N + N^{2/3}\mu_i| - \log |N^{2/3}\mu_i| - \delta_N \mu_i^{-1} \right|.$$ 

On the other hand, still on $\mathcal{E}$, for any $i \leq k$ and all sufficiently large $N$, the first logarithm is non-negative and no larger than $\log(3\tilde{\sigma}_N + C_1)$; the second one is no larger in absolute value than $|\log c_1 + \log(\sigma_N + C_1)|$; and the last term on the right hand side of the above display is no larger in absolute value than $2\tilde{\sigma}_N/c_1$. Each of these bounds is $O(\ell_N^{1/2})$.

Hence, overall on $\mathcal{E}$,

$$\left| \sum_{i=1}^{N} d_i \right| \leq \delta_N^2 \sum_{i=1}^{N} \mu_i^{-2} + C_2 \ell_N^{1/2} \quad (23)$$

for some constant $C_2 = C_2(\varepsilon, \gamma)$. Since $\mathbb{P}(\mathcal{E}) \geq 1 - \varepsilon$, the latter inequality holds for sufficiently large $N$, with probability at least $1 - \varepsilon$.

This yields (20) and completes the proof. \qed

4. **Proofs for Gaussian ensembles.**

4.1. **Proofs from Section 2.** This section implements the three steps of the analysis, described in Section 2, that lead to Theorem 2.
4.1.1. Preliminaries. For $p \geq 1$, denote by $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ the $p$-norm of a random variable $X$.

Sub-gaussianity. We say that a centred random variable $X$ belongs to the sub-gamma family $SG(v, u)$ for $v, u > 0$, if

$$\log \mathbb{E}e^{tX} \leq \frac{t^2v}{2(1-tu)}, \quad \forall t : |t| < \frac{1}{u}. \quad (24)$$

If $X \in SG(v, u)$ then $X \in SG(v', u')$ for each $v' \geq v, u' \geq u$. For arbitrary $c \in \mathbb{R}$, we have $cX \in SG(c^2v_X, |c|u_X)$. If $X \in SG(v_X, u_X)$ and $Y \in SG(v_Y, u_Y)$ are independent, then $X + Y \in SG(v_X + v_Y, \max\{u_X, u_Y\})$. If $X \sim \mathcal{N}(0, 1)$, then $X \in SG(1, 0)$, and if $X \sim \chi^2(d) - d$, then $X \in SG(2d, 2)$. We refer to chapter 2.4 of [12] for this last result.

The recurrence parameters. We often view the deterministic sequences $r_i, m_i, \gamma_i, \delta_i, g_i$ etc. as discretizations of functions evaluated at $x_i = (i - 1)/(N\theta^2_N)$, with step size $\Delta_N = 1/(N\theta^2_N)$. For example, $r_i = r(x_i)$ with $r(x) = 1 + \sqrt{1 - x}$ for $x \in [0, 1)$, where this function is concave decreasing.

The operator $T$. Iterating (16) yields

$$L_i = (T\xi)_i = \xi_i + \gamma_i\xi_{i-1} + \cdots + \gamma_i\cdots\gamma_2\xi_1, \quad i \geq 2,$$

$$L_1 = (T\xi)_1 = \xi_1. \quad (25)$$

The generalized exponential moving average used in (25) is a linear map $T = T_N : \mathbb{R}^N \rightarrow \mathbb{R}^N$ formally defined by

$$Ta_i = (Ta)_i = \begin{cases} a_i + \gamma_i a_{i-1} + \cdots + \gamma_i \cdots \gamma_2 a_1 & i \geq 2 \\ a_1 & i = 1. \end{cases} \quad \forall a \in \mathbb{R}^N.$$ 

The corresponding matrix $T = (T_{ij})$ is lower triangular, with entries

$$T_{ij} = \gamma_i \cdot \delta_{i+1}, \quad \delta_{ij} = \begin{cases} \gamma_i \gamma_{i-1} \cdots \gamma_{j} & i \geq j \\ 1 & i = j - 1 \\ 0 & i < j - 1. \end{cases} \quad (26)$$

Resummation yields

$$\sum_{i=1}^{N} Ta_i = \sum_{j=1}^{N} g_{j+1} a_j, \quad (27)$$

where $g_j = (T^*1)_{j-1}$ is given by

$$g_j = 1 + \gamma_j + \cdots + \gamma_N \cdots \gamma_j = 1 + \sum_{i=j}^{N} \gamma_i \cdot \delta_{ij} \quad \forall j \leq N. \quad (28)$$

for $2 \leq j \leq N$ and $g_{N+1} = 1$. Since $\gamma_i$ is increasing, we have

$$|Ta_i| \leq \frac{1}{1 - \gamma_i} \max_{j \leq i} |a_j|, \quad (29)$$

which simplifies to $Ta_i \leq (1 - \gamma_i)^{-1}a_i$ for an increasing sequence $a_i \geq 0$. If $T_i$ satisfies a recurrence

$$T_i = \gamma_i T_{i-1} + a_i, \quad T_1 = a_1 \quad (30)$$

for $i = 2, \ldots, N$. If $\gamma_i > 0$ for all $i$, then the recurrence is clearly decreasing, and $T_i$ is concave decreasing. If $\gamma_i < 0$ for some $i$, then $T_i$ is concave increasing, and $T_i$ is concave increasing.
then $T_i = (Ta)_i$ and bound (29) applies.

**Estimates by integrals.** With $\Delta_N = 1/(N\theta_N^2)$, if $f(x)$ is increasing, then

$$
\sum_{i=a}^{b} f(x_i) \Delta_N \leq \int_{x_a}^{x_b + \Delta_N} f(x) \, dx.
$$

(31)

In the corresponding lower bound the integral has limits $x_a - \Delta_N$ and $x_b$. In particular,

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{(r_i - 1)^{\beta}} < \theta_N^2 \int_{0}^{x_a + \Delta_N} \frac{dx}{(r(x) - 1)^{\beta}} = \theta_N^2 \int_{0}^{\theta_N^2} \frac{dx}{(1 - x)^{3/2}}
$$

(32)

$$
\leq \begin{cases} 
C_\beta \theta_N^{2}\beta^{-1/2}N^{(\beta - 2)/3} & \text{if } \beta \in (0, 2) \\
C_\beta w_N^{-1/2}N^{(\beta - 2)/3} & \text{if } \beta > 2,
\end{cases}
$$

for sufficiently large $N$ with $C_\beta = 2|\beta - 2|^{-1}$ and $w_N = \sigma_N/2$. Alternatively, the error bound for the trapezoid rule, in integral form, states

$$
\left| \sum_{i=a}^{b} f(x_i) \Delta_N - \int_{x_a}^{x_b} f(x) \, dx \right| \leq \frac{\Delta_N}{8} \int_{x_a}^{x_b} |f''(x)| \, dx + \frac{\Delta_N}{2} \left[ |f(x_a)| + |f(x_b)| \right]
$$

(33)

$$
= \varepsilon N_1(f) + \varepsilon N_2(f).
$$

4.1.2. **Step one: CLT for $\sum_{i=1}^{N} L_i$.** Equation (27) applied to $L_i = T \xi_i$ yields

$$
\sum_{i=1}^{N} L_i = \sum_{i=1}^{N} g_{i+1} \xi_i,
$$

where $g_i$ is as in (28), and we recall that $\xi_i$ are independent zero-mean random variables. Lyapunov’s CLT implies that

$$
\frac{\sum_{i=1}^{N} L_i}{\sqrt{\sum_{i=1}^{N} g_{i+1}^2 \xi_i^2}} \xrightarrow{d} N(0, 1)
$$

as long as

$$
\sum_{i=1}^{N} g_{i+1}^4 \xi_i^4 / \left( \sum_{i=1}^{N} g_{i+1}^2 \xi_i^2 \right)^2 \to 0.
$$

(34)

Let us establish (34), and find $\sum_{i=1}^{N} g_{i+1}^2 \xi_i^2$.

The proofs of the following two lemmas are given in Subsections A.2.1 and A.2.2.

**Lemma 5.** For any integer $q \geq 1$ and all sufficiently large $N$, there exist constants $c_2, C_q > 0$, such that for all $1 \leq i \leq N$,

$$
\xi_i^{2q} \leq C_q \alpha^q N^{-q} \quad \text{and} \quad \xi_i^2 \geq c_2 \alpha N^{-1}
$$

**Lemma 6.** Let $w_N = \sigma_N/2$ so that $\theta_N = 1 + N^{-2/3} w_N$. Suppose that $(\log \log N)^2 \ll w_N \ll (\log N)^2$. Then, for all $1 \leq i \leq N - N^{1/3}$, any $k > 0$, and all sufficiently large $N$,

$$
g_i \geq \frac{r_i}{2(r_i - 1)} \left( 1 - \log^{-k} N \right).
$$

(35)
Further, for all $1 \leq i \leq N$ and all sufficiently large $N$, 

\begin{equation}
 g_i < \frac{r_i}{2(r_i - 1)} \left( 1 + w_N^{-3/2} \right).
\end{equation}

There is also a trivial bound, sharper than (36) for $N - N^{1/3} \leq i \leq N + 1$:

\begin{equation}
 g_i \leq N - i + 2.
\end{equation}

Let $n_0 = N - \lfloor N^{1/3} \rfloor - 1$. These lemmas yield, along with Riemann sum bounds like (31), (32)

\[ \sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2 \geq \frac{1}{N\theta_N^2} \sum_{i=2}^{n_0} \frac{1}{(r_i - 1)^2} > \int_{x_n^0}^{1} \frac{dx}{1-x} = -\log(1-x_{n_0}) \geq \log N. \]

For the last inequality, note that $1 - x_{n_0} = 1 - \theta_N^{-2} + O(N^{-2/3})$. On the other hand,

\begin{equation}
 1 - \theta_N^{-2} = 2w_N N^{-2/3} + O(w_N^2 N^{-4/3}).
\end{equation}

**Remark.** The logarithmic growth of $\sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2$ is a consequence of our choosing $\theta_N$ local to one. Had it been separated from one, the asymptotic variance of $\sum_{i=1}^{N} L_i$ would be constant. This agrees well with the fact that linear spectral statistics without singularities local to one. Had it been separated from one, the asymptotic variance of $N \mathbf{E} \xi_i$ would be constant. This agrees well with the fact that linear spectral statistics without singularities close to the edge of the semi-circle law do not need scaling for the convergence to normality.

Similarly, from (32) and (36)

\[ \sum_{i=1}^{N} g_{i+1}^4 \mathbf{E} \xi_i^4 \leq \frac{1}{N^2} \sum_{i=2}^{n_0} \frac{1}{(r_i - 1)^4} \leq \frac{1}{N^{1/3} w_N}. \]

Hence,

\[ \sum_{i=1}^{N} g_{i+1}^4 \mathbf{E} \xi_i^4 / \left( \sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2 \right)^2 \leq \frac{1}{N^{1/3} w_N \log^2 N} \to 0, \]

which establishes the Lyapunov condition (34). Let us now approximate $\sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2$.

Since, as we have just shown, $\sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2 \geq \log N$, we will tolerate approximation errors of magnitude $o(\log N)$. The following lemma is established in Subsection A.2.3.

**Lemma 7.** Under assumptions of Lemma 6 for all $1 \leq i \leq N$,

\[ \mathbf{E} \xi_i^2 = \frac{2\alpha}{N\theta_N^2 r_i^3} (1 + \varepsilon_i), \quad |\varepsilon_i| < \frac{1}{N(r_i - 1)} \leq N^{-2/3}. \]

Combining this lemma with Lemma 6 yields

\[ g_{i+1}^2 \mathbf{E} \xi_i^2 = \begin{cases} \frac{\alpha}{2N\theta_N^2 r_i^3} \left[ 1 + O(w_N^{-3/2}) \right] & 1 \leq i \leq N - N^{1/3} \\ O(N^{-1}(N - i + 1)^2) & N - N^{1/3} \leq i \leq N \end{cases} \]

Over the second range, the sum $\sum_{i=1}^{N} g_{i+1}^2 \mathbf{E} \xi_i^2 = O(1)$, which will be negligible. Over the first range, introduce $f(x) = 1/[(r(x) - 1)^2 r(x)] = 1/[(1 - x)(1 + \sqrt{1 - x})]$. Monotonicity of $r(x)$ yields

\begin{equation}
 f(x_i) \leq \frac{r_{i+1}^2}{(r_{i+1} - 1)^2 r_{i+1}^3} \leq f(x_{i+1}).
\end{equation}
Apply the trapezoidal rule bounds (33) with \(|f''(x)| \lesssim (1-x)^{-3}\), and \(\epsilon_{N1}(f) = O(N^{-2+4/3})\), while \(\epsilon_{N2}(f) = O(N^{-1+2/3})\). With \(x_a = 0\) and \(x_b = (N - [N^{1/3}])/N\theta_N^2 = \theta_N^{-2} + O(N^{-2/3})\),

\[
\frac{1}{N\theta_N^2} \sum_{i=a}^{b} f(x_i) = \int_{1-x_b}^{1-x_a} \frac{dx}{(1+x)^2} + O(N^{-1/3}) = 2\rho(x_b) + O(N^{-1/3}),
\]

where \(\rho(x) = \log \left[ \frac{1}{2} \left( 1 + \frac{1}{\sqrt{1-x}} \right) \right]\). Jumps of 1 in \(a, b\) to cover the two sides of (39) do not alter the approximation. In the range of interest, \(0 \leq \rho(x) \leq (1-x)^{-1} \leq (1-\theta_N^{-2})^{-1} = O(w_N^{-1} N^{2/3})\), so that \(\rho(x_b) = \rho(\theta_N^{-2}) + O(w_N^{-1})\). In summary,

\[
\sum_{i=1}^{N} \xi_i^2 = [\alpha \rho(\theta_N^{-2}) + O(w_N^{-1})] [1 + O(w_N^{-3/2})] + O(1)
\]

Recalling that \(\tilde{\tau}_N^2 = \alpha \rho(\theta_N^{-2})\), we have established the following theorem.

**THEOREM 8.** If \(\theta_N = 1 + N^{-2/3} w_N\) with \((\log \log N)^2 \ll w_N \ll (\log N)^2\). Then,

\[
\tilde{\tau}_N^{-1} \sum_{i=1}^{N} L_i \sim N(0, 1).
\]

### 4.1.3. Step 2a: Uniform bound on \(L_i\).

**LEMMA 9.** If \(\theta_N = 1 + N^{-2/3} w_N\) with \((\log \log N)^2 \ll w_N \ll (\log N)^2\). Then,

\[
\max_{1 \leq i \leq N} |L_i| = o_P \left( N^{-1/3} \right).
\]

To this end, we show that \(\xi_i\) and \(L_i\) are sub-gamma variables and apply exponential tail inequalities. The proof of the following lemma is given in appendix A.2.4.

**LEMMA 10.** For any \(1 \leq 1 \leq N\), \(\xi_i \in SG(v_i, u_i)\) and \(L_i \in SG(v_{Li}, u_{Li})\) with

\[
v_i = \frac{2\alpha}{N\theta_N^2 r_i^2}, \quad v_{Li} = \frac{\alpha}{2N\theta_N^2 (r_i - 1)} \quad \text{and} \quad u_i = u_{Li} = \frac{\alpha}{N\theta_N^2 r_i^2}.
\]

**PROOF OF LEMMA 9.** Suppose that \(L \in SG(v, u)\) and that \(1 > 2vt > ut\), as will happen in our uses below. Then

\[
P(|L| > 2\sqrt{2vt}) \leq P(|L| > \sqrt{2vt} + ut) \leq e^{-t},
\]

where the second inequality, valid for \(t > 0\), is essentially the display prior to Theorem 2.3 in [12]. When applied to \(L = L_i\), the bound

\[
B_i(t) = 2\sqrt{2vt} \leq \sqrt{t/N} (1 - x_i)^{-1/4}
\]

increases with \(i\), and for \(i \leq n_0 = N - [N^{1/3} \log^{2+\eta} N]\), it is of order \(\sqrt{T} N^{-1/3} \log^{-1/2-\eta/4} N\). We may take a union bound over such \(i\) in (41) by replacing \(t \leftarrow \log N\) because \(B_{n_0}(t + \log N) = o(N^{-1/3})\). More precisely, with \(\eta = 1\) and \(t = 5 \log \log N + \log N\), there exists an absolute constant \(C_1\) such that

\[
P \left( \max_{1 \leq i \leq n_0} |L_i| > \frac{C_1}{N^{1/3} \log^{1/4} N} \right) \leq \frac{2}{\log \log N}.
\]
For $i$ close to $N$, however, this approach fails since $B_N(t) \asymp \sqrt{t}N^{-1/3}w_N^{-1/4}$, which is no longer $o(N^{-1/3})$ unless $t = O(\log \log N)$.

Instead, we use a simple form of chaining plus a version of the Kolmogorov maximal inequality. We pick indices $n_0 < n_1 < \ldots < n_K = N$ and use a bound of the form

\begin{equation}
P\left(\max_{i > n_0} |L_i| > \epsilon_1 + \epsilon_2\right) \leq P\left(\max_{0 \leq k < K} |L_{n_k}| > \epsilon_1\right) + \sum_{k=0}^{K-1} P\left(\max_{n_k < j \leq n_{k+1}} |L_j - L_{n_k}| > \epsilon_2\right).
\end{equation}

Iterating the relation $L_i = \xi_i + \gamma_i L_{i-1}$ and recalling definition (26) of $\gamma_{j:i}$, we have for $i < j$

\[ L_j = \gamma_{j:i+1} L_i + \sum_{k=i+1}^{j} \gamma_{j:k+1} \xi_k, \]

Since $\gamma_{j:i+1} < 1$, we then have

\[ L_j - L_i < L_j/\gamma_{j:i+1} - L_i =: L_{i,j}, \]

where

\begin{equation}
L_{i,j} = \sum_{k=i+1}^{j} \xi_k/\gamma_{k:i+1}
\end{equation}

are partial sums of independent random variables. Set $i = n_k$. We then have

\begin{equation}
P\left(\max_{n_k < j \leq n_{k+1}} |L_j - L_{n_k}| > 4\epsilon\right) \leq P\left(\max_{n_k < j \leq n_{k+1}} |L_{n_k,j}| > 4\epsilon\right) \leq 4 \max_{n_k < j \leq n_{k+1}} P(|L_{n_k,j}| > \epsilon),
\end{equation}

where the second inequality uses a maximal inequality, Theorem 1 from [19].

Now choose $K = \lfloor \log^5 N \rfloor$ and for $k = 1, \ldots, K$, let $n_k$ be the closest integer to $n_0 + kN^{1/3}\log^{-2} N$. For these intervals the products $\gamma_{j:i+1}$ are not too small: in Subsection A.2.5 we prove

\begin{lemma}
Under assumptions of Lemma 9, if $N - N^{1/3}\log^3 N \leq i < j \leq N$ and $j - i \leq N^{1/3}\log^{-2} N + 1$ then for large $N$

\[ \gamma_{j:i+1} \geq 1/2. \]
\end{lemma}

For $j \in (n_k, n_{k+1}]$ we therefore have $\gamma_{j:n_k+1} \geq 1/2$, and so from (44) and lemma 10

\[ L_{n_k,j} \in SG\left(\frac{8\alpha + 1}{N\theta^2/3} \log^2 N, \frac{2\alpha}{N\theta^2/3} \log^2 N\right). \]

For some absolute constant $C_2$ and large $N$, the tail bound (41) then implies, with $t = 10\log \log N$,

\[ \max_{n_k < j \leq n_{k+1}} P\left(|L_{n_k,j}| > \frac{C_2\sqrt{\log \log N}}{N^{1/3}\log^3 N}\right) \leq \frac{2}{\log^{10} N}. \]

With the same $t$ and recalling $B_{n_k}(t) \leq B_{N}(t) \asymp \sqrt{t}N^{-1/3}w_N^{-1/2}$, we can find $C_3$ so that

\[ P\left(|L_{n_k}| > \frac{C_3\sqrt{\log \log N}}{N^{1/3}w_N^{1/2}}\right) \leq \frac{2}{\log^{10} N}. \]
For $i > n_0$, use the last two bounds and the maximal inequality (45), and for $i \leq n_0$ recall (42). We conclude that on an event of probability at least $1 - 12 \log^{-5} N$ we have

$$\max_{1 \leq i \leq N} |L_i| \leq C_4 N^{-1/3} \varepsilon_N$$

where $\varepsilon_N = \max(\log^{-1/4} N, \sqrt{\log \log N / w_N}) \leq 1/\sqrt{\log \log N}$ under the assumptions on $w_N$. This completes the proof of Lemma 9. \hfill \Box

4.1.4. *Step 2b: Uniform bound on $R_i$*. We show that $R_i$ is close enough to $L_i$ so that an analogous uniform bound holds.

**Lemma 12.** Under the assumptions of Lemma 9,

$$\max_{1 \leq i \leq N} |R_i| = o_P \left( N^{-1/3} \right).$$

The starting point for analysis of the nonlinear process $R_i$ is the perturbation representation $R_i = L_i + (T \varepsilon_i)$. Note that decomposition (15) expresses $\varepsilon_i$ in the form $\varepsilon_i = \varepsilon(R_i, \beta_i, \gamma_i, \delta_i)$. Consider a Winsorized version of $R_i$

$$\tilde{R}_i = \phi_{N^{-1/3}/2} (R_i), \quad \phi_u(x) = \begin{cases} -u & x < -u \\ x & |x| \leq u \\ u & x > u, \end{cases}$$

and create a modified series from $\tilde{R}_{i-1}$:

$$\tilde{\varepsilon}_i = \varepsilon(\tilde{R}_{i-1}; \beta_i, \gamma_i, \delta_i)$$

$$(47) \quad \tilde{R}_i = L_i + T \tilde{\varepsilon}_i.$$  

We will show

$$\max_{1 \leq i \leq N} |\tilde{R}_i| = o_P \left( N^{-1/3} \right).$$

A key observation is that on the event $\tilde{R}_N = \{ \max_{1 \leq i \leq N} |\tilde{R}_i| \leq N^{-1/3}/2 \}$, we have $\tilde{R}_i = R_i$ for $i = 1, \ldots, N$. Indeed, $\tilde{\varepsilon}_1 = 0$ since $\varepsilon_1 = 0$, and $\tilde{R}_1 = L_1 = R_1$ so $|R_1| = |\tilde{R}_1| \leq N^{-1/3}/2$ on $\tilde{R}_N$ and hence $\tilde{R}_1 = R_1$. Then (47) implies $\tilde{\varepsilon}_2 = \varepsilon_2$, so $\tilde{R}_2 = R_2$ and so again $|R_2| = |\tilde{R}_2| \leq N^{-1/3}/2$, and so $\tilde{R}_2 = R_2$. Hence $\tilde{\varepsilon}_3 = \varepsilon_3$ and so on, so that eventually $\tilde{R}_i = R_i = \tilde{R}_i$ for $i = 1, \ldots, N$.

But this observation and (48) imply Lemma 12, since for any $\epsilon > 0$

$$P(\|R\|_\infty \leq \epsilon N^{-1/3}) \geq P(\|R\|_\infty \leq \epsilon N^{-1/3}, \tilde{R}_N) = P(\|\tilde{R}\|_\infty \leq \epsilon N^{-1/3}) \to 1.$$  

We now outline the proof of (48), leaving more technical details to appendix A.2.6. By lemma 9, it is sufficient to prove that $\max_{1 \leq i \leq N} |T \tilde{\varepsilon}_i| = o_P(N^{-1/3})$. The terms $\tilde{\varepsilon}_i$ may be decomposed by rewriting a Winsorized version of (15) as follows, after defining $\tilde{R}_{i-1}^{(1)} = R_{i-1} - (R_{i-1}/(1 - R_{i-1}))$:

$$\tilde{\varepsilon}_i = -\delta_i + \beta_i \tilde{R}_{i-1}^{(1)} + \tilde{R}_{i-1}^{(1)} (\gamma_i \tilde{R}_{i-1}^2 - \delta_i) + \gamma_i \tilde{R}_{i-1}^2, \quad i \geq 1.$$  

Let $\tilde{\varepsilon}_i^m$, $\tilde{\varepsilon}_i^g$ and $\tilde{\varepsilon}_i^q$ respectively denote the last three terms. Using linearity of $T$, we have

$$T \tilde{\varepsilon}_i = -T \delta_i + T \tilde{\varepsilon}_i^m + T \tilde{\varepsilon}_i^g + T \tilde{\varepsilon}_i^q.$$
It is relatively straightforward to establish sufficiently tight bounds on the terms $T\delta_i$, $T\bar{\varepsilon}_i$, $T\bar{\varepsilon}_i^0$ using (29) (see appendix A.2.6). For $T\bar{\varepsilon}_i^m$, note that $\bar{\varepsilon}_i^m = \beta_i \bar{R}_{i-1}^{(1)}$ is a martingale difference, since $\beta_i$ has mean 0 and is independent of $\bar{R}_{i-1}^{(1)}$. The term $T\bar{\varepsilon}_i^m$ can be viewed as a special case of the quantity

$$T(\beta R)_i = \sum_{j=1}^{i} \gamma_i j + 1 R_j \beta_j,$$

with $R_j$ measurable in the sigma-field generated by $\alpha_1, \beta_1, \ldots, \alpha_{i-1}, \beta_{i-1}$ and $\|R_j\|_p \leq \rho_N$. This is a sum $\sum_{j=1}^{i} X_j$ of martingale differences with $p$-th moments, and the Marcinkiewicz-Zygmund-type inequality of [45, Theorem 2.1] says that $\|\sum_{j=1}^{i} X_j\|^p \leq (p - 1) \sum_{j=1}^{i} \|X_j\|^p$.

In appendix A.2.6, we use this inequality together with the Markov inequality and the union bound to show that there exists $C > 0$ such that for all sufficiently large $N$, with probability at least $1 - 1/N$,

$$\max_i |T\bar{\varepsilon}_i^m| \leq CN^{-2/3} \log^{3/2} N.$$

Since the right hand side is obviously $o(N^{-1/3})$, this finishes the proof of (48).

### 4.1.5. Step 3: Linear approximation for $\log |M_N|$.

Recall that

$$\log |M_N| = \sum_{i=1}^{N} \log |1 - R_i|.$$ 

Since $\max_i |R_i| = o_P(N^{-1/3})$, we have a uniform Taylor’s approximation

$$\log |1 - R_i| = -R_i - R_i^2 / 2 + o_P(N^{-1}).$$

Summing up,

$$\log |M_N| = \sum_{i=1}^{N} (-R_i - R_i^2 / 2) + o_P(1).$$

(51)

In the rest of this subsection, our goal is to show that we can replace each term $-R_i - R_i^2 / 2$ with the linear process $L_i$, with inclusion of a deterministic shift. To be precise, we will show that

$$\sum_{i=1}^{N} (-R_i - R_i^2 / 2) + \sum_{i=1}^{N} L_i = \frac{1 - \alpha}{6} \log N + O_P(\log \log N).$$

(52)

Similarly to (50), we have

$$T\varepsilon_i = -T\delta_i + T\bar{\varepsilon}_i^m + T\bar{\varepsilon}_i^s + T\bar{\varepsilon}_i^q,$$

where with the notation $\bar{R}_{i-1}^{(1)} = R_{i-1} / (1 - R_j)$, we have

$$\bar{\varepsilon}_i^m = \beta_i \bar{R}_{i-1}^{(1)}, \quad \bar{\varepsilon}_i^s = \bar{R}_{i-1}^{(1)} (\gamma_i R_{i-1}^2 - \delta_i), \quad \bar{\varepsilon}_i^q = \gamma_i R_{i-1}^2.$$

The decomposition (53) leads to

$$\sum R_i = \sum L_i + T\bar{\varepsilon}_i^m + T\bar{\varepsilon}_i^s + T\bar{\varepsilon}_i^q - T\delta_i.$$

We will see that $\sum T\bar{\varepsilon}_i^m + T\bar{\varepsilon}_i^s = O_P(1)$. Note however that $\bar{\varepsilon}_i^q = \gamma_i R_{i-1}^2$ are positive, and will contribute to the deterministic shift. We therefore further decompose $T\bar{\varepsilon}_i^q$ using

$$\bar{\varepsilon}_i^q = \bar{\varepsilon}_i^{q\Delta} + (\bar{\varepsilon}_i^{qL} - \bar{\varepsilon}_i^{qE}) + \bar{\varepsilon}_i^{qE},$$
where the quadratic “difference”, “Linear approximation” and “Expectation” terms are respectively given by
\[
\begin{align*}
\epsilon_{q^d} &= \gamma_i(R^2_{i-1} - L^2_{i-1}) \\
\epsilon_{qL} &= \gamma_i L^2_{i-1} \\
\epsilon_{qE} &= \gamma_i E L^2_{i-1}.
\end{align*}
\]

An analysis similar to one we used to prove Lemma 12 leads to the following lemma. The proof is rather technical and is given in Subsection A.2.7.

**Lemma 13.** Under the assumptions of Lemma 9,
\[
\begin{align*}
\sum_i T\epsilon_i^m + T\epsilon_i^s + T\epsilon_i^{q^d} &= O_P(1), \\
\sum_i R_i^2 &= O_P(1), \\
\sum_i T\epsilon_i^{qL} - T\epsilon_i^{qE} &= o_P(1), \\
\sum_i T\epsilon_i^{qE} &= \alpha \sum_i T\delta_i + o_P(1).
\end{align*}
\]

Combining the results of lemma 13, we arrive at
\[
\sum_i R_i + R_i^2 / 2 = \sum_i L_i + (\alpha - 1) \sum_i T\delta_i + O_P(1).
\]

Remarkably, for \(\alpha = 1\), this will be the end of the proof. When \(\alpha \neq 1\), the remaining sum \((\alpha - 1) \sum_i T\delta_i\) results in an additional shift. The proof of the following lemma is postponed to Section A.2.8.

**Lemma 14.** It holds, for large enough \(N\),
\[
\sum_{i=1}^{N} T\delta_i = \frac{1}{6} \log N + O(\log \log N).
\]

Now the CLT for \(\log |M_N|\) follows from Theorem 8.

**Corollary 15.** Under the assumptions of Theorem 8,
\[
\frac{\log |M_N| + \frac{\alpha - 1}{6} \log N}{\tau_N} \xrightarrow{d} N(0, 1).
\]

4.1.6. CLT for \(\log |D_N|\). From (7) and (11), \(D_N = M_N \theta_N N \prod_{i=1}^{N} r_i\). Hence
\[
\begin{align*}
\log |D_N| &= \log |M_N| + N \log \theta_N + \sum_{i=1}^{N} \log (1 + \sqrt{1 - x_i}).
\end{align*}
\]

In the trapezoidal approximation to \(\Delta_N \sum_{i=1}^{N} \log(1 + \sqrt{1 - x_i})\) we have \(\varepsilon_{N^2}(f) = O(\Delta_N)\) and, since \(f''(x) \propto (1 - x)^{-3/2}\) for \(0 < x < 1\), also \(\varepsilon_{N^1}(f) = O(N^{1/3} \Delta_N^2)\). So from (33)
\[
\begin{align*}
\sum_{i=1}^{N} \log(1 + \sqrt{1 - x_i}) &= N \theta_N^2 \int_{1-x_N}^{1} \log(1 + \sqrt{u}) du + O(1) \\
&= \frac{1}{2} N \theta_N^2 - \frac{2}{3} (2w_N)^{3/2} + O(1).
\end{align*}
\]
At the second line we used \( \int_0^1 \log(1 + \sqrt{u}) \, du = \frac{1}{2} \) and, with \( a_N = 1 - x_N = 2w_NN^{-2/3} + O(N^{-1}) \), also \( \int_0^{a_N} \log(1 + \sqrt{u}) \, du = \frac{2}{3}a_N^{3/2} + O(a_N^2) \).

Using this in (58) together with (38) and

\[ N \log \theta_N = w_NN^{1/3} + O(w_N^2N^{-1/3}) \]

yields, recalling that \( 2w_N = \sigma_N \),

\[ \log |D_N| = \log |M_N| + \frac{1}{2}N + \sigma_NN^{1/3} - \frac{2}{3}\sigma_N^{3/2} + O(1). \]

Using Corollary 15, we obtain Theorem 2.

4.2. Proofs from Section 3.

4.2.1. Preliminaries. Correlation functions. Let \( P_N(x_1, \ldots, x_N) \) be a joint density of unordered eigenvalues \( l_1, \ldots, l_N \) of \( \text{G}(\text{U}/\text{O})E \) (scaled so that \( \max l_i \) is close to 2 for large \( N \)).

Following [48], the \( k \)-point correlation function is defined as

\[ R_k(x_1, \ldots, x_k) = \frac{N!}{(N-k)!} \int \cdots \int P_N(x_1, \ldots, x_N) \, dx_{k+1} \cdots dx_N. \]

Note that this is not a probability density: it has total integral \( N!/(N-k)! \).

For any integrable function \( F(x_1, \ldots, x_k) \), we have

\[ \mathbf{E} F(l_1, \ldots, l_k) = \frac{(N-k)!}{N!} \int \cdots \int F(x_1, \ldots, x_k) R_k(x_1, \ldots, x_k) \, dx_1 \cdots dx_k. \]

Write \( \rho_N(\lambda) = \rho_{N, \alpha}(\lambda) = N^{-1}R_1(\lambda) \) for the normalized one-point correlation function, interpreted as the “mean density” of the eigenvalues. The expected value of a linear spectral statistic can be written as

\[ \mathbf{E} \left[ N^{-1} \sum_{i=1}^N f(l_i) \right] = \mathbf{E} \left[ N^{-1} \sum_{i=1}^N f(l_i) \right] = \int f(\lambda) \rho_{N, \alpha}(\lambda) \, d\lambda. \]

A key tool in approximating such expectations will be a uniform bound, due to Gömez and Tikhomirov, for the deviation of the one-point function in GUE from the semi-circle density \( p_{SC}(x) = (2\pi)^{-1/2}4 - x^21_{|x| \leq 2} \). Indeed, [23, Theorem 1.2] show the existence of positive absolute constants \( a, A \) such that for all \( |x| \leq 2 - aN^{-2/3} \),

\[ |\rho_N(x) - p_{SC}(x)| \leq \frac{A}{N(4 - x^2)}. \]

Determinantal correlation functions imply the following elementary variance bound, which we prove in appendix A.2.9.

**Lemma 16.** For eigenvalues from GUE,

\[ \text{Var} \left[ N^{-1} \sum_{i=1}^N f(l_i) \right] \leq N^{-1} \int f^2(x) \rho_N(x) \, dx. \]

**Remark.** For the usual linear statistic, with \( f \) not depending on \( N \) and analytic in a neighborhood of \([-2, 2]\), this is a terrible bound since then \( \text{Var} \left[ \sum_{i=1}^N f(l_i) \right] = O(1) \). But in our critical case settings, it seems to give the right order, and will become useful below.
Edge bounds. For both GUE and GOE, we have
\begin{equation}
\rho_{N,\alpha}(2 + sN^{-2/3}) \leq \begin{cases} 
CN^{-1/3}e^{-2s} & s > -1 \\
CN^{-1/3}|s|^{1/2} & -N^{2/3} - \varepsilon < s \leq -1,
\end{cases}
\end{equation}
for any \(0 < \varepsilon < 2/3\). The bounds for GUE follow directly from the Götze-Tikhomirov bounds and Tracy-Widom asymptotics of the Hermite functions at the edge. For GOE the one-point function \(\rho_{N,2}\) differs from \(\rho_{N,1}\) by a term involving integrals of scaled Hermite functions, and this can again be analyzed by bounds on Hermite polynomials. These bounds are implicit in [28] and [23], but for the reader’s convenience some discussion appears in appendix A.3.

4.2.2. Gaussian Non-concentration. Let us introduce new notation
\begin{equation}
\sigma_N = (\log N)^{O(\log \log N)}.
\end{equation}

\textbf{Lemma 17.} Suppose that \(W_N\) is a matrix drawn from either GOE or GUE, divided by \(\sqrt{N}\) to have support of the limiting spectral distribution on \([-2,2]\). Suppose also that \(|E - 2| \leq N^{-2/3}\sigma_N\). Then for each \(c_0 > 0\) and each \(d \in (0,c_0)\), we have for \(N > N(d)\),
\begin{equation}
P(\min_j |\lambda_j - E| \leq N^{-2/3-c_0}) \leq N^{-d}.
\end{equation}

\textbf{Proof.} The lemma is an immediate consequence of the bounds
\begin{equation}
\rho_{N,\alpha}(E) \lesssim N^{-1/3}\sigma_N^{1/2}, \quad \alpha = 1,2,
\end{equation}
holding uniformly for all \(E : |E - 2| \leq N^{-2/3}\sigma_N\). Bounds (66) follow directly from (63).

Indeed, for a matrix \(M\) and set \(I\), let the number of eigenvalues of \(M\) in \(I\) be \(N_M(I)\). Set \(I = [E - N^{-2/3-c_0}, E + N^{-2/3-c_0}]\). We have
\begin{equation}
P(\min_j |\lambda_j - E| \leq N^{-2/3-c_0}) = P(N_{W_N}(I) \geq 1)
\end{equation}
\begin{align*}
\leq EN_{W_N}(I) = N \int_I \rho_{N,\alpha}(E) \, dE 
\lesssim 2\sigma_N^{1/2}N^{-c_0} \leq N^{-d}
\end{align*}
for each \(d < c_0\) and \(N\) large. \(\square\)

4.2.3. Proof of Proposition 3. First, we prove the proposition for GUE case \(\alpha = 1\). To establish Proposition 3 via (62), we will calculate the expectation of the truncated statistics
\begin{equation}
L_{IN} = \frac{1}{N} \sum_{j=1}^{N} f_c^l(\lambda_j), \quad l = 1,2,
\end{equation}
where
\begin{equation}
f_c(\lambda) = \frac{1}{E - \lambda} 1\{|E - \lambda| > cN^{-2/3}\}.
\end{equation}
The truncation makes the function integrable with respect to the density \(\rho_N(x)\).

The truncation is typically harmless: if there are no eigenvalues near \(E\), more precisely if \(N_W(E - cN^{-2/3}, E + cN^{-2/3}) = 0\), then
\begin{equation}
\sum_{j=1}^{N} f_c^l(\lambda_j) = \sum_{j=1}^{N} (E - \lambda_j)^{-l}.
\end{equation}
Lemma 4 assures that for \(\varepsilon > 0\) small, there exists \(c = c(\varepsilon, \gamma)\) small so that equality holds with probability at least \(1 - \varepsilon\). Therefore, it suffices to show that Proposition 3 holds with \(N^{-1} \sum (E - \lambda_j)^{-l}\) replaced by \(L_{IN}(c)\) for each \(c > 0\) fixed.

The main work, contained in the next lemma, is to control the expected values of \(L_{IN}\).
LEMMA 18. Suppose that $\sigma_N$ satisfies (17). Then for each $c > 0$ we have

$$
\mathbb{E}L_{1N} = \mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{N} f^l_c(\lambda_j)\right) = \begin{cases} 
1 + O\left((1 + |\sigma_N|^{1/2})N^{-1/3}\right), & l = 1 \\
O(N^{1/3}), & l = 2.
\end{cases}
$$

Lemma 18 and the variance bound (62) quickly yield Proposition 3. Indeed, for $L_{2N} > 0$, Proposition 3 holds since $L_{2N} = O_P(\mathbb{E}L_{2N})$. For $L_{1N}$, bound (62) implies

$$
\text{Var}(L_{1N}) \leq N^{-1}\mathbb{E}L_{2N} = O(N^{-2/3}),
$$

and we conclude Proposition 3 from $L_{1N} - \mathbb{E}L_{1N} = O_P(\sqrt{\text{Var}(L_{1N})})$.

PROOF OF LEMMA 18 (FOR GUE). First let us bound the error of replacing $\rho_N$ with $\rho_{SC}$ in the integral

$$
\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} f^l_c(\lambda_i)\right) = \int f^l_c(\lambda)\rho_N(\lambda)d\lambda.
$$

Abbreviate $\epsilon_N = N^{-2/3}\sigma_N$, $\delta_N = aN^{-2/3}$, and decompose $\mathbb{R}$ into $I_N = [-2 + \delta_N, 2 - \delta_N]$ along with $J_N = (2 - \delta_N, \infty)$ and $J_N^- = (-\infty, -2 + \delta_N)$, and write $g = f^l_c$. Then,

$$
\int g\rho_N - g\rho_{SC} = \int_{I_N} g(\rho_N - \rho_{SC}) + \int_{J_N \cup J_N^-} g\rho_N - \int_{J_N \cup J_N^-} g\rho_{SC}.
$$

First, we have by (63),

$$
\int_{J_N} |g|\rho_N \leq (\sup_{J_N} |g|) \int_{J_N} \rho_N \lesssim N^{2l/3 - 2/3} \int_{-\infty}^\infty \rho_N(2 + sN^{-2/3})ds \\
\lesssim N^{2l/3 - 1} \int_{-\infty}^\infty e^{-2s}ds \lesssim N^{2l/3 - 1}.
$$

Similar bounds hold for the integrals over $J_N$ and for those with respect to $\rho_{SC}$. For the middle interval, we use the Götze and Tikhomirov bound (61),

$$
\int_{I_N} |g(\rho_N - \rho_{SC})| \lesssim \frac{1}{N} \int_{2^{-\delta_N}}^{2^{-\delta_N}} \frac{|f^l_c(\lambda)|}{4 - \lambda^2}d\lambda.
$$

Observe that on $[-2 + \delta_N, 0]$, we have $0 \leq f^l_c(\lambda) \leq 1$. Therefore,

$$
\frac{1}{N} \int_{-2 + \delta_N}^{0} \frac{|f^l_c(\lambda)|}{4 - \lambda^2}d\lambda \leq \frac{1}{N} \int_{-2 + \delta_N}^{0} \frac{d\lambda}{2 + \lambda} = O(N^{-1} \log \delta_N) = o(N^{-1/3}).
$$

To deal with the remaining part of the integral, make the change of variable $\lambda = 2 - uN^{-2/3}$ and note that $f^l_c(\lambda) = (u + \sigma_N)^{-1}N^{2/3}$ except for $|u + \sigma_N| \leq c$, where it vanishes. Thus

$$
\int_{0}^{2^{-\delta_N}} \frac{|f^l_c(\lambda)|}{4 - \lambda^2}d\lambda \leq \int_{0}^{2^{-\delta_N}} \frac{|f^l_c(\lambda)|}{2 - \lambda}d\lambda
$$

$$
= N^{2l/3} \int_{u}^{2N^{2l/3}} \frac{1}{u + \sigma_N} \frac{d\lambda}{2 - \lambda} \leq C N^{2l/3},
$$

where, for example, we may take $C = C_{\gamma,c} = c^{-1} \int_{a}^{\gamma+1} u^{-1}d\gamma + \int_{\gamma+1}^{\infty} (u - \gamma)^{-1}u^{-1}d\gamma$.

So we have established the following approximation

$$
\int f^l_c(\lambda)\rho_N(\lambda)d\lambda = \int f^l_c(\lambda)\rho_{SC}(\lambda)d\lambda + O\left(N^{2l/3 - 1}\right),
$$
and it remains to analyze the integral with respect to the semi-circle density.

Let \( m(z) = \int (\lambda - z)^{-1} p_{SC}(\lambda) \, d\lambda = \left( -z + \sqrt{z^2 - 4} \right) / 2 \) denote the Stieltjes transform of \( p_{SC}(\lambda) \). When \( \sigma_N \geq c \), we simply have, since \( E = 2 + \epsilon_N = 2 + \sigma_N N^{-2/3} \),

\[
\int f_\epsilon^l(\lambda) p_{SC}(\lambda) \, d\lambda = \left. \frac{1}{E - \lambda} \right|_{-2}^{2} \, p_{SC}(\lambda) \, d\lambda = \begin{cases} 
-m(2 + \epsilon_N) = 1 + O(\epsilon_N^{1/2}) & l = 1 \\
-m'(2 + \epsilon_N) = O(\epsilon_N^{-1/2}) & l = 2.
\end{cases}
\]

When \( \sigma_N < c \), note that \( f_\epsilon(\lambda) = 0 \) for \( \lambda \in [\lambda_-, \lambda_+] \) with \( \lambda_\pm = 2 + N^{-2/3}(\sigma_N \pm c) \), and in particular \( \lambda_- < 2 \). The square-root decay of \( p_{SC} \) near 2 implies that

\[
\int_{\lambda_-}^{\lambda_+} f_\epsilon(\lambda) p_{SC}(\lambda) \, d\lambda \leq c^{-l} N^{2l/3} \int_{\lambda_-}^{\lambda_+} p_{SC}(\lambda) \, d\lambda \lesssim N^{2l/3 - 1}.
\]

Consider now \( l = 1 \). The change of variable \( \lambda = \lambda_- - N^{-2/3} x = E - N^{-2/3} (x + c) \) yields, along with \( p_{SC} \lesssim \sqrt{2 - \lambda} \),

\[
\int_{-2}^{\lambda_-} f_\epsilon(\lambda) - \frac{1}{2 - \lambda} p_{SC}(\lambda) \, d\lambda \lesssim |\sigma_N| N^{-2/3} \int_{-2}^{\lambda_-} \frac{d\lambda}{(E - \lambda) \sqrt{2 - \lambda}} 
\leq |\sigma_N| N^{-1/3} \int_{0}^{\infty} \frac{dx}{(x + c) \sqrt{x}} \lesssim N^{-1/3}.
\]

The same bound for \( p_{SC}(\lambda) \) also gives

\[
\int_{-2}^{\lambda_-} \frac{p_{SC}(\lambda)}{2 - \lambda} \, d\lambda = -m(2) - \int_{-2}^{2} \frac{p_{SC}(\lambda)}{2 - \lambda} \, d\lambda = 1 + O(N^{-1/3}).
\]

Combining the last three displays yields, for \(-\gamma \leq \sigma_N < c\),

\[
\int_{-2}^{2} f_\epsilon(\lambda) p_{SC}(\lambda) \, d\lambda = 1 + O_{\epsilon, \gamma}(N^{-1/3}).
\]

We turn to \( l = 2 \), still with \( \sigma_N < c \). Again setting \( \lambda = \lambda_- - N^{-2/3} x \), we have

\[
\int_{-2}^{\lambda_-} f_\epsilon^2(\lambda) p_{SC}(\lambda) \, d\lambda \lesssim \int_{-2}^{\lambda_-} (2 - \lambda + N^{-2/3} \sigma_N)^{-2} \sqrt{2 - \lambda} \, d\lambda 
\lesssim N^{1/3} \int_{0}^{\infty} (x + c)^{-2} (x + c - \sigma_N)^{1/2} \, dx \lesssim N^{1/3}.
\]

Together with (67) this shows that \( \int f_\epsilon^2(\lambda) p_{SC}(\lambda) \, d\lambda = O(N^{1/3}) \) and completes the proof.

This finishes our proof of Proposition 3 for the GUE case.

For the GOE case, the proposition follows from the following theorem. A proof of this theorem can be found in Section A.4.

**Theorem 19.** Let \( M_N^{\text{SC}} \) and \( M_N^{\text{RE}} \) be \( N \times N \) (unscaled) GUE and GOE matrices, respectively. Suppose that \( f_N \) is a series of functions such that

\[
f_N \left( M_N^{\text{SC}} \right) = a_N + O_P(b_N),
\]

for some sequences \( a_N \) and \( b_N \). Then,

\[
f_N \left( M_N^{\text{RE}} \right) = a_N + O_P(b_N + \text{TV}(f_N)),
\]

where \( \text{TV}(f_N) \) is the total variation of \( f_N \).

Indeed, the theorem and the fact that scaling of the argument does not change the total variation of a function yield the equivalents of Lemmas 16 and 18 for GOE. These equivalents, combined with the anticoncentration bound of Lemma 4 imply Proposition 3 for GOE.
5. Extension to Wigner matrices.

5.1. Lindeberg swapping formalism for asymptotically flat $Q$.

Definitions. A Wigner matrix is an Hermitian $N \times N$ matrix $W_N = (\xi_{ij} / \sqrt{N})$ satisfying
(i) the upper-triangular components $\{\text{Re}\, \xi_{ij}, \text{Im}\, \xi_{ij}\}_{i<j}$ and $\{\xi_{ii}\}$ are independent random variables with mean zero,
(ii) $E|\xi_{ij}|^2 = 1$ for $i \neq j$ and $E\xi_{ii}^2 \leq B$ for some absolute constant $B$;
(iii) a moment bound uniform in $N$: for all $p \in \mathbb{Z}_{>0}$, there is a constant $C_p$ such that

\begin{align*}
E|\text{Re}\, \xi_{ij}|^p, E|\text{Im}\, \xi_{ij}|^p \leq C_p.
\end{align*}

This definition is standard, e.g. [10, Def 2.2], except that we also require independence of $\text{Re}\, \xi_{ij}$ and $\text{Im}\, \xi_{ij}$ to simplify our swapping arguments. Condition (ii) allows for zero variances on the diagonal, as in the SSK model of [31].

The moments of two Wigner matrices $W_N, W'_N$ match to order $m$ if for integer $0 < a \leq m$

\begin{align*}
E(\text{Re}\, \xi_{ij})^a = E(\text{Re}\, \xi'_{ij})^a, \quad E(\text{Im}\, \xi_{ij})^a = E(\text{Im}\, \xi'_{ij})^a
\end{align*}

for all $1 \leq i < j \leq N$. Note that this constrains only the off-diagonal entries. The diagonal entries already match to order one by assumption, which is all that we need.

An event sequence $A_N$ holds with high probability if there exists a $d > 0$ such that

\begin{align*}
P(A_N^c) \lesssim N^{-d}.
\end{align*}

An event $B_N$ holds with overwhelming probability (w.o.p.) if, for all $A > 0$,

\begin{align*}
P(B_N^c) \lesssim N^{-A}.
\end{align*}

If $X_N \lesssim c_N$ w.o.p. and there are constants $C_0, C_2$ such that eventually $c_N \geq N^{-C_0}$ and $EX_N \lesssim N^{C_2}$, then $EX_N \lesssim c_N$. [For proof, see e.g. [10, Lemma 7.1].] Here and later “$X_N \lesssim c_N$ w.o.p.” means that there exists $C$ such that event $X_N \leq Cc_N$ holds w.o.p. Similarly for statements like $X_N = O(c_N)$ w.o.p.

In Proposition 20 and its consequence Proposition 21, we make the swapping argument explicit for abstract $Q$ satisfying generic asymptotic ‘flatness’ derivative bounds. In the next subsection we assemble tools – resolvent perturbation and local law – with the goal of establishing, in Proposition 24, the necessary flatness bounds for some specific choices of $Q$ needed for our later applications.

Fix $c_0 > 0$ and set $\|F\|_{c_0} = \sup\{|F(t)|, |t| \leq N^{c_0}\}$. Let $\delta_N \to 0$ in such a way that $\delta_N \gtrsim N^{-c_1}$ for some $c_1 > 0$. Let $Q$ be a function on $N \times N$ Hermitian/symmetric matrices taking values in $[0, 1]$. Let Wigner matrices $W_N, W'_N$ be given and define $Q_{\gamma}(t) = Q(W(t))$ as in Section 1.2. We say that $Q$ satisfies condition F or $F(\delta_N)$ if for all $\gamma$ and $1 \leq k \leq 4$ we have w.o.p. that

\begin{align*}
\|Q_{\gamma}^{(k)}\|_{c_0} \lesssim N^{-\frac{k}{2}} \delta_N.
\end{align*}

PROPOSITION 20. Let $W_N, W'_N$ be Wigner matrices whose moments match to third order. Let $c_0, c_1 > 0$ be fixed and for each $j = 1, \ldots, m$, let $Q_j : \mathbb{C}^{N \times N} \to [0, 1]$ satisfy condition $F(\delta_{j,N})$. If $Q = \prod_{j=1}^{m} Q_j$, then,

\begin{align*}
EQ(W_N) - EQ(W'_N) \lesssim \max_{j=1,\ldots,m} \delta_{j,N}.
\end{align*}
PROOF. Consider first the case \( m = 1 \). We set \( \Delta_{\gamma i} = Q(W^{(i)}) - Q(W_0) \), and decompose
\[
E Q(W_N) - E Q(W'_N) = \sum_{\gamma} E (\Delta_{\gamma 0} - \Delta_{\gamma 1}).
\]

Let \( E_N = E_N(W_0') \) denote the overwhelming probability event \((F)\) and then introduce ‘good’
events \( G_N = E_N \cap \{ |\xi^{(i)}| \leq N^{c_0} \} \). Let \( A \) be a fixed constant such that \( N^{2-A} \lesssim \delta_N \). Using
boundedness of \( Q \) and the moment bound \((68)\), with \( p \) chosen so that \( p c_0 > A \), we have
\[
(70) \quad E (\Delta_{\gamma 0} - \Delta_{\gamma 1}) = E (\Delta_{\gamma 0} 1(G_N)) - E (\Delta_{\gamma 1} 1(G_{N1})) + O(N^{-A}).
\]

As before, set \( Q_{\gamma}(t) = Q(W^{(i)}_t) \), so that \( \Delta_{\gamma i} = Q_{\gamma}(\xi^{(i)}) - Q_{\gamma}(0) \). By Taylor expansion,
\[
\Delta_{\gamma i} = \sum_{j=1}^{k-1} \frac{1}{j!} Q^{(j)}(0)(\xi^{(i)})^j + \frac{1}{k!} Q^{(k)}(\xi^{(i)})^k,
\]
for some \( \xi^* \) with \( |\xi^*| \leq |\xi^{(i)}| \). Both \( Q_{\gamma}(t) \) and event \( E_N \) are independent of \( \xi^{(i)} \), so
\[
E (Q^{(j)}(0)(\xi^{(i)})^j 1(G_{Ni})) = E (Q^{(j)}(0) 1(E_N)) E ([\xi^{(i)}]^j | |\xi^{(i)}| \leq N^{c_0}] [\xi^{(i)}]^j | (\xi^{(i)})^k] + O(N^{-j/2} \delta_N \cdot N^{-A}).
\]

where we used the fact that \( E [|\xi^{(i)}|^j 1(|\xi^{(i)}| > N^{c_0})] \leq C_p N^{-c_0} (p-j) = O(N^{-A}) \) for suitable
\( p \), as follows from the Markov inequality and \((68)\). For the remainder, on event \( G_{Ni} \) we also
have \( |Q^{(k)}(\xi^{(i)})| \leq \|Q^{(k)}\|_{c_0} \lesssim N^{-k/2} \delta_N \), and hence
\[
|E (Q^{(k)}(\xi^{*})(\xi^{(i)})^k 1(G_{Ni}))| \lesssim N^{-k/2} \delta_N.
\]

Summarizing, we have
\[
E (\Delta_{\gamma 0} 1(G_{Ni})) = \sum_{j=1}^{k-1} \frac{1}{j!} E (Q^{(j)}(0) 1(E_N)) E ([\xi^{(i)}]^j) + O(N^{-k/2} \delta_N + N^{-A}).
\]

Choose \( k = k(\gamma) \) so that \( E (\xi^{(i)})^j = E (\xi^{(0)})^j \) for \( 1 \leq j \leq k - 1 \). Then the sums cancel and
\( (70) \) yields
\[
E (\Delta_{\gamma 0} - \Delta_{\gamma 1}) = O(N^{-k/2} \delta_N + N^{-A}).
\]

For the \( O(N^2) \) off-diagonal terms, moment matching to third order allows \( k(\gamma) = 4 \), while
for the \( N \) diagonal terms, we take \( k(\gamma) = 2 \), since then only \( E \xi^{(i)} = 0 \). Summing over all \( \gamma \), we obtain
\[
(71) \quad E Q(W_N) - E Q(W'_N) = O(\delta_N + N^{2-A}) = O(\delta_N)
\]
from the choice of \( A \).

For \( m > 1 \), apply the product rule, use \((F)\) and \( \|Q_{j,\gamma}\|_{c_0} \leq 1 \):
\[
\|Q^{(k)}\|_{c_0} \lesssim \sum_{\ell_1 + \cdots + \ell_m = k} \binom{k}{\ell_1, \ldots, \ell_m} \prod_{1 \leq j \leq m} N^{-\frac{\ell_j}{2}} \delta_j, N \lesssim N^{-\frac{k}{2}} \max_{j=1, \ldots, m} \delta_j, N,
\]
Thus \( Q \) satisfies \( F(\max_j \delta_j, N) \) and the result follows from \((71)\). \( \square \)

We use Proposition \( 20 \) to formulate a criterion that allows joint convergence in distribution
of vector functions of \( W_N \) to be transferred to the corresponding functions of \( W'_N \). A proof
of the following proposition is in appendix B.1.
Proposition 21. Let $W_N, W'_N$ be Wigner matrices whose moments match up to third order. Let $\xi_N = \xi_N(W_N)$ and $\xi'_N = \xi_N(W'_N)$ both be $\mathbb{R}^m$ valued random vectors. Suppose that $\xi_N \xrightarrow{d} \xi$, and that each component $\xi_j$ of the limit has a continuous distribution function.

Let $\eta_N \to 0$ be given, and suppose that for each $1 \leq j \leq m$ and $s \in \mathbb{R}$ there exists a function $Q_j(\cdot, s)$ satisfying condition $F(\delta_{j,N})$ such that for $W = W_N, W'_N$, w.o.p.

$$1\{\xi_{Nj}(W) \leq s - \eta_N\} \leq Q_j(W, s) \leq 1\{\xi_{Nj}(W) \leq s + \eta_N\}. \tag{72}$$

Then we also have (joint) convergence $\xi'_N \xrightarrow{d} \xi$.

5.2. Flatness for Stieltjes functionals.

5.2.1. Resolvent perturbation: deterministic bounds. We recall and modify some bounds of [46] on stability of Hermitian matrices with respect to perturbation in one or two entries, using Ward’s identity to improve the bounds at the edge.

Let $M_0$ be a Hermitian $N \times N$ matrix, $z = E + i\eta \in \mathbb{C}_+$ and $V$ an elementary matrix as defined after (9). Set $M_t = M_0 + tN^{-1/2}V$ and $R_t = R_t(z) = (M_t - z)^{-1}$, and $s_t(z) = N^{-1} \text{tr} R_t(z)$. Recall from [46] the definitions of the matrix norms $\|A\|_{(q,p)}$, and in particular

$$\|A\|_{(\infty,1)} = \max_{1 \leq i,j \leq N} |A_{ij}|, \quad \|A\|_{(\infty,2)} = \max_{i} \left( \sum_j |A_{ij}|^2 \right)^{1/2}.$$ 

Note also that if $V$ is an elementary matrix, then

$$|\text{tr}(AV)| = |\text{tr}(VA)| \leq 2\|A\|_{(\infty,1)} \tag{73}$$

$$\|AVB\|_{(\infty,1)} \leq 2\|A\|_{(\infty,1)} \|V\|_{(\infty,1)} \|B\|_{(\infty,1)}. \tag{74}$$

Let $\kappa_N(z, t) = |t|N^{-1/2}\|R_t\|_{(\infty,1)}$. Lemma 12 of [46] says that if $\kappa_N(z, t) \to 0$ as $N \to \infty$, then for large $N$

$$R_{t+u} = R_t + \sum_{j=1}^{\infty} \left( \frac{-u}{\sqrt{N}} \right)^j (R_tV)^j R_t, \tag{75}$$

with the right side being absolutely convergent. In addition, for $1 \leq p \leq \infty$,

$$\|R_t\|_{(\infty,p)} \leq \|R_0\|_{(\infty,p)} \exp\{2|t|N^{-1/2}\|R_0\|_{(\infty,1)}\}. \tag{76}$$

Here the factor 2 arises from the use of $\|V\|_{(1,\infty)} \leq 2$ in the proof of Lemma 12 of [46]. The same bound holds with the roles of $R_0$ and $R_t$ reversed.

Expansion (75) allows evaluation of $t$-derivatives of $s_t(z)$. Indeed

$$\partial_t^j s_t(z) = j!N^{-j/2}c_j(z, t) \tag{77}$$

$$c_j(z, t) = (-1)^j N^{-1} \text{tr}((R_tV)^j R_t). \tag{78}$$

The following variant of [46, Proposition 13] yields uniform bounds on $c_j$ in terms of $\|\text{Im} R\|_{\infty} = \max_{1 \leq i,j \leq N} |\text{Im} R_{i,j}|$, which allows tighter bounds near the edge.

Proposition 22. Let $c_0$ and $A$ be small and positive. and define

$$S_{c}(A) = \{ z = E + i\eta \in \mathbb{C} : |E - 2| \leq N^{-2/3} + A, \eta > N^{-2/3} - A \}$$

$$\kappa_N = \sup_{|t| \leq N^\alpha, z \in S_{c}(A)} |t|\|R_0\|_{(\infty,1)}/\sqrt{N}. \tag{79}$$

Then for $z \in S_{c}(A)$ and $|t| \leq N^{c_0}$,

$$|c_j(z, t)| \leq (N\eta)^{-1/2} e^{2(j+1)\kappa_N} \|R_0\|_{(\infty,1)}^{j-1} \|\text{Im} R_0\|_{\infty}. \tag{80}$$
PROOF. From the cyclic property of traces, then (73) and (74), we have
\[ |\text{tr}((R_tV)^j R_t)| = |\text{tr}(V(R_tV)^{j-1}R_t^2)| \leq 2\|(R_tV)^{j-1}R_t^2\|_{(\infty,1)} \leq 2\|R_t\|_{(\infty,1)}^{-1}\|R_t^2\|_{(\infty,1)}. \]
The Ward identity, e.g. [10, eq. (3.6)] says that
\[ \sum_j |R_{ij}|^2 = \eta^{-1} \text{Im } R_{ii} \]
is valid for any resolvent matrix \( R = (W - E - i\eta)^{-1} \) with \( \eta \neq 0 \) and Hermitian (or symmetric) \( W \). For \( \eta > 0 \), we have
\[ \| R_t \|_{(\infty,2)}^2 = \max_i \sum_j |R_{ij}|^2 = \eta^{-1} \| \text{Im } R \|_{\infty}. \]

If \( B \) is a normal matrix, (i.e. \( B^* B = BB^* \)), then
\[ \| AB \|_{(\infty,1)} \leq \| A \|_{(\infty,2)} \| B \|_{(\infty,2)}. \]
This uses the Cauchy-Schwarz bound \( |(AB)_{ij}|^2 \leq \sum_k |A_{ik}|^2 \sum_k |B_{kj}|^2 \), since \( B \) normal implies \( \sum_k |B_{kj}|^2 = \sum_k |B_{jk}|^2 \leq \| B \|^2_{(\infty,2)}. \)
The resolvent of a Hermitian matrix is normal, so from (82), (76), and then (81) we have
\[ \| R_t^2 \|_{(\infty,1)} \leq \| R_t \|_{(\infty,2)}^2 \leq e^{2\kappa_N} \| R_0 \|_{(\infty,2)}^2 = \eta^{-1} e^{2\kappa_N} \| \text{Im } R_0 \|_{\infty}. \]
Combine the last display with the first of the proof and then refer to (76) to bound \( \| R_t \|_{(\infty,1)} \leq e^{2\kappa_N} \| R_0 \|_{(\infty,1)} \) to arrive at (80). \( \square \)

5.2.2. Local law. We will need the local law for Wigner matrices and some of its important consequences, in particular at the spectral edge.

PROPOSITION 23. Let \( W_N \) be a Wigner matrix.
(i) (local law) Let \( R(z) = (W_N - zI)^{-1} \) denote the resolvent matrix and \( s_{sc}(z) \) the Stieltjes transform of the semicircle law. Fix \( \tau > 0 \) small. For each \( \epsilon > 0 \), we have w.o.p.\n\[ R_{ij} = s_{sc}(z) \delta_{ij} + O(N^\epsilon \Psi(z)), \]
uniformly for \( z \in \mathcal{S}(\tau) = \{ E + i\eta : |E| < \tau^{-1}, N^{-\frac{1+\tau}{4}} \leq \eta \leq \tau^{-1} \} \) and \( i, j = 1, \ldots, N \), where
\[ \Psi(z) = \sqrt{\frac{|\text{Im } s_{sc}(z)|}{N\eta}} + \frac{1}{N\eta}. \]
(ii) (semi-circle law on small scales) For each \( \epsilon > 0 \), we have w.o.p. that
\[ N_{W_N}(I) = N \int_I \rho_{sc}(dx) + O(N^\epsilon), \]
uniformly for all intervals \( I \subset \mathbb{R} \), where \( N_{W_N}(I) \) denotes the number of eigenvalues of \( W_N \) in \( I \) and \( \rho_{sc} \) denotes the semi-circle law.
(iii) (at the edge.) Let \( A > 0 \) be small and fixed, and let \( \mathcal{S}_e(A) \) be the edge domain (79). For each \( \epsilon > 0 \) and uniformly for \( z = E + i\eta \in \mathcal{S}_e(A) \), we have w.o.p.
\[ \| R \|_{(\infty,1)} \lesssim 1 \wedge \eta^{-1}, \quad \| \text{Im } R \|_{\infty} \lesssim (\eta^{1/2} + N^{-1/3+\epsilon+A}) \wedge \eta^{-1}. \]
Let \( W_0 = W - \xi N^{-1/2} V \) with \( V \) an elementary matrix and \( \xi \) satisfying moment bounds (68). Set \( R_0 = (W_0 - zI)^{-1}. \) Then the bounds (83) apply to \( R_0 \) also.
Remark. For clarity, we emphasize that these are simultaneous high probability bounds for all $z$ in the indicated ranges. For example, then w.o.p.
\[
\sup_{z \in \mathbb{S}(A)} (1 \vee \text{Im } z) \| R(z) \|_{(\infty,1)} \lesssim 1.
\]
Such statements follow from the $N^2$-Lipschitz continuity of $R_{ij}(z)$, $s_{sc}(z)$ and of the right side bounds over the indicated ranges, c.f. e.g. [10, Remark 2.7].

Proof. For (i) and (ii), see e.g. [10, Theorems 2.6, 2.8]. We turn to (iii). Basic bounds on $s_{sc}(z)$, e.g. [17, Lemma 6.2], establish for $\eta > 0$, $|E| \leq 10$ and $\kappa = |E| - 2$ that
\[
|s_{sc}(z)| \leq 1, \quad \text{Im } s_{sc}(z) \lesssim \sqrt{\kappa + \eta}.
\]
For $N^{-2/3-A} \leq \eta \leq 1$, we have $\Psi(z) \lesssim (N\eta)^{-1/2} \leq N^{-1/6+A/2}$ and so from the local law $\| R \|_{(\infty,1)} \lesssim 1$. For $\eta \geq 1$, just use the elementary bound $|R_{jk}| \leq \eta^{-1}$ arising from the spectral decomposition
\[
R_{jk}(E + i\eta) = \sum_{l=1}^{N} u_l(j) u_l^*(k) / (\lambda_l - E - i\eta),
\]
where $u_l(j)$ denotes the $j$-th component of the eigenvector $u_l$ corresponding to $\lambda_l(W_N)$.

For $(\text{Im } R)_{jj}$ we exploit the improved bounds on $\text{Im } s_{sc}$ at the edge. Since $\kappa \leq N^{-2/3+A}$, we have $\text{Im } s_{sc} \lesssim N^{-1/3+A/2} + \eta^{1/2}$ and $N\eta \geq N^{1/3-A}$, and conclude
\[
\Psi(z) \lesssim \frac{N^{-1/6+A/4}}{\sqrt{N\eta}} + \frac{1}{\sqrt{N\eta^{1/2}}} + \frac{1}{N\eta} \lesssim N^{-1/3+A}.
\]
Hence, the second part of (83) follows from the local law.

Turning to $R_0$, we put $\Delta = W - W_0 = N^{-1/2}\xi V$ and use the resolvent identity $R_0 = R + R\Delta R + R_0(\Delta R)^2$. Write $\| \cdot \|_*$ for $\| \cdot \|_{(\infty,1)}$. Even for $R_0$, the bound $\| R_0 \|_* \leq \eta^{-1}$ follows from (84) as before. So to conclude the rest of (83) for $R_0$, it suffices to show that w.o.p. $\| R_0 - R \|_* \lesssim N^{-1/3+\epsilon+A}$ for $N^{-2/3-A} \leq \eta \leq 1$.

We have the trivial bound $\| R_0 \|_* \lesssim \eta^{-1} \leq N^{2/3+A}$. Since (68) implies that $|\xi| \leq N^{\epsilon/2}$ w.o.p., we have that $\| \Delta \|_* \lesssim N^{-1/2+\epsilon/2}$, and along with $\| R \|_* \lesssim 1$, and bound (74) for elementary matrices, we find that w.o.p. both
\[
\| R\Delta R \|_* \lesssim N^{-1/2+\epsilon/2}, \quad \| R_0(\Delta R)^2 \|_* \lesssim N^{2/3+A-1+\epsilon} \lesssim N^{-1/3+\epsilon+A}.
\]

5.2.3. Stieltjes functionals. We return to establishing flatness condition (F) for certain functionals $Q = \tilde{G} \circ g$. Let $W$ be an Hermitian matrix and $s_{W}(z)$ its empirical Stieltjes transform. In the following proposition, we consider examples of *Stieltjes functionals* $g(W) = \Lambda(s_W)$ for some continuous linear functional $\Lambda$ acting on functions holomorphic on $\mathbb{C}^+$. The first two of these examples will be used in the next subsection to extend the non-concentration property for the eigenvalues of $G(U/O)E$ matrices (Lemma 17) to Wigner matrices and, using this, to extend the log determinant CLT to Wigner matrices. The last two examples are key to the analysis of the SSK model in the companion paper [27]. There, we need to extend results on the $k$-th largest eigenvalue and the trace of the inverse powers of $z - W_N$ from $G(U/O)E$ to Wigner matrices.

**Proposition 24.** Let $W$ be a Wigner matrix. Let $\epsilon > 0$ and $0 < c_0 < \frac{1}{2}$ and let $E \in \mathbb{R}$ be such that $|E - 2| \lesssim N^{-2/3+A}$.

For each of the following statistics, define functions $g: \mathbb{C}^{N \times N} \to \mathbb{R}$, $G: \mathbb{R} \to \mathbb{R}$ and a sequence $\delta_N$ according to the following specifications, in each case for $1 \leq j \leq 4$:
1. **Log-determinant:** with $\gamma_N = N^{-2/3-\epsilon}$,

\[ g(W) = N \int_{\gamma_N}^{N^{100}} \text{Im} s_W(E + i\eta) \, d\eta, \quad \|G^{(j)}\|_\infty \leq (\log N)^{-j/4}, \quad \delta_N = (\log N)^{-1/4}. \]

2. **Eigenvalue counting:** with $\eta = N^{-2/3-9\epsilon}$ and $E, i = 1, 2$ such that $|E_i - 2| \leq N^{-2/3+10\epsilon}$,

\[ g(W) = \frac{N}{\pi} \int_{E_1}^{E_2} \text{Im} s_W(x + i\eta) \, dx, \quad \|G^{(j)}\|_\infty \leq (\log N)^{Cj}, \quad \delta_N = N^{-\frac{1}{4} + O(\epsilon)}. \]

3. **Inverse moments:** with $\eta = N^{-2/3-\epsilon}$ and $l \in \mathbb{Z}_+$,

\[ g(W) = N^{-\frac{2}{3}l+1} \text{Re} s_W^{(l-1)}(E + i\eta), \quad \|G^{(j)}\|_\infty \leq (\log N)^{Cj}, \quad \delta_N = N^{-\frac{1}{4} + O(\epsilon)}. \]

In each of the cases listed above, the corresponding function $Q = G \circ g$ satisfies the condition of eq. (F). That is, for $1 \leq k \leq 4$, it follows w.o.p. that

\[ \|Q^{(k)}\|_{c_0} \lesssim N^{-\frac{k}{2}} \delta_N. \]

**Proof.** Define $g^\gamma(t) = g(W_t^\gamma)$ so that $Q_t^\gamma(t) = G(g_\gamma(t))$. In order to bound $Q^{(k)}_t(t)$ we start with bounds for $\partial_t^k g^\gamma$. Recalling (10), we have $g^\gamma(t) = \Lambda(s^\gamma_t)$. Standard results on differentiation of integrals and then (77) imply that

\[ \partial_t^k g^\gamma(t) = \Lambda(\partial_t^k s^\gamma_t) = j! N^{-j/2} \Lambda(c^\gamma_j(\cdot, t)), \]

where from (78) and (10))

\[ c^\gamma_j(z, t) = (-1)^j N^{-1} \text{tr}((R_t^3 V_t)^j R_t^3). \]

Hence, to bound $\|\partial_t^k g^\gamma\|$, it suffices to use bounds on the coefficients $c^\gamma_j$. We will omit the superscript $\gamma$ to simplify notations. From Propositions 22 and 23, for fixed $A > 0$,

\begin{equation}
N|c_j(E + i\eta, t)| \lesssim \begin{cases} 
\eta^{-1/2} + N^{-1/3+c\epsilon} \eta^{-1} & N^{-2/3-A} \leq \eta \leq 1 \\
\eta^{-j-1} & \eta \geq 1,
\end{cases}
\end{equation}

uniformly in $|t| \leq N^{ca}$. Note that there is no dependence on $j$ for $\eta \leq 1$.

In the log-determinant case, we have that

\[ \Lambda(f) = \int_{\gamma_N}^{N^{100}} N \text{Im} f(y + i\eta) \, d\eta. \]

Evaluated at $c_j(\cdot, t)$, we use (85) with $A = 2\epsilon$ to obtain

\[ \int_{\gamma_N}^{N^{100}} N |c_j(E + i\eta, t)| \, d\eta \lesssim \int_{\gamma_N}^{1} (\eta^{-1/2} + N^{-1/3+c\epsilon} \eta^{-1}) \, d\eta + \int_{1}^{N^{100}} \eta^{-j-1} \, d\eta \lesssim 1. \]

For the remaining integrals, we need only the following consequence of of eq. (85).

\begin{equation}
N|c_j(E + i\eta, t)| \lesssim N^{\frac{j}{4} + c\epsilon + 2A}.
\end{equation}

Set $A = 10\epsilon$ for the eigenvalue counting case. This yields

\[ \int_{E_1}^{E_2} \frac{N}{\pi} |c_j(y + i\eta, t)| \, dy \lesssim N^{\frac{j}{4} + c\epsilon + 2A} N^{-\frac{3}{2} + \epsilon} = N^{-\frac{1}{4} + O(\epsilon)}. \]
For inverse moments, we have \( \Lambda(c_j(t)) = N^{-2l/3+1}\Re c_j^{(l-1)}(E + i\eta) \). Let \( \Gamma \) be a contour of radius \( N^{-\frac{4}{d}-2\epsilon} \) around \( E + i\eta \). In this way, each \( c_j \) is analytic on the interior of \( \Gamma \), and so we use Cauchy’s integral formula and (86) with \( A = 2\epsilon \) to see that

\[
N^{-\frac{4}{d}+1}|c_j^{(l-1)}(E + i\eta)| \leq \frac{(l-1)!}{2\pi i} \int_{\Gamma} \frac{|c_j(w)|}{N^{\frac{2}{d}|w - E - i\eta|}}|dw| \lesssim N^{-\frac{4}{d}+O(\epsilon)}.
\]

In sum suppose that, for sequences \( a_N \) and \( b_N \) such that \( a_N b_N \to 0 \), we have w.o.p.

\[
\|\partial_t^j g^\gamma(t)\|_{c_0} \lesssim N^{-\frac{2}{d}} a_N, \quad \|G^{(j)}\|_{\infty} \lesssim b_N^j.
\]

In the proof so far, we have seen that the above conditions hold with the following values of \( a_N \) and \( b_N \) for some constant \( C \):

1. Log-determinant: \( a_N = 1, \quad b_N = (\log N)^{-1/4} \).
2. Eigenvalue counting: \( a_N = N^{-\frac{4}{d}+O(\epsilon)}, \quad b_N = (\log N)^C \).
3. Inverse moments: \( a_N = N^{-\frac{4}{d}+O(\epsilon)}, \quad b_N = (\log N)^C \).

We apply Faà di Bruno’s formula to compute bounds for \( \partial_t^k (G \circ g^\gamma)(t) \). Let \( \mathcal{M}_k = \{ m \in \mathbb{Z}_{\geq 0}^k : \sum_{j=1}^k j m_j = k \} \), so that \( m_+ = m_1 + \cdots + m_k \geq 1 \) for each \( m \in \mathcal{M}_k \). Then for certain combinatorial constants \( C_{km} \) we have that, uniformly in \( |t| \leq N^{c_0} \),

\[
|\partial_t^k (G \circ g^\gamma)(t)| \lesssim \sum_{m \in \mathcal{M}_k} C_{km}|G^{(m_+)n}(g^\gamma(t))| \cdot \prod_{j=1}^k |g^{(j)}(t)|^{m_j}
\]

\[
\lesssim \sum_{m \in \mathcal{M}_k} C_{km} b_N^m \prod_{j=1}^k N^{-\frac{j}{d}} a_N^m = N^{-\frac{k}{d}} \sum_{m \in \mathcal{M}_k} C_{km} (a_N b_N)^m \lesssim N^{-\frac{k}{d}} a_N b_N,
\]

Hence, the conclusion in each case follows with \( \delta_N = a_N b_N \). \( \square \)

5.3. Concluding the extension to Wigner matrices.

5.3.1. Wigner Non-concentration.

**Proposition 25.** Let \( W_N \) be a Wigner matrix whose off-diagonal moments match GOE or GUE to third order. Call its eigenvalues \( \lambda_1, \ldots, \lambda_N \). Let \( E \in \mathbb{R} \) be such that \( |E - 2| \lesssim N^{-\frac{2}{d}} \sigma_N \), with \( \sigma_N = (\log N)^{O(\log \log N)} \). Then there exists a \( c_1 \) such that, for each \( c_0 \in (0, c_1) \), there exists \( d > 0 \) such that, for \( N \) large,

\[
P(\min_{j=1, \ldots, N} |\lambda_j' - E| \leq N^{-\frac{2}{d} - c_0}) \leq N^{-d}.
\]

**Proof.** Define the eigenvalue counting function \( N_W(E_1, E_2) = \# \{ j : E_1 \leq \lambda_j(W) \leq E_2 \} \). The event in (87) has the form \( N_W(E_1, E_2) \geq 1 \). The first step is to approximate this using the Stieltjes transform.

Let \( \epsilon = 2c_0 \) and define \( \ell = \frac{1}{2} N^{-\frac{2}{d} - \epsilon} \), and \( \eta = N^{-\frac{2}{d} - 9\epsilon} \). Let \( E_1, E_2 \in \mathbb{R} \) be such that \( |E_1 - 2|, |E_2 - 2| \lesssim N^{-\frac{2}{d}} \sigma_N \) and \( E_2 - E_1 \geq 2\ell \).

A suitable approximation is given by Corollary 17.3 of [17] (based on the local law and eigenvalue rigidity\(^1\)), which we apply twice with \( E = E_1 \) and \( E_2 \) respectively. Subtracting

\(^1\)The definition of Wigner matrices in [17] is slightly different from ours, so its proof needs minor adjustments to take the difference into account. Specifically, one needs to use suitable versions of the local law and eigenvalue rigidity (theorems 2.6 and 2.9 from [10]). We refer the interested reader to section A4 of [27] for details.
the latter bounds from the former, this yields w.o.p. that
\[
\frac{N}{\pi} \int_{E_1 + \ell}^{E_1 - \ell} \text{Im} \ s_W(y + i\eta) \, dy - 2N^{-\epsilon} \leq \mathcal{N}_W(E_1, E_2) \leq \frac{N}{\pi} \int_{E_1 - \ell}^{E_1 + \ell} \text{Im} \ s_W(y + i\eta) \, dy + 2N^{-\epsilon}.
\]

Let \( E^\pm = E \pm 2N^{-\frac{2}{3} - \epsilon_0} \), and define the function
\[
g(W) = \frac{N}{\pi} \int_{E^- + \ell}^{E^- - \ell} \text{Im} \ s_W(y + i\eta) \, dy,
\]
Applying these bounds with \((E_1, E_2) = (E^-, E^+)\) and \((E^- + 2\ell, E^+ - 2\ell)\), we conclude that, w.o.p.,
\[
\mathcal{N}_W(E^- + 2\ell, E^+ - 2\ell) - 2N^{-\epsilon} \leq g(W) \leq \mathcal{N}_W(E^-, E^+) + 2N^{-\epsilon}.
\]

Let \( G \) be a smooth increasing function such that
\[
G(x) = \begin{cases} 1 & \text{if } x \geq 2/3, \\ 0 & \text{if } x \leq 1/3. \end{cases}
\]
Taking \( Q = G \circ g \) and applying \( G \) to each side of eq. (88), we then have that, w.o.p.,
\[
1 \{ \mathcal{N}_W(E^- + 2\ell, E^+ - 2\ell) \geq 1 \} \leq Q(W) \leq 1 \{ \mathcal{N}_W(E^-, E^+) \geq 1 \}.
\]

Now we can use Propositions 20 and 24(2) to compare \( Q(W_N) \) with \( Q(W_N) \), for \( W_N \) drawn from \( \text{GOE} \) with eigenvalues \( \lambda_j \). For any \( A > 0 \), we have
\[
P(\min_j |\lambda_j - E| \leq 2N^{-\frac{4}{3} - \epsilon_0} - 2\ell) = P\{ \mathcal{N}_{W_N}(E^- + 2\ell, E^+ - 2\ell) \geq 1 \}
\leq E Q(W_N) + O(N^{-A})
\leq E Q(W_N) + O(N^{-\frac{4}{3} + O(\epsilon)})
\leq P(\min_j |\lambda_j - E| \leq 2N^{-\frac{4}{3} - \epsilon_0}) + O(N^{-\frac{4}{3} + O(\epsilon)})
\leq \frac{1}{2} N^{-d} + O(N^{-1/3 + O(\epsilon)}) \leq N^{-d}
\]
At the last line we applied the non-concentration bound for \( \text{GOE} \), Section 4.2.2.

For \( N \) large, we have \( 2N^{-\frac{2}{3} - \epsilon_0} - 2\ell \geq N^{-2/3 - \epsilon_0} \) and so the final bound (87) follows from these inequalities.

The next lemma shows that non-concentration implies control of inverse power sums at around their typical magnitude. The proof is by standard dyadic decomposition (see appendix B.2).

**Lemma 26.** Let \( \{ \lambda_j \} \) be the eigenvalues of a Wigner matrix \( W_N \) whose off-diagonal moments match GOE or GUE to third order. Suppose that \( |E - 2| \leq N^{-2/3} \sigma_N \). Then there exist constants \( \{ C_r \} \) such that for each \( \epsilon > 0 \) small, with high probability we have
\[
S_r(E) := \sum_{j=1}^{N} \frac{1}{|\lambda_j - E|^r} \leq \begin{cases} C_1 N & \text{if } r = 1 \\ C_r N^{2r/3 + (r+1)\epsilon} & \text{if } r \geq 2 \end{cases}
\]
The bounds also hold for \( S_r(E') \) uniformly in \( |E' - E| \leq \delta/2 \) with \( \delta = N^{-2/3 - \epsilon} \), by increasing \( C_r \) to \( 2^r C_r \).
5.3.2. Log-deteminant. We derive the central limit theorem for the log-deteminant
\[ \tilde{L}_N(W_N) = \tau_N^{-1}(L_N(W_N) - \mu_N) \xrightarrow{d} \mathcal{N}(0, 1). \]

for a Wigner matrix $W_N'$ and $E = E_N = 2 + \sigma_N N^{-2/3}$. Recall the scaling constants $\mu_N, \tau_N$ from (3). Let $W_N$ be drawn from (scaled) GOE or GUE. From Theorem 1, which we have already established for the Gaussian ensembles, we have

\[ \tilde{L}_N(W_N) = \tau_N^{-1}(L_N(W_N) - \mu_N) \xrightarrow{d} \mathcal{N}(0, 1). \]

**Proposition 27 (Log determinant CLT).** Let $W_N'$ be a Wigner matrix whose off-diagonal moments match GOE or GUE to third order. Let $E = E_N = 2 + \sigma_N N^{-2/3}$ with

\[-\gamma \leq \sigma_N \ll \log^2 N.\]

Then

\[ \tau_N^{-1}(\log|\det(W_N' - E)| - \mu_N) \xrightarrow{d} \mathcal{N}(0, 1). \]

**Proof.** To rewrite the log-deteminant in terms of an integral of the Stieltjes transform, note that $\frac{d}{d\eta} \log|\lambda - E - i\eta| = \text{Im}[(\lambda - E - i\eta)^{-1}]$, which yields [46, eq. (46)]

\[ L_N(W) = \log|\det(W - E - iN^{100})| - N \int_0^{N^{100}} \text{Im} s_W(E + i\eta) \, d\eta. \]

The uniform moment bounds (68) imply that

\[ \log|\det(W - E - iN^{100})| = N \log(N^{100}) + O_P(N^{-50}). \]

Moreover, for each $\epsilon > 0$ small, if we take $\gamma_N = N^{-\frac{2}{3}-2\epsilon}$, then non-concentration implies that the contribution to the integral from $\eta \leq \gamma_N$ is negligible. Indeed, with $\lambda_j = \lambda_j(W_N)$,

\[ N \text{Im} s_W(E + i\eta) = \eta \sum_{j=1}^{N} \frac{1}{(\lambda_j - E)^2 + \eta^2} \leq \eta \sum_{j=1}^{N} \frac{1}{(\lambda_j - E)^2}. \]

By lemma 26 we thus have

\[ \left| N \int_0^{\gamma_N} \text{Im} s_W(E + i\eta) \, d\eta \right| \leq \frac{1}{2} \gamma_N^2 S_2(E) = O_P(\gamma_N^2 N^{\frac{4}{3}+3\epsilon}) = o_P(1). \]

To summarize, if we define the Stieltjes functional

\[ g(W) = N \int_{\gamma_N}^{N^{100}} \text{Im} s_W(E + i\eta) \, d\eta, \]

set $\bar{\mu}_N = \mu_N + N \log(N^{100})$ and define

\[ \xi_N(W) = \tau_N^{-1}(g(W_N) - \bar{\mu}_N), \]

then we have shown that $\tilde{L}_N(W_N) = \xi_N(W_N) + o_P(1)$.

We carry out the Lindeberg swapping with $g(W)$. Let $H : \mathbb{R} \rightarrow [0, 1]$ be a smooth decreasing function such that

\[ H(x) = \begin{cases} 
1 & \text{if } x \leq -\eta_N \\
0 & \text{if } x \geq \eta_N.
\end{cases} \]

For $s \in \mathbb{R}$ define $G_s(x) = H(\tau_N^{-1}(x - \bar{\mu}_N) - s)$. One verifies that

\[ 1\{\xi_N(W) \leq s - \eta_N\} \leq G_s(g(W)) \leq 1\{\xi_N(W) \leq s + \eta_N\}. \]
Setting $Q(W, s) = G_s(g(W))$, we obtain bound (72).

Observe that $\|G_s^{(j)}\|_\infty \lesssim (\tau_N \eta_N)^{-j} \lesssim (\log N)^{-j/4}$ if we choose $\eta_N = \tau_N^{-1/2}$. Then Proposition 24 (1) implies that $Q(\cdot, s)$ satisfy condition F with $\delta_N = (\log N)^{-1/4}$. From Proposition 21 we conclude that $\xi_N(W_N')$ and hence $\tilde{L}(W_N')$ have the same limiting distribution as $\xi_N(W_N)$ and $\tilde{L}_N(W_N)$. Thus the validity of Theorem 1 for Gaussian ensembles implies eq. (90). □

6. The spiked case. In this section, we consider deformed Wigner matrices

$$W_{h,N} = W_N + h\nu\nu^*,$$

where $\nu$ is an arbitrary deterministic vector from $\mathbb{R}^N$ ($\alpha = 2$) or $\mathbb{C}^N$ ($\alpha = 1$) with unit norm, $\|\nu\| = 1$ and $h \neq 1$ is a fixed non-critical spike. Since $\nu$ is arbitrary, a version of the isotropic local law of Knowles and Yin [30] plays a key role.

**Proposition 28.** Let $W_N$ be a Wigner matrix whose off-diagonal moments match GOE or GUE to third order. We have:

(i) (isotropic local law) Fix $\tau > 0$. Then for each $\varepsilon > 0$, we have w.o.p.

$$\nu^* R(z) \nu = s_{sc}(z) \nu^* \nu + O(N^{\varepsilon}\Psi(z))$$

uniformly for $z \in S(\tau)$ and deterministic vectors $\nu, \nu$ of unit Euclidean length in $\mathbb{C}^N$.

(ii) (isotropic delocalization) Let $u^{(j)}$ be the $j$-th principal normalized eigenvector of $W_N$. Then, for each $\varepsilon > 0$, we have w.o.p.

$$\max_j |\nu^* u^{(j)}|^2 = O(N^{\varepsilon-1})$$

uniformly for normalized deterministic vectors $\nu \in \mathbb{C}^N$.

The proof is a direct modification, summarized in appendix B.3, of those of Case A of Theorems 2.2 and 2.5 of [30], established for Wigner matrices with sub-exponential entries whose third moments match those of GUE/GOE. In contrast, we allow for arbitrary bounded variance profile along the main diagonal and assume bounded moments (68) instead of the sub-exponentiality.

**Remark.** While [11] proves the isotropic law for generalized Wigner matrices (without moment matching to GUE), the assumptions on variances exclude our setting in which the diagonal variances follow arbitrary profiles.

**Proposition 29 (Spiked log determinant CLT).** Let $W_{h,N}$ be a non-critically spiked Wigner matrix as defined above. Let $E = E_N = 2 + \sigma_N^{-2}/3$ with $-\gamma \leq \sigma_N \ll \log^2 N$. Then

$$\tau_N^{-1}(\log|\det(W_{h,N} - E)| - \mu_N) \overset{d}{\to} N(0, 1).$$

**Proof.** Note that, for any $E$ such that $W_N - E$ is invertible, we have

$$\log|\det(W_{h,N} - E)| = \log|\det(W_N - E)| + \log|1 + h\nu^*(W_N - E)^{-1}\nu|.$$

To transfer to $W_{h,N}$ the CLT established for $W_N$ in Proposition 27, it is sufficient to prove that

$$\tau_N^{-1}\log|1 + h\nu^*(W_N - E)^{-1}\nu| \overset{P}{\to} 0.$$
Fix $\epsilon > 0$ small, and let $z = E + i\eta$ with $\eta = N^{-1/3-3\epsilon}$. Since $R(z) = (W_N - z)^{-1}$, we have
\[ v^* R(E)v = s_{sc}(z) + (v^* R(z)v - s_{sc}(z)) + v^* (R(E) - G(z))v. \]

Now use spectral decomposition (84), the non-concentration bound of Lemma 26 and finally isotropic delocalization (93) to conclude that with high probability,
\[
|v^* (R(E) - R(z))v| \leq \sum_j |v^* u^{(j)}|^2 \left| \frac{1}{E - \lambda_j} - \frac{1}{E + i\eta - \lambda_j} \right|
\leq \eta \max_j |v^* u^{(j)}|^2 \sum_j \frac{1}{(E - \lambda_j)^2}
\leq \eta \max_j |v^* u^{(j)}|^2 N^{4/3+\epsilon} = o_P(1).
\]

Further, for such $\eta$, we have $\Psi(z) \lesssim (N\eta)^{-1/2}$ and so (93) yields
\[ v^* R(z)v - s_{sc}(z) = o_P(1). \]

Finally, $s_{sc}(z) = -1 + O(|E - 2|^{1/2} + \eta^{1/2}) = -1 + o(1)$, (e.g. [17, Lemma 6.2]), and so
\[ v^* R(E)v = -1 + o_P(1). \]

Therefore (95) holds, since for any fixed $h \neq 1$,
\[ \log |1 + hv^*(W_N - E)^{-1}v| = \log |1 - h| + o_P(1) = O_P(1). \]

**Remark.** For the Gaussian ensembles and $\sigma_N$ slowly diverging to infinity, proposition 29 can be extended to the critical case $h = 1$ using the tri-diagonal representations of the spiked GUE/GOE. In such an extension, an extra shift $-\frac{1}{2}\log N$ will ensure the convergence of the normalized log determinant to the standard normal distribution (see appendix B.4).

**Acknowledgments.** The authors would like to thank Ofer Zeitouni for drawing our attention to references [4] and [14].

**Funding.** The first and fourth authors were supported in part by NSF grant DMS 1811614.

**REFERENCES**

[1] Albrecht, J. T., Chan, C. P. and Edelman, A. (2009). Sturm sequences and random eigenvalue distributions. *Foundations of Computational Mathematics* 9 461–483.

[2] Anderson, G. and Qi, S. L. (1997). A monotonicity property of the gamma function. *Proceedings of the American Mathematical Society* 125 3355–3362.

[3] Anderson, G. W., Guionnet, A. and Zeitouni, O. (2010). *An Introduction to Random Matrices*. Cambridge university press.

[4] Augeri, F., Butez, R. and Zeitouni, O. (2020). A CLT for the characteristic polynomial of random Jacobi matrices, and the G/βE models for $\beta > 0$. *arXiv preprint arXiv:2011.06870v2*.

[5] Bai, Z. D. and Silverstein, J. (2009). *Spectral Analysis of Large Dimensional Random Matrices*, Second ed. Springer, New York.

[6] Baik, J., Ben Arous, G. and Pêché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability* 33 1643–1697.

[7] Baik, J. and Lee, J. O. (2016). Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. *Journal of Statistical Physics* 165 185–224.

[8] Baik, J. and Lee, J. O. (2017). Fluctuations of the Free Energy of the Spherical Sherrington–Kirkpatrick Model with Ferromagnetic Interaction. *Annales Henri Poincare* 18 1867–1917.
[9] Baik, J., Lee, O. J. and Wu, H. (2018). Ferromagnetic to Paramagnetic Transition in Spherical Spin Glass. *Journal of Statistical Physics* **173** 1484–1522.

[10] Benaych-Georges, F. and Knowles, A. (2018). Lectures on the local semicircle law for Wigner matrices. arXiv:1601.04055v4.

[11] Błaszczyszyn, B., Knowles, A., Yau, H.-T. and Yin, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electronic Journal of Probability* **19** 1–53.

[12] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

[13] Bourgade, P. and Mody, K. (2019). Gaussian fluctuations of the determinant of Wigner matrices. *Electronic Journal of Probability* **24** 1–28.

[14] Bourgade, P., Mody, K. and Pain, M. (2021). Optimal local law and central limit theorem for β-ensembles. arXiv preprint arXiv:2103.06841v2.

[15] Duy, T. K. (2017). Distributions of the determinants of Gaussian beta ensembles. In *2023 Spectral and Scattering Theory and Related Topics* 77–85. RIMS Kokyuroku.

[16] Erdélyi, A. (1960). Asymptotic solutions of differential equations with transition points or singularities. *J. Mathematical Phys.* **1** 16–20. MR111915

[17] Erdös, L. and Yau, H.-T. (2017). A dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics* **28**. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI.

[18] Erdös, L., Yau, H.-T. and Yin, J. (2012). Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields* **154** 341–407.

[19] Etemadi, N. (1985). On some classical results in probability theory. *Sankhyā A: The Indian Journal of Statistics, Series A* 215–221.

[20] Forrester, P. J. (2010). *Log-Gases and Random Matrices*. Princeton University Press.

[21] Forrester, P. J. and Rains, E. M. (2001). Inter-relationships between orthogonal, unitary and symplectic matrix ensembles. In *Random Matrix Models and Their Applications: Mathematical Sciences Research Institute publications*, **40** 171–207. Cambridge University Press.

[22] Fyodorov, Y. V., Khoruzhenko, B. A., Simm, N. J. et al. (2016). Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian Unitary Ensemble. *The Annals of Probability* **44** 2980–3031.

[23] Götzé, F. and Tikhomirov, A. (2005). The rate of convergence for spectra of GUE and LUE matrix ensembles. *Central European Journal of Mathematics* **3** 666–704.

[24] Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*, seventh ed. Elsevier Inc.

[25] James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statistics* **25** 40–75. MR0060779 (15,726b)

[26] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Mathematical Journal* **91** 151–204.

[27] Johnstone, I. M., Klochkov, Y., Onatski, A. and Pavlyshyn, D. (2021). Spin glass to paramagnetic transition in spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Manuscript in preparation*.

[28] Johnstone, I. M. and Ma, Z. (2012). Fast approach to the Tracy-Widom law at the edge of GOE and GUE. *The Annals of Applied Probability* **22** 1962–1988.

[29] Johnstone, I. M. and Onatski, A. (2020). Testing in high-dimensional spiked models. *Annals of Statistics* **48** 1231–1254.

[30] Knowles, A. and Yin, J. (2013). The Isotropic Semicircle Law and Deformation of Wigner Matrices. *Communications on Pure and Applied Mathematics* **LXVI** 1663–1749.

[31] Kosterlitz, J. M., Thouless, D. J. and Jones, R. C. (1976). Spherical model of a spin-glass. *Physical Review Letters* **36** 1217–1220.

[32] Kravchuk, Y. V. (2007). Correlations of the characteristic polynomials in the Gaussian Unitary Ensemble or a singular Hankel determinant. *Duke Mathematical Journal* **139** 581–619.

[33] Lambert, G. and Paquette, E. (2020). Strong approximation of Gaussian β-ensemble characteristic polynomials: the hyperbolic regime. arXiv preprint arXiv:2001.09042.

[34] Lambert, G. and Paquette, E. (2020). Strong approximation of Gaussian β-ensemble characteristic polynomials: the edge regime and the stochastic Airy function. arXiv preprint arXiv:2009.05003.

[35] Landou, B. and Soose, P. (2020). Applications of mesoscopic CLTs in random matrix theory. *Ann. Appl. Probab.* **30** 2769–2795. MR4187127

[36] Li, Y., Schnelli, K. and Xu, Y. (2017). Central limit theorem for mesoscopic eigenvalue statistics of deformed Wigner matrices and sample covariance matrices. arXiv:1909.12821v2 1–36.

[37] Maida, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.* **12** 1131–1150. MR2336602
[38] MEHTA, M. L. (1991). Random matrices, Second ed. Academic Press, Inc., Boston, MA. MR1083764
[39] MEHTA, M. L. (2004). Random matrices, third ed. Pure and Applied Mathematics (Amsterdam) 142. Elsevier/Academic Press, Amsterdam. MR2129906
[40] MUIRHEAD, R. J. (1982). Aspects of multivariate statistical theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR652932
[41] OLVER, F. W. J. (1974). Asymptotics and Special Functions. Academic Press.
[42] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. Annals of Statistics 41 1204–1231.
[43] PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population Structure and Eigenanalysis. PLoS Genet 2 e190.
[44] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. Probability Theory and Related Fields 134 127–173.
[45] RIO, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. Journal of Theoretical Probability 22 146–163.
[46] TAO, T. and VU, V. (2012). A central limit theorem for the determinant of a Wigner matrix. Advances in Mathematics 231 74–101.
[47] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. Communications in Mathematical Physics 177 727–754.
[48] TRACY, C. A. and WIDOM, H. (1998). Correlation functions, cluster functions, and spacing distributions for random matrices. Journal of Statistical Physics 92 809–835.
[49] TROTTER, H. F. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner’s semi-circle law and a theorem of Kac, Murdock, and Szegö. Advances in mathematics 54 67–82.
[50] WIGNER, E. P. (1965). Distribution laws for the roots of a random Hermitian matrix. Statistical Theories of Spectra: Fluctuations 446–461.
[51] YAO, J. and BAI, Z. (2005). On the convergence of the spectral empirical process of Wigner matrices. Bernoulli 11 1059–1092.
Supplementary Material for “An edge CLT for the log determinant of Wigner matrices” by I.M. Johnstone, Y. Klochkov, A. Onatski, and D. Pavlyshyn

Abstract: The Supplementary Material contains relatively more technical proofs of the main paper. It consists of Appendix A, which corresponds to the proofs for Gaussian ensembles, and Appendix B, which corresponds to proofs for Wigner extension. To help the reader to navigate the Supplement, we start from a Table of Contents that encompasses the main text. The Table’s references to the Supplement’s sections contain short descriptions of the content of the sections.

CONTENTS

1 Introduction ................................................................. 1
  1.1 Two motivating applications ....................................... 2
  1.2 Outline of approach ............................................... 4
  1.3 Related work ....................................................... 7
  1.4 Organization and some notation ................................... 8
2 GβE: The CLT slightly away from the edge ........................... 8
3 G(U/O)E: All the way to the edge ..................................... 10
4 Proofs for Gaussian ensembles ........................................ 11
  4.1 Proofs from Section 2 ............................................. 11
    4.1.1 Preliminaries .................................................. 12
    4.1.2 Step one: CLT for \( \sum_{i=1}^{N} L_i \) ......................... 13
    4.1.3 Step 2a: Uniform bound on \( L_i \) .......................... 15
    4.1.4 Step 2b: Uniform bound on \( R_i \) .......................... 17
    4.1.5 Step 3: Linear approximation for \( \log |M_N| \) ............... 18
    4.1.6 CLT for \( \log |D_N| \) ........................................ 19
  4.2 Proofs from Section 3 ............................................. 20
    4.2.1 Preliminaries .................................................. 20
    4.2.2 Gaussian Non-concentration ................................ 21
    4.2.3 Proof of Proposition 3 ....................................... 21
5 Extension to Wigner matrices ........................................ 24
  5.1 Lindeberg swapping formalism for asymptotically flat \( Q \) ....... 24
  5.2 Flatness for Stieltjes functionals ................................ 26
    5.2.1 Resolvent perturbation: deterministic bounds ............... 26
    5.2.2 Local law .................................................... 27
    5.2.3 Stieltjes functionals ....................................... 28
  5.3 Concluding the extension to Wigner matrices ................. 30
    5.3.1 Wigner Non-concentration ................................ 30
    5.3.2 Log-determinant ........................................... 32
6 The spiked case ......................................................... 33
Acknowledgments .................................................................. 34
Funding ............................................................................. 34
References ......................................................................... 34
A Proofs for the Gaussian ensembles .................................... 38
  A.1 Proof of Lemma 4 (about relative location of \( \lambda_i \) and \( E \)) .... 38
  A.2 Proofs of technical lemmas from Section 4 ...................... 39
    A.2.1 Proof of Lemma 5 (bounds on even moments of \( \xi_i \)) ....... 39
    A.2.2 Proof of Lemma 6 (bounds on \( g_i \), used to verify the Lyapunov condition) .... 39
A.2.3 Proof of Lemma 7 (improved bounds on $E\xi_i^2$) ............................................. 40
A.2.4 Proof of Lemma 10 ($\xi_i$ and $L_i$ are sub-gamma) ............................................. 40
A.2.5 Proof of Lemma 11 (products $\gamma_{j,i+1}$ are not too small) ..................................... 41
A.2.6 Proof of equation (48) (about $|R_0|$ being small) from Lemma 12 ............... 41
A.2.7 Proof of Lemma 13 (bounds on components of $\sum (R_i + R_i^2 / 2)$) ........... 43
A.2.8 Proof of Lemma 14 (asymptotics of $\sum T_\delta$) .................................................. 47
A.2.9 Proof of Lemma 16 (a variance bound on linear spectral statistics of GUE) .......... 47

A.3 Discussion of edge bounds for one-point functions (63) ........................................ 48
A.4 Proof of Theorem 19 (linking functions of GUE and GOE) ..................................... 50

B Proofs for Wigner extension and Section 6 ................................................................. 52
B.1 Proof of Proposition 21 (about convergence of $\xi_n$ implying convergence of $\xi_n'$) ................................................................. 52
B.2 Proof of Lemma 26 (bounds on inverse power sums for Wigner matrices) ............ 53
B.3 Proof of proposition 28 (isotropic local law and delocalization) ......................... 53
B.4 Extension of spiked CLT to the critical case for G(U/O)E ................................. 60

APPENDIX A: PROOFS FOR THE GAUSSIAN ENSEMBLES

A.1 Proof of Lemma 4 (about relative location of $\lambda_i$ and $E$). Let $\lambda_1 \geq \cdots \geq \lambda_N$ be eigenvalues of $W = W_N$ and define the eigenvalue counting function $N_W(I) = \# \{j : \lambda_j \in I\}$. The first bound in (21) follows from Tracy-Widom convergence of $\lambda_1$.

For the second bound, set $I_+^k = [2 - \gamma N^{-2/3}, \infty)$ and note that

$$P(\lambda_k > E) \leq P(\lambda_k > 2 - \gamma N^{-2/3}) = P(N_W(I_+^k) > k - 1)).$$

Using the tail bound (63) for the one-point function,

$$\mathbb{E}N_W(I_+^k) = \int_{2 - \gamma N^{-2/3}}^{\infty} N \rho_N(x) \, dx = \int_{-\gamma}^{\infty} N^{1/3} \rho_N(2 + sN^{-2/3}) \, ds \leq C_1(\gamma) \int_{-\gamma}^{\infty} e^{-2s} \, ds \leq C_2(\gamma).$$

The last two displays and Markov’s inequality yield $P(\lambda_k > E) \leq C_2(\gamma)/(k - 1) < \epsilon$ so long as $k = k(\epsilon, \gamma)$ is sufficiently large.

For the first bound in (22), set $x_{jN} = N^{2/3}(\lambda_j - 2)$. Using (21), choose $k = k(\epsilon/2, \gamma + 1)$ so that the event $\mathcal{E}_{jN} = \{x_{jN} < -\gamma - 1\}$ has probability at least $1 - \epsilon/2$. On $\mathcal{E}_{jN}$, for $j \geq k$ we have $x_{jN} - \sigma_N \leq x_{kN} + \gamma < -1$.

Consider now $j < k$. Since the $j$th Tracy-Widom law $F_j$ (of type 2 or 1 for the GUE or GOE case, respectively) has a continuous distribution function, weak convergence of $x_{jN}$ also implies that $P\{x_{jN} \leq x\} \to F_j(x)$ uniformly in $x$. Since $F_j$ is uniformly continuous, we can choose $c_1$ small so that for large $N$ and each $j < k$,

$$P\{x_{jN} \in [\sigma_N - c_1, \sigma_N + c_1]\} \leq \epsilon/2k.$$

Let $\mathcal{E}_{jN}^c$ be the corresponding event. On the event $\cap_{j \leq k} \mathcal{E}_{jN}$, which has probability at least $1 - \epsilon$, we have $\min_{1 \leq j \leq N} N^{2/3}|\lambda_j - E| = \min_{j \geq k} |x_{jN} - \sigma_N| \geq c_1$, and so the first bound in (22) is proved.

For the second bound in (22), since $N^{2/3}|\lambda_j - E| \leq |x_{jN}| + |\sigma_N|$, we have

$$P\left(\max_{j \leq k} N^{2/3}|\lambda_j - E| > C_1 + |\sigma_N|\right) \leq P\left(\max_{j \leq k} |x_{jN}| > C_1\right) \leq P(x_{1N} > C_1) + P(x_{kN} < -C_1).$$

Again using Tracy-Widom convergence of $x_{1N}$ and $x_{kN}$, the right side can be made less than $\epsilon$ for large $N$ by choosing $C_1$ large.
A.2. Proofs of technical lemmas from Section 4.

A.2.1. Proof of Lemma 5 (bounds on even moments of \( \xi_i \)). The existence of \( C_\gamma \) follows from the fact that \( \xi_i \) are sub-gamma random variables (see Definition (24) and equation (102)) and from equation (2.7) of [12]. The lower bound on \( E\xi_i^2 \) follows from \( m_i r_i = (i - 1)/N\theta_N^2 \).

\[
\mathbb{E}\xi_i^2 = \frac{\alpha}{N\theta_N^2 r_i} + \frac{\alpha m_i r_i}{N\theta_N^2 r_i^2} = \frac{\alpha}{N\theta_N^2 r_i^2} \left[ 1 + \frac{i - 1}{N\theta_N^2 r_i^2} \right].
\]

A.2.2. Proof of Lemma 6 (bounds on \( g_i \), used to verify the Lyapunov condition). Since \( \gamma_i \) is increasing with \( i \), we have

\[
g_i > 1 + \gamma_i + \ldots + \gamma_i^{N-i+1} = \frac{1 - \gamma_i^{N-i+2}}{1 - \gamma_i}.
\]

On the other hand, for all sufficiently large \( N \)

\[
\gamma_i < m_i \leq 1 - \sqrt{1 - \frac{N}{N\theta_N^2}} = 1 - \sqrt{1 - \theta_N^{-2}} < 1 - N^{-1/3}w_N^{1/2}.
\]

Hence, for \( i \leq N - N^{1/3} \), any \( k > 0 \), and all sufficiently large \( N \),

\[
\gamma_i^{N-i+2} < \left( 1 - N^{-1/3}w_N^{1/2} \right)^{N^{1/3}} < e^{-w_N^{1/2}} < e^{-k \log \log N} = \log^{-k} N.
\]

The lower bound (35) follows from this and the fact that \( 1 - \gamma_i = 2 (r_i - 1)/r_i \).

The elementary bound (37) follows from the definition of \( g_i \). It shows that the lower bound (35) fails for \( i > N - N^\alpha \) for any \( \alpha < 1/3 \).

For the upper bound, we seek a value \( \kappa \) for which the inequalities \((1 - \gamma_i) g_i \leq 1 + \kappa \) may be established by induction for \( i = N, N - 1, \ldots, 1 \). The initial step holds for any \( \kappa \geq 0 \), since \( g_N = 1 + \gamma_N \) implies that \((1 + \gamma_N) g_N = 1 - \gamma_N^2 \). Assuming \((1 - \gamma_i) g_i \leq 1 + \kappa \) and using the recursion \( g_{i-1} = \gamma_{i-1} g_i + 1 \), we have

\[
(1 - \gamma_{i-1}) g_{i-1} \leq (1 - \gamma_{i-1}) \left[ \frac{\gamma_i}{1 - \gamma_i} (1 + \kappa) + 1 \right],
\]

and so the induction step works at least so long as

\[
\frac{1 - \gamma_{i-1}}{1 - \gamma_i} \gamma_{i-1} (1 + \kappa) - \gamma_{i-1} \leq \kappa
\]

for \( 2 \leq i \leq N \). On rearrangement this condition becomes

\[
\gamma_{i-1} (\gamma_i - \gamma_{i-1}) \leq \kappa (1 - \gamma_i - \gamma_{i-1} + \gamma_i^2).
\]

Both sides of the inequality are monotone in such a way that we need only work with \( i = N \). Indeed, write \( \gamma_i = \gamma(x_i) \) and note that \( \gamma(x) = (1 - R)/(1 + R) \) and \( \gamma'(x) = R^{-1}(1 + R)^{-2} \) are increasing, where \( R(x) = \sqrt{1 - x} \). Also set \( R_N = R(x_N) = \sqrt{1 - x_N} \). We then have both

\[
\gamma_{i-1} (\gamma_i - \gamma_{i-1}) \leq \gamma_N \gamma_N' \Delta_N \leq \Delta_N / \left[ R_N (1 + R_N)^2 \right]
\]

and

\[
1 - \gamma_i - \gamma_{i-1} + \gamma_i^2 = (1 - \gamma_{i-1})^2 - (\gamma_i - \gamma_{i-1}) \geq (1 - \gamma_N)^2 - \gamma_N' \Delta_N = (4R_N^2 - \Delta_N) / \left[ R_N (1 + R_N)^2 \right].
\]
From these displays and (97), we see that any \( \kappa \) larger than \( \Delta_N/(4R_N^3 - \Delta_N) \) suffices for the induction. Noting that for large enough \( N \) we have \( R_N > \sqrt{w_N N^{-1/3}} \) and
\[
4R_N^3/\Delta_N - 1 > 4w_N^{3/2}N^{-1/2} - 1 > 3w_N^{3/2},
\]
we conclude that we may certainly take \( \kappa = w_N^{-3/2} \) in the induction and in our upper bound.

A.2.3. Proof of Lemma 7 (improved bounds on \( E\xi_i^2 \)). Using (96) and the identities \( m_ir_i - m_{i-1}r_{i-1} = 1/N\theta_N^2 \) and \( m_{i-1} + r_{i-1} = 2 \) we have
\[
\alpha^{-1}N\theta_N^2r_i^2E\xi_i^2 = 1 + \frac{mi}{r_i^2} = 1 + \frac{mi}{r_i^2} + \frac{1}{N\theta_N^2r_{i-1}^2} = \frac{2}{r_i} + \frac{1}{N\theta_N^2r_{i-1}^2}.
\]
Appealing to the monotonicity of \( r_i \in [1,2] \), we get
\[
\frac{1}{2\alpha}N\theta_N^2r_i^3E\xi_i^2 - 1 = \frac{r_i - r_{i-1}}{r_i} + \frac{r_i}{2N\theta_N^2r_{i-1}^2} \leq \frac{r_i}{2N\theta_N^2r_{i-1}^2} \leq \frac{1}{2N} \tag{98}
\]
On the other hand,
\[
1 - \frac{1}{2\alpha}N\theta_N^2r_i^3E\xi_i^2 = \frac{r_i - r_i}{r_i} - \frac{r_i}{2N\theta_N^2r_{i-1}^2} \leq \frac{r_i - r_i}{r_i} \leq \frac{1}{2N\theta_N^2(r_i - 1)}, \tag{99}
\]
where we used
\[
\frac{r_i - r_i}{r_i} = \frac{-\Delta_N r_i(x*)}{r_i} \leq \frac{\Delta_N}{2(r_i - 1)}. \tag{100}
\]
Inequalities (98) and (99) establish the required bound on \( |\varepsilon_i| \). The \( N^{-2/3} \) bound follows from (107).

A.2.4. Proof of Lemma 10 (\( \xi_i \) and \( L_i \) are sub-gamma). Recall the definition (24). Straightforward calculations using the definitions of \( \alpha_i \) and \( \beta_i \) and the monotonicity properties of \( SG(v,u) \), keeping in mind that \( \beta_1 = 0 \), show that
\[
\alpha_i \in SG \left( \frac{\alpha}{N\theta_N^2r_i^2}, 0 \right), \quad \beta_i \in SG \left( \frac{\alpha m_i}{N\theta_N^2r_i^2}, \frac{\alpha}{N\theta_N^2r_i^2} \right).
\]
To bound the moments of \( \beta_i \), we use the following lemma.

**Lemma 30.** Suppose \( X \in SG(v,v) \) with \( v \leq 1/2 \). Then for any \( p > 2 \),
\[
\|X\|^2_p \leq 8vp^2.
\]

**Proof of Lemma 30.** First, closely following the proof of Theorem 2.3 of [12], we obtain the following inequality
\[
\|X\|^p_p \leq p2^{p-1} \left( (2v)^{p/2} \Gamma(p/2) + (2v)^p \Gamma(p) \right).
\]
Since for any \( x > 1, \Gamma(x) \leq x^{x-1} \) (see [2]), and \( (2v)^p \leq (2v)^{p/2} \), we get
\[
\|X\|^p_p \leq \left( 2^{p-2} + 2^{3p/2 - 1} \right) \cdot \frac{v^{p/2} \leq 2^{3p/2} \cdot v^{p/2}}{p^p}.
\]
\( \Box \)
Since $m_i/(r_i r_{i-1}^2) < 1/(r_i r_{i-1}) < 1$, we have $\beta_i \in SG(\alpha/N\theta_N^2, \alpha/N\theta_N^2)$ and so, from lemma 30,

\begin{equation}
\|\beta_i\|_p^2 \leq \frac{8\alpha p^2}{N\theta_N^2} \lesssim \frac{p^2}{N}.
\end{equation}

In addition,

\begin{equation}
\xi_i = \alpha_i + \beta_i \in SG\left(\frac{\alpha}{N\theta_N^2 r_i^2} + \frac{\alpha m_i}{N\theta_N^2 r_i r_{i-1}}, \frac{\alpha}{N\theta_N^2 r_i^2}\right) \in SG\left(\frac{2\alpha}{N\theta_N^2 r_i^2}, \frac{2\alpha}{N\theta_N^2 r_i^2}\right) \equiv SG(v_i, u_i).
\end{equation}

The latter inclusion follows from the facts that $r_{i-1} > r_i$ and $r_i + m_i = 2$.

Next, we use the identity (25) that expresses $L_i$ as a weighted sum of $\xi_j$. For $i = 1$, the inclusion $L_i \in SG(v_{L_i}, u_{L_i})$ follows from the identity $L_1 = \xi_1$ and the observation that $v_{L_1} \geq v_1$ and $u_{L_1} = u_i$. Further, for any $1 < i \leq N$ we have from (25) and (102), $L_i \in SG(\tilde{v}_{L_i}, \tilde{u}_{L_i})$ with

\[ \tilde{v}_{L_i} \leq v_i + \sum_{j=0}^{i-2} \gamma_i^2 \gamma_{i-j}^2 v_{i-j-1}. \]

Since $\gamma_i$ is increasing in $i$, this yields

\[ \tilde{v}_{L_i} \leq v_i (1 + \gamma_i^2 + \gamma_i^4 + ... ) = \frac{2\alpha}{N\theta_N^2 r_i^2 (1 - \gamma_i^2)} < \frac{\alpha}{2N\theta_N^2 r_i^2} = v_{L_i}, \]

where we used

\begin{equation}
(1 - \gamma^2)^{-1} = r^2/(4R) < (r - 1)^{-1}
\end{equation}

with $R = r - 1$, and $r > 1$. For $\tilde{u}_{L_i}$, we have $\tilde{u}_{L_i} \leq \max_{j \leq i} u_j = \alpha \left( N\theta_N^2 r_i^2 \right)^{1/2}$ because $r_i$ is decreasing in $i$. Hence, the inclusion $L_i \in SG(v_{L_i}, u_{L_i})$ follows by the monotonicity of the sub-gamma family.

A.2.5. Proof of Lemma 11 (products $\gamma_{j;i+1}$ are not too small). Let $n_0 = N - \left[ N^{1/3} \log^3 N \right]$. Since $x_{n_0} = n_0/N\theta_N^2 \geq \theta_N^2 - (\log^3 N)/(N^{2/3})$, we have $\epsilon_N = 1 - x_{n_0} \leq 2 (\log^3 N)/N^{2/3}$ for large $N$ and $w_N \ll \log^2 N$. Recalling that for $\epsilon \in (0, \frac{1}{2})$, we have both $\log(1 - \epsilon) \geq -2\epsilon$ and $-\log(1 + \epsilon) \geq -\epsilon$ and noting that $\log(\gamma(x)) = \log(1 - \sqrt{1 - x}) - \log(1 + \sqrt{1 - x})$, we get

\[ \log \gamma(x_{n_0}) \geq -3\sqrt{\epsilon_N} \geq -3\sqrt{2(\log^3 N)/N^{1/3}}. \]

Since $\gamma_j$ and $x_j$ are increasing in $j$, we have for large $N$

\[ \log \gamma_{j;i+1} \geq (j - i) \log \gamma_{i+1} \geq (j - i) \log \gamma(x_{n_0}) \geq -C \log^{-1/2} N \geq \log \frac{1}{2}. \]

A.2.6. Proof of equation (48) (about $|\tilde{R}_i|$ being small) from Lemma 12. Recall the decomposition

\[ T \tilde{e}_i = -T \delta_i + T \tilde{e}_i^m + T \tilde{e}_i^a + T \tilde{e}_i^q. \]

We start from the term $T \tilde{e}_i^m$. As explained in the proof of lemma 12 in the main text, this term can be viewed as a sum $\sum_{j=1}^i \gamma_{j;i+1} R_{j;i-1}^{(1)} \beta_j = \sum_{j=1}^i X_j$ of martingale differences, and the Marcinkiewicz-Zygmund-type inequality of [45, Theorem 2.1] says that $\| \sum_{j=1}^i X_j \|_p^2 \leq (p - 1) \| X_j \|_p^2$. 

Appealing to (101) we obtain

\begin{equation}
\| \sum_{j=1}^{i} \gamma_{i} \beta_{j} \|_{p}^{2} \leq \frac{p-1}{1-\gamma_{i}} \max_{j \leq i} \| \tilde{R}_{j-1}^{(1)} \|_{p}^{2} \| \beta_{j} \|_{p}^{2} \leq \frac{\alpha p^{3}}{N(r_{i} - 1)} \max_{j \leq i} \| \tilde{R}_{j-1}^{(1)} \|_{p}^{2}.
\end{equation}

To obtain the latter inequality, we also used

\begin{equation}
\frac{1}{1-\gamma_{i}} \leq \frac{r_{i}}{r_{i} - m_{i}} = \frac{r_{i}}{2(r_{i} - 1)} < \frac{1}{r_{i} - 1}.
\end{equation}

By construction \( |\tilde{R}_{i}| \leq N^{-1/3}/2 \) and for \( N \) sufficiently large \((1 - N^{-1/3}/2)^{-1} \leq 4/3 \) so that \( R_{i}^{(1)} \leq N^{-1/3} \). For the Winsorized “martingale” terms, then, from (104)

\begin{equation}
\| T_{\varepsilon_{i}}^{\text{n}} \|_{p}^{2} \leq \frac{\alpha p^{3}}{N^{5/3}(r_{i} - 1)} \leq \frac{\alpha p^{3}}{N^{4/3}}.
\end{equation}

For the last inequality, we used

\begin{equation}
\frac{1}{r_{i} - 1} \leq \frac{1}{r_{N+1} - 1} = (1 - \theta_{N}^{-2})^{-1/2} = \frac{N^{1/3}}{\sqrt{2N}} (1 + o(1)).
\end{equation}

For any random variable \( P(|X| \geq \varepsilon \|X\|_{2 \log N}) \leq e^{-2\log N} = N^{-2} \), so taking the union bound over \( i = 1, \ldots, N \), we find that there exists \( C > 0 \) such that for all sufficiently large \( N \), with probability at least \( 1 - 1/N \),

\[ \max_{i} |T_{\varepsilon_{i}}^{\text{n}}| \leq CN^{-2/3} \log^{3/2} N. \]

Turning to \( T_{\delta_{i}} \), one finds that \( 1/r(x) \) is convex increasing, with \((1/r)' = -r'/r^{2} \), and so for \( i \geq 2 \)

\begin{equation}
0 < \delta_{i} = m_{i} \left( \frac{1}{r_{i}} - \frac{1}{r_{i-1}} \right) \leq \Delta_{N} \frac{-\gamma r_{i}'}{r_{i}} (x_{i}) = \frac{\Delta_{N}}{2} \frac{m_{i}}{r_{i}^{2}/(r_{i} - 1)} \leq \frac{\Delta_{N}}{2(r_{i} - 1)}.
\end{equation}

Furthermore, \( \delta_{i} \) is increasing, so that from (29) followed by (108), then (105) and (107),

\begin{equation}
T_{\delta_{i}} \leq \frac{\delta_{i}}{1 - \gamma_{i}} \leq \frac{1}{2N\theta_{N}^{2}/(r_{i} - 1)^{2}} = O(N^{-1/3}w_{N}^{-1}).
\end{equation}

For the “small” term \( \varepsilon_{i}^{s} \), we have \( \varepsilon_{i}^{s} \leq (1 - \tilde{R}_{i-1})^{-1} (\gamma_{i} |\tilde{R}_{i-1}|^{3} + \delta_{i} |\tilde{R}_{i-1}|) \leq \frac{4}{3} [(8N)^{-1} + \delta_{i} N^{-1/3}] \). From (29) and (105), we have

\begin{equation}
|T_{\varepsilon_{i}}^{s}| \leq \frac{1}{1 - \gamma_{i}} \left[ \frac{1}{6N} + \frac{4\delta_{i}}{3N^{1/3}} \right] \leq \frac{1}{N(r_{i} - 1)}
\end{equation}

for large \( N \), since from (108) and (107) \( \delta_{i} N^{-1/3} \leq \Delta_{N} [N^{1/3}(r_{i} - 1)]^{-1} = O(N^{-1}w_{N}^{-1/2}) \). In particular, again from (107), \( \max_{i \leq N} |T_{\varepsilon_{i}}^{s}| = O(N^{-2/3}w_{N}^{-1/2}) \).

Finally, since the “quadratic” term \( \varepsilon_{i}^{q} \leq \tilde{R}_{i-1}^{2} \), we have from (29)

\begin{equation}
|T_{\varepsilon_{i}}^{q}| \leq \frac{1}{4N^{2/3}(1 - \gamma_{i})} = O(N^{-1/3}w_{N}^{-1/2}).
\end{equation}

Taking into account the representation \( \tilde{R}_{i} = L_{i} + T_{\varepsilon_{i}} \) and result (46) of Lemma 9, we see that there exists \( C > 0 \) such that on an event of probability at least \( 1 - o(1) \),

\[ \max_{i \leq N} |\tilde{R}_{i}| \leq CN^{-1/3}/\sqrt{\log \log N}, \]

which establishes (48) of Lemma 12.
A.2.7. Proof of Lemma 13 (bounds on components of $\sum (R_i + R_i^2/2)$). That $|\sum_i^N T \varepsilon_i^m| \leq N^{-1} \sum_1^N (r_i - 1)^{-1} = O(1)$ on a set $\mathcal{R}_N$ of probability $1 - o(1)$ follows already from (110) and (32).

To show that $\sum_i^N T \varepsilon_i^m = O_P(1)$ it would in principle suffice to show that $\|T \varepsilon_i^m\|_1$ is summable. Lemma 12 controls $\{R_i\}$ on $\mathcal{R}_N$, but this is not enough for moments of $R_i$ since $P(\mathcal{R}_N^c) \approx O(\log^{-k} N)$. Instead we define a recursive mutilation $\hat{R}_i$ for $R_i$ with two properties: (a) that $\hat{R}_i = R_i$ for $i = 1, \ldots, N$ on $\mathcal{R}_N$, and (b) for all such $i$,

$$\|\hat{R}_i\|_4 \leq 4\alpha^{1/2} N^{-1/2} (r_i - 1)^{-1/2}. \quad (112)$$

To define $\{\hat{R}_i\}$, recall that the process $\{R_i\}$ satisfies

$$R_i = L_i - T \delta_i + T \varepsilon_i^m + T \varepsilon_i^q. \quad (113)$$

We set $\hat{R}_1 = L_1 = R_1$. For $i \geq 2$, given $\{\alpha_i, \beta_i, \gamma_i, \delta_i\}_1^i$ and $\{R_j, \hat{R}_j\}_1^{i-1}$, set

$$\hat{R}_i = L_i - T \delta_i + T \varepsilon_i^m + \phi_{N(r_i - 1)}(T \varepsilon_i^q) + T \varepsilon_i^q, \quad (113)$$

with the modifications

$$\varepsilon_m = \beta_i \hat{R}_{i-1}/[1 - \phi_1/2(\hat{R}_{i-1})], \quad \varepsilon^q = \gamma_i \phi_{N(r_i - 1)}(\hat{R}_{i-1}) \hat{R}_{i-1}. \quad (113)$$

Property (a) is verified by chasing definitions: let $\mathcal{H}_i = \{\varepsilon_j^m = \varepsilon_j^m, \varepsilon_j^q = \varepsilon_j^q \text{ for } j = 1, \ldots, i \} \cap \{||T \varepsilon_i^m|| \leq 2/N(r_i - 1)\}$. By definition event $\mathcal{H}_i$ implies $\hat{R}_j = R_j$ for $j \leq i$. One checks by induction that $\mathcal{H}_i$ holds on $\mathcal{R}_N$ ($\mathcal{H}_1$ is always true), and so $\{\hat{R}_j\}_1^N = \{R_j\}_1^N$.

We turn to property (b). By Lemma 10 and Theorem 2.3 of [12],

$$\|L_i\|_4 \leq 2 \left( \frac{4\alpha}{N \theta^2_N (r_i - 1)} \right)^2 + 4! \left( \frac{4\alpha}{N \theta^2_N r_i^2} \right)^4 \leq \frac{33\alpha^2}{N^2 \theta^4_N (r_i - 1)^2} \quad (114)$$

for all sufficiently large $N$. Hence

$$\|L_i\|_4 \leq \frac{3\alpha^{1/2}}{\theta_N \sqrt{N (r_i - 1)}}. \quad (114)$$

To bound the norm of $T \varepsilon_i^m$, observe that since $\beta_i$ and $\hat{R}_{i-1}$ are independent, $\|\varepsilon_i^m\|_p \leq 2\|\hat{R}_{i-1}\|_p \|\beta_i\|_p$. Since $\beta_i \in SG\left(\frac{\alpha}{\theta^2_N}, \frac{\alpha}{\theta^2_N}\right)$, we have $\|\beta_i\|_p \leq \frac{8\alpha^2}{\theta^2_N}$ by Lemma 30, and so

$$\|\varepsilon_i^m\|_p \leq \frac{2^{5/2}\alpha^{1/2}}{\theta_N \sqrt{N}} \|\hat{R}_{i-1}\|_p. \quad (114)$$

By the Marcinkiewicz-Zygmund type inequality (see Theorem 2.1 from Rio (2009)), and using $(1 - \gamma_i)^{-1} \leq (r_i - 1)^{-1} = o\left(N^{1/3}\right)$,

$$\|T \varepsilon_i^m\|_4 \leq 3^{1/2} \|\varepsilon_i^m\|_p \leq \frac{2^{9/2}\alpha^{1/2}}{\theta_N \sqrt{N} (1 - \gamma_i^2)} \max_{j \leq i-1} \|\hat{R}_j\|_4 = o\left(N^{-1/3}\right) \max_{j \leq i-1} \|\hat{R}_j\|_4. \quad (114)$$

Next, by definition,

$$\|\phi_{N(r_i - 1)}(T \varepsilon_i^q)\|_4 \leq \frac{2}{N (r_i - 1)}.$$

And finally, we have $\|\varepsilon_i^q\|_4 \leq N^{-1/3} \|\hat{R}_{i-1}\|_4$. Therefore, from (29)

$$\|T \varepsilon_i^q\|_4 \leq N^{-1/3} \frac{1}{1 - \gamma_i} \max \|\hat{R}_j\|_4 = o(1) \max \|\hat{R}_j\|_4. \quad (114)$$
Along with (109), all this sums up to

\[ \| \hat{R}_i \|_4 \leq \frac{3 \alpha^{1/2}(1 + o(1))}{\sqrt{N} (r_i - 1)} + \frac{1}{2N (r_i - 1)^2} + o(N^{-1/3}) \max_{j \leq i - 1} \| \hat{R}_j \|_4 + \frac{2}{N (r_i - 1)} + o(1) \max_{j \leq i - 1} \| \hat{R}_j \|_4, \]

which implies

\[ \| \hat{R}_i \|_4 \leq \frac{1}{\sqrt{N} (r_i - 1)} (3 \alpha^{1/2} + o(1)) + o(1) \max_{j \leq i - 1} \| \hat{R}_j \|_4. \]

Hence, by induction (and for sufficiently large \( N \)), we conclude that (112) and property (b) hold.

Returning to \( \sum T \hat{\varepsilon}_i^m \), the Marcinkiewicz-Zygmund type bound (114) and then (112) show that

\[ \| T \hat{\varepsilon}_i^m \|_4 \leq \frac{C \alpha^{1/2}}{\sqrt{N} (1 - \gamma_i)} \leq \frac{\alpha^{1/2}}{N(r_i - 1)}, \]

From (32), \( \sum \| T \hat{\varepsilon}_i^m \|_4 \leq \sum \| T \hat{\varepsilon}_i^m \|_4 \leq C \theta_R = O(1) \). Therefore \( \sum_{1}^{N} T \hat{\varepsilon}_i^m = O_P(1) \) and the same is then true for \( \sum_{1}^{N} T \hat{\varepsilon}_i^m \) as they are equal on \( R_N \), a set of probability \( 1 - o(1) \).

To finally show that \( \sum T \hat{\varepsilon}_i^q \leq O_P(1) \), we first define \( \hat{\varepsilon}_i^q = \gamma_i (\hat{R}_{i-1}^2 - L_{i-1}^2) \), noting that \( \{ \hat{\varepsilon}_i^q \}_{1}^{N} = \{ \varepsilon_i^q \}_{1}^{N} \) on \( R_N \). We first use (113) to bound \( \| \hat{R}_i - L_i \|_2 \). Since \( \| \hat{\varepsilon}_i^q \|_2 \leq \| \hat{R}_{i-1}^2 \|_2 \) and \( \| \hat{R}_{i-1} \|_4 \), we have from (112) and (105) that

\[ \| T \hat{\varepsilon}_i^q \|_2 \leq \frac{1}{1 - \gamma_i} \max_{j \leq i} \| \hat{\varepsilon}_i^q \|_2 \leq \frac{16 \alpha}{N(r_i - 1)^2}. \]

Combining this with (109), (115), and the trivial bounds \( \| \phi_{2N} - \phi_{1} \|_2 \leq 2N^{-1} (r_i - 1)^{-1} \) and \( (r_i - 1)^{-1} \leq (r_i - 1)^{-2} \), we get

\[ \| \hat{R}_i - L_i \|_2 \leq C \alpha N^{-1} (r_i - 1)^{-2}. \]

Since both \( \| \hat{R}_i \|_2 \) and \( \| L_i \|_2 \) are \( O(\alpha^{1/2} N^{-1/2} (r_i - 1)^{-1/2}) \), we have

\[ \| \hat{\varepsilon}_i^q \|_1 \leq \| \hat{R}_{i-1} - L_{i-1} \|_2 \| \hat{R}_{i-1} + L_{i-1} \|_2 \leq C \alpha^{3/2} N^{-3/2} (r_i - 1)^{-5/2}, \]

and from (29), we get \( \| T \hat{\varepsilon}_i^q \|_1 \leq C \alpha^{3/2} N^{-3/2} (r_i - 1)^{-7/2} \). Now appealing to (32) with \( \beta = 7/2 \),

\[ \sum_{1}^{N} \| T \hat{\varepsilon}_i^q \|_1 \leq C \alpha^{3/2} N^{-1/2} w_N^{-3/4} N^{1/2} = O(w_N^{-3/4}), \]

which establishes that \( \sum T \hat{\varepsilon}_i^q = O_P(1) \) and completes the proof of (54).

The bound for \( \sum R_i^2 \) follows from (112). Indeed,

\[ \sum_{i=1}^{N} \| \hat{R}_i \|_2 \leq \sum_{i=1}^{N} \| \hat{R}_i \|_2^2 \leq \sum_{i=1}^{N} \| \hat{R}_i \|_4^2 \leq \sum_{i=1}^{N} \frac{16 \alpha}{N(r_i - 1)} = O(1), \]

Therefore, \( \sum_{i=1}^{N} \hat{R}_i^2 = O_P(1) \) and so is the sum of \( R_i^2 \), which establishes (55).
We turn to the proof of (56) and (57). Using (26), then \( \gamma_{i:j+1}\gamma_j = \gamma_{i:j} \), and (28), we have

\[
\sum_{i=1}^{N} T \varepsilon_i^{qL} = \sum_{i=1}^{N} \sum_{j \leq i} \gamma_{i:j+1} \gamma_j L_{j-1} = \sum_{j=1}^{N-1} (g_{j+1} - 1)L_j^2
\]

\[
= \xi^T T^T GT \xi,
\]

with \( \xi = (\xi_1, \ldots, \xi_{N-1})^T \) and matrices \( G = \text{diag}(g_2 - 1, \ldots, g_N - 1) \), and \( T = T_{N-1} \).

The rescaled vector \( x = (\xi_1 / \sigma_1, \ldots, \xi_{N-1} / \sigma_{N-1})^T \), \( \sigma_i^2 = \mathbb{E} \varepsilon_i^2 \) has independent components with common variance. We use a variance bound for quadratic forms in such variables [5, Lemma 2.26], namely \( \text{Var} x^T Ax \leq \nu_4 \|A\|_{\text{HS}}^2 \), where in our case \( \nu_4 = \max \mathbb{E} x_i^4 \leq C_2/c_2 \) by Lemma 5 and \( A = DT^T GT D \), with \( D = \text{diag} (\sigma_i) \). We have

\[
\text{Var} [\sum_i T (\varepsilon_i^{qL} - \varepsilon_i^{qE})] = \text{Var} (\xi^T T^T GT \xi) = \text{Var} (x^T Ax) \lesssim \|A\|_{\text{HS}}^2.
\]

Again by Lemma 5, \( \max \sigma_i^2 \leq C_1 \alpha / N \), and so

\[
\|A\|_{\text{HS}} \leq C_1 \alpha N^{-1} \|T\|_{\text{op}} \|GT\|_{\text{HS}}.
\]

Decomposing \( T \) into a sum of sub-diagonal matrices, we have by the triangle inequality

\[
\|T\|_{\text{op}} \leq 1 + \gamma_{N-1} + \cdots + \gamma_{N-1} \cdots \gamma_2 \leq \frac{1}{1 - \gamma_{N-1}} = O(N^{1/3} w_N^{-1/2}).
\]

We also have,

\[
\|GT\|_{\text{HS}}^2 = \sum_{i=1}^{N-1} (g_{i+1} - 1)^2 (1 + \gamma_i^2 + \cdots + \gamma_i^2 \cdots \gamma_2^2) \leq \sum_{i=1}^{N-1} (g_{i+1} - 1)^2 \frac{1}{1 - \gamma_i^2}.
\]

By Lemma 6, for large enough \( N \), each \( g_i \geq 1 \) satisfies \( g_i - 1 < (r_i - 1)^{-1} \). Then from (105) and (32),

\[
\frac{1}{N} \|GT\|_{\text{HS}}^2 < \sum_{i=2}^{N} \frac{1}{N(r_i - 1)^2} (1 - \gamma_i^2) < \sum_{i=2}^{N} \frac{1}{N(r_i - 1)^3} = O(N^{1/3} w_N^{-1/2}).
\]

Summing up, we get

\[
\|A\|_{\text{HS}} \leq C_1 \alpha N^{-1} O(N^{1/3} w_N^{-1/2}) O(N^{2/3} w_N^{-1/4}) = O(\omega_N^{-3/4}),
\]

which suffices to establish (56).

It remains to show (57).

**Lemma 31.** There exists \( C > 0 \) such that for all \( 1 \leq i \leq N \),

\[
|\gamma_i \mathbf{E} L_i^2 - \alpha \delta_i| < \frac{C \alpha}{N^2 (r_i - 1)^3}.
\]

The lemma implies (57), for using (29), then (105) and (32), we have

\[
\sum_i T \varepsilon_i^{qE} - \alpha T \delta_i \leq \sum_{i=1}^{N} \frac{1}{1 - \gamma_i^2} \frac{C \alpha}{N^2 (r_i - 1)^4} \leq \frac{1}{N^2} \sum_{i=1}^{N} \frac{1}{(r_i - 1)^3} = O(\omega_N^{-3/2}).
\]

**Proof of Lemma 31.** Recall that \( L_i = \xi_i + \gamma_i L_{i-1} \). Since \( \xi_i \) and \( L_{i-1} \) are independent, \( \mathbf{E} L_i^2 \) satisfies the recursion

\[
(116) \quad \mathbf{E} L_i^2 = \mathbf{E} \xi_i^2 + \gamma_i^2 \mathbf{E} L_{i-1}^2, \quad i \geq 1.
\]
The idea is to use $E \xi_i^2 \approx (\alpha \Delta_N)(2/r_i^2)$ to show that an approximate solution is $EL_i^2 \approx s_i$, with

$$s_i = \frac{\alpha}{2 r_i(r_i-1)} \Delta_N,$$

and to use the equality, which we prove at the end of this section,

$$\delta_i = \gamma_i s_i / \alpha + O(N^{-2}(r_i - 1)^{-3}).$$

The other key ingredient is the identity

$$\frac{2}{r_i^3} = \frac{1}{2 r_i(r_i-1)} - \frac{\gamma_i^2}{2r_i(r_i-1)},$$

which, recalling (103), follows from $1 - \gamma^2 = 4R/r^2$ and $R = r - 1$.

In detail, from Lemma 7 we have

$$E \xi_i^2 = (\alpha \Delta_N)(2/r_i^2) + O(N^{-2}(r_i - 1)^{-1})$$

$$= s_i - \gamma_i^2 s_i + O(N^{-2}(r_i - 1)^{-1})$$

$$= s_i - \gamma_i^2 s_{i-1} + \eta_i, \quad |\eta_i| = O(N^{-2}(r_i - 1)^{-3}).$$

since the function $\rho(x) = 1/r(x)(r(x) - 1)$ with $r(x) = 1 + \sqrt{1-x}$, $x \in (0, 1)$ has $|\rho'(x)| \leq (r(x) - 1)^{-3}$, and so $0 \leq s_i - s_{i-1} \leq \Delta_N^2(r_i - 1)^{-3}$.

Putting this into (116), we obtain

$$EL_i^2 - s_i = \gamma_i^2(EL_{i-1}^2 - s_{i-1}) + \eta_i,$$

whose solution, from (30), is $EL_i^2 - s_i = (\tilde{T} \eta_i)_i$, where the linear operator $\tilde{T}$ is defined as $T$ but with $\gamma$ replaced by $\gamma^2$. Now bound (103) and an analogue of bound (29) imply that

$$|EL_i^2 - s_i| = |\tilde{T} \eta_i| \leq \frac{1}{\Gamma - \gamma_i^2} \max |\eta_i| = O(N^{-2}(r_i - 1)^{-4}).$$

To complete the proof, simply write

$$\alpha^{-1} \gamma_i(EL_{i-1}^2 - \delta_i = \alpha^{-1} \gamma_i(s_{i-1} - s_i) + O(N^{-2}(r_i - 1)^{-4}) = O(N^{-2}(r_i - 1)^{-4}).$$

It remains to establish (117). Similarly to (108), we have

$$\delta_i \geq m_i \Delta_N \frac{-r'}{r^2} (x_{i-1}) = \frac{\Delta_N}{2} \frac{m_i}{r_i^2(r_i-1) - 1}.$$

Since $f(r) = r^{-2}(r - 1)^{-1}$ is convex decreasing in $r \in (1, 2)$ with $d f / dr = -(3r - 2)/r^3(r-1)^2$, we have

$$\frac{1}{r_i^2(r_i - 1)} - \frac{1}{r_{i-1}^2(r_{i-1} - 1)} \leq \frac{3r_i - 2}{r_i^3(r_i - 1)^2} \frac{r_{i-1} - r_i}{r_{i-1}^2(r_{i-1} - 1)} \leq \frac{2\Delta_N}{(r_i - 1)^3},$$

where we have used (100). Hence, the difference between the upper and lower bounds in (108) and (118) is no larger than $\Delta_N^2/(r_i - 1)^3$ and

$$\delta_i = \frac{\Delta_N}{2} \frac{\gamma_i}{r_i(r_i - 1)} + O(N^{-2}(r_i - 1)^{-3}).$$
A.2.8. **Proof of Lemma 14 (asymptotics of \( \sum T\delta_i \)).** Resummation (27) gives \( \sum_{i=1}^{N} g_{i+1} \delta_i = \sum_{i=2}^{N} g_{i+1} \delta_i \). From (108), (118) and Lemma 6 we have bounds

\[
d(x_{i-1}) \Delta_N \leq \delta_i \leq d(x_i) \Delta_N, \quad \text{for } 2 \leq i \leq N,
\]

\[
g_i \leq g(x_i) (1 + O(w_N^{-3/2})), \quad \text{for } 2 \leq i \leq N,
\]

\[
g_i \geq g(x_i) (1 + O(w_N^{-3/2})), \quad \text{for } 2 \leq i \leq N - N^{1/3},
\]

along with the definitions

\[
d(x) = d = \frac{\gamma}{2r(r-1)}, \quad g(x) = g = \frac{r}{2(r-1)}, \quad r = 1 + \sqrt{1-x}.
\]

Combining these, and using monotonicity of \( \gamma \) and \( r \), we get

\[
(gd)(x_{i-1}) \Delta_N (1 + O(w_N^{-3/2})) \leq g_{i+1} \delta_i \leq (gd)(x_{i+1}) \Delta_N (1 + O(w_N^{-3/2}))
\]

for \( 2 \leq i \leq N - N^{1/3} \) in the lower bound and \( 2 \leq i \leq N \) in the upper.

Since \( 1 - \gamma = 2(r-1)/r \), we can decompose

\[
gd = \frac{\gamma}{4(r-1)^2} = \frac{1}{4(r-1)^2} - \frac{1}{2r(r-1)} = f_1 - f_2,
\]

say, and then observe that \( f_2(x) \propto (1-x)^{-1/2} \) is integrable for \( x \in [0,1] \), so that the sums \( \sum f_2(x_i) \Delta_N = O(1) \) can be ignored. On the other hand, the integral of \( f_1(x) \) is

\[
I(x_a, x_b) = \frac{1}{4} \int_{x_a}^{x_b} \frac{dx}{1-x} = \frac{1}{4} \log \left( \frac{1-x_a}{1-x_b} \right).
\]

Let \( x^{(0)} = a_0 \Delta_N, x^{(1)} = (N - [N^{1/3}] + a_1) \Delta_N, \) and \( x^{(2)} = (N + a_2) \Delta_N \) – the right choices of fixed small integers \( a_i \) legitimize bounds like (31), but make negligible contributions. Indeed, \( x^{(0)} = O(N^{-1}) \), and from (38),

\[
1 - x^{(1)} = (2w_N + 1)N^{-2/3}(1 + O(N^{-1/3})), \quad 1 - x^{(2)} = 2w_N N^{-2/3}(1 + O(N^{-1/3})),
\]

so that

\[
I(x^{(0)}, x^{(1)}) = \frac{1}{6} \log N + O(\log w_N).
\]

The remaining contribution is negligible: \( I(x^{(1)}, x^{(2)}) = O(w_N^{-1}) + O(N^{-1/3}) \), and so to finish

\[
\sum T\delta_i = \sum g_{i+1} \delta_i = [I(x^{(0)}, x^{(1)}) + O(1)][1 + O(w_N^{-3/2})] = \frac{1}{6} \log N + O(\log \log N).
\]

A.2.9. **Proof of Lemma 16 (a variance bound on linear spectral statistics of GUE).** With determinantal structure (such as GUE), we have [(48), (1.2)]

\[
R_k(x_1, \ldots, x_k) = \det(K_N(x_i, x_j))_{i,j=1,\ldots,k},
\]

where the kernel \( K_N(x, y) = \sum_{k=0}^{N-1} \phi_k(x) \phi_k(y) \) with \( \{\phi_k(x)\} \) obtained by orthonormalizing \( \{x^k e^{-N x^2/4}\} \). We then have in particular

\[
R_1(x) = K_N(x, x), \quad R_2(x, y) = R_1(x)R_2(y) - K_N^2(x, y).
\]
Furthermore, let \( J = \mathbb{E} \left[ N^{-1} \sum_{i=1}^{N} f(l_i) \right]^2 \). Expanding, we have

\[
J = N^{-2} \mathbb{E} \left[ \sum_{i=1}^{N} f^2(l_i) \right] + N^{-2} \mathbb{E} \left[ \sum_{i \neq j} f(l_i) f(l_j) \right] = N^{-1} \mathbb{E} f^2(l_i) + N^{-2} \times N(N-1) \mathbb{E} \left[ f(l_i) f(l_j) \right].
\]

Now apply (59) and then (120) to get

\[
J = N^{-2} \int f^2(x) R_1(x) dx + N^{-2} \int \int f(x) f(y) R_2(x, y) dxdy
\]

\[
- N^{-1} \int f^2(x) \rho_N(x) dx + \left[ \int f(x) \rho_N(x) dx \right]^2 - N^{-2} \int \int f(x) f(y) K^2_N(x, y) dxdy.
\]

Dropping the last term and recalling (60), we obtain (62) which establishes lemma 16.

**A.3. Discussion of edge bounds for one-point functions (63).** Let \( H_N(x) = e^{x^2} \left( -\frac{d}{dx} \right)^N e^{-x^2} \) be the Hermite polynomials. The corresponding orthonormal Hermite functions,

\[
\varphi_N(x) = c_N e^{-x^2/2} H_N(x), \quad c_N = (2^N N! \sqrt{\pi})^{-1/2},
\]

are even/odd as \( N \) is even/odd. Consequently \( I_N = \int \varphi_N(x) dx \) vanishes for \( N \) odd, and from a calculation with generating functions, or [24] 7.373.2,

\[
\sigma_{2m} = \sum_{m=0}^{N-1} \varphi_m^2(x),
\]

\[
\sigma_{2m+1} = \sum_{m=0}^{N-1} \varphi_m \varphi_{m+1},
\]

\[
\sigma_{2m+2} = \sigma_{2m} + (N/2)^{1/2} \varphi_{N-1}(x) \varphi_N(x) + I_{N-1}^{-1} \varphi_{N-1}(x) \chi_{N-1}^\delta.
\]

Note these forms have total mass \( N \). To recover the forms of interest to us, on scale \([-2, 2] \) with total mass 1, we use

\[
\rho_{N, \alpha}(y) = \frac{1}{\sqrt{2N}} \sigma_{N, \alpha}(\sqrt{N/2} y).
\]

Following [47], introduce \( \varphi = (N/2)^{1/4} \varphi_N \) and \( \psi = (N/2)^{1/4} \varphi_N \). We have a useful integral representation (see e.g. equation (57) in [47])

\[
\sigma_{N,1}(x) = 2 \int_0^\infty \varphi(x+z) \psi(x+z) dz.
\]

Further, observe from (121) that both \( I_\varphi = \int \varphi \) and \( I_\psi = \int \psi \) converge to \( \sqrt{2} \) for large even and odd values of \( N \) respectively. We get

\[
\sigma_{N,2}(x) - \sigma_{N,1}(x) = \psi(x) \left[ I_\varphi - 2 I_\varphi - I^{-1}_\psi \chi_{N-1}^\delta \right].
\]
Now we turn to bounds for scaled Hermite functions near the bulk edge. Set $\tau_N = N^{-1/6}/\sqrt{2}$ and define

$$\varphi_\tau(s) = \varphi(x), \quad \psi_\tau(s) = \psi(x), \quad x = \sqrt{2N} + s\tau_N$$

The following bounds are essentially established in [41, p. 403] and [28]

**Proposition 32.** Fix $0 < \varepsilon < 2/3$. Then for large $N$, uniformly in the indicated ranges

$$\tau_N \varphi_\tau(s), \quad \tau_N \psi_\tau(s) = \begin{cases} O(e^{-s}) & s \geq 0 \\ O\left(\frac{1 + |s|}{N^{1/4}}\right) & -N^{2/3-\varepsilon} < s \leq 0. \end{cases}$$

We discuss the proof below. Taking it as given for now, observe then that

$$\sigma_{N,1}(x) = 2\tau_N \int_s^\infty \varphi_\tau(y)\psi_\tau(y) dy.$$ 

Combining this with Proposition 32, we obtain

$$\sigma_{N,1}(x) = \begin{cases} O\left(\frac{\sqrt{N}}{e^{2s}}\right) & s \geq 0 \\ O\left(\frac{1 + \sqrt{s}}{N^{1/2}}\right) & -N^{2/3-\varepsilon} < s \leq 0. \end{cases}$$

From (122), we have $\rho_{N,\alpha}(2 + sN^{-2/3}) = (2N)^{-1/2}O_N\left(\sqrt{2N} + s\tau_N\right)$ and since $(2N)^{-1/2}\tau_N^{-1} = N^{-1/3}$, the claim (63) for $\alpha = 1$ follows from this.

For $\alpha = 2$, observe that

$$\int_x^\infty \varphi(x') dx' = \int_s^\infty \tau_N \varphi_\tau(s') ds' = \begin{cases} O\left(e^{-s}\right) & s \geq 0 \\ O\left(\frac{1 + |s|}{N^{1/4}}\right) & -N^{2/3-\varepsilon} < s \leq 0. \end{cases}$$

Combining this with Proposition 32 applied to $\psi_\tau(s)$ in (124), we arrive at

$$(\sigma_{N,2} - \sigma_{N,1})(x) = -\psi_\tau(s) \left[ \int_s^\infty \tau_N \varphi_\tau + O(1) \right] = \begin{cases} O\left(\frac{\sqrt{N}}{e^{2s}}\right) & s \geq 0 \\ O\left(\frac{1 + |s|}{N^{1/2}}\right) & -N^{2/3-\varepsilon} < s \leq 0. \end{cases}$$

This implies that (63) holds for $\alpha = 2$ as well.

**Discussion of proof of Proposition 32.** Proposition 32 is based on analysis of the second order differential equation satisfied by Hermite functions $\varphi_N$ using the Liouville-Green transform around the turning point at the upper edge. This is detailed in [16] (attributed to Skovgaard), and given as an example of Theorem 11.3.1. in [41, Ex 4.2, 4.3 p 403]. This example was also worked out in detail (for another purpose) in [28], JM12 below. Although the focus there was on $s > -c$, we indicate how the analysis also extends to much larger ranges of negative $s$. We focus here on the bound for $\varphi_N$; for $\varphi_{N-1}$ it is essentially the same, see JM12.

Rescaling the $x$-axis via $x = \sqrt{2N} + 1,\xi$, and setting $w_N(\xi) = \varphi_N(x)$, the Liouville-Green transform introduces new independent and dependent variables $\zeta$ and $W = \xi^{1/2}w_N$. The transform $W$ approximately satisfies the Airy equation $W''(\zeta) = \kappa_N^2 W(\zeta)$ with $\kappa_N = 2N + 1$, and it is shown that $\varphi_N$ is approximated by the (recessive) solution $Ai(\kappa_N^{2/3}\zeta)$ with explicit error bounds. Indeed, cf [28, (71)], with $r(\xi) = [\hat{\xi}(\xi)/\hat{\zeta}(1)]^{-1/2}$,

$$\tau_N \varphi(x) = (N/2)^{1/4}\tau_N \varphi_N(x) = \sqrt{2}r(\xi)\{Ai(\kappa_N^{2/3}\zeta) + e_2(\xi, \kappa_N)\}.$$ (125)
The function $\zeta(\xi)$ is increasing and $C^2$ on $(0, \infty)$ [41, p 391], with $\zeta(\xi)$ non-negative and bounded. The arguments leading to (78) and (85) in JM12 show that for $|s| \geq N^{2/3-\epsilon}$,

\begin{equation}
(126) \quad r(\xi) \leq 1 + O(N^{-\epsilon}), \quad \kappa^{2/3} = s(1 + O(N^{-\epsilon})).
\end{equation}

To describe error bound even in the oscillatory region of $\text{Ai}$, [41] introduces continuous and positive functions $E \geq 1$ and $M \leq 1$ such that $|\text{Ai}(x)| \leq M(x)/E(x)$ and satisfying

\begin{equation}
(127) \quad E(x) \sim \sqrt{2} e^{(2/3)x^{3/2}}, \quad M(x) \sim \pi^{-1/2} (1 + |x|)^{-1/4},
\end{equation}

the former as $x \to +\infty$, the latter as $|x| \to \infty$. In the Hermite case it follows from [41, p. 403] that $|e_2(\xi, \kappa_N)| \leq N^{-1}(M/E)(\kappa_N^{2/3} \zeta)$. From (125) and boundedness of $r(\xi)$, it follows that

\[ \tau_N |\varphi(x)| \leq C(M/E)(\kappa_N^{2/3} \zeta) \leq \begin{cases} C E^{-1}(\kappa_N^{2/3} \zeta) \leq Ce^{-2s} & s > 0 \\ CM(\kappa_N^{2/3} \zeta) \leq C(1 + |s|)^{-1/4} & -N^{2/3-\epsilon} < s \leq 0, \end{cases} \]

where the first bound follows from (127) and JM12, Lemma 2 and the second from (126) and (127).

A.4. Proof of Theorem 19 (linking functions of GUE and GOE). The main engine of this result is an identity stated in [21], which relates the eigenvalues of a GUE to the eigenvalues of two independent GOEs. In particular, we use it in the following lemma.

**Lemma 33.** Let $M_N^C$ be an $N \times N$ GUE, and let $f$ be a function of bounded variation with total variation $TV(f)$. If $M_N^R$, $\tilde{M}_N^R$ are two independent GOEs, then

\begin{equation}
(128) \quad f(M_N^C) \overset{d}{=} \frac{1}{2} \left( f(M_N^R) + f(\tilde{M}_N^R) \right) + X_N,
\end{equation}

where $|X_N| \leq TV(f)$, and $\overset{d}{=}$ denotes equality in distribution.

**Proof.** Let $M_N^R$, $\tilde{M}_{N+1}^R$ be independent $N \times N$ and $(N+1) \times (N+1)$ GOEs. Call the eigenvalues of $M_N^R$ and $\tilde{M}_{N+1}^R$ $\{\lambda_i\}_{i=1}^N$ and $\{\lambda_i\}_{i=1}^{N+1}$, respectively. Further, denote the combined set of eigenvalues $\{\lambda_i\}_{i=1}^N \cup \{\lambda_i\}_{i=1}^{N+1}$ as $\Lambda^+$, and enumerate its elements in decreasing order

\[ \Lambda^+ = \{\lambda_1^+ \geq \ldots \geq \lambda_2^+ \} \]

Theorem 5.2 of [21] implies that the even elements of this set are equal in distribution to the eigenvalues of an $N \times N$ GUE.

Thus, if $M_N^C$ is an $N \times N$ GUE, we have

\[ f(M_N^C) \overset{d}{=} \sum_{i=1}^N f(\lambda_{2i}^+) \]

\[ = \frac{1}{2} \left( \sum_{j=1}^{2N+1} f(\lambda_j^+) + \sum_{i=1}^N [f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+)] - f(\lambda_{2N+1}^+) \right) \]

\[ = \frac{1}{2} \left( f(W_N^R) + f(\tilde{W}_{N+1}^R) - f(\lambda_{2N+1}^+) + \sum_{i=1}^N [f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+)] \right). \]
Notice that, since \( \lambda_j^+ \) are ordered, we have

\[
\left| \sum_{i=1}^{N} \left[ f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+) \right] \right| \leq \text{TV}(f).
\]

Further, let \( \tilde{M}_N^{R} \) be the principal submatrix of \( \tilde{M}_{N+1}^{R} \), which is thus independent and equal in distribution to \( M_N^{R} \). If we let \( \tilde{\mu}_1, \ldots, \tilde{\mu}_N \) be the eigenvalues of \( \tilde{M}_N^{R} \), then Cauchy’s interlacing theorem yields

\[
\tilde{\lambda}_1 \geq \tilde{\mu}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_N \geq \tilde{\mu}_N \geq \tilde{\lambda}_{N+1},
\]

and so we have

\[
\left| f(M_{N+1}^{R}) - f(\lambda_{2N+1}^+) - f(M_N^{R}) \right| = \left| \sum_{i=1}^{N} f(\tilde{\lambda}_i) - \sum_{i=1}^{N} f(\tilde{\mu}_i) + (f(\tilde{\lambda}_{N+1}) - f(\lambda_{2N+1}^+)) \right|
\leq \sum_{i=1}^{N} |f(\tilde{\lambda}_i) - f(\tilde{\mu}_i)| + |f(\tilde{\lambda}_{N+1}) - f(\lambda_{2N+1}^+)|
\leq \text{TV}(f).
\]

We conclude that (128) holds.

An immediate useful corollary is as follows.

**Corollary 34.** Under assumptions of Lemma 33,

\[
\mathbf{E} f(M_{N}^{R}) = \mathbf{E} f(M_{N}^{C}) + O(\text{TV}(f)),
\]

\[
\text{Var} f(M_{N}^{R}) \leq \text{Var} f(M_{N}^{C}) + 2\text{TV}^2(f).
\]

**Remark.** Notice that corollary 34 also holds for scaled Gaussian matrices \( W_N^{R/C} = M_N^{R/C}/\sqrt{N} \), since \( f(W_N^{R/C}) = g(M_N^{R/C}) \) for \( g(\lambda) = f(\lambda/\sqrt{N}) \), which satisfy \( \text{TV}(f) = \text{TV}(g) \).

However, to finish proving Theorem 19 in its generality, we require the following technical lemma about tightness.

**Lemma 35.** Let \( X_N, Y_N \) be iid sequences of random variables such that \( X_N + Y_N \) is tight. Then \( X_N \) (and thus also \( Y_N \)) is tight.

**Proof.** For any constant \( K \), we have

\[
\mathbf{P}(X_N > K) = \mathbf{P}(X_N > K, Y_N > K)^{1/2} \leq \mathbf{P}(|X_N + Y_N| > K)^{1/2},
\]

and similarly,

\[
\mathbf{P}(X_N < -K) \leq \mathbf{P}(|X_N + Y_N| > K)^{1/2},
\]

which yield

\[
\sup_N \mathbf{P}(|X_N| > K) \leq 2 \sup_N \mathbf{P}(|X_N + Y_N| > K)^{1/2}.
\]

The right hand side of the latter inequality can be made arbitrarily small, by the tightness of \( X_N + Y_N \). \( \square \)
With all these results in hand, we are ready to complete the proof of Theorem 19. We have

\[
\left| \frac{f_N(W_N^b) - a_N}{b_N + TV(f_N)} + \frac{f_N(W_N^b) - a_N}{b_N + TV(f_N)} \right| = 2 \left| \frac{(f_N(W_N^b) + f_N(W_N^b))/2 - a_N}{b_N + TV(f_N)} \right| \\
\leq 2 \left| \frac{(f_N(W_N^b) + f_N(W_N^b))/2 + X_N - a_N}{b_N} \right| + 2 \left| \frac{X_N}{TV(f_N)} \right|.
\]

The first term in the latter sum is tight by assumption, whereas the second term is no larger than 2. But since the two terms on the left hand side of (129) are iid, Lemma 35 yields that they must be tight, and so

\[
f_N(W_N^b) = a_N + O_P(b_N + TV(f_N)).
\]

APPENDIX B: PROOFS FOR WIGNER EXTENSION AND SECTION 6

B.1. Proof of Proposition 21 (about convergence of \( \xi_N \) implying convergence of \( \xi'_N \)).

Let us first show that \( \xi_{Nj}(W_N') \overset{d}{\to} \xi_j \) marginally for each \( j \). Fix some \( \epsilon > 0 \). Then, for large enough \( N \),

\[
\mathbb{P}(\xi_{Nj}(W_N') \leq s) \leq \mathbb{P}(\xi_{Nj}(W_N') \leq s + \epsilon - \eta_N) \leq \mathbb{E}Q_j(W_N', s + \epsilon) + O(N^{-A}),
\]

Similarly, for \( N \) large

\[
\mathbb{E}Q_j(W_N', s + \epsilon) \leq \mathbb{P}(\xi_{Nj}(W_N') \leq s + 2\epsilon + O(N^{-A}) \leq \mathbb{P}(\xi_j \leq s + 2\epsilon) + O_s(\epsilon),
\]

where the last inequality follows from the convergence \( \xi_{Nj}(W_N') \overset{d}{\to} \xi_j \). Since \( Q_j(\cdot, s + \epsilon) \) satisfies condition \( F(\delta_{j,N}) \), we have by Proposition 20,

\[
\mathbb{E}Q_j(W_N', s + \epsilon) \leq \mathbb{E}Q_j(W_N, s + \epsilon) + O_s(\epsilon).
\]

We therefore obtain for \( N \) large,

\[
\mathbb{P}(\xi_{Nj}(W_N') \leq s) \leq \mathbb{P}(\xi_j \leq s + 2\epsilon) + O_s(\epsilon).
\]

Similarly, we can obtain a lower bound

\[
\mathbb{P}(\xi_{Nj}(W_N') \leq s) \geq \mathbb{P}(\xi_j \leq s - 2\epsilon) + O_s(\epsilon).
\]

Since \( \epsilon \) can be chosen arbitrarily small and \( \xi_j \) has continuous distribution, it follows that \( \xi_{Nj}(W_N') \overset{d}{\to} \xi_j \).

Now let \( \delta_N = \max_j \delta_{j,N} \). It suffices to show that for each \( s = (s_j) \)

\[
\mathbb{P}(\xi_N \leq s - \eta_N) \leq \mathbb{P}(\xi'_N \leq s + \eta_N) + O(\delta_N) \quad \text{and}
\]

\[
\mathbb{P}(\xi_N \leq s - \eta_N) \leq \mathbb{P}(\xi_N \leq s + \eta_N) + O(\delta_N).
\]

Indeed, we then have

\[
|\mathbb{P}(\xi'_N \leq s) - \mathbb{P}(\xi_N \leq s)| \leq \sum_j \mathbb{P}(|\xi_{Nj} - s_j| \leq \eta_N) + \sum_j \mathbb{P}(|\xi'_N - s_j| \leq \eta_N) + O(\delta_N) \to 0
\]

because each \( \xi_{Nj}, \xi'_{Nj} \) has a continuous limiting distribution function.
We verify inequality (130). For each \( A > 0 \) large, we have from (72) for \( W_N \), then Proposition 20 and then (72) again, now for \( W_N' \), that
\[
\mathbb{P}(\xi_N \leq s - \eta_N) \leq \mathbb{E} \prod_j Q_j(W_N, s_j) + O(N^{-A}) \\
\leq \mathbb{E} \prod_j Q_j(W_N', s_j) + O(\delta_N) \leq \mathbb{P}(\xi_N' \leq s + \eta_N) + O(\delta_N).
\]
Inequality (131) follows similarly.

B.2. Proof of Lemma 26 (bounds on inverse power sums for Wigner matrices). Let \( \delta = N^{-2/3-\epsilon} \) and \( A_N = \{\min_j |E - \lambda_j| > \delta\} \): by Proposition 25 this event has probability at least \( 1 - N^{-\epsilon/2} \). We will work on event \( A_N \), and show that there the claims hold w.o.p. On \( A_N \) the interval \( I_0 = [E - \delta, E + \delta] \) contains no eigenvalues. Consider the ‘coronae’ defined by \( I_k = \{x \in \mathbb{R} : 2^{k-1}\delta < |x - E| \leq 2^k\delta\} \) for \( 1 \leq k \leq k' = \min\{k : E - 2^k\delta \leq 1\} \), and add two half-infinite intervals \( I_{-1} \) and \( I_{k'+1} \) to obtain a disjoint cover of \( \mathbb{R} \). We may then bound (on event \( A_N \))
\[
S_r(E) \leq \sum_{k=1}^{k'} N_{W_N}(I_k) + \frac{N}{(2^k\delta)^r}.
\]
(132)
The semicircle density is bounded by \( \sqrt{2 - x^2} 1_{x \leq 2} \) and so \( \rho_{sc}([2 - a, 2 + b]) \leq a^{3/2} \). The lower endpoint of \( I_k \) is \( E - 2^k\delta \geq 2 - 2^k\delta - \sigma_N N^{-2/3} \). Since \( \sigma_N^{3/2} \leq N^\epsilon \) for large \( N \),
\[
\rho_{sc}(I_k) \leq \sqrt{2} \left( (2^k\delta)^{3/2} + N^{\epsilon-1} \right).
\]
Proposition 23 (ii) says that, with overwhelming probability, simultaneously for all \( k \leq k' = O(\log N) \), we have
\[
N_{W_N}(I_k) \leq N \rho_{sc}(I_k) + O(N^\epsilon) \leq \sqrt{2}N(2^k\delta)^{3/2} + CN^\epsilon.
\]
Putting this into (132) and noting that \( 2^k\delta \in [\frac{1}{2}, 3] \) we obtain w.o.p.
\[
S_r(E) \leq 2r^{+1/2}N \sum_{k=1}^{k'} (2^k\delta)^{3/2-r} + CN^\epsilon \delta^{-r} + 2^r N.
\]
The sum may be bounded using
\[
N \delta^{3/2-r} \sum_{k=1}^{k'} 2^{3/2-r}k^k \leq \begin{cases} 2^{1/2}N & \text{if } r = 1 \\ 4N^{-3\epsilon/2}\delta^{-r} & \text{if } r \geq 2. \end{cases}
\]
Observe that \( N^\epsilon \delta^{-r} = N^{(2/3+\epsilon)r+\epsilon} \). For \( r = 1 \), this is \( o(N^{-1}) \) and so \( S_1(E) \leq C_1 N \) on \( A_N \) w.o.p. For \( r \geq 2 \) this is the dominant term, so that \( S_r(E) \leq C_r N^{2r/3+(r+1)} \).

The bounds also hold for \( S_r(E') \) uniformly in \( |E' - E| \leq \delta/2 \), by increasing \( C_r \) to \( 2^r C_r \). Indeed, for such \( E' \), on event \( A_N \) we have \( |\lambda_j - E'| \geq \frac{1}{2} |\lambda_j - E| \) for all \( j \).

B.3. Proof of proposition 28 (isotropic local law and delocalization). We will modify the proofs of Theorems 2.2 and 2.5 of [30] to reach two main goals: replace their sub-exponential assumption on the entries of Wigner matrices by the uniform moment bound (68), and allow for an arbitrary variance profile along the Wigner diagonal. We are going to reformulate \( \zeta \)-high probability bounds used in [30] in terms of weaker polynomial bounds.
For this, it will be convenient to use the concept of stochastic domination, as defined in [10, def. 2.5]:

**Definition of stochastic domination.** Let

\[ X = \left( X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)} \right), \quad Y = \left( Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)} \right) \]

be two families of non-negative random variables, where \( U^{(N)} \) is a possibly \( N \)-dependent parameter set. We say that \( X \) is stochastically dominated by \( Y \), uniformly in \( u \), and write \( X \prec Y \), if for all (small) \( \epsilon > 0 \) and (large) \( A > 0 \) we have

\[ \sup_{u \in U^{(N)}} P \left[ X^{(N)}(u) > N^\epsilon Y^{(N)}(u) \right] \leq N^{-A} \]

for large enough \( N \geq N_0(\epsilon, A) \). If for some complex family \( X \) we have \( |X| \prec Y \), we also write \( X = O_{\prec}(Y) \).

With this definition, we are ready to point out necessary changes to the proofs of Theorems 2.2 and 2.5 of [30]. We use mostly the same notation, and refer the reader to [30] for definitions. For example, the resolvent matrix will be denoted as \( G(z) \) instead of \( R(z) \) as in the main body of our paper. We use numbering (KY3.xx) for formula (3.xxa) for a formula here which is a stochastic dominance analog of (KY3.xx).

**Section KY3.** To accommodate our setting, proposition KY3.1 should be reformulated as follows.

**Proposition 36.** Fix any \( \tau > 0, \ 0 < \epsilon < \tau, \) and \( n > 0 \). Then under assumptions of proposition 28, for all deterministic normalized \( v, w \in \mathbb{C}^N \) and all \( z \in S(\tau) \),

\[ (133) \quad E |G_{vw}(z) - s_{sc}(z)v^*w|^n \leq (N^\epsilon \Psi(z))^n \]

for all sufficiently large \( N \geq N_0(\tau, \epsilon, n) \).

The above proposition implies eq. (92). Indeed, let \( A \) be large and take \( n = A/\epsilon \). Then by Markov’s inequality

\[ P(|G_{vw}(z) - s_{sc}(z)v^*w| \geq N^{2\epsilon} \Psi(z)) \leq \frac{E |G_{vw}(z) - s_{sc}(z)v^*w|^n}{N^{\epsilon n}(N^\epsilon \Psi(z))^n} \leq N^{-cn} = N^{-A} \]

for sufficiently large \( N \), which in our notations means that

\[ v^*R(z)w = s_{sc}(z)v^*w + O(N^\epsilon \Psi(z)) \]

w.o.p., as required.

**Subsection KY3.1.** Replace Lemma KY3.5 by the following one.

**Lemma 37.** Let \( a_1, \ldots, a_N \) be independent random variables with zero mean and all moments bounded uniformly in \( N \). Then for any deterministic complex numbers \( A_i \), we have

\[ \left| \sum_{i=1}^{N} A_i a_i \right| < \left( \sum_{i=1}^{N} |A_i|^2 \right)^{1/2} . \]

A proof of this lemma is almost identical to the proof of Lemma 8.2 in [18], so we omit it. Further, replace Theorems KY3.6 and KY3.7 by Theorems 2.6 and 2.9 (respectively) from [10].

**Subsection KY3.2.** We need to reformulate the statements formulated in terms of \( \zeta \)-high probability using the notion of stochastic domination. In particular, lemma KY3.8 should be reformulated as follows.
**Lemma 38.** Fix $\tau > 0$. Then,

$$|G_{vi}(z)| + |G_{iv}(z)| + |G_{vi}(z)| + |G_{iv}(z)| \prec \sqrt{\frac{\text{Im}G_{vv}(z)}{N\eta}} + |v_i|$$

for all $z \in S(\tau)$.

In [30]'s proof of the lemma, replace the first display by

$$|G_{vi}(z)| \prec \left(\frac{1}{N} \sum_{k} |G_{vk}^{(i)}|^2 \right)^{1/2},$$

which holds by lemma 37. Further, change inequality (KY3.19) to $|G_{ii}| \prec 1$ (which follows from Theorem 2.6 of [10], and then, in all of the remaining displayed inequalities change $\leq$ to $\prec$. Note that so far, we have not used any information about the diagonal variance profile. This information will be used in the next subsection.

**Subsections KY3.3-KY3.4** To manage the modification of the proof to cover stochastic dominance and weaker conditions on diagonal moments, it is convenient to somewhat reorganize the material in Sections KY3.3 and 3.4, along with ideas from Case 1 of KY4.1. At the cost of some duplication of text from KY, we thus write out this part of the proof in relatively self-contained form.

Given a resolvent matrix $G(z)$ as in KY Th 2.2 we consider three cases of linear functionals $L_\nu G(z)$ and corresponding control functions $\Pi(z)$:

$$L_\nu G = \begin{cases} 
\text{Re} G_{vv} - \text{Re} s_{sc} & \Psi \\
\text{Im} G_{vv} - \text{Im} s_{sc} & \Phi \\
\text{Im} G_{vv} & \Pi 
\end{cases} \quad \Pi = \begin{cases} 
\Psi \\
\Phi.
\end{cases}$$

Fix $\tau > 0$, $0 < \epsilon < \tau$ and even $n \geq 2$. We seek to prove inequalities\(^2\)

\[(3.20a) \quad E(L_\nu G(z))^n \leq (N^{\epsilon/2} \Pi(z))^n,\]

for all $z \in S(\tau)$. The first two choices for $L_\nu G$ together yield Proposition 36, while the third is needed for an intermediate step in the proof.

We verify that (3.20a) holds when $H_0$ (matrix $W_N$ in the notations of previous sections) is a GOE/GUE matrix. In that case $E(L_\nu G(z))^n = E(L_{e_1} G(z))^n$ by unitary invariance. From the entrywise local law and (KY3.4), we have for $z \in S(\tau)$ that

$$\text{Im} G_{11}(z) \prec \Phi(z), \quad |G_{11}(z) - s_{sc}(z)| \prec \Psi(z),$$

so that $L_\nu G(z) \prec \Pi(z)$. Since $\Psi(z) \gtrsim N^{-1/2}$ and $E(L_\nu G(z))^n \leq N^p(n)$ from the rough bound $|G_{11}(z)| \leq \eta^{-1} \leq N$, [10, Lemma 7.1] implies that $E(L_\nu G(z))^n \prec \Pi(z)^n$, which yields (3.20a) for GOE/GUE.

“From now on we work on the product space generated by the Wigner matrix $H = (N^{-1/2} W_{ij})_{i,j}$ and the GOE/GUE matrix $(N^{-1/2} V_{ij})_{i,j}$. We fix a bijective ordering map on the index set of the independent matrix elements,

\[(KY3.21) \quad \varphi : \{(i,j) : 1 \leq i \leq j \leq N\} \rightarrow \{1, \ldots, \gamma_{\max}\} \quad \text{where} \quad \gamma_{\max} := \frac{N(N+1)}{2},\]

\(^2\)For $L_\nu G = \text{Im} G_{vv}$, Lemma KY3.9 uses $\Phi$, but the bound with $\Psi$ is better and allows a more uniform treatment.
and denote by \( H_\gamma = (h_{ij}^\gamma), \gamma = 0, \ldots, \gamma_{\text{max}}, \) the Wigner matrix with upper-triangular entries defined by

\[
h_{ij}^\gamma = \begin{cases} 
N^{-1/2}W_{ij} & \text{if } \varphi(i,j) \leq \gamma, \\
N^{-1/2}V_{ij} & \text{otherwise}.
\end{cases}
\]

In particular, \( H_0 \) is a GOE/GUE matrix and \( H_{\gamma_{\text{max}}} = H. \)

Let \( E^{(ij)} \) denote the matrix whose matrix elements are given by \( E_{kl}^{(ij)} = \delta_{ik}\delta_{jl}. \) Fix \( \gamma \geq 1 \) and let \( (a,b) \) be determined by \( \varphi(a,b) = \gamma. \) We shall compare \( H_{\gamma-1} \) with \( H_\gamma \) for each \( \gamma \) and then sum up the differences. Note that the matrices \( H_{\gamma-1} \) and \( H_\gamma \) differ only in the entries \( (a,b) \) and \( (b,a) \) and they can be written as

\[
\text{(KY3.22)} \quad H_{\gamma-1} = Q + N^{-1/2}V \quad \text{where} \quad V := V_{ab} E^{(ab)} + 1(a \neq b)V_{ba} E^{(ba)},
\]

and

\[
\text{(KY3.23)} \quad H_\gamma = Q + N^{-1/2}W \quad \text{where} \quad W := W_{ab} E^{(ab)} + 1(a \neq b)W_{ba} E^{(ba)},
\]

here the matrix \( Q \) satisfies \( Q_{ab} = Q_{ba} = 0. \)

Next, we introduce the Green functions

\[
\text{(KY3.24a)} \quad E(L_\nu G^{\gamma_{\text{max}}} - E(L_\nu G^0)^n = \sum_{\gamma=1}^{\gamma_{\text{max}}} (X_\gamma - X_{\gamma-1}),
\]

where, since \( n \) is even, \( X_\gamma = E(L_\nu G^\gamma)^n \geq 0. \) Note that in the \( R,S,T \) notation, \( X_\gamma = E(L_\nu T)^n \) and \( X_{\gamma-1} = E(L_\nu S)^n. \)

For any \( K \in \mathbb{N} \) we have the resolvent expansions

\[
\text{(KY3.25)} \quad S = R + \sum_{k=1}^{K-1} N^{-k/2}(-RV)^k R + N^{-K/2}(-RV)^K S,
\]

\[
\text{(KY3.26)} \quad R = S + \sum_{k=1}^{K-1} N^{-k/2}(SV)^k S + N^{-K/2}(SV)^K R.
\]

With \( K = 4 \) in (KY3.26), using the entrywise local law for the Wigner matrix \( S, \) and the rough bound \( \| R \| \leq \eta^{-1} \) to estimate the remainder term in (KY3.26), and recalling \( (68) \) instead of (KY2.1), we find

\[
\text{(KY3.27a)} \quad |R_{ij} - \delta_{ij}s_{sc}| \prec |S_{ij} - \delta_{ij}s_{sc}| + N^{-1/2} \prec \Psi. \]

There are trivial changes in (KY3.28) - (KY3.30), for later use we record

\[
\text{(KY3.29a)} \quad |R_{\nu a}| \prec \sqrt{\frac{\text{Im}S_{\nu\nu}}{N\eta}} + \Psi + |v_a|
\]

\[
\text{(KY3.30a)} \quad |S_{\nu\nu} - R_{\nu\nu}| \prec N^{-1/2}\left( \frac{\text{Im}S_{\nu\nu}}{N\eta} + |v_a|^2 + |v_b|^2 \right).
\]
We now apply (KY3.25) with $K = 4$ and introduce the notation $S - R = \sum_{k=1}^{4} Y_k$, whereby $Y_k$ has $k$ factors $V$. Then $L_vS - L_vR = \sum_{k=1}^{4} \hat{L}_v Y_k$, where $\hat{L}_v Y$ is either $\text{Re} Y_{\nu\nu}$ or $\text{Im} Y_{\nu\nu}$, since the terms involving $s_{\nu\nu}(z)$, if present, cancel. We expand the difference

\begin{equation}
(L_vS)^n - (L_vR)^n = \sum_{m=1}^{n} \binom{n}{m} (L_vS - L_vR)^m (L_vR)^{n-m} = \sum_{m=1}^{n} (L_vR)^{n-m} \sum_{k=m}^{4m} A_{m,k}
\end{equation}

where

$$A_{m,k} = \binom{n}{m} \sum_{k_1, \ldots, k_m = 1}^{4} 1(k_1 + \cdots + k_m = k) \prod_{i=1}^{m} \hat{L}_v Y_{k_i}.$$ 

Thus $A_{m,k}$ collects all terms with $k$ factors $V_{ab}$ or $V_{ba}$, and hence is of order $N^{-k/2}$. Break the sum in (134) in two so that $k \leq k_\gamma - 1$ and $k \geq k_\gamma$. The value $k_\gamma$ is chosen so that the first $k_\gamma - 1$ moments of $V_{ab}$ and $W_{ab}$ are the same, meaning that for any $t_1, t_2 \in \mathbb{N}$ s.t. $t_1 + t_2 < k_\gamma$, we have $\mathbb{E} \left( V_{ab}^{t_1} V_{ab}^{t_2} \right) = \mathbb{E} \left( W_{ab}^{t_1} W_{ab}^{t_2} \right)$. Thus we take $k_\gamma = 4$ for $a \neq b$ and $k_\gamma = 2$ when $a = b$. We obtain

\begin{equation}
\text{(KY3.32)} \quad (L_vS)^n - (L_vR)^n = \sum_{m=1}^{n} A_{m,k} + \sum_{m=1}^{n} A'_{m,k} = A_{\gamma} + A'_{\gamma},
\end{equation}

where, for example,

$$A'_{m,k} = (L_vR)^{n-m} \sum_{k=k_\gamma, k_\gamma m} A_{m,k}.$$ 

Thus $\mathbb{E} A_{\gamma}$ depends on the randomness only through $Q$ and the first $k_\gamma - 1$ moments of $V_{ab}$. Consequently $\mathbb{E} A_{\gamma}$ equals the corresponding term in the expansion (KY3.32) of $\mathbb{E} (L_vT)^n - \mathbb{E} (L_vR)^n$.

For the higher order terms, the analog of the key inequality proved by KY has the form

\begin{equation}
\text{(3.33a)} \quad |\mathbb{E} A'_{\gamma}| \leq \frac{\mathcal{E}_{ab}}{N^{\gamma/2}} \left[ \mathbb{E} (L_vS)^n + (N^\gamma \Pi)^n \right].
\end{equation}

The factor $\mathcal{E}_{ab} = \mathcal{E}(v_a, v_b, N)$ will be detailed below; for now we simply need that $\varepsilon_{\gamma} := N^{-\gamma/2} \mathcal{E}_{ab} \geq 0$ satisfies $\sum_{\gamma} \varepsilon_{\gamma} \leq \frac{1}{2}$.

Before proving (3.33a), we show how it implies (3.20a). Repeating the derivation of (KY3.32) for $T$ instead of $S$, using that the first $k_\gamma - 1$ moments of $V_{ab}$ and $W_{ab}$ are the same, and using (3.33a) and its analog with $S$ replaced by $T$, we find

$$X_{\gamma} - X_{\gamma-1} \leq \varepsilon_{\gamma}(X_{\gamma} + X_{\gamma-1} + 2\Pi_n),$$

for $1 \leq \gamma \leq \gamma_{\max} = N(N + 1)/2$ and with $\Pi_n = (N^\gamma \Pi)^n$. Rewriting this and making the abbreviation $r_{\gamma} = (1 - \varepsilon_{\gamma})^{-1}(1 + \varepsilon_{\gamma}) \geq 1$, we therefore find that

$$X_{\gamma} \leq r_{\gamma}(X_{\gamma-1} + 2\varepsilon_{\gamma} \Pi_n).$$

Since (3.20a) holds for GOE/GUE, we have the initial estimate $X_0 \leq \Pi_n$, and find on iteration that

$$X_{\gamma} \leq \left( \prod_{j=1}^{\gamma} r_j \right) \left( 1 + 2 \sum_{j=1}^{\gamma} \varepsilon_j \right) \Pi_n$$

\[3\] KY omit the factor 2, but since there are $O(N^2)$ inequalities, it should perhaps be tracked explicitly.
Using \( \log(1 + x) \leq x \) and \( \sum \gamma \epsilon_\gamma \leq \frac{1}{2} \) we find that \( \log(\prod r_\gamma) \leq 4 \sum \epsilon_\gamma \leq 2 \) which implies that

\[
\mathbb{E}(L_v G)^n = X_{\max} \leq C \Pi_n,
\]

which is (3.20a). [We may take \( C = 2e^2 \).]

We turn to the proof of (3.33a). The key step is a high-probability bound for \( \prod_{i=1}^m \hat{L}_v Y_{k_i} \), see (135) and (138) below. Consider first a term \((\hat{Y}_{k})_{\text{vv}}\), which up to sign is given by \( N^{-k/2}[(RV)^k \hat{S}]_{\text{vv}} \), where \( \hat{S} \) is short for \( R \) if \( k \leq 3 \) and for \( S \) if \( k = 4 \). Multiplying out gives

\[
[(RV)^k \hat{S}]_{\text{vv}} = \sum R_{\text{va}_1} V_{a_1 b_1} R_{b_1 a_2} \cdots V_{a_m b_m} \hat{S}_{b_m v},
\]

where the sum is over all choices of \( a_1, b_1, \ldots, a_k, b_k \) such that \( \{a_\ell, b_\ell\} = \{a, b\} \) for each \( \ell \). The number of terms in the sum is bounded: \( 2^k \leq 2^{4n} \) if \( a \neq b \) and \( 1 \) if \( a = b \). Let

\[
s_i = \#\text{matrix elements } R_{\text{va}_1}, R_{\text{av}}, S_{\text{va}}, S_{\text{av}},
\]
in that term when \( k = k_i \), and write \( t_i \) for the corresponding number when \( a \) is replaced with \( b \). We have \( s_i, t_i \in \{0, 1, 2\} \) and \( s_i + t_i = 2 \).

Now abbreviate some terms used by KY:

\[
\mathcal{R}_{\text{v}, a} = |R_{\text{va}}| + |R_{\text{av}}| + |S_{\text{va}}| + |S_{\text{av}}|,
\]
ditto for \( \mathcal{R}_{\text{v}, b} \). Since \( |V_{\text{ab}}| \ll 1 \) from (68) and \( |R_{\text{ij}}| \ll 1 \) from (3.27a) and the entrywise local law, we obtain

\[
|(\hat{Y}_{k_i})_{\text{vv}}| = N^{-k_i/2} |[(RV)^k \hat{S}]_{\text{vv}}| \ll N^{-k_i/2} \sum_{s_i=0}^{2} \mathcal{R}_{\text{v}, a}^s \mathcal{R}_{\text{v}, b}^t,
\]

with the sum reducing to \( \mathcal{R}_{\text{v}, a}^2 \) if \( a = b \). Since \( |\hat{L}_v Y_{k_i}| \leq |(Y_{k_i})_{\text{vv}}| \), with \( k = \sum_i k_i \) and similarly for \( s \) and \( t \), we then get

\[
(135) \quad \prod_{i=1}^m \hat{L}_v Y_{k_i} \ll N^{-k/2} \sum_{s=0}^{2m} \mathcal{R}_{\text{v}, a}^s \mathcal{R}_{\text{v}, b}^t,
\]

for \( a \neq b \), with the conditions

\[
\text{(KY3.35) } s + t = 2m, \quad k \geq \max\{s, t\},
\]

while if \( a = b \) in (135) the sum is replaced simply by \( \mathcal{R}_{\text{v}, a}^{2m} \).

Using lemma 38 and (2.9a) we get the bound

\[
\mathcal{R}_{\text{v}, a} \ll \sqrt{\frac{\text{Im} S_{\text{vv}}}{N \eta}} + |v_a| =: \sqrt{x} + |v_a|.
\]

Using KY Lemma 3.10,

\[
(\sqrt{x} + |v_a|^s) \ll (\sqrt{x})^s + |v_a|^s \leq (x + N^{-1/2})^{s/2} (1 + N^{s/4} |v_a|^s).
\]

On \( S(\tau) \), we have \( \Psi \leq N^{-\tau/2} \) and (KY3.5) implies that \( x + N^{-1/2} \lesssim \Omega_\psi \), with the intermediate control term

\[
\Omega_\psi = \Omega_\psi(z) = \frac{\text{Im} S_{\text{vv}}}{N \eta} + \Psi.
\]
We arrive at the bound
\[ R_{v,a}^s \prec \Omega_v^{t/2} (1 + N^{s/4} |v_a|^s). \]

Now recall that \( k \geq k_v \) in the higher order term \( A_{v}^s \). In the off-diagonal case, this means that both \( k \geq \max\{4, m\} \) and \( k \geq \max\{s, t\} \geq (s + t)/2 \) from (KY3.35) and so
\[ N^{-k/2} R_{v,a}^t R_{v,b}^t \leq [N^{-\max\{1,s/4\}} R_{v,a}^s][N^{-\max\{1,t/4\}} R_{v,a}^t]. \]

In the diagonal case, \( k \geq \max\{2, m\} \), and we have just \( N^{-k/2} R_{v,a}^{2m} \leq N^{-\max\{1,m/2\}} R_{v,a}^{2m} \).

Checking cases, one sees that for \( s \geq 0 \),
\[ N^{-\max\{1,s/4\}} (1 + N^{s/4} |v_a|^s) \leq 2 N^{-1} + N^{-1/2} |v_a| + |v_a|^2 =: S_a. \]

From (3.35), (136), and recalling that \( s + t = 2m \), we obtain
\[ N^{-k/2} R_{v,a}^s R_{v,b}^t \prec \Omega_v^m S_a S_b \quad \text{for } a \neq b \]
and \( N^{-k/2} R_{v,a}^{2m} \prec \Omega_v^m S_a \) for \( a = b \). As these bounds are uniform in the relevant \( s, t \), (135) implies that
\[ \prod_{i=1}^{m} \hat{L}_v Y_k \prec E_{ab} \Omega_v^m, \]
with \( E_{ab} = S_a S_b \) for \( a \neq b \) and \( = S_a \) if \( a = b \).

The number of terms in \( A_{m,k} \) is crudely bounded by \( \binom{n}{m} 4^m \leq C_n \) and so from (138)
\[ \sum_{k=k_v, \sqrt{m}}^{4m} |A_{m,k}| \prec E_{ab} \Omega_v^m = E_{ab} N^{-me} (N\gamma \Omega_v)^m. \]

Using \( x^n - y^n \leq (x + y)^n \), we have
\[ A_{v,m} \prec E_{ab} N^{-me} |L_v R|^{n-m} (N\gamma \Omega_v)^m \leq E_{ab} N^{-me} [|L_v R| + N\gamma \Omega_v]^n. \]

From (3.30a) and (KY3.5),
\[ |L_v R - L_v S| \leq |R_{vv} - S_{vv}| \prec \frac{\text{Im} S_{vv}}{N \eta} + N^{-1/2} \lesssim \Omega_v, \]
and so, using KY Lemma 3.10,
\[ ||L_v R| + N\gamma \Omega_v|^n \prec ||L_v S| + N\gamma \Omega_v|^n \prec (L_v S)^n + (N\gamma \Omega_v)^n. \]

But \( \Omega_v^n \prec \left( \frac{\text{Im} S_{vv}}{N \eta} \right)^n + \Psi^n \), which establishes
\[ |A_{v,m}| \prec E_{ab} N^{-me} \left[ (L_v S)^n + \left( \frac{N\gamma \text{Im} S_{vv}}{N \eta} \right)^n + (N\gamma \Psi)^n \right]. \]

To turn this into a bound on expectations, we use [10, Lemma 7.1]: the right side is certainly larger than \( N^{-2-nr-n/2} \), and so it suffices to check that \( E|A'_v|^2 \leq N C_2 \) for some constant \( C_2 \). Note first that (68) implies that \( E|V_{ab}|^k \leq C_k \). Using this and the deterministic bounds \( ||R||, ||S|| \leq N \), we find successively by rough bounds that
\[ E \prod_{i=1}^{m} (Y_k)_{vv}^2 \leq (2N)^{k+2m} C_{2k}, \]
and that

\[ \mathbf{E}A_{\gamma m}^2 \leq (CN)^{Cn}. \]

Thus we may take expectations to conclude that for \( N \geq N(\epsilon, \tau, n) \), and for \( 1 \leq m \leq n \),

\[ |\mathbf{E}A_{\gamma m}'| \leq \mathcal{E}_{ab}N^{\epsilon/4-\epsilon m} \left[ \mathbf{E}(L_\nu S)^n + \mathbf{E} \left( \frac{N^\epsilon \operatorname{Im} S_{\nu \nu}}{N \eta} \right)^n + (N^\epsilon \Phi^n) \right]. \]

At this point we focus on the specific case \( L_\nu S = \operatorname{Im} S_{\nu \nu} \). On \( \mathbf{S}(\tau) \) we have \( N \eta \geq N^\tau > N^\epsilon \) and so in this case the previous display along with \( \psi \lesssim \Phi \) implies that,

\[ \mathbf{E}A_{\gamma m}' \leq \mathcal{E}_{ab}N^{\epsilon/4-\epsilon m} \left[ \mathbf{E}(\operatorname{Im} S_{\nu \nu})^n + (N^\epsilon \Phi^n) \right]. \]

Summing over \( m \) we obtain (3.33a) and hence (3.20a) for \( L_\nu S = \operatorname{Im} S_{\nu \nu} \).

For the remaining cases of \( L_\nu \), we now use the bound (3.20a) just established for \( \operatorname{Im} S_{\nu \nu} \) in \( \mathbf{E}A_{\gamma m}' \), along with (KY3.4) to bound

\[ \mathbf{E} \left( \frac{\operatorname{Im} S_{\nu \nu}}{N \eta} \right)^n \prec \left( \frac{\Phi}{N \eta} \right)^n \prec \psi^{2n} \prec \psi^n, \]

so that

\[ |\mathbf{E}A_{\gamma m}'| \leq \mathcal{E}_{ab}N^{\epsilon/4-\epsilon m/4} \left[ \mathbf{E}(L_\nu S)^n + (N^\epsilon \Psi^n) \right], \]

and hence (3.20a) and (3.33a) follow for the remaining cases of \( L_\nu \) as well.

It remains to prove (93). The proof is immediate and very similar to that of (KY2.14). Using (92), we obtain

\[ \eta^{-1}|v^*u(j)|^2 \leq \sum_i \frac{\eta|v^*u(i)|^2}{(\lambda_j - \lambda_i)^2 + \eta^2} = \operatorname{Im} G_{\nu \nu}(\lambda_j + i\eta) < 1 \]

uniformly in \( N^{-1+\tau} \leq \eta \leq \tau^{-1} \) for any positive \( \tau \). This yields (93). In this derivation, we implicitly used Remark KY2.4 that the overwhelming probability bounds on \( v^*G(z)w - s_{sc}(z)v^*w \) hold simultaneously for all \( z \in \mathbf{S}(\tau) \).

### B.4. Extension of spiked CLT to the critical case for G(U/O)E

Let \( \hat{W}_{h,N} = W_N + h \nu \nu^* \) be the spiked (scaled) GUE or GOE with spike \( h \in [0, \infty) \) along the direction \( \nu \in \mathbb{C}^N \) for GUE and \( \nu \in \mathbb{R}^N \) for GOE, and \( \|\nu\| = 1 \) in both cases. Since the joint distribution of the elements of \( W_N \) is invariant with respect to transformations \( W_N \rightarrow UW_NU^* \), where \( U \) is any unitary matrix for GUE and any orthogonal matrix for GOE, the joint distribution of the eigenvalues of \( W_{h,N} \) does not depend on the exact value of vector \( \nu \). Therefore, without loss of generality, it will be convenient to set \( \nu = (0, ..., 0, 1)^t \).

For GUE and GOE, the tridiagonalization algorithm does not change the bottom right value, see proof of Proposition 7 in [46]. Hence, the analogue of (5) for \( \hat{W}_{h,N} \) is

\[
\sqrt{N}\hat{W}_{h,N} = \begin{pmatrix}
 a_1 & b_1 \\
 b_1 & \ddots & \ddots \\
 \ddots & \ddots & a_{N-1} & b_{N-1} \\
 b_{N-1} & a_N + \sqrt{Nh} \\
\end{pmatrix} = \sqrt{N}\hat{W}_N + \sqrt{Nh}\nu\nu^*.
\]

Recall the definition of sequence \( R_2, \ldots, R_N \) for the tridiagonal \( \hat{W}_N \). Since \( \hat{W}_{h,N} \) differs only in the lower right element, it will have the corresponding ratio sequence
$R_2, \ldots, R_{N-1}, R_{h,N}$, with the only difference coming from the fact that the very last step of recursion (13) now becomes

$$R_{h,N} = \alpha_N + \frac{h}{\theta_N r_N} - \gamma_N + \frac{\gamma_N + \beta_N - \delta_N}{1 - R_{N-1}} = R_N + \frac{h}{\theta_N r_N}.$$  

Therefore, denoting $D_{h,N} = \det (W_{h,N} - 2\theta_N)$, we get that

$$\log |D_{h,N}| = \log |D_N| + \log \left| \frac{1 - R_{h,N}}{1 - R_N} \right| = \log |D_N| + \log \left| 1 - \frac{h}{\theta_N r_N} \right|.$$  

From (11), (38) and Lemma 12,

$$\theta_N = 1 + \frac{1}{2} \sigma_N N^{-2/3}, \quad r_N = 1 + \sigma_N^{1/2} N^{-1/3}, \quad R_N = o_P(N^{-1/3}),$$

which implies that for fixed $h$,

$$\log \left| 1 - \frac{h}{\theta_N r_N} \right| = \begin{cases} \log |1 - h| + o_P(\sigma_N^{1/2} N^{-1/3}) & h \neq 1 \\ -\frac{1}{3} \log N + \frac{1}{2} \log \sigma_N + o_P(1) & h = 1. \end{cases}$$

Hence there is an extra shift $-\frac{1}{3} \log N$ when $h = 1$. Combining Theorem 2 and (141), we have the following theorem.

**Proposition 39.** Let $D_{h,N}$ be the determinant of $W_{h,N} - 2\theta_N$, where $2\theta_N = E = 2 + N^{-2/3} \sigma_N$ with $(\log \log N)^2 \ll \sigma_N \ll (\log N)^2$. Then,

$$\left( \log |D_{h,N}| - \mu_N + 1_{\{h=1\}} \frac{1}{3} \log N \right) / \tilde{\tau}_N \overset{d}{\rightarrow} N(0,1).$$

Note that, for $h \neq 1$, this proposition is generalized by proposition 29 to Wigner matrices and a wider range for the local singularity parameter $-\gamma \leq \sigma_N \ll \log^2 N$. 

