TRANSITION FROM TRACY–WIDOM TO GAUSSIAN FLUCTUATIONS OF EXTREMAL EIGENVALUES OF SPARSE ERDŐS–RÉNYI GRAPHS

BY JIAOYANG HUANG[*], BENJAMIN LANDON[**] AND HORNG-TZER YAU[†]

Department of Mathematics, Harvard University, [*]jiaoyang@math.harvard.edu; [**]landon@math.harvard.edu;
[†]htyau@math.harvard.edu

We consider the statistics of the extreme eigenvalues of sparse random matrices, a class of random matrices that includes the normalized adjacency matrices of the Erdős–Rényi graph $G(N, p)$. Tracy–Widom fluctuations of the extreme eigenvalues for $p \gg N^{-2/3}$ was proved in (*Probab. Theory Related Fields* **171** (2018) 543–616; *Comm. Math. Phys.* **314** (2012) 587–640). We prove that there is a crossover in the behavior of the extreme eigenvalues at $p \sim N^{-2/3}$. In the case that $N^{-7/9} \ll p \ll N^{-2/3}$, we prove that the extreme eigenvalues have asymptotically Gaussian fluctuations. Under a mean zero condition and when $p = CN^{-2/3}$, we find that the fluctuations of the extreme eigenvalues are given by a combination of the Gaussian and the Tracy–Widom distribution. These results show that the eigenvalues at the edge of the spectrum of sparse Erdős–Rényi graphs are less rigid than those of random d-regular graphs (Bauerschmidt et al. (2019)) of the same average degree.

1. Introduction. In this work, we study the statistics of eigenvalues at the edge of the spectrum of sparse random matrices. A natural example is the adjacency matrix of the Erdős–Rényi graph $G(N, p)$, which is the random undirected graph on N vertices in which each edge appears independently with probability p.

Introduced in [30, 33], the Erdős–Rényi graph $G(N, p)$ has numerous applications in graph theory, network theory, mathematical physics and combinatorics. For further information, we refer the reader to the monographs [9, 39]. Many interesting properties of graphs are revealed by the eigenvalues and eigenvectors of their adjacency matrices. Such phenomena and the applications have been intensively investigated for over half a century. To mention some, we refer the readers to the books [11, 13] for a general discussion on spectral graph theory, the survey article [37] for the connection between eigenvalues and expansion properties of graphs and the articles [14, 15, 49, 50, 52, 54, 57–59] on the applications of eigenvalues and eigenvectors in various algorithms, that is, combinatorial optimization, spectral partitioning and clustering.

The adjacency matrices of Erdős–Rényi graphs have typically pN nonzero entries in each column and are sparse if $p \ll 1$. When p is of constant order, the Erdős–Rényi matrix is essentially a Wigner matrix (up to a nonzero mean of the matrix entries). When $p \to 0$ as $N \to \infty$, the law of the matrix entries is highly concentrated at 0, and the Erdős–Rényi matrix can be viewed as a singular Wigner matrix. The singular nature of this ensemble can be expressed by the fact that the kth moment of a matrix entry (in the scaling that the bulk of the eigenvalues lie in an interval of order 1) decays like ($k \geq 2$)

$$(1.1) \qquad N^{-1}(pN)^{-(k-2)/2}.$$

When $p \ll 1$, this decay in k is much slower than the $N^{-k/2}$ case of Wigner matrices, and is the main source of difficulties in studying sparse ensembles with random matrix methods.

The class of random matrices whose moments decay like (1.1) were introduced in the works [17, 20] as a natural generalization of the sparse Erdős–Rényi graph and encompass many other sparse ensembles. This is the class we study in this work.

The global statistics of the eigenvalues of the Erdős–Rényi graph are well understood. The empirical eigenvalue distribution converges to the semicircle distribution provided $p \gg 1/N$, which follows from Wigner's original proof. It was proven in [62], the spectral norm of normalized adjacency matrices of Erdős–Rényi graphs converges to 2 if $p \gg \log^4 N/N$. The sharp estimates were derived in [6, 7]; it was proven that the spectral norm converges to 2 if and only if $p \gg \log N/N$. For the global eigenvalue fluctuations, it was proven in [53] that the linear statistics, after normalizing by $p^{1/2}$, converge to a Gaussian random variable.

A three-step dynamical approach to the local statistics of random matrices has been developed in a series of papers [3, 10, 19, 23–29, 42, 43]. This strategy is as follows:

1. Establish a local semicircle law controlling the number of eigenvalues in windows of size $\log(N)^C/N$.

2. Analyze the local ergodicity of Dyson Brownian motion to obtain universality after adding a small Gaussian noise to the ensemble.

3. A density argument comparing a general matrix to one with a small Gaussian component.

This approach has been used to prove the bulk universality of a wide variety of random matrix ensembles, for example, the sparse random matrices considered here [17, 20, 38] or random matrices with correlated entries [2, 12, 21, 41]. A different approach which proves universality for some class of random matrices was developed in [61].

Universality for the edge statistics of Wigner matrices (the statement that the distribution of the extremal eigenvalues converge to the Tracy–Widon law) was first established by the moment method [56] under certain symmetry assumptions on the distribution of the matrix elements. The moment method was further developed in [32, 51] and [55]. A different approach to edge universality for Wigner matrices based on the direct comparison with corresponding Gaussian ensembles was developed in [29, 60]. Edge universality for sparse ensembles was proven first in the regime $p \gg N^{-1/3}$ in the works [17, 20] and then extended to the regime $p \gg N^{-2/3}$ in [47].

One of the key obstacles in the proof of the edge universality for sparse ensembles is the lack of an optimal rigidity estimate at the edge of the spectrum, which is an overwhelming probability a priori bound on the distance of the extremal eigenvalues from the spectral edge. An important component of the work of [47] is establishing such optimal rigidity estimates at the edge; furthermore, for sparse ensembles they calculate a deterministic correction to the usual semicircle spectral edge ± 2. This correction is larger than the $N^{-2/3}$ Tracy–Widom fluctuations and is therefore necessary for their proof of universality.

In our first main result, we show that a rigidity result can no longer hold on the scale $N^{-2/3}$ in the regime $N^{-7/9} \ll p \ll N^{-2/3}$. We find a contribution to the edge fluctuations of order $(p^{1/2}N)^{-1}$ which dominates the usual Tracy–Widom scale $N^{-2/3}$ for p in this regime; moreover, we find that these fluctuations are asymptotically Gaussian. Comparing this to the result of [47], we see a transition in the behavior at $p = N^{-2/3}$ from the Tracy–Widom distribution to the Gaussian distribution.

Our proof is based on constructing a higher order self-consistent equation for the Stieltjes transform of the empirical eigenvalue distributions. Self-consistent equations have been widely used in the random matrix theory [1, 34, 36, 48]. The leading order component of our self-consistent equation corresponds to the semicircle law. In the work [29], Lemma 4.1, it was noted when computing the higher moment of the trace of the Green's function, averaging over indices gives rise to a cancellation of fluctuations, which gives the optimal error

estimates for the self-consistent equation. The "fluctuation averaging lemma" was further developed in the random band matrices setting [16, 18]. The works [45–47] develop a systematic approach analyzing the self-consistent equation around the spectrum edge. Higher order corrections to the self-consistent equations were also constructed in [4, 47]. The works [5, 31] calculate higher order corrections to the expected density of states for various random matrix ensembles.

In the current setting, our derivation of a recursive moment estimate for the normalized trace follows the recent work of [47]. However, we construct an explicit *random* measure from carefully chosen observables of the random matrix which approximates the Stieltjes transform m_N of the empirical measure of the random matrix. We prove a rigidity result with respect to this random measure, and then show that the edge of this measure has asymptotic Gaussian fluctuations of a larger order than the rigidity estimate, implying the Gaussian fluctuations for the edge eigenvalues themselves.

Our second main result concerns sparse random matrices H with centered entries, for example, $\mathbb{P}[H_{ij} = \pm 1/\sqrt{2Np}] = p/2$ and $\mathbb{P}[H_{ij} = 0] = 1 - p$. When $p = CN^{-2/3}$, we find that the distribution of the extreme eigenvalues converges to an independent sum of a Gaussian and Tracy–Widom random variable. The Gaussian fluctuation we find in both the case $p \ll N^{-2/3}$ and $p = CN^{-2/3}$ is an expansion in the support of the density of states—the correction to the smallest eigenvalue is the same as to the largest eigenvalue, but with the opposite sign. We exhibit this correction as a specific extensive quantity involving the matrix elements of H. If one subtracts this quantity, one finds the usual Tracy–Widom fluctuations down to $p \gg N^{-7/9}$. For example, we show that the gap $\lambda_2 - \lambda_1$ is still given asymptotically by the gap of the GOE, on the usual $N^{-2/3}$ scale.

In order to exhibit Tracy–Widom fluctuations, we compare sparse ensembles to Gaussian divisible ensembles, that is, a sparse ensemble with a GOE component. In [47], the local law allowed for the comparison of sparse ensembles directly to the GOE. When $p \ll N^{-2/3}$, it does not appear possible to compare directly to the GOE and so instead we use a Gaussian divisible ensemble with a small $o(1)$ GOE component.

Edge universality for such ensembles is established in [44]. This work proves a version of Dyson's conjecture on the local ergodicity of Dyson Brownian motion for edge statistics. That is, for wide classes of initial data, the edge statistics of DBM coincides with the GOE, and moreover, [44] finds the optimal time to equilibrium $t \sim N^{-1/3}$ for sufficiently regular initial data.

Organization. We define the model and present the main results in the rest of Section 1. In Section 2, we obtain the optimal edge rigidity estimates in the regime $p \gg N^{-7/9}$. It implies that in the regime $N^{-7/9} \ll p \ll N^{-2/3}$, the extreme eigenvalues have Gaussian fluctuation. In Section 3, we recall the results from [44] for edge universality of Gaussian divisible ensembles. In Section 4, we analyze the Green's function to compare a sparse ensemble to a Gaussian divisible ensemble and establish our results about Tracy–Widom fluctuations.

Conventions. We use C to represent large universal constants, and c small universal constants, which may be different from line by line. Let $Y \geq 0$. We write $X \ll Y$ or $Y \gg X$, if there exists a small exponent $\mathfrak{c} > 0$ such that $X \leq N^{-\mathfrak{c}}Y$ for $N \geq N(\mathfrak{c})$ large enough. We write $X \lesssim Y$ or $Y \gtrsim X$ if there exists a constant $C > 0$, such that $X \leq CY$. We write $X \asymp Y$ if there exists a constant $C > 0$ such that $Y/C \leq X \leq Y/C$. We say an event Ω holds with overwhelming probability, if for any $D > 0$, $\mathbb{P}(\Omega) \geq 1 - N^{-D}$ for $N \geq N(D)$ large enough.

1.1. *Sparse random matrices.*

In this section, we introduce the class of sparse random matrices that we consider. This class was introduced in [17, 20] and we repeat the discussion appearing there.

The Erdős–Rényi graph is the undirected random graph in which each edge appears with probability p. It is notationally convenient to replace the parameter p with q defined through

$$(1.2) \qquad\qquad p = q^2/N.$$

We allow q to depend on N. We denote by A the adjacency matrix of the Erdős–Rényi graph. The adjacency matrix A is an $N \times N$ symmetric matrix whose entries a_{ij} above the main diagonal are independent and distributed according to

$$(1.3) \qquad\qquad a_{ij} = \begin{cases} 1 & \text{with probability } q^2/N, \\ 0 & \text{with probability } 1 - q^2/N. \end{cases}$$

We extract the mean of each entry and rescale the matrix so that the limiting eigenvalue distribution is supported on $[-2, 2]$. We introduce the matrix H by

$$(1.4) \qquad\qquad H := \frac{A - q^2 |e\rangle\langle e|}{q\sqrt{1 - q^2/N}},$$

where e is the unit vector

$$(1.5) \qquad\qquad e = (1, \ldots, 1)^T / \sqrt{N}.$$

It is easy to check that the matrix elements of H have mean zero $\mathbb{E}[h_{ij}] = 0$, variance $\mathbb{E}[h_{ij}^2] = 1/N$ and satisfy the moment bounds

$$(1.6) \qquad \begin{aligned} \mathbb{E}[h_{ij}^k] &= \frac{1}{Nq^{k-2}}\left[\left(1 - \frac{q^2}{N}\right)^{-k/2+1}\left(\left(1 - \frac{q^2}{N}\right)^{k-1} + (-1)^k\left(\frac{q^2}{N}\right)^{k-1}\right)\right] \\ &= \frac{\Omega(1)}{Nq^{k-2}}, \end{aligned}$$

for $k \geq 2$. This motivates the following definition.

DEFINITION 1.1 (Sparse random matrices). We assume that $H = (h_{ij})$ is an $N \times N$ random matrix whose entries are real and independent up to the symmetry constraint $h_{ij} = h_{ji}$. We further assume that (h_{ij}) have the same moments, satisfying $\mathbb{E}[h_{ij}] = 0$, $\mathbb{E}[h_{ij}^2] = 1/N$ and that for any $k \geq 2$, the kth cumulant of h_{ij} is given by

$$(1.7) \qquad\qquad \frac{(k-1)!\mathcal{C}_k}{Nq^{k-2}},$$

where $q = q(N)$ is the sparsity parameter, such that $0 < q \lesssim \sqrt{N}$. For $\mathcal{C}_k$, we make the following assumptions:

(1) $|\mathcal{C}_k| \leq C_k$ for some constant $C_k > 0$.
(2) $\mathcal{C}_4 \geq c$.

REMARK 1.2. The lower bound, $\mathcal{C}_4 \geq c$, ensures that the scaling by q for the ensemble H is "correct." Otherwise, we can always rescale q to make $\mathcal{C}_4 \asymp 1$.

We denote the eigenvalues of H by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, and the Green's function of H by

$$G(z) := (H - z)^{-1}.$$

The Stieltjes transform of the empirical eigenvalue distribution is denoted by

$$m_N(z) := \frac{1}{N} \operatorname{Tr} G(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z}.$$

For N-dependent random (or deterministic) variables A and B, we say B stochastically dominate A, if for any $\varepsilon > 0$ and $D > 0$, then

(1.8)
$$\mathbb{P}(A \geq N^{\varepsilon} B) \leq N^{-D},$$

for $N \geq N(\varepsilon, D)$ large enough, and we write $A \prec B$ or $A = \mathcal{O}_{\prec}(B)$.

We define the following quantity, which governs the fluctuation of the extreme eigenvalues

(1.9)
$$\mathcal{X} := \frac{1}{N} \sum_{ij} \left(h_{ij}^2 - \frac{1}{N} \right).$$

If H is the normalized adjacency matrix of Erdős–Rényi graphs, $\mathcal{X}$ characterizes the fluctuation of the total number of edges. $\mathcal{X}$ is of size $\mathcal{O}_{\prec}(\sqrt{N}q)$. Moreover, by the central limit theorem, the fluctuations of $\sqrt{N}q\mathcal{X}$ are asymptotically Gaussian with mean zero and variance $6\mathcal{C}_4$.

1.2. *Main results.* We first recall the strong entrywise local semicircle law for sparse random matrices from [20].

THEOREM 1.3 ([20], Theorem 2.8). *Let H be as in Definition 1.1 and $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$. Let $\mathfrak{b} > 0$ be a large constant. Then uniformly for any $z = E + i\eta$ such that $-\mathfrak{b} \leq E \leq \mathfrak{b}$ and $1/N \ll \eta \leq \mathfrak{b}$, we have*

(1.10)
$$\max_{i,j} |G_{ij}(z) - \delta_{ij} m_{\mathrm{sc}}(z)| \prec \left(\frac{1}{q} + \sqrt{\frac{\operatorname{Im} m_{\mathrm{sc}}(z)}{N\eta}} + \frac{1}{N\eta} \right),$$

where $G(z)$ is the Green's function of the matrix H, and $m_{\mathrm{sc}}(z)$ is the Stieltjes transform of the semicircle distribution.

Our first main results, Theorem 1.4 and Corollary 1.5 concern the behavior of the extremal eigenvalues of H and A in the regime $N^{1/9} \ll q \ll N^{1/6}$. Specifically, we prove that the extremal have Gaussian fluctuations governed by the quantity $\mathcal{X}$ as defined in (1.9).

THEOREM 1.4. *Let H be as in Definition 1.1 and $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$. Let $\mathcal{X}$ be as defined in (1.9) and denote the eigenvalues of H by $\lambda_1, \lambda_2, \ldots, \lambda_N$. There exists a deterministic constant L depending on q and the cumulants $\mathcal{C}_2, \mathcal{C}_3, \ldots$ of h_{ij} (as defined in Proposition 2.5) so that the following holds. Let $\mathfrak{c} > 0$ and an integer $k \geq 1$ be given. With overwhelming probability, we have, for $1 \leq i \leq k$,*

(1.11)
$$|\lambda_i - L - \mathcal{X}| \leq N^{\mathfrak{c}} \left(\frac{1}{q^6} + \frac{1}{N^{2/3}} \right)$$

and

(1.12)
$$|\lambda_{N-i} + L + \mathcal{X}| \leq N^{\mathfrak{c}} \left(\frac{1}{q^6} + \frac{1}{N^{2/3}} \right).$$

As an easy corollary of Theorem 1.4, we derive asymptotic Gaussian fluctuations for the second largest eigenvalue of the Erdős–Rényi graph in the regime $N^{-7/9} \ll p \ll N^{-2/3}$.

COROLLARY 1.5. *Let A be the adjacency matrix of the Erdős–Rényi graph $G(N, p)$. We denote its eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$. Then for $N^{-7/9} \ll p \ll N^{-2/3}$, the second largest eigenvalue of A converges weakly to a Gaussian random variable as $N \to \infty$,*

(1.13)
$$\sqrt{N} \left(\mu_2 - \sqrt{Np} \left(2 + (Np)^{-1} - \frac{5}{4}(Np)^{-2} \right) \right) \to \mathcal{N}(0, 1).$$

Here, $\mathcal{N}(0, 1)$ is the standard Gaussian random variable.

The following theorem concerns Tracy–Widom fluctuations for sparse random matrices as defined in Definition 1.1. In the case $q = cN^{1/6}$, the limiting distribution of the extremal eigenvalues is given by an independent sum of a Tracy–Widom and Gaussian random variable. In the case $N^{1/6} \ll q \ll N^{1/6}$, we recover the Tracy–Widom fluctuations after subtracting $\mathcal{X}$.

THEOREM 1.6. *Let H be as in Definition 1.1 and $N^{1/9} \ll q \leq CN^{1/6}$. Let $\mathcal{X}$ be as defined in (1.9) and denote the eigenvalues of H by $\lambda_1, \lambda_2, \ldots, \lambda_N$. Let $F : \mathbb{R}^k \to \mathbb{R}$ be a bounded test function with bounded derivatives. There is a universal constant $\mathfrak{c} > 0$ so that*

$$
(1.14) \quad
\begin{aligned}
&\mathbb{E}_H\big[F(N^{2/3}(\lambda_1 - L - \mathcal{X}), \ldots, N^{2/3}(\lambda_k - L - \mathcal{X})\big] \\
&= \mathbb{E}_{\mathrm{GOE}}\big[F\big(N^{2/3}(\mu_1 - 2), \ldots, N^{2/3}(\mu_k - 2)\big)\big] + \mathcal{O}(N^{-\mathfrak{c}}).
\end{aligned}
$$

The second expectation is with respect to a GOE matrix with eigenvalues $\mu_1, \mu_2, \ldots, \mu_N$. In the case $q = CN^{1/6}$, we have

$$
(1.15) \quad
\begin{aligned}
&\mathbb{E}_H\big[F(N^{2/3}(\lambda_1 - L), \ldots, N^{2/3}(\lambda_k - L)\big] \\
&= \mathbb{E}_{\mathrm{GOE}, H}[F\big(N^{2/3}(\mu_1 - 2 + \mathcal{X}), \ldots, N^{2/3}(\mu_k - 2 + \mathcal{X})\big) + \mathcal{O}(N^{-\mathfrak{c}}).
\end{aligned}
$$

In the second expectation, H is independent of the GOE matrix and its eigenvalues μ_i. Analogous results hold for the smallest eigenvalues.

REMARK 1.7. In the regime of Theorem 1.6, that is, $N^{1/9} \ll q \leq CN^{1/6}$, the leading fluctuation of λ_1 is given by $\mathcal{X}$, which is of order $\mathcal{O}(1/\sqrt{N}q)$, and the next order fluctuation is given by the Tracy–Widom fluctuation, which is of order $\mathcal{O}(1/N^{2/3})$. In the sparser regime, that is, $q \leq CN^{1/9}$, we suspect that there will be higher order fluctuations of sizes $\mathcal{O}(1/\sqrt{N}q^3), \mathcal{O}(1/\sqrt{N}q^5), \ldots$, which depend on subgraph counting quantities.

2. Edge rigidity for sparse random matrices. In this section, we prove the following rigidity estimates for sparse random matrices in the vicinity of the spectral edges. The following proposition states that the Stieltjes transform of the empirical eigenvalue density is close to the Stieltjes transform of a *random* measure $\tilde{\rho}_\infty$, which is explicitly constructed in Proposition 2.6.

Fix an arbitrarily small constant $\mathfrak{a}$ and denote the shifted spectral domain

$$
(2.1) \quad
\begin{aligned}
\mathcal{D} = \Big\{ z = \kappa + i\eta \in \mathbb{C}_+ : |\kappa| \leq 1, 0 \leq \eta \leq 1, \\
|\kappa| + \eta \geq N^{\mathfrak{a}}\Big(\frac{1}{q^3 N^{1/2}} + \frac{1}{q^3 N\eta} + \frac{1}{(N\eta)^2}\Big)\Big\}.
\end{aligned}
$$

THEOREM 2.1. *Let H be as in Definition 1.1 and $q \geq N^{\mathfrak{c}}$ with arbitrarily small $\mathfrak{c} > 0$. Let $m_N(z)$ be the Stieltjes transform of its eigenvalue density. There exists a deterministic constant L (as defined in Proposition 2.5), and an explicit random symmetric measure $\tilde{\rho}_\infty$ (as defined in Proposition 2.6) with Stieltjes transform $\tilde{m}_\infty$ such that the following holds: fix $\tilde{z} = L + \mathcal{X} + z$, where $z = \kappa + i\eta$ and $\mathcal{X}$ is defined in (1.9), then with overwhelming probability, uniformly for any $z \in \mathcal{D}$, we have:*

- *If $\kappa \geq 0$,*

$$
(2.2) \quad
\begin{aligned}
&\big|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big| \\
&\prec \frac{1}{\sqrt{|\kappa| + \eta}}\Big(\frac{1}{N\eta^{1/2}} + \frac{1}{N^{1/2}q^{3/2}} + \frac{1}{(N\eta)^2} + \frac{1}{q^3 N\eta}\Big).
\end{aligned}
$$

- *If $\kappa \leq 0$,*

$$(2.3) \qquad \left| m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z}) \right| \prec \frac{1}{N\eta} + \frac{1}{(N\eta)^{1/2} q^{3/2}}.$$

The analogous statement holds for $\tilde{z} = -L - \mathcal{X} + z$.

We postpone the proof of Theorem 2.1 to Sections 2.2, 2.3 and 2.4.

2.1. *Proofs of Theorem 1.4 and Corollary 1.5.* In this section, we prove Theorem 1.4 and Corollary 1.5 which follow easily from Theorem 2.1.

PROOF OF THEOREM 1.4. We only prove (1.11), as (1.12) follows from the same argument. Note that the spectral edge of the random measure $\tilde{\rho}_\infty$ satisfies (2.20). By fixing a sufficiently small $\mathfrak{c} > 0$ and taking

$$(2.4) \qquad \eta = \frac{1}{N^{2/3}}, \qquad \kappa \geq N^{\mathfrak{c}}\left(\frac{1}{q^6} + \frac{1}{N^{2/3}}\right),$$

in (2.2), we get that

$$
\begin{aligned}
(2.5) \qquad |m_N(\tilde{z})| &= \mathcal{O}\left(\frac{\eta}{\sqrt{\kappa + \eta}}\right) \\
&\quad + \mathcal{O}_{\prec}\left(\frac{1}{\sqrt{|\kappa| + \eta}}\left(\frac{1}{N\eta^{1/2}} + \frac{1}{N^{1/2} q^{3/2}} + \frac{1}{(N\eta)^2} + \frac{1}{q^3 N\eta}\right)\right) \\
&\ll \frac{1}{N\eta},
\end{aligned}
$$

with overwhelming probability, where we used (2.21). The equation (2.5) implies that there is no eigenvalue in the interval $[L + \mathcal{X} + \kappa - \eta, L + \mathcal{X} + \kappa + \eta]$. Otherwise, if there was an eigenvalue, λ_i in the spectral window $[L + \mathcal{X} + \kappa - \eta, L + \mathcal{X} + \kappa + \eta]$, we would have that

$$(2.6) \qquad \mathrm{Im}[m_N(\tilde{z})] \geq \frac{1}{N}\,\mathrm{Im}\left[\frac{1}{\lambda_i - (L + \mathcal{X} + \kappa + i\eta)}\right] \geq \frac{1}{2N\eta},$$

contradicting (2.5). It follows that with overwhelming probability we have

$$(2.7) \qquad \lambda_1 - L - \mathcal{X} \leq N^{\mathfrak{c}}\left(\frac{1}{q^6} + \frac{1}{N^{2/3}}\right).$$

For the lower bound, by the same argument as in [22], Lemma B.1, using the Helffer–Sjöstrand formula, we have that with overwhelming probability

$$
\begin{aligned}
(2.8) \qquad \#\{i : \lambda_i \in (L + \mathcal{X} - \kappa, \infty))\} &= \int_{L+\mathcal{X}-\kappa}^{\infty} \rho_\infty(x)\,dx + o(\kappa^{3/2}) \\
&= \frac{2}{3\pi}\kappa^{3/2} + o(\kappa^{3/2}),
\end{aligned}
$$

where

$$(2.9) \qquad \kappa = N^{\mathfrak{c}}\left(\frac{1}{q^6} + \frac{1}{N^{2/3}}\right).$$

The claim (1.11) follows from (2.7) and (2.8) after taking $\mathfrak{c} > 0$ sufficiently small. $\qquad\square$

PROOF OF COROLLARY 1.5. Consider the following normalized adjacency matrix:

$$(2.10) \qquad H := \frac{A - q^2 |e\rangle\langle e|}{q\sqrt{1 - q^2/N}},$$

where $q = (Np)^{1/2}$. We denote the eigenvalues of H by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. It follows from Theorem 1.4 that for any fixed index i,

$$(2.11) \qquad \sqrt{N}q(\lambda_i - L) \to \mathcal{N}(0, 1), \quad L = 2 + \frac{1}{q^2} - \frac{5}{4q^4}.$$

By the Cauchy interlacing theorem,

$$(2.12) \qquad \sqrt{1 - q^2/N}\lambda_2 \leq \mu_2 \leq q\sqrt{1 - q^2/N}\lambda_1,$$

and the claim (1.13) follows by combining (2.11) and (2.12). $\quad\square$

2.2. Construction of the higher order self-consistent equation.

In this section, we construct the higher order self-consistent equations for sparse random matrices. We also derive some useful estimates on the equilibrium measure and its Stieltjes transform.

PROPOSITION 2.2. *Let H be as in Definition 1.1, $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$, and fix a large constant $\mathfrak{b} > 0$. Let $m_N(z)$ be the Stieltjes transform of the empirical eigenvalue density of H. There exists a polynomial (the higher order self-consistent equation)*

$$(2.13) \qquad P_0(z, m) = 1 + zm + Q(m),$$

where

$$(2.14) \qquad Q(m) = m^2 + \frac{1}{q^2}(a_2 m^4 + a_3 m^6 + \cdots),$$

is an even polynomial (of degree at most $\mathcal{O}(1/\mathfrak{e})$) of m with coefficients $a_2, a_3, \ldots = \mathcal{O}(1)$ that depend on q and the cumulants $\mathcal{C}_2, \mathcal{C}_3, \ldots$ of h_{ij}, such that uniformly for any $z = E + i\eta$ with $-\mathfrak{b} \leq E \leq \mathfrak{b}$, $1/N \ll \eta \leq \mathfrak{b}$, we have

$$(2.15) \qquad \mathbb{E}\big[P_0\big(z, m_N(z)\big)\big] \prec \frac{\mathbb{E}[\mathrm{Im}[m_N(z)]]}{N\eta}.$$

The proof uses the cumulant expansion to compute the expectations, similar to the derivation of Stein's lemma. The cumulant expansion was used in [5, 35, 47] to study the spectrum of random matrices. We postpone the proof to the end of this section.

REMARK 2.3. The same idea was first used in [4] to derive the higher order self-consistent equation for the adjacency matrices of d-regular graphs.

REMARK 2.4. The first few terms of P_0 are given by

$$(2.16) \qquad P_0(z, m) = 1 + zm + m^2 + \frac{6\mathcal{C}_4}{q^2}m^4 + \frac{120\mathcal{C}_6}{q^4}m^6 + \mathcal{O}\left(\frac{1}{q^6}\right).$$

We recall $\mathcal{X}$ as defined in (1.9), which characterizes the fluctuation of the number of edges if H is the normalized adjacency matrix of the Erdős–Rényi graph. We define the random polynomial $P(z, m)$ as

$$(2.17) \qquad P(z, m) = P_0(z, m) + \mathcal{X}m^2.$$

The solutions $m_\infty(z)$ of $P(z, m_\infty(z)) = 0$ and $\tilde{m}_\infty(z)$ of $P(z, \tilde{m}_\infty(z)) = 0$ define holomorphic functions from the upper half-plane $\mathbb{C}_+$ to itself. It turns out that they are Stieltjes transforms of probability measures ρ_∞ and $\tilde{\rho}_\infty$, respectively. The following lemmas collect some properties of $m_\infty(z)$, $\tilde{m}_\infty(z)$ and the measures ρ_∞ and $\tilde{\rho}_\infty$.

PROPOSITION 2.5. *There exists a deterministic algebraic function $m_\infty : \mathbb{C}_+ \to \mathbb{C}_+$, which depends on the cumulants $\mathcal{C}_2, \mathcal{C}_3, \ldots$ of h_{ij}, such that the following hold:*

1. *m_∞ is a solution of the polynomial equation, $P_0(z, m_\infty(z)) = 0$.*
2. *m_∞ is the Stieltjes transform of a deterministic symmetric probability measure ρ_∞. Moreover, $\operatorname{supp} \rho_\infty = [-L, L]$, where L depends on the coefficients of P_0, and ρ_∞ is strictly positive on $(-L, L)$ and has square root behavior at the edge.*
3. *In a neighborhood of the spectrum, that is, $[-L, L]$, we have the following estimate on the imaginary part of m_∞:*

$$(2.18) \qquad \operatorname{Im}[m_\infty(E + i\eta)] \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } E \in [-L, L], \\ \eta/\sqrt{\kappa + \eta} & \text{if } E \notin [-L, L], \end{cases}$$

and

$$(2.19) \qquad |\partial_2 P_0(z, m_\infty(z))| \asymp \sqrt{\kappa + \eta}, \quad \partial_2^2 P_0(z, m_\infty(z)) = 1 + \mathcal{O}(1/q^2),$$

where $\kappa = \operatorname{dist}(\operatorname{Re}[z], \{-L, L\})$.

PROPOSITION 2.6. *There exists an algebraic function $\tilde{m}_\infty : \mathbb{C}_+ \to \mathbb{C}_+$, which depends on the random quantity $\mathcal{X}$ as defined in (1.9), such that the following hold:*

1. *$\tilde{m}_\infty$ is a solution of the polynomial equation, $P(z, \tilde{m}_\infty(z)) = 0$.*
2. *$\tilde{m}_\infty$ is the Stieltjes transform of a random symmetric probability measure $\tilde{\rho}_\infty$. Moreover, $\operatorname{supp} \tilde{\rho}_\infty = [-\tilde{L}, \tilde{L}]$ and ρ_∞ is strictly positive on $(-\tilde{L}, \tilde{L})$ and has square root behavior at the edge.*
3. *$\tilde{L}$ has Gaussian fluctuation, explicitly*

$$(2.20) \qquad \tilde{L} = L + \mathcal{X} + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N}q^3}\right).$$

In a neighborhood of the spectrum, that is, $[-\tilde{L}, \tilde{L}]$, we have the following estimate on the imaginary part of $\tilde{m}_\infty$:

$$(2.21) \qquad \operatorname{Im}[\tilde{m}_\infty(E + i\eta)] \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } E \in [-\tilde{L}, \tilde{L}], \\ \eta/\sqrt{\kappa + \eta} & \text{if } E \notin [-\tilde{L}, \tilde{L}], \end{cases}$$

and

$$(2.22) \qquad |\partial_2 P(z, \tilde{m}_\infty(z))| \asymp \sqrt{\kappa + \eta}, \quad \partial_2^2 P(z, \tilde{m}_\infty(z)) = 1 + \mathcal{O}(1/q^2),$$

where $\kappa = \operatorname{dist}(\operatorname{Re}[z], \{-\tilde{L}, \tilde{L}\})$.

The proofs of Propositions 2.5 and 2.6 follow the same proof as [47], Lemma 4.1. For completeness, we sketch the proofs in Appendix B.

Before proving Proposition 2.2, we collect some estimates which will be used repeatedly in the rest of the paper.

PROPOSITION 2.7. *Let H be as in Definition 1.1, $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$, and fix a large constant $\mathfrak{b} > 0$. Let $G(z)$ be the Green's function of H and $m_N(z)$ the Stieltjes transform of the eigenvalue density of H. For any indices $1 \leq i, j \leq N$, we have*

$$(2.23) \qquad h_{ij} \prec \frac{1}{q}.$$

Uniformly, for any $z = E + i\eta$ such that $-\mathfrak{b} \leq E \leq \mathfrak{b}$, $1/N \ll \eta \leq \mathfrak{b}$, we have

$$(2.24) \qquad \sum_k |G_{ik}(z)G_{kj}(z)| \prec \frac{\mathrm{Im}[m_N(z)]}{\eta},$$

$$(2.25) \qquad |\partial_{ij}m_N(z)| \prec \frac{\mathrm{Im}[m_N(z)]}{N\eta}, \qquad \frac{1}{N}|\partial_z m_N(z)| \prec \frac{\mathrm{Im}[m_N(z)]}{N\eta},$$

where ∂_{ij} is the derivative with respect to h_{ij}.

PROOF. The bound on h_{ij} follows from Markov's equality. For (2.24), by the Cauchy–Schwarz inequality and the Ward identity (A.1), we have

$$(2.26) \qquad \begin{aligned} \left| \sum_k G_{ik}G_{kj} \right| &\leq \frac{1}{2}\sum_k [|G_{kj}|^2 + |G_{ik}|^2] \\ &= \frac{\mathrm{Im}[G_{jj}] + \mathrm{Im}[G_{ii}]}{2\eta} \prec \frac{\mathrm{Im}[m_N]}{\eta}, \end{aligned}$$

where in the last inequality we used (A.3). The estimates in (2.25) follow from (2.24) and the calculations,

$$(2.27) \qquad |\partial_{ij}m_N| = \frac{2 - \delta_{ij}}{N} \left| \sum_k G_{ik}G_{kj} \right| \prec \frac{\mathrm{Im}[m_N]}{N\eta}$$

and

$$(2.28) \qquad \frac{1}{N}|\partial_z m_N| = \frac{1}{N^2} \left| \sum_{ij} G_{ij}G_{ji} \right| \prec \frac{\mathrm{Im}[m_N]}{N\eta}. \qquad \square$$

The following proposition allows us to estimate the expectation of certain polynomials of entries of the Green's function.

PROPOSITION 2.8. *Let H be as in Definition 1.1, $q \geq N^{\mathfrak{c}}$ with arbitrarily small $\mathfrak{c} > 0$, and fix a large constant $\mathfrak{b} > 0$. Let $G(z)$ be the Green's function of the matrix H. Let $F(G(z))$ be a function of the following form*:

$$(2.29) \qquad F(G(z)) = \prod_{k=1}^d \left(\frac{1}{N}\sum_{i=1}^N G_{ii}^{s_k}(z) \right),$$

with fixed exponents $s_1, s_2, \ldots, s_k$. Uniformly, for any $z = E + i\eta$ such that $-\mathfrak{b} \leq E \leq \mathfrak{b}$, $1/N \ll \eta \leq \mathfrak{b}$, we have

$$(2.30) \qquad \left| \frac{1}{N^2}\mathbb{E}\left[\sum_{ij} G_{ij}G_{ii}^{r_1}G_{jj}^{r_2}F(G) \right] \right| \prec \frac{\mathbb{E}[\mathrm{Im}[m_N]]}{N\eta}.$$

PROOF. For the sum of diagonal entries, we have the following trivial bound:

$$(2.31) \qquad \left| \frac{1}{N^2}\sum_{i=j} G_{ij}G_{ii}^{r_1}G_{jj}^{r_2}F(G) \right| \prec \frac{1}{N}.$$

For the off-diagonal terms, we use the following identities:

$$(2.32) \qquad G_{ij} = G_{ii}\sum_{k\neq i} h_{ik}G_{kj}^{(i)}, \qquad G_{kj}^{(i)} = G_{kj} - \frac{G_{ki}G_{ij}}{G_{ii}},$$

and the cumulant expansion. We take a large constant l such that $q^\ell \geq N^{10}$,

$$
\begin{aligned}
&\frac{1}{N^2}\mathbb{E}\Big[\sum_{ij} G_{ij}G_{ii}^{r_1}G_{jj}^{r_2}F(G)\Big] \\
&= \frac{1}{N^2}\sum_{ijk}\mathbb{E}[h_{ik}G_{kj}^{(i)}G_{ii}^{r_1+1}G_{jj}^{r_2}F(G)] \\
&= \sum_{p=1}^{\ell}\frac{\mathcal{C}_{p+1}}{N^3 q^{p-1}}\sum_{ijk}\mathbb{E}[G_{kj}^{(i)}\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))] + \mathcal{O}\Big(\frac{1}{N^3}\Big) \\
&= \sum_{p=1}^{\ell}\frac{\mathcal{C}_{p+1}}{N^3 q^{p-1}}\sum_{ijk}\mathbb{E}[G_{kj}\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))] + \mathcal{O}_{\prec}\Big(\frac{\mathbb{E}[\operatorname{Im}[m_N]]}{N\eta}\Big).
\end{aligned}
$$
(2.33)

We consider terms in the Leibniz expansion of $\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))$. If such a term contains at least one off-diagonal term, then the resulting expression $G_{kj}\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))$ contains at least two off-diagonal terms. We can estimate it using (2.24), which gives an error $\mathcal{O}_{\prec}(\operatorname{Im}[m_N]/N\eta)$. For example, if A_{ijk} is a term in the expansion of $\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))$ that contains the off-diagonal term G_{ik}, then

$$
\begin{aligned}
\Big|\frac{\mathcal{C}_{p+1}}{N^3 q^{p-1}}\sum_{ijk}\mathbb{E}[G_{kj}A_{ijk}]\Big| &\prec \frac{1}{N^3}\sum_{ijk}\mathbb{E}[|G_{kj}G_{ik}|] \\
&\leq \frac{1}{N^3}\sum_{ijk}\mathbb{E}[|G_{kj}|^2+|G_{ik}|^2] \\
&\prec \frac{\mathbb{E}[\operatorname{Im}[m_N]]}{N\eta}.
\end{aligned}
$$
(2.34)

Thus,

$$
\begin{aligned}
&\sum_{p\geq 1}\frac{\mathcal{C}_{p+1}}{N^3 q^{p-1}}\sum_{ijk}\mathbb{E}[G_{kj}\partial_{ik}^p(G_{ii}^{r_1+1}G_{jj}^{r_2}F(G))] \\
&= \sum_{p\geq 1}\frac{(r_1+1)\mathcal{C}_{p+1}}{N^3 q^{p-1}}\sum_{ijk}\mathbb{E}[G_{kj}(\partial_{ik}^p G_{ii}^{r_1+1})G_{jj}^{r_2}F(G)] \\
&\quad + \mathcal{O}_{\prec}\Big(\frac{\mathbb{E}[\operatorname{Im}[m_N]]}{N\eta}\Big).
\end{aligned}
$$
(2.35)

Note that when derivatives hit the $F(G)$ term they generically contain off-diagonal terms or are of lower cardinality. If p is an odd number, then $(\partial_{ik}^p G_{ii}^{r_1+1})$ must contain off-diagonal terms and so we dispense with this case. If p is an even number, then

$$
\partial_{ik}^p G_{ii}^{r_1+1} = p!\binom{p/2+r_1}{p/2}G_{ii}^{p/2+r_1+1}G_{kk}^{p/2}
$$
(2.36)
$$
+ \text{ terms with off-diagonal terms.}
$$

For terms in $(\partial_{ik}^p G_{ii}^{r_1+1})$ which contains off-diagonal terms, we can estimate them in the same way as (2.34), which gives an error $\mathcal{O}_{\prec}(\operatorname{Im}[m_N]/N\eta)$. Therefore, by plugging (2.36)

into (2.35), we get

$$
\begin{aligned}
\frac{1}{N^2}&\mathbb{E}\Big[\sum_{ij} G_{ij} G_{ii}^{r_1} G_{jj}^{r_2} F(G)\Big] \\
&= \sum_{p\geq 1, 2|p} \frac{p! \mathcal{C}_{p+1}}{N^2 q^{p-1}} \binom{p/2+r_1}{p/2} \\
&\quad \times \sum_{jk} \mathbb{E}\Big[G_{kj} G_{kk}^{p/2} G_{jj}^{r_2} \Big(\frac{1}{N}\sum_i G_{ii}^{r_1+p/2+1}\Big) F(G)\Big] \\
&\quad + \mathcal{O}_{\prec}\Big(\frac{\mathbb{E}[\mathrm{Im}[m_N]]}{N\eta}\Big).
\end{aligned}
$$

(2.37)

The terms on the right-hand side of (2.37) are in the same form as the one on the left-hand side of (2.37), except that there are some factors of $1/q$ in front of them. For each term on the right-hand side of (2.37), we can repeat the procedure above, each time gaining at least a factor of q^{-1} for any terms that contain only diagonal Green's function elements. After finitely many steps, all the errors are bounded by $\mathcal{O}_{\prec}(N^{-1})$, and the claim follows. □

PROOF OF PROPOSITION 2.2. The polynomial $P_0(z, m)$ is constructed in a way such that (2.15) is satisfied. More precisely, we compute the following expectation using the cumulant expansion:

$$
\begin{aligned}
\mathbb{E}[1+zm_N] &= \frac{1}{N}\sum_{ij} \mathbb{E}[h_{ij} G_{ij}] \\
&= \frac{1}{N}\sum_{ij}\sum_{p=1}^{\ell} \frac{\mathcal{C}_{p+1}}{N q^{p-1}} \mathbb{E}[\partial_{ij}^p G_{ij}] + \mathcal{O}\Big(\frac{1}{q^\ell}\Big),
\end{aligned}
$$

(2.38)

where ℓ is large enough such that $q^\ell \geq N^{10}$. For the derivatives of the resolvent entries G_{ij}, three kinds of terms might occur: the terms containing more than one (counting multiplicity) off-diagonal entries, the terms containing one (counting multiplicity) off-diagonal entry, or the terms containing no off-diagonal entry.

For those terms containing more than one off-diagonal entry, using (2.24), we have the following trivial bound:

$$
(2.39) \qquad \Big|\frac{1}{N^2}\sum_{ij} \mathbb{E}[G_{ij}^{r_1} G_{ii}^{r_2} G_{jj}^{r_3}]\Big| \prec \frac{1}{N^2}\sum_{ij} \mathbb{E}[|G_{ij}|^2] \prec \frac{\mathbb{E}[\mathrm{Im}[m_N]]}{N\eta}.
$$

For those terms containing one off-diagonal entry, thanks to Proposition 2.8, we have the same estimate

$$
(2.40) \qquad \Big|\frac{1}{N^2}\sum_{ij} \mathbb{E}[G_{ij} G_{ii}^{r_1} G_{jj}^{r_2}]\Big| \prec \frac{\mathbb{E}[\mathrm{Im}[m_N]]}{N\eta}.
$$

Finally, we analyze those terms containing no off-diagonal entry, say $G_{ii}^{r_1} G_{jj}^{r_2}$. If $r_1, r_2 \leq 1$, then its average can be written in terms of m_N, that is,

$$
(2.41) \qquad \frac{1}{N^2}\sum_{ij} G_{ii}^{r_1} G_{jj}^{r_2} = m_N^{r_1+r_2}.
$$

Otherwise, we assume $r_1 \geq 2$. Thanks to the following identities:

$$G_{ii}^{r_1} G_{jj}^{r_2} + z G_{kk} G_{ii}^{r_1} G_{jj}^{r_2} = \left(\sum_m h_{km} G_{km} \right) G_{ii}^{r_1} G_{jj}^{r_2},$$

(2.42)

$$G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2} + z G_{kk} G_{ii}^{r_1} G_{jj}^{r_2} = \left(\sum_m h_{im} G_{im} \right) G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2},$$

by taking difference, we get

$$\mathbb{E}[G_{ii}^{r_1} G_{jj}^{r_2}] = \mathbb{E}[G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}] + \sum_m \mathbb{E}[h_{km} G_{km} G_{ii}^{r_1} G_{jj}^{r_2}]$$

$$- \sum_m \mathbb{E}[h_{im} G_{im} G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}]$$

(2.43)

$$= \mathbb{E}[G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}] + \sum_m \sum_{p=1}^{\ell} \frac{\mathcal{C}_{p+1}}{N q^{p-1}} (\mathbb{E}[\partial_{km}^p (G_{km} G_{ii}^{r_1} G_{jj}^{r_2})]$$

$$- \mathbb{E}[\partial_{im}^p (G_{im} G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2})]) + \mathcal{O}\left(\frac{1}{q^\ell} \right).$$

We consider first the leading term in the cumulant expansion $p = 2$. The crucial point is that the only terms resulting from the Leibniz expansion of the derivatives which do not contain at least two off-diagonal entries directly cancel:

$$= \frac{\mathcal{C}_2}{N} (\mathbb{E}[\partial_{km} (G_{km} G_{ii}^{r_1} G_{jj}^{r_2})] - \mathbb{E}[\partial_{im} (G_{im} G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2})])$$

$$= \frac{\mathcal{C}_2}{N} (\mathbb{E}[\partial_{km} (G_{km}) G_{ii}^{r_1} G_{jj}^{r_2}] - \mathbb{E}[\partial_{im} (G_{im}) G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}])$$

$$+ \text{ at least two off-diagonal terms}$$

$$= \frac{\mathcal{C}_2}{N} (\mathbb{E}[(G_{ii} G_{mm}) G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}] - \mathbb{E}[(G_{kk} G_{mm}) G_{ii}^{r_1} G_{jj}^{r_2}])$$

$$+ \text{ at least two off-diagonal terms}$$

$$= \text{ terms with at least two off-diagonal terms.}$$

The terms with at least two off-diagonal terms can be bounded as in (2.39). Therefore, averaging over i, j, k, m, we obtained that

$$\frac{1}{N^2} \sum_{ij} \mathbb{E}[G_{ii}^{r_1} G_{jj}^{r_2}] = \frac{1}{N^3} \sum_{ijk} \mathbb{E}[G_{kk} G_{ii}^{r_1-1} G_{jj}^{r_2}]$$

(2.44)

$$+ \text{ higher order terms}$$

$$+ \mathcal{O}_{\prec}\left(\frac{\mathbb{E}[\text{Im}[m_N(z)]]}{N\eta} \right),$$

where those higher order terms refer to terms of size $\mathcal{O}(q^{-1})$. In this way, we introduced a new resolvent entry G_{kk}, and the exponent of G_{ii} was reduced by one. By repeating this process, we can reduce all the exponents to one. The average of the resulting expression can be written in terms of m_N, that is, $m_N^{r_1+r_2}$. We can repeat the above procedure to the higher order terms, until the error is of order $\mathcal{O}_{\prec}(N^{-1})$. In this way, we obtain a polynomial $Q(m_N)$ such that

(2.45)
$$\mathbb{E}[1 + z m_N] = -\mathbb{E}[Q(m_N)] + \mathcal{O}_{\prec}\left(\frac{\mathbb{E}[\text{Im}[m_N]]}{N\eta} \right). \qquad \square$$

2.3. *Moments of the higher order self-consistent equation.* In this section, we compute higher moments of the self-consistent equation near the spectral edges $\pm \tilde{L}$. This gives us a recursive moment estimate for the Stieltjes transform $m_N(z)$. The rigidity estimates follow from a careful analysis of the recursive moment estimate and an iteration argument.

Note that the polynomial we consider is random and depends on $\mathcal{X}$. We will need to use the deterministic bounds of the random quantity in (2.21), and so we have to consider the Stieltjes transform m_N at the random spectral argument $\tilde{z} = \pm(L + \mathcal{X}) + z$, where $z \in \mathcal{D}$. We recall the shifted spectral domain from (2.1):

$$(2.46) \quad \mathcal{D} = \left\{ z = \kappa + i\eta \in \mathbb{C}_+ : |\kappa| \le 1, 0 \le \eta \le 1, \right.$$
$$\left. |\kappa| + \eta \ge N^{\mathfrak{a}} \left(\frac{1}{q^3 N^{1/2}} + \frac{1}{q^3 N\eta} + \frac{1}{(N\eta)^2} \right) \right\}.$$

PROPOSITION 2.9. *Let H be as in Definition 1.1 and $q \ge N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$. Let $m_N(z)$ be the Stieltjes transform of its eigenvalue density. We recall L as defined in Proposition 2.5, and $\tilde{\rho}_\infty$ with its Stieltjes transform $\tilde{m}_\infty$ as defined in Proposition 2.6. Fix $\tilde{z} = L + \mathcal{X} + z$, where $z = \kappa + i\eta \in \mathcal{D}$ and $\mathcal{X}$ is defined in (1.9). We have the estimate*

$$\mathbb{E} \big[\big| P(\tilde{z}, \tilde{m}_N(\tilde{z})) \big|^{2r} \big]$$
$$(2.47) \quad \prec \max_{s_1 + s_2 \ge 1} \mathbb{E} \left[\left\{ \left(\frac{1}{q^3} + \frac{1}{N\eta} \right) \frac{\text{Im}[m_N(\tilde{z})] |\partial_2 P(\tilde{z}, m_N(\tilde{z}))|}{N\eta} \right\}^{s_1/2} \right.$$
$$\left. \times \left(\frac{\text{Im}[m_N(\tilde{z})]}{N\eta} \right)^{s_2} \big| P(\tilde{z}, m_N(\tilde{z})) \big|^{2r - s_1 - s_2} \right].$$

The analogous statement holds for $\tilde{z} = -L - \mathcal{X} + z$.

Before proving Proposition 2.9, we first derive some estimates of the derivatives of the polynomial P and the Green's function G with respect to the matrix entries of H.

PROPOSITION 2.10. *Using the same notation as in Proposition 2.9, for the derivatives of $P(\tilde{z}, m_N(\tilde{z}))$, we have*

$$(2.48) \quad \big| \partial_{ij}^p P(\tilde{z}, m_N(\tilde{z})) \big| \prec \big| \partial_2 P(\tilde{z}, m_N(\tilde{z})) \big| \frac{\text{Im}[m_N(\tilde{z})]}{N\eta}.$$

For the derivatives of $G_{ij}(\tilde{z})$, we have

$$(2.49) \quad \partial_{ij}^p G_{ij}(\tilde{z}) = (\partial_{ij}^p G_{ij})(\tilde{z}) + \mathcal{O}_\prec \left(\frac{\text{Im}[m_N(\tilde{z})]}{N\eta} \right).$$

PROOF. In this proof, we write $P = P(\tilde{z}, m_N(\tilde{z}))$ and $\partial_2 P = (\partial_2 P)(\tilde{z}, m_N(\tilde{z}))$. We recall Q as defined in (2.14). For the derivative of P,

$$\partial_{ij} P = \partial_{ij} \big(1 + \tilde{z} m_N(\tilde{z}) + Q(m_N(\tilde{z})) + \mathcal{X} m_N^2(\tilde{z}) \big)$$
$$= \partial_2 P \partial_{ij}(m_N(\tilde{z})) + (\partial_{ij} \tilde{z}) m_N(\tilde{z}) + (\partial_{ij} \mathcal{X}) m_N^2(\tilde{z})$$
$$(2.50) \quad = \partial_2 P (\partial_{ij} m_N)(\tilde{z}) + \frac{2 - \delta_{ij}}{N} h_{ij} \big(m_N(\tilde{z}) + m_N^2(\tilde{z}) + \partial_2 P \partial_z m_N(\tilde{z}) \big)$$
$$\prec |\partial_2 P| \frac{\text{Im}[m_N(\tilde{z})]}{N\eta},$$

where we used (2.25). Similarly, for the higher order derivatives of P, we have

$$(2.51) \qquad \left|\partial_{ij}^p P(\tilde{z}, m_N(\tilde{z}))\right| \prec |\partial_2 P| \frac{\mathrm{Im}[m_N(\tilde{z})]}{N\eta}.$$

For the derivative of $G_{ij}(\tilde{z})$, we have

$$
\begin{aligned}
(2.52) \qquad \partial_{ij} G_{ij}(\tilde{z}) &= \partial_{ij}(G_{ij})(\tilde{z}) + (\partial_z G_{ij})(\tilde{z})\partial_{ij}\tilde{z} \\
&= \partial_{ij}(G_{ij})(\tilde{z}) + \frac{2-\delta_{ij}}{N} h_{ij} \sum_k G_{ik}(\tilde{z})G_{kj}(\tilde{z}) \\
&= \partial_{ij}(G_{ij})(\tilde{z}) + \mathcal{O}_{\prec}\left(\frac{\mathrm{Im}[m_N(\tilde{z})]}{N\eta}\right),
\end{aligned}
$$

where in the last equality we used (2.24). The higher order derivatives of $G_{ij}(\tilde{z})$ can be estimated in the same way. $\square$

PROOF OF PROPOSITION 2.9. In this proof, we write, for simplicity of notation, $P = P(\tilde{z}, m_N(\tilde{z}))$, $G = G(\tilde{z})$, $m_N = m_N(\tilde{z})$, $P' = P'(\tilde{z}, m_N(\tilde{z})) = \partial_2 P(\tilde{z}, m_N(\tilde{z}))$, $\partial_{ij}^p P = \partial_{ij}^p(P(\tilde{z}, m_N(\tilde{z})))$, $\partial_{ij}^p G = \partial_{ij}^p(G(\tilde{z}))$, $\partial_{ij}^p m_N = \partial_{ij}^p(m_N(\tilde{z}))$, $D_{ij}^p G = (\partial_{ij}^p G)(\tilde{z})$ and $D_{ij}^p m_N = (\partial_{ij}^p m_N)(\tilde{z})$.

The starting point is the following identity:

$$(2.53) \qquad 1 + \tilde{z} m_N(\tilde{z}) = \sum_{ij} h_{ij} G_{ij}(\tilde{z}).$$

Using the cumulant expansion, we can write the moment of $P(\tilde{z}, m_N(\tilde{z}))$ as

$$
\begin{aligned}
(2.54) \qquad \mathbb{E}\left[\left|P(\tilde{z}, m_N(\tilde{z}))\right|^{2r}\right] &= \mathbb{E}\left[Q P^{r-1}\bar{P}^r\right] + \mathbb{E}\left[\mathcal{X} m_N^2 P^{r-1}\bar{P}^r\right] \\
&\quad + \frac{1}{N}\mathbb{E}\left[\sum_{ij} h_{ij} G_{ij}(\tilde{z}) P^{r-1}\bar{P}^r\right] \\
&= \mathbb{E}\left[Q P^{r-1}\bar{P}^r\right] + \mathbb{E}\left[\mathcal{X} m_N^2 P^{r-1}\bar{P}^r\right] \\
&\quad + \frac{1}{N}\sum_{ij}\sum_{p=1}^{\ell} \frac{\mathcal{C}_{p+1}}{Nq^{p-1}}\mathbb{E}\left[\partial_{ij}^p(G_{ij}P^{r-1}\bar{P}^r)\right] \\
&\quad + \mathcal{O}\left(\frac{1}{q^\ell}\right),
\end{aligned}
$$

where ℓ is large enough such that $q^\ell \geq N^{10}$ (say). In the following, we estimate the sum on the right-hand side of (2.54). For $p \neq 2, 3$, we write that

$$(2.55) \qquad \mathbb{E}\left[\partial_{ij}^p(G_{ij}P^{r-1}\bar{P}^r)\right] = \mathbb{E}\left[D_{ij}^p(G_{ij})P^{r-1}\bar{P}^r\right] + \text{other terms},$$

and show the leading order term cancels with $\mathbb{E}[Q P^{r-1}\bar{P}^r]$ from our construction of Q, and the "other terms" are negligible. For $p = 2, 3$, we write

$$
\begin{aligned}
(2.56) \qquad \mathbb{E}\left[\partial_{ij}^p(G_{ij}P^{r-1}\bar{P}^r)\right] &= \mathbb{E}\left[D_{ij}^p(G_{ij})P^{r-1}\bar{P}^r\right] \\
&\quad - p\mathbb{E}\left[G_{ii}G_{jj}\partial_{ij}^{p-1}(P^{r-1}\bar{P}^r)\right] + \text{other terms},
\end{aligned}
$$

and the "other terms" are negligible. We will show that the leading order contribution of the first term on the right-hand side cancels with $\mathbb{E}[Q P^{r-1}\bar{P}^r]$, and that of the second term cancels with the fluctuation term $\mathbb{E}[\mathcal{X} m_N^2 P^{r-1}\bar{P}^r]$ in (2.54).

For the first- and second-order terms in (2.54), that is, $p = 1, 2$, we have the following estimates. Their proofs depend on a careful analysis of the expansion, we postpone them after the proof of Proposition 2.9.

CLAIM 2.11. *The first-order term in* (2.54), *that is*, $p = 1$, *satisfies*

$$
\frac{1}{N^2} \sum_{ij} \mathbb{E}[\partial_{ij}(G_{ij} P^{r-1} \bar{P}^r)]
$$

(2.57)
$$
= -\mathbb{E}[m_N^2 P^{r-1} \bar{P}^r]
$$
$$
+ \mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{(N\eta)^2} |P'||P|^{2r-2} \right] + \max_{s \geq 2} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right] \right).
$$

CLAIM 2.12. *The second-order term in* (2.54), *that is*, $p = 2$, *satisfies*

$$
\frac{1}{N} \sum_{ij} \frac{\mathcal{C}_3}{Nq} \mathbb{E}[\partial_{ij}^2 (G_{ij} P^{r-1} \bar{P}^r)]
$$

(2.58)
$$
\prec \max_{s=0,1} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta} \right)^s \left(\frac{\operatorname{Im}[m_N]|P'|}{(N\eta)^2} \right) |P|^{2r-2-s} \right]
$$
$$
+ \max_{s \geq 2} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right].
$$

For the third-order term in (2.54), that is, $p = 3$,

$$
\frac{\mathcal{C}_4}{N^2 q^2} \sum_{ij} \mathbb{E}[\partial_{ij}^3 (G_{ij} P^{r-1} \bar{P}^r)]
$$

(2.59)
$$
= \frac{3\mathcal{C}_4}{N^2 q^2} \sum_{ij} \mathbb{E}[\partial_{ij}(G_{ij}) \partial_{ij}^2 (P^{r-1} \bar{P}^r)] + \frac{\mathcal{C}_4}{N^2 q^2} \sum_{ij} \mathbb{E}[\partial_{ij}^3 (G_{ij}) P^{r-1} \bar{P}^r]
$$
$$
+ \mathcal{O}_{\prec}\left(\left(\frac{1}{q^3} + \frac{1}{N\eta} \right) \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]|P'|}{N\eta} \right)^s |P|^{2r-1-s} \right] \right),
$$

where we used that $\max_{i \neq j} |G_{ij}| \prec 1/q + 1/\sqrt{N\eta}$.

For the higher order terms in (2.54), that is, $p \geq 4$,

$$
\frac{\mathcal{C}_{p+1}}{N^2 q^{p-1}} \sum_{ij} \mathbb{E}[\partial_{ij}^p (G_{ij} P^{r-1} \bar{P}^r)]
$$

(2.60)
$$
= \frac{\mathcal{C}_{p+1}}{N^2 q^{p-1}} \sum_{ij} \mathbb{E}[\partial_{ij}^p (G_{ij}) P^{r-1} \bar{P}^r]
$$
$$
+ \mathcal{O}_{\prec}\left(\frac{1}{q^3} \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]|P'|}{N\eta} \right)^s |P|^{2r-1-s} \right] \right).
$$

We combine all the estimates (2.57), (2.58), (2.59) and (2.60) together, and use (2.49)

$$
\mathbb{E}[|P|^{2r}] = \mathbb{E}[(Q - m_N^2) P^{r-1} \bar{P}^r] + \sum_{p \geq 3} \sum_{ij} \frac{\mathcal{C}_{p+1}}{Nq^{p-1}} \mathbb{E}[(D_{ij}^p G_{ij}) P^{r-1} \bar{P}^r]
$$
$$
+ \left(\mathbb{E}[\mathcal{X} m_N^2 P^{r-1} \bar{P}^r] - \sum_{ij} \frac{2\mathcal{C}_3}{N^2 q} \mathbb{E}[G_{ii} G_{jj} \partial_{ij}(P^{r-1} \bar{P}^r)] \right)
$$

$$(2.61) \qquad -\frac{3\mathcal{C}_4}{N^2 q^2} \sum_{ij} \mathbb{E}[G_{ii} G_{jj} \partial_{ij}^2 (P^{r-1} \bar{P}^r)] \Big)$$

$$+ \mathcal{O}_{\prec}\Bigg(\Big(\frac{1}{q^3} + \frac{1}{N\eta}\Big) \max_{s \geq 1} \mathbb{E}\Bigg[\Big(\frac{\mathrm{Im}[m_N]|P'|}{N\eta}\Big)^s |P|^{2r-1-s}\Bigg]$$

$$+ \max_{s \geq 1} \mathbb{E}\Bigg[\Big(\frac{\mathrm{Im}[m_N]}{N\eta}\Big)^s |P|^{2r-s}\Bigg]\Bigg).$$

In the following, we show that the second line is negligible, and then Proposition 2.9 will shortly follow. We will soon see that there is a cancellation between those terms from $p = 3$ and the random term $\mathcal{X} m_N^2 P^{r-1} \bar{P}^r$, which is the reason we use the random polynomial $P(z, m)$ instead of $P_0(z, m)$. We calculate the $\mathcal{X}$ term:

$$\mathbb{E}[\mathcal{X} m_N^2 P^{r-1} \bar{P}^r] = \frac{1}{N} \sum_{ij} \mathbb{E}[h_{ij}^2 m_N^2 P^{r-1} \bar{P}^r] - \frac{1}{N^2} \sum_{ij} \mathbb{E}[m_N^2 P^{r-1} \bar{P}^r]$$

$$= \frac{1}{N} \sum_{ij} \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N q^{p-1}} \mathbb{E}[\partial_{ij}^p (h_{ij} m_N^2 P^{r-1} \bar{P}^r)]$$

$$(2.62) \qquad\qquad - \frac{1}{N^2} \sum_{ij} \mathbb{E}[m_N^2 P^{r-1} \bar{P}^r]$$

$$= \frac{1}{N} \sum_{ij} \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N q^{p-1}} \mathbb{E}[h_{ij} \partial_{ij}^p (m_N^2 P^{r-1} \bar{P}^r)]$$

$$+ \frac{1}{N} \sum_{ij} \sum_{p \geq 2} \frac{p \mathcal{C}_{p+1}}{N q^{p-1}} \mathbb{E}[\partial_{ij}(h_{ij}) \partial_{ij}^{p-1} (m_N^2 P^{r-1} \bar{P}^r)].$$

For the first term in (2.62), we use the cumulant expansion again,

$$\sum_{ij} \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N^2 q^{p-1}} \mathbb{E}[h_{ij} \partial_{ij}^p (m_N^2 P^{r-1} \bar{P}^r)]$$

$$(2.63) \qquad = \sum_{ij} \sum_{p, p' \geq 1} \frac{\mathcal{C}_{p+1} \mathcal{C}_{p'+1}}{N^3 q^{p+p'-2}} \mathbb{E}[\partial_{ij}^{p+p'} (m_N^2 P^{r-1} \bar{P}^r)]$$

$$= \mathcal{O}_{\prec}\Bigg(\max_{s \geq 2} \mathbb{E}\Bigg[\Big(\frac{\mathrm{Im}[m_N]}{N\eta}\Big)^s |P|^{2r-s}\Bigg]\Bigg).$$

For the second term in (2.62), we have

$$\sum_{ij} \sum_{p \geq 2} \frac{p \mathcal{C}_{p+1}}{N^2 q^{p-1}} \mathbb{E}[\partial_{ij}^{p-1} (m_N^2 P^{r-1} \bar{P}^r)]$$

$$= \sum_{ij} \sum_{p \geq 2} \frac{p \mathcal{C}_{p+1}}{N^2 q^{p-1}} \mathbb{E}[m_N^2 \partial_{ij}^{p-1} (P^{r-1} \bar{P}^r)]$$

$$+ \mathcal{O}_{\prec}\Bigg(\max_{s \geq 2} \mathbb{E}\Bigg[\Big(\frac{\mathrm{Im}[m_N]}{N\eta}\Big)^s |P|^{2r-s}\Bigg]\Bigg)$$

$$(2.64) \qquad = \sum_{ij} \frac{2\mathcal{C}_3}{N^2 q} \mathbb{E}[m_N^2 \partial_{ij} (P^{r-1} \bar{P}^r)] + \sum_{ij} \frac{3\mathcal{C}_4}{N^2 q^2} \mathbb{E}[m_N^2 \partial_{ij}^2 (P^{r-1} \bar{P}^r)]$$

$$+ \mathcal{O}_{\prec}\left(\frac{1}{q^3} \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]|P'|}{N\eta}\right)^s |P|^{2r-1-s}\right]\right.$$

$$\left. + \max_{s \geq 2} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta}\right)^s |P|^{2r-s}\right]\right).$$

Therefore, the second line in (2.61) simplifies to

$$\frac{2\mathcal{C}_3}{N^2 q} \sum_{ij} \mathbb{E}[(m_N^2 - G_{ii}G_{jj})\partial_{ij}(P^{r-1}\bar{P}^r)]$$

$$+ \frac{3\mathcal{C}_4}{N^2 q^2} \sum_{ij} \mathbb{E}[(m_N^2 - G_{ii}G_{jj})\partial_{ij}^2(P^{r-1}\bar{P}^r)]$$

$$+ \mathcal{O}_{\prec}\left(\frac{1}{q^3} \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]|P'|}{N\eta}\right)^s |P|^{2r-1-s}\right]\right.$$

(2.65)
$$\left. + \max_{s \geq 2} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta}\right)^s |P|^{2r-s}\right]\right)$$

$$= \frac{2\mathcal{C}_3}{N^2 q} \sum_{ij} \mathbb{E}[(m_N^2 - G_{ii}G_{jj})\partial_{ij}(P^{r-1}\bar{P}^r)]$$

$$+ \mathcal{O}\left(\max_{s \geq 2} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta}\right)^s |P|^{2r-s}\right]\right)$$

$$+ \mathcal{O}_{\prec}\left(\left(\frac{1}{q^3} + \frac{1}{N\eta}\right) \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]|P'|}{N\eta}\right)^s |P|^{2r-1-s}\right]\right),$$

where we used that $|G_{ii} - m_N| \prec 1/q + 1/\sqrt{N\eta}$. For the first term on the right-hand side of (2.65), notice that

$$m_N^2 - G_{ii}G_{jj} = (m_N - G_{ii})m_N + (m_N - G_{jj})m_N$$

(2.66)
$$- (m_N - G_{ii})(m_N - G_{jj})$$

$$= (m_N - G_{ii})m_N + (m_N - G_{jj})m_N + \mathcal{O}_{\prec}\left(\frac{1}{q^2} + \frac{1}{N\eta}\right).$$

Thus the first term in the right-hand side of (2.65) simplifies to

$$\frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{jj})m_N(\partial_{ij}P)P^{r-2}\bar{P}^r]$$

$$+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{jj})m_N(\partial_{ij}\bar{P})|P|^{2r-2}]$$

(2.67)
$$+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{ii})m_N(\partial_{ij}P)P^{r-2}\bar{P}^r]$$

$$+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{ii})m_N(\partial_{ij}\bar{P})|P|^{2r-2}]$$

$$+ \mathcal{O}_{\prec}\left(\left(\frac{1}{q^3} + \frac{1}{N\eta}\right)\mathbb{E}\left[\frac{\mathrm{Im}[m_N]|P'|}{N\eta}|P|^{2r-2}\right]\right).$$

By symmetry, we only need to estimate the first term in (2.67), as the other terms can be estimated in the same way. Using (2.50), we have

$$\frac{1}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{jj}) m_N (\partial_{ij} P) P^{r-2} \bar{P}^r]$$

(2.68)
$$= \frac{1}{N^2 q} \sum_{ij} \mathbb{E}[(m_N - G_{jj}) m_N P'(D_{ij} m_N) P^{r-2} \bar{P}^r]$$

$$+ \frac{1}{N^2 q} \sum_{ij} \frac{2 - \delta_{ij}}{N}$$

$$\times \mathbb{E}[h_{ij}(m_N - G_{jj}) m_N (P' \partial_z m_N + m_N^2 + m_N) P^{r-2} \bar{P}^r].$$

We have that $D_{ij} m_N = \sum_k G_{ki} G_{kj}/N$. Using the identity $G_{ki} = \sum_{\ell \neq i} h_{i\ell} G_{\ell k}^{(i)} G_{ii}$, we can write the first term in (2.68) as

$$\frac{1}{N^3 q} \sum_{ijk} \mathbb{E}[G_{ki} G_{kj} (m_N - G_{jj}) m_N P' P^{r-2} \bar{P}^r]$$

$$= \frac{1}{N^3 q} \sum_{ijk} \sum_{\ell \neq i} \mathbb{E}[h_{i\ell} G_{\ell k}^{(i)} G_{ii} G_{kj} (m_N - G_{jj}) m_N P' P^{r-2} \bar{P}^r]$$

(2.69)
$$= \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N^4 q^p} \sum_{ijk} \sum_{\ell \neq i} \mathbb{E}[G_{\ell k}^{(i)} \partial_{i\ell}^p (G_{ii} G_{kj} (m_N - G_{jj}) m_N P' P^{r-2} \bar{P}^r)]$$

$$= \frac{\mathcal{O}(1)}{N^4 q} \sum_{ijk} \sum_{\ell \neq i} \mathbb{E}[G_{\ell k}^{(i)} G_{ii} G_{kj} (m_N - G_{jj}) m_N P' (\partial_{i\ell} P) P^{r-3} \bar{P}^r)]$$

$$+ \frac{\mathcal{O}(1)}{N^4 q} \sum_{ijk} \sum_{\ell \neq i} \mathbb{E}[G_{\ell k}^{(i)} G_{ii} G_{kj} (m_N - G_{jj}) m_N P' (\partial_{i\ell} \bar{P}) P^{r-2} \bar{P}^{r-1})]$$

$$+ \mathcal{O}_{\prec} \left(\max_{s \geq 1} \mathbb{E}\left[\left(\frac{1}{q^3} + \frac{1}{N\eta} \right) \left(\frac{\mathrm{Im}[m_N] |P'|}{N\eta} \right)^s |P|^{2r-1-s} \right] \right).$$

For the first term on the right-hand side of (2.69), we have

$$\left| \frac{1}{N^4 q} \sum_{ijk} \sum_{\ell \neq i} G_{\ell k}^{(i)} G_{ii} G_{kj} (m_N - G_{jj}) (\partial_{i\ell} P) \right|$$

(2.70)
$$\leq \frac{1}{N^4 q} \sum_{ij} |G_{ii}| \left| \sum_{\ell \neq i, k} (\partial_{i\ell} P) G_{\ell k}^{(i)} (G_{kj}(m_N - G_{jj})) \right|$$

$$\prec \frac{\mathrm{Im}[m_N] |P'|}{(N\eta)^2} \sqrt{\frac{\mathrm{Im}[m_N]}{N\eta}} \left(\frac{1}{q^2} + \frac{1}{N\eta} \right),$$

where we used that $\|G^{(i)}\|_{L_2 \to L_2} \leq 1/\eta$, $|\partial_{i\ell} P| \prec \mathrm{Im}[m_N] |P'|/N\eta$, $|m_N - G_{jj}| \prec 1/q + 1/\sqrt{N\eta}$, and $\sum_k |G_{kj}|^2 \leq \mathrm{Im}[m_N]/\eta$. The same estimate holds for the second term on the right-hand side of (2.69). Combining the above estimates, it follows that the first term in

(2.68) is bounded by

$$
\begin{aligned}
\text{(2.71)} \quad \mathcal{O}_{\prec}\bigg(&\max_{s\geq 1} \mathbb{E}\bigg[\bigg(\frac{1}{q^3}+\frac{1}{N\eta}\bigg)\bigg(\frac{\operatorname{Im}[m_N]|P'|}{N\eta}\bigg)^s|P|^{2r-1-s}\bigg] \\
&+\mathbb{E}\bigg[\frac{\operatorname{Im}[m_N]|P'|}{(N\eta)^2}\sqrt{\frac{\operatorname{Im}[m_N]|P'|}{N\eta}}\bigg(\frac{1}{q^2}+\frac{1}{N\eta}\bigg)|P|^{2r-3}\bigg]\bigg).
\end{aligned}
$$

For the second term on right-hand side of (2.68), we use the cumulant expansion again,

$$
\begin{aligned}
\text{(2.72)} \quad &\frac{1}{N^2 q}\sum_{ij}\frac{2-\delta_{ij}}{N}\mathbb{E}[h_{ij}(m_N-G_{jj})m_N(P'\partial_z m_N+m_N^2+m_N)P^{r-2}\bar{P}^r] \\
&=\sum_{p\geq 1}\frac{\mathcal{C}_{p+1}}{N^3 q^p}\sum_{ij}\frac{2-\delta_{ij}}{N} \\
&\quad \times \mathbb{E}[\partial_{ij}^p((m_N-G_{jj})m_N(P'\partial_z m_N+m_N^2+m_N)P^{r-2}\bar{P}^r)] \\
&=\mathcal{O}_{\prec}\bigg(\max_{s\geq 2}\mathbb{E}\bigg[\bigg(\frac{\operatorname{Im}[m_N]}{N\eta}\bigg)^s|P|^{2r-s}\bigg]\bigg).
\end{aligned}
$$

The estimates (2.61), (2.65), (2.67), (2.68), (2.71) and (2.72) all together lead to the following estimate:

$$
\begin{aligned}
\text{(2.73)} \quad \mathbb{E}[|P|^{2r}]=&\mathbb{E}[(Q-m_N^2)P^{r-1}\bar{P}^r]+\sum_{p\geq 3}\sum_{ij}\frac{\mathcal{C}_{p+1}}{Nq^{p-1}}\mathbb{E}[(D_{ij}^p G_{ij})P^{r-1}\bar{P}^r] \\
&+\mathcal{O}_{\prec}\bigg(\max_{s_1+s_2\geq 1}\mathbb{E}\bigg[\bigg(\bigg(\frac{1}{q^3}+\frac{1}{N\eta}\bigg)\bigg(\frac{\operatorname{Im}[m_N]|P'|}{N\eta}\bigg)\bigg)^{s_1/2} \\
&\times\bigg(\frac{\operatorname{Im}[m_N]}{N\eta}\bigg)^{s_2}|P|^{2r-s_1-s_2}\bigg]\bigg).
\end{aligned}
$$

We have left the first line in (2.73). This is estimated similarly to Proposition 2.2. More precisely, we repeat the procedure as in Proposition 2.2, by repeatedly using the cumulant expansion. Derivatives hitting $\tilde{z}$ give an extra copy of $\operatorname{Im}[m_N]/N\eta$ (see (2.49)) and so the resulting term can be bounded by

$$
\text{(2.74)} \quad \mathcal{O}_{\prec}\bigg(\max_{s\geq 1}\mathbb{E}\bigg[\bigg(\frac{\operatorname{Im}[m_N]}{N\eta}\bigg)^s|P|^{2r-s}\bigg]\bigg).
$$

If the derivative hits P or $\bar{P}$, we get a factor of $\operatorname{Im}[m_N]|P'|/qN\eta$, and the resulting term can be bounded by

$$
\text{(2.75)} \quad \mathcal{O}_{\prec}\bigg(\max_{s\geq 0}\mathbb{E}\bigg[\frac{\operatorname{Im}[m_N]|P'|}{q^3 N\eta}\bigg(\frac{\operatorname{Im}[m_N]}{N\eta}\bigg)^s|P|^{2r-s-2}\bigg]\bigg).
$$

If the derivative never hits $\tilde{z}$, P, $\bar{P}$, these terms will cancel with Q. Proposition 2.9 follows. $\qquad\square$

PROOF OF CLAIM 2.11. We rewrite the left-hand side of (2.57) as

$$
\begin{aligned}
\text{(2.76)} \quad \frac{1}{N^2}\sum_{ij}\mathbb{E}[\partial_{ij}(G_{ij}P^{r-1}\bar{P}^r)]=&\frac{1}{N^2}\sum_{ij}\mathbb{E}[\partial_{ij}(G_{ij})P^{r-1}\bar{P}^r] \\
&+\frac{1}{N^2}\sum_{ij}\mathbb{E}[G_{ij}\partial_{ij}(P^{r-1}\bar{P}^r)].
\end{aligned}
$$

For the first term in (2.76), thanks to (2.49),

$$
\begin{aligned}
\frac{1}{N^2} &\sum_{ij} \mathbb{E}[\partial_{ij}(G_{ij}) P^{r-1} \bar{P}^r] \\
&= -\frac{1}{N^2} \sum_{ij} \mathbb{E}[G_{ii} G_{jj} P^{r-1} \bar{P}^r] - \frac{1}{N^2} \sum_{ij} \mathbb{E}[G_{ij} G_{ij} P^{r-1} \bar{P}^r] \\
&\quad + \mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{N\eta} |P|^{2r-1} \right] \right) \\
&= -\mathbb{E}[m_N^2 P^{r-1} \bar{P}^r] + \mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{N\eta} |P|^{2r-1} \right] \right),
\end{aligned}
\tag{2.77}
$$

where we used (2.24). For the second term in (2.76), we use Proposition 2.10. We write the second term in (2.76) as follows:

$$
\begin{aligned}
\frac{1}{N^2} &\sum_{ij} \mathbb{E}[G_{ij} \partial_{ij}(P^{r-1} \bar{P}^r)] \\
&= \frac{\mathcal{O}(1)}{N^2} \sum_{ij} \mathbb{E}[G_{ij}(\partial_{ij} P) P^{r-2} \bar{P}^r] + \frac{\mathcal{O}(1)}{N^2} \sum_{ij} \mathbb{E}[G_{ij} \partial_{ij}(\bar{P}) |P|^{2r-2}].
\end{aligned}
\tag{2.78}
$$

Using (2.50), the first term in (2.78) is

$$
\begin{aligned}
\frac{\mathcal{O}(1)}{N^2} &\sum_{ij} \mathbb{E}[G_{ij}(D_{ij} m_N) P' P^{r-2} \bar{P}^r] \\
&\quad + \frac{\mathcal{O}(1)}{N^3} \sum_{ij} \mathbb{E}[(2 - \delta_{ij}) h_{ij} G_{ij} (m_N^2 + P' \partial_z m_N + m_N) P^{r-2} \bar{P}^r] \\
&= \frac{\mathcal{O}(1)}{N^3} \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N q^{p-1}} \\
&\quad \times \sum_{ij} \mathbb{E}[(2 - \delta_{ij}) \partial_{ij}^p (G_{ij}(m_N^2 + P' \partial_z m_N + m_N) P^{r-2} \bar{P}^r)] \\
&\quad + \mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{(N\eta)^2} |P'| |P|^{2r-2} \right] \right) \\
&= \mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{(N\eta)^2} |P'| |P|^{2r-2} \right] + \max_{s \geq 2} \mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right] \right),
\end{aligned}
\tag{2.79}
$$

where in the first term we used

$$
\left| \frac{1}{N^2} \sum_{ij} G_{ij}(D_{ij} m_N) \right| = \left| \frac{1}{N^3} \operatorname{Tr} G^3 \right| \leq \frac{1}{N^3} \operatorname{Tr} |G|^3 \leq \frac{\operatorname{Im}[m_N]}{(N\eta)^2},
\tag{2.80}
$$

and we used (2.51) for the second term. The same estimates hold for the second term in (2.78). The discussion above together implies the claimed estimate (2.57). $\quad\square$

PROOF OF CLAIM 2.12. We rewrite the left-hand side of (2.58) as

$$\frac{1}{N} \sum_{ij} \frac{\mathcal{C}_3}{Nq} \mathbb{E}[\partial_{ij}^2 (G_{ij} P^{r-1} \bar{P}^r)]$$

$$= \frac{1}{N} \sum_{ij} \frac{\mathcal{C}_3}{Nq} \mathbb{E}[\partial_{ij}^2 (G_{ij}) P^{r-1} \bar{P}^r]$$

(2.81)

$$+ \frac{1}{N} \sum_{ij} \frac{2\mathcal{C}_3}{Nq} \mathbb{E}[\partial_{ij} (G_{ij}) \partial_{ij} (P^{r-1} \bar{P}^r)]$$

$$+ \frac{1}{N} \sum_{ij} \frac{\mathcal{C}_3}{Nq} \mathbb{E}[G_{ij} \partial_{ij}^2 (P^{r-1} \bar{P}^r)].$$

For the first term in (2.81), thanks to Proposition 2.10 and (2.24), we have

(2.82)

$$\frac{1}{N} \sum_{ij} \frac{\mathcal{C}_3}{Nq} \mathbb{E}[\partial_{ij}^2 (G_{ij}) P^{r-1} \bar{P}^r] = \frac{1}{N} \sum_{ij} \frac{6\mathcal{C}_3}{Nq} \mathbb{E}[G_{ij} G_{ii} G_{jj} P^{r-1} \bar{P}^r]$$

$$+ \mathcal{O}_\prec \left(\mathbb{E}\left[\frac{\mathrm{Im}[m_N]}{N\eta} |P|^{2r-1} \right] \right).$$

We use the identities $G_{ij} = \sum_{k \neq i} G_{ii} h_{ik} G_{kj}^{(i)}$, and $G_{kj}^{(i)} = G_{kj} - G_{ki} G_{ji} / G_{ii}$. Then by the cumulant expansion, we have

$$\frac{1}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij} G_{ii} G_{jj} P^{r-1} \bar{P}^r]$$

$$= \frac{1}{N^2 q} \sum_{ij} \mathbb{E}\left[\sum_{k \neq i} h_{ik} G_{kj}^{(i)} G_{ii}^2 G_{jj} P^{r-1} \bar{P}^r \right]$$

$$= \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N^3 q^p} \sum_{ij} \mathbb{E}\left[\sum_{k \neq i} G_{kj}^{(i)} \partial_{ik}^p (G_{ii}^2 G_{jj} P^{r-1} \bar{P}^r) \right]$$

(2.83)

$$= \sum_{p \geq 1} \frac{\mathcal{C}_{p+1}}{N^3 q^p} \sum_{ij} \mathbb{E}\left[\sum_{k \neq i} G_{kj} G_{ii}^2 G_{jj} \partial_{ik}^p (P^{r-1} \bar{P}^r)) \right]$$

$$+ \mathcal{O}_\prec \left(\max_{s \geq 1} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right] \right)$$

$$= \mathcal{O}_\prec \left(\max_{s \geq 1} \mathbb{E}\left[\frac{1}{N\eta} \left(\frac{\mathrm{Im}[m_N] |P'|}{N\eta} \right)^s |P|^{2r-1-s} \right] \right.$$

$$\left. + \max_{s \geq 1} \mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right] \right),$$

where we used (2.24) in the third line and that $\|G\|_{L_2 \to L_2} \leq 1/\eta$, and hence

(2.84)

$$\left| \frac{1}{N^2} \mathbb{E}\left[\sum_{kj} G_{jj} G_{kj} \partial_{ik}^p (P^{r-1} \bar{P}^r) \right] \right|$$

$$\prec \max_{s \geq 1} \mathbb{E}\left[\frac{1}{N\eta} \left(\frac{\mathrm{Im}[m_N] |P'|}{N\eta} \right)^s |P|^{2r-s-1} \right].$$

For the second term in (2.81), using (2.49), it is given by

$$(2.85) \qquad -\sum_{ij} \frac{2\mathcal{C}_3}{N^2 q} \mathbb{E}[G_{ii} G_{jj} \partial_{ij}(P^{r-1}\bar{P}^r)] + \mathcal{O}_{\prec}\left(\mathbb{E}\left[\left(\frac{\mathrm{Im}[m_N]}{N\eta}\right)^2 |P|^{2r-2}\right]\right).$$

For the last term in (2.81), it is given by

$$
\frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(\partial^2_{ij}P)P^{r-2}\bar{P}^r)] + \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(\partial_{ij}P)^2 P^{r-3}\bar{P}^r)]
$$

$$(2.86)$$

$$
+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(\partial^2_{ij}\bar{P})^2 P^{r-1}\bar{P}^{r-2})]
$$

$$
+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(\partial^2_{ij}\bar{P})|P|^{2r-2}]
$$

$$
+ \frac{\mathcal{O}(1)}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(\partial_{ij}P)(\partial_{ij}\bar{P})P^{r-2}\bar{P}^{r-1}].
$$

We estimate the first and second terms; the other terms can be estimated in the same way. Using (2.50), (2.51) and (2.25), we can write the first term in (2.86) as

$$
\frac{1}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}\partial_{ij}(P'D_{ij}m_N + (2-\delta_{ij})h_{ij}
$$

$$
\times (m_N^2 + m_N + P'\partial_z m_N)/N)P^{r-2}\bar{P}^r]
$$

$$(2.87) \qquad = \frac{1}{N^2 q} \sum_{ij} \mathbb{E}[G_{ij}(P'(D^2_{ij}m_N + 2D_{ij}\partial_z m_N \partial_{ij}\tilde{z}) + P''(D_{ij}m_N)^2
$$

$$
+ (2-\delta_{ij})(m_N^2 + m_N + P'\partial_z m_N)/N)P^{r-2}\bar{P}^r]
$$

$$
+ \mathcal{O}_{\prec}\left(\mathbb{E}\left[\left(\left(\frac{\mathrm{Im}[m_N]}{N\eta}\right)^2 + \frac{\mathrm{Im}[m_N]|P'|}{(N\eta)^2}\right)|P|^{2r-2}\right]\right).
$$

We estimate the first term on the right-hand side of (2.87) in two parts. For the first part,

$$
\left|\frac{1}{N^2 q} \sum_{ij} G_{ij} P'(D^2_{ij}m_N)\right|
$$

$$
\leq \frac{\mathcal{O}(1)|P'|}{N^3 q}\left|\sum_{ijk} G_{ij}(G_{kj}G_{ii}G_{jk} + G_{ki}G_{ji}G_{jk})\right|
$$

$$(2.88)$$

$$
\leq \frac{\mathcal{O}(1)|P'|}{N^3 q}\left|\sum_{ij} G_{ii}G_{ij} \sum_{k} G^2_{kj}\right|
$$

$$
+ \frac{\mathcal{O}(1)|P'|}{N^3 q} \sum_{ijk}|G^2_{ij}| \sum_{k}|G_{ki}G_{jk}|
$$

$$
\prec \frac{\mathrm{Im}[m_N]|P'|}{q(N\eta)^2},
$$

where we used that $\|G\|_{L_2 \to L_2} \le 1/\eta$, and (2.24). For the second part, we have

$$
(2.89) \quad \frac{1}{N^2 q} \left| \sum_{ij} G_{ij} P' D_{ij} \partial_z m_N \partial_{ij} \tilde{z} \right| \le \frac{\mathcal{O}(1)}{N^4 q} \left| \sum_{ij} P' G_{ij} (G^3)_{ij} (2 - \delta_{ij}) h_{ij} \right|
$$

$$
\prec \frac{|P'|}{N^4 q^2} \sum_{ij} |(G^3)_{ij}| \prec \frac{\mathrm{Im}[m_N] |P'|}{(N\eta)^2}.
$$

For the second term on the right-hand side of (2.87), we have

$$
(2.90) \quad \left| \frac{1}{N^2 q} \sum_{ij} G_{ij} D_{ij} m_N D_{ij} m_N \right| \le \frac{1}{N^3 q} \sum_i \left| \sum_{jk} (G_{ij} D_{ij} m_N) G_{jk} (G_{ki}) \right|
$$

$$
\prec \frac{(\mathrm{Im}[m_N])^2}{q(N\eta)^3},
$$

where we used that $\|G\|_{L_2 \to L_2} \le 1/\eta$, $\sum_i |G_{ki}|^2 \prec \mathrm{Im}[m_N]/\eta$, and $\sum_i |G_{ij} D_{ij} \times m_N|^2 \prec (\mathrm{Im}[m_N])^3 / N\eta^2$. For the last term in (2.87), since $\|G\|_{L_2 \to L_2} \le 1/\eta$, we have

$$
(2.91) \quad \left| \frac{1}{N^2} \sum_{ij} G_{ij} \right| \le \frac{1}{N\eta}.
$$

It follows that the first term in (2.86) can be bounded by

$$
(2.92) \quad \mathcal{O}_\prec \left(\mathbb{E} \left[\left(\left(\frac{\mathrm{Im}[m_N]}{N\eta} \right)^2 + \frac{\mathrm{Im}[m_N] |P'|}{(N\eta)^2} \right) |P|^{2r-2} \right] \right).
$$

The second term in (2.86) can be rewritten as

$$
\frac{1}{N^2 q} \sum_{ij} \mathbb{E} \left[G_{ij} (\partial_{ij} P)^2 P^{r-3} \bar{P}^r) \right]
$$

$$
+ \frac{1}{N^2 q} \sum_{ij} \mathbb{E} \left[G_{ij} (P' D_{ij} m_N)(P' D_{ij} m_N) P^{r-3} \bar{P}^r) \right]
$$

$$
(2.93) \quad = \frac{1}{N^2 q} \sum_{ij} \frac{2 - \delta_{ij}}{N} \mathbb{E} \left[h_{ij} G_{ij} (P' \partial_z m_N + m_N^2 + m_N) \right.
$$

$$
\times \left. (P' D_{ij} m_N + \partial_{ij} P) P^{r-3} \bar{P}^r) \right].
$$

For the first term in (2.93), we use the estimate (2.90), and it follows that

$$
(2.94) \quad \left| \frac{1}{N^2 q} \sum_{ij} \mathbb{E} \left[G_{ij} (P' D_{ij} m_N)(P' D_{ij} m_N) P^{r-3} \bar{P}^r) \right] \right|
$$

$$
\prec \mathbb{E} \left[\frac{\mathrm{Im}[m_N]}{N\eta} \left(\frac{\mathrm{Im}[m_N] |P'|}{(N\eta)^2} \right) |P|^{2r-3} \right].
$$

For the second term in (2.93), we again use the cumulant expansion and find that it is given by

$$
\frac{\mathcal{C}_{p+1}}{N^3 q^{p-1}} \sum_{p \ge 1} \sum_{ij} \frac{2 - \delta_{ij}}{N} \mathbb{E} \left[\partial_{ij}^p (G_{ij} \right.
$$

$$
(2.95) \quad \times (P' \partial_z m_N + m_N^2 + m_N)(P' D_{ij} m_N + \partial_{ij} P) P^{r-3} \bar{P}^r)) \right]
$$

$$
= \mathcal{O}_\prec \left(\max_{s \ge 3} \mathbb{E} \left[\left(\frac{\mathrm{Im}[m_N]}{N\eta} \right)^s |P|^{2r-s} \right] \right).
$$

It follows that the second term in (2.86) is bounded by

$$
\mathcal{O}_{\prec}\left(\mathbb{E}\left[\frac{\operatorname{Im}[m_N]}{N\eta}\left(\frac{\operatorname{Im}[m_N]|P'|}{(N\eta)^2}\right)|P|^{2r-3}\right]\right.
$$
(2.96)
$$
\left.+\max_{s\geq 3}\mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta}\right)^s|P|^{2r-s}\right]\right).
$$

By combining the estimates (2.92) and (2.96), the last term in (2.81) can be bounded by

$$
\mathcal{O}_{\prec}\left(\max_{s=0,1}\mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta}\right)^s\left(\frac{\operatorname{Im}[m_N]|P'|}{(N\eta)^2}\right)|P|^{2r-2-s}\right]\right.
$$
(2.97)
$$
\left.+\max_{s\geq 2}\mathbb{E}\left[\left(\frac{\operatorname{Im}[m_N]}{N\eta}\right)^s|P|^{2r-s}\right]\right).
$$

The discussion above implies the claimed estimate (2.58). $\quad\square$

2.4. *Proof of Theorem* 2.1. We only analyze the behavior of the Stieltjes transform m_N in the region close to the right edge $\tilde{L}$. The analysis for m_N in the region close to the left edge can be done in the same way. Recall the shifted spectral domain $\mathcal{D}$ from (2.1). The following stability lemma is a variation of the usual stability lemma, for example, [8], Lemma 4.5.

LEMMA 2.13. *Let H be as in Definition* 1.1 *and $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$. There exists a constant $\varepsilon > 0$ such that the following holds. Suppose that $\delta : \mathcal{D} \to \mathbb{R}$ is a function so that*

$$
(2.98)\qquad\qquad |P(L+\mathcal{X}+z, m_N(L+\mathcal{X}+z))| \leq \delta(z).
$$

Suppose that $N^{-2} \leq \delta(z) \leq \varepsilon$ for $z \in \mathcal{D}$, that δ is Lipschitz continuous with Lipschitz constant N and moreover that for each fixed κ the function $\eta \mapsto \delta(\kappa + i\eta)$ is nonincreasing for $\eta > 0$. Then

$$
(2.99)\qquad |m_N(L+\mathcal{X}+z)-\tilde{m}_\infty(L+\mathcal{X}+z)| = \mathcal{O}\left(\frac{\delta(z)}{\sqrt{|\kappa|+\eta+\delta(z)}}\right),
$$

where the implicit constant is independent of z and N.

PROOF. We abbreviate $\tilde{z} = L + \mathcal{X} + z$. From (2.22), we have $|\partial_2 P(\tilde{z}, \tilde{m}_\infty(\tilde{z}))| \asymp \sqrt{|\kappa|+\eta}$, and $\partial_2^2 P(\tilde{z}, \tilde{m}_\infty(\tilde{z})) = 1 + \mathcal{O}(1/q^2)$. By a Taylor expansion, we have

$$
-P(\tilde{z}, m_N(\tilde{z})) + \partial_2 P(\tilde{z}, \tilde{m}_\infty(\tilde{z}))(m_N(\tilde{z})-\tilde{m}_\infty(\tilde{z}))
$$
(2.100)
$$
+(1+o(1))(m_N(\tilde{z})-\tilde{m}_\infty(\tilde{z}))^2 = 0.
$$

We abbreviate $\mathcal{R}(\tilde{z}) := P(\tilde{z}, m_N(\tilde{z}))$. There exists $a(\tilde{z}) \asymp \sqrt{|\kappa|+\eta}$ and $b(\tilde{z}) \asymp 1$, such that

$$
(2.101)\qquad \mathcal{R}(\tilde{z}) = a(\tilde{z})(m_N(\tilde{z})-\tilde{m}_\infty(\tilde{z})) + b(\tilde{z})(m_N(\tilde{z})-\tilde{m}_\infty(\tilde{z}))^2.
$$

With (2.101) in hand, Lemma 2.13 follows by a continuity argument essentially the same as [8], Lemma 4.5. $\quad\square$

Before proving Theorem 2.1, we first prove a weaker estimate of the Stieltjes transform $m_N(\tilde{z})$.

PROPOSITION 2.14. *Let H be as in Definition 1.1 and $q \geq N^{\mathfrak{e}}$ with arbitrarily small $\mathfrak{e} > 0$. Let $m_N(z)$ be the Stieltjes transform of its eigenvalue density, and $\tilde{m}_\infty(z)$ as defined in Proposition 2.6. Uniformly, for any $z = \kappa + i\eta \in \mathcal{D}$ as defined in (2.46), letting $\tilde{z} = L + \mathcal{X} + z$, we have*

$$\left| m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z}) \right| \prec \sqrt{|\kappa| + \eta}. \tag{2.102}$$

PROOF. For $|\kappa| + \eta \gg 1/q^3\sqrt{N}$, and $\kappa \leq 1$, by Proposition 2.6, we have

$$\mathrm{Im}\big[\tilde{m}_\infty(\tilde{z})\big] \asymp \Phi(z) := \begin{cases} \sqrt{|\kappa| + \eta} & \kappa \leq 0, \\ \eta/\sqrt{|\kappa| + \eta} & \kappa \geq 0 \end{cases} \tag{2.103}$$

and

$$\big|\partial_2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big)\big| \asymp \sqrt{|\kappa| + \eta}. \tag{2.104}$$

We denote

$$\Lambda_N(z) := \big| m_N(L + \mathcal{X} + z) - \tilde{m}_\infty(L + \mathcal{X} + z)\big|. \tag{2.105}$$

Then we have

$$\mathrm{Im}\big[m_N(\tilde{z})\big] \lesssim \Phi(z) + \Lambda_N(z), \tag{2.106}$$

and by Proposition 2.6

$$\begin{aligned} \partial_2 P\big(\tilde{z}, m_N(\tilde{z})\big) &= \partial_2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big) + \mathcal{O}\big(\big|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big|\big) \\ &= \mathcal{O}\big(\sqrt{|\kappa| + \eta} + \Lambda_N(z)\big). \end{aligned} \tag{2.107}$$

By Hölder's inequality, we obtain from Proposition 2.9,

$$\begin{aligned} \mathbb{E}\big[\big|P\big(\tilde{z}, m_N(\tilde{z})\big)\big|^{2r}\big] &\prec \frac{1}{(N\eta)^r}\Big(\frac{1}{q^3} + \frac{1}{N\eta}\Big)^r \mathbb{E}\big[\Lambda_N(z)^{2r} \\ &\quad + \big(|\kappa| + \eta\big)^{r/2}\big(\Phi(z)^r + \Lambda_N(z)^r\big)\big]. \end{aligned} \tag{2.108}$$

With overwhelming probability, we have the following Taylor expansion:

$$\begin{aligned} P\big(\tilde{z}, m_N(\tilde{z})\big) &= P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big) + \partial_2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big)\big(m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big) \\ &\quad + \frac{\partial_2^2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big) + o(1)}{2}\big(m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big)^2 \\ &= \partial_2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big)\big(m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big) \\ &\quad + \big(1 + o(1)\big)\big(m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big)^2, \end{aligned} \tag{2.109}$$

where we used that $\partial_2^2 P\big(\tilde{z}, \tilde{m}_\infty(\tilde{z})\big) = 2 + \mathcal{O}(1/q^2)$ and $\Lambda_N(z) \ll 1$ with overwhelming probability. Rearranging the last equation and using the definition of $\Lambda_N(z)$, we have arrived at

$$\Lambda_N(z)^2 \lesssim \Lambda_N(z)\sqrt{|\kappa| + \eta} + \big|P\big(\tilde{z}, m_N(\tilde{z})\big)\big|, \tag{2.110}$$

and thus

$$\mathbb{E}\big[\Lambda_N(z)^{4r}\big] \lesssim \big(|\kappa| + \eta\big)^r \mathbb{E}\big[\Lambda_N(z)^{2r}\big] + \mathbb{E}\big[\big|P\big(\tilde{z}, m_N(\tilde{z})\big)\big|^{2r}\big]. \tag{2.111}$$

We replace $\mathbb{E}[|P(\tilde{z}, m_N(\tilde{z}))|^{2r}]$ in (2.111) by (2.108). Moreover, on the domain $\mathcal{D}$, we have the estimates $N\eta\sqrt{|\kappa| + \eta} \gg 1$ and $N\eta(|\kappa| + \eta)q^3 \gg 1$, and so,

$$\mathbb{E}\big[\Lambda_N(z)^{4r}\big] \prec \big(|\kappa| + \eta\big)^{2r}, \tag{2.112}$$

and by Markov's inequality we get $\Lambda_N(z) \prec \sqrt{|\kappa| + \eta}$. $\quad\square$

PROOF OF THEOREM 2.1.   We assume that there exists some deterministic control parameter $\Lambda(z)$ such that the prior estimate holds

(2.113) $$|m_N(L + \mathcal{X} + z) - \tilde{m}_\infty(L + \mathcal{X} + z)| \prec \Lambda(z) \lesssim \sqrt{|\kappa| + \eta}.$$

Since $\Phi(z) \gtrsim \sqrt{|\kappa| + \eta}$ and $\Lambda(z) \prec \sqrt{|\kappa| + \eta}$, (2.108) simplifies to

(2.114) $$|P(\tilde{z}, m_N(\tilde{z}))| \prec \left(\frac{1}{q^{3/2}\sqrt{N\eta}} + \frac{1}{N\eta} \right) ((\Lambda(z) + \Phi(z))\sqrt{|\kappa| + \eta})^{1/2}.$$

If $\kappa \geq 0$, then $\Phi(z) = \eta/\sqrt{|\kappa| + \eta}$, and (2.114) simplifies to

(2.115)
$$|P(\tilde{z}, m_N(\tilde{z}))| \prec \frac{1}{N\eta^{1/2}} + \frac{1}{N^{1/2}q^{3/2}} + \frac{(|\kappa| + \eta)^{1/4}\Lambda^{1/2}}{N\eta}$$
$$+ \frac{(|\kappa| + \eta)^{1/4}\Lambda^{1/2}}{q^{3/2}(N\eta)^{1/2}}.$$

Thanks to Lemma 2.13, by taking $\delta(z)$ the right-hand side of (2.115) times N^ϵ, we have

(2.116)
$$|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})| \prec \frac{1}{\sqrt{|\kappa| + \eta}} \left(\frac{1}{N\eta^{1/2}} + \frac{1}{N^{1/2}q^{3/2}} \right.$$
$$\left. + \frac{(|\kappa| + \eta)^{1/4}\Lambda^{1/2}}{N\eta} + \frac{(|\kappa| + \eta)^{1/4}\Lambda^{1/2}}{q^{3/2}(N\eta)^{1/2}} \right).$$

By iterating (2.116), we get

(2.117)
$$|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})|$$
$$\prec \frac{1}{\sqrt{|\kappa| + \eta}} \left(\frac{1}{N\eta^{1/2}} + \frac{1}{N^{1/2}q^{3/2}} + \frac{1}{(N\eta)^2} + \frac{1}{q^3 N\eta} \right).$$

This completes the proof of (2.2).

If $\kappa \leq 0$, then $\Phi(z) = \sqrt{|\kappa| + \eta}$ and $\Lambda(z) \prec \sqrt{|\kappa| + \eta}$, (2.114) simplifies to

(2.118) $$|P(\tilde{z}, m_N(\tilde{z}))| \prec \frac{(|\kappa| + \eta)^{1/2}}{N\eta} + \frac{(|\kappa| + \eta)^{1/2}}{(N\eta)^{1/2}q^{3/2}}.$$

It follows from Lemma 2.13, by taking $\delta(z)$ the right-hand side of (2.118) times N^ϵ, we have

(2.119) $$|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})| \prec \frac{1}{N\eta} + \frac{1}{(N\eta)^{1/2}q^{3/2}}.$$

This completes the proof of (2.3).   □

3. Edge statistics of $H(t)$.   Let H be as in Definition 1.1. In this section, we assume that $q \gg N^{1/9}$, and consider the ensemble

(3.1) $$H(t) := e^{-t/2}H + (1 - e^{-t})^{1/2}W,$$

where $H(0) = H$ and W is an independent GOE matrix. The random matrix $H(t)$ also satisfies the properties of Definition 1.1. Thanks to Propositions 2.5 and 2.6, we can construct a random probability measure for $H(t)$, which is supported on $[-\tilde{L}_t, \tilde{L}_t]$, where

(3.2) $$\tilde{L}_t = L_t + \mathcal{X}_t + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N}q^3}\right), \quad \mathcal{X}_t = \frac{1}{N}\sum_{ij}\left(h_{ij}^2(t) - \frac{1}{N}\right).$$

We take $t \gg N^{-1/3}$ as

$$(3.3) \qquad t = \frac{N^{\mathfrak{d}}}{N^{1/3}},$$

for a $0 < \mathfrak{d} < 1/3$ to be determined. We denote $\mathcal{A}$ to be the set of sparse random matrices H, such that (2.2) and (2.3) hold at edges $\pm\tilde{L}$:

$$(3.4) \qquad \mathcal{A} := \big\{ H : (2.2) \text{ and } (2.3) \text{ hold around edges } \pm(L + \mathcal{X}) \big\}.$$

By Theorem 2.1, we know that the event $\mathcal{A}$ holds with probability $\mathbb{P}(\mathcal{A}) \geq 1 - N^{-D}$ for any $D \geq 0$. We denote the empirical eigenvalue distribution of H by μ_H. We denote the eigenvalues of $H(t)$ as $\lambda_1(t), \lambda_2(t), \ldots, \lambda_N(t)$.

In this section, we prove the following theorem, which states that the fluctuations of extreme eigenvalues of $H(t)$ are given by a combination of the Tracy–Widom distribution and the Gaussian distribution.

THEOREM 3.1. *Let H be as in Definition 1.1 with $q \gg N^{1/9}$. Let $H(t)$ be as in (3.1), with eigenvalues denoted by $\lambda_1(t), \lambda_2(t), \ldots, \lambda_N(t)$, and $t = N^{-1/3+\mathfrak{d}}$. Let $k \geq 1$ and $F : \mathbb{R}^k \to \mathbb{R}$ be a bounded test function with bounded derivatives. There is a universal constant $\mathfrak{c} > 0$ depending on $\mathfrak{d}$, so that for any $H \in \mathcal{A}$ as defined in (3.4) we have*

$$(3.5) \qquad \begin{aligned} &\mathbb{E}_W\big[F(N^{2/3}(\lambda_1(t) - L_t - \mathcal{X}), \ldots, N^{2/3}(\lambda_k(t) - L_t - \mathcal{X})|H\big] \\ &= \mathbb{E}_{\text{GOE}}[F(N^{2/3}(\mu_1 - 2), \ldots, N^{2/3}(\mu_k - 2)) + \mathcal{O}(N^{-\mathfrak{c}}), \end{aligned}$$

where the expectation on the right-hand side is with respect to a GOE matrix with eigenvalues denoted by μ_i. Moreover, one can also put the $\mathcal{X}$ on the right-hand side of (3.5),

$$(3.6) \qquad \begin{aligned} &\mathbb{E}_{H,W}\big[F(N^{2/3}(\lambda_1(t) - L_t), \ldots, N^{2/3}(\lambda_k(t) - L_t)\big] \\ &= \mathbb{E}_{H,\text{GOE}}[F(N^{2/3}(\mu_1 - 2 + \mathcal{X}), \ldots, N^{2/3}(\mu_k - 2 + \mathcal{X}))] + \mathcal{O}(N^{-\mathfrak{c}}), \end{aligned}$$

where expectation on the right-hand is with respect to a GOE matrix with eigenvalues μ_i, and a sparse random matrix H (note that on the right-hand side, H is independent of the GOE and only enters the expectation through $\mathcal{X}$).

Take $\eta_* = N^{-2/3+\mathfrak{d}/2}$, and $\tilde{z} = L + \mathcal{X} + z$, where $z = \kappa + i\eta$. For any $H \in \mathcal{A}$, from the defining relations of $\mathcal{A}$, that is, (2.2) and (2.3), we have

$$(3.7) \qquad \big|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big| \ll \frac{\eta}{\sqrt{\kappa + \eta}},$$

for $0 \leq \kappa \leq 1$ and $\eta_* \leq \eta \leq \mathfrak{b}$, and

$$(3.8) \qquad \big|m_N(\tilde{z}) - \tilde{m}_\infty(\tilde{z})\big| \ll \sqrt{|\kappa| + \eta},$$

for $-1 \leq \kappa \leq 0$ and $\eta_* \leq \eta \leq \mathfrak{b}$. Hence, combining with estimates (2.21), we get

$$(3.9) \qquad \text{Im}\big[m_N(\tilde{z})\big] \asymp \text{Im}\big[\tilde{m}_\infty(\tilde{z})\big], \qquad \eta_* \leq \eta \leq \mathfrak{b}, \qquad |\kappa| \leq 1,$$

for $H \in \mathcal{A}$. Moreover, in the regime $q \gg N^{-1/9}$, (2.2) also implies (1.11) such that $\lambda_1(0) - \tilde{L} \leq N^{-2/3+\mathfrak{d}/2}$. Hence, H is $\eta*$-regular in the sense of [44], Definition 2.1, and the result of [44], Theorem 2.2, applies for $t = N^{-1/3+\mathfrak{d}}$ as above. This result gives the limiting distribution of the extreme eigenvalues of H_t. The result involves scaling parameters coming from the free convolution, which we must now define.

We denote $\rho_{\text{sc}}(x)$ the semicircle law which is the limit eigenvalue density of a Gaussian orthogonal ensemble W. The semicircle law at time t, that is, the limit eigenvalue density

of $(1 - e^{-t})^{1/2} W$, is given by $(1 - e^{-t})^{-1/2} \rho_{\text{sc}}((1 - e^{-t})^{-1/2} x) \, dx$. We denote the free convolution of the empirical eigenvalue density of $e^{-t/2} H$, with the semicircle law at time t by $\hat{m}_{t,\text{fc}}$, and the free convolution of $e^{-t/2} \tilde{\rho}_\infty(e^{t/2} x)$ (as defined in Proposition 2.6) by $\tilde{m}_{t,\text{fc}}$. They satisfy the functional equations

$$(3.10) \qquad \tilde{m}_{t,\text{fc}}(z) = \int \frac{\tilde{\rho}_\infty(x) \, dx}{e^{-t/2} x - \tilde{\xi}(z)}, \qquad \hat{m}_{t,\text{fc}}(z) = \int \frac{d\mu_H(x)}{e^{-t/2} x - \hat{\xi}(z)},$$

where

$$(3.11) \qquad \tilde{\xi}(z) := z + (1 - e^{-t}) \tilde{m}_{t,\text{fc}}(z), \qquad \hat{\xi}(z) := z + (1 - e^{-t}) \hat{m}_{t,\text{fc}}(z).$$

For $t = N^{-1/3+\mathfrak{d}}$, these measures have densities which are supported on a single interval, with square root behavior at the edges. The edges are defined as follows. Let $\tilde{\xi}_+$ and $\hat{\xi}_+$ be the largest real solutions to

$$(3.12) \qquad \begin{aligned} 1 &= (1 - e^{-t}) \int \frac{\tilde{\rho}_\infty(x) \, dx}{(e^{-t/2} x - \tilde{\xi}_+)^2}, \\ 1 &= (1 - e^{-t}) \int \frac{d\mu_H(x)}{(e^{-t/2} x - \hat{\xi}_+)^2}. \end{aligned}$$

Then the edges $\tilde{E}_+$ and $\hat{E}_+$ of the free convolution with Stieltjes transforms $\tilde{m}_{t,\text{fc}}(z)$ and $\hat{m}_{t,\text{fc}}(z)$ are defined by $\tilde{\xi}_+ = \tilde{\xi}(\tilde{E}_+)$, $\hat{\xi}_+ = \hat{\xi}(\hat{E}_+)$, respectively. We introduce the scaling parameters $\hat{\gamma}_0$ and $\tilde{\gamma}_0$ as follows:

$$(3.13) \qquad \begin{aligned} \tilde{\gamma}_0 &:= \left(-(1 - e^{-t})^3 \int \frac{\tilde{\rho}_\infty(x) \, dx}{(e^{-t/2} - \tilde{\xi}_+)^3} \right)^{-1/3}, \\ \hat{\gamma}_0 &:= \left(-(1 - e^{-t})^3 \int \frac{d\mu_H(x)}{(e^{-t/2} - \hat{\xi}_+)^3} \right)^{-1/3}. \end{aligned}$$

The main result of [44], Theorem 2.2, states that for any η_*-regular H, and smooth test function $F : \mathbb{R}^k \mapsto \mathbb{R}$, there exists a universal constant $\mathfrak{c} > 0$ depending only on $\mathfrak{d} > 0$ as above, such that

$$(3.14) \qquad \begin{aligned} &\mathbb{E}_W \big[F\big(N^{2/3} \hat{\gamma}_0(\lambda_1(t) - \hat{E}_+), \ldots, N^{2/3} \hat{\gamma}_0(\lambda_k(t) - \hat{E}_+) \big) | H \big] \\ &= \mathbb{E}_{\text{GOE}} \big[F\big(N^{2/3}(\mu_1 - 2), \ldots, N^{2/3}(\mu_k - 2) \big) \big] + \mathcal{O}(N^{-\mathfrak{c}}), \end{aligned}$$

where the expectation on the right-hand side is with respect to a Gaussian orthogonal ensemble with largest few eigenvalues $\mu_1, \mu_2, \ldots, \mu_k$. The parameters $\hat{\gamma}_0$ and $\hat{E}_+$ depend on H. The following two propositions show that they are close to $\tilde{\gamma}_0$ and $\tilde{E}_+$, which we can calculate explicitly. The proofs are deferred to Appendix C.

PROPOSITION 3.2.   *Under the assumptions of Theorem* 3.1, *we have*

$$(3.15) \qquad e^{t/2} \hat{\xi}_+ - \tilde{L} \asymp t^2, \qquad e^{t/2} \tilde{\xi}_+ - \tilde{L} \asymp t^2,$$

and the estimate

$$(3.16) \qquad |\hat{\xi}_+ - \tilde{\xi}_+| \prec t^2 \left(\frac{1}{N^{1/2} q^{3/2} t^2} + \frac{1}{N t^3} \right).$$

For the edges of the free convolution law, we have

$$(3.17) \qquad |\hat{E}_+ - \tilde{E}_+| \prec \frac{1}{N^{1/2} q^{3/2}} + \frac{1}{N t}.$$

For the scaling parameter, we have

$$(3.18) \quad |\hat{\gamma}_0 - \tilde{\gamma}_0| = (1 - e^{-t})^3 e^{3t/2} \left| \int \frac{d\mu_H(x)}{(x - e^{t/2}\hat{\xi}_+)^3} - \int \frac{\tilde{\rho}_\infty(x)\,dx}{(x - e^{t/2}\tilde{\xi}_+)^3} \right|$$

$$\prec \frac{1}{Nt^3} + \frac{1}{N^{1/2}q^{3/2}t^2}.$$

PROPOSITION 3.3. *Under the assumptions of Theorem* 3.1, *there exists a universal constant* $\mathfrak{c} > 0$, *such that*

$$(3.19) \quad \tilde{E}_+ = 2 + \mathcal{X} + \frac{6\mathcal{C}_4}{q^2}(1 - 2t) + \frac{120\mathcal{C}_6}{q^4} - \frac{81\mathcal{C}_4^2}{q^4} + \mathcal{O}_\prec(N^{-2/3-\mathfrak{c}})$$

$$= L_t + \mathcal{X} + \mathcal{O}_\prec(N^{-2/3-\mathfrak{c}})$$

$$= \tilde{L}_t + \mathcal{O}_\prec(N^{-2/3-\mathfrak{c}})$$

and

$$(3.20) \quad \tilde{\gamma}_0 = 1 + \mathcal{O}_\prec(N^{-\mathfrak{c}}).$$

Thanks to Propositions 3.2 and 3.3, we can replace $\hat{\gamma}_0$ and $\hat{E}_+$ in (3.14) by $L_t + \mathcal{X}$ and 1 respectively, which gives an error of size $\mathcal{O}(N^{-\mathfrak{c}})$. Thus, we have

$$(3.21) \quad \mathbb{E}_W\big[F\big(N^{2/3}(\lambda_1(t) - L_t - \mathcal{X}), \ldots, N^{2/3}(\lambda_k(t) - L_t - \mathcal{X})\big)|H\big]$$

$$= \mathbb{E}_{\text{GOE}}\big[F\big(N^{2/3}(\mu_1 - 2), \ldots, N^{2/3}(\mu_k - 2)\big)\big] + \mathcal{O}(N^{-\mathfrak{c}}).$$

This completes the proof of (3.5).

In (3.5), we can take $F(x) = F_1(x + N^{2/3}\mathcal{X})$ for any smooth function F_1, as the estimates only depend on $\|F'\|_\infty$ and $\|F\|_\infty$. Taking expectation over H, we then see that

$$(3.22) \quad \mathbb{E}_{H,W}\big[F(N^{2/3}(\lambda_1(t) - L_t), \ldots, N^{2/3}(\lambda_k(t) - L_t)\big]$$

$$= \mathbb{E}_{H,\text{GOE}}\big[F\big(N^{2/3}(\mu_1 - 2 + \mathcal{X}), \ldots, N^{2/3}(\mu_k - 2 + \mathcal{X})\big)\big]$$

$$+ \mathcal{O}(N^{-\mathfrak{c}}).$$

This yields Theorem 3.1.

4. Comparison: Proof of Theorem 1.6. We recall $H(t)$ from (3.1). For simplicity of notation, in this section we denote the Stieltjes transform of the eigenvalue density of $H(t)$ as $m_t(z)$, and the Stieltjes transform of the eigenvalue density of H as $m(z)$. In this section, we prove the following theorem, which states that for $t = N^{-1/3+\mathfrak{d}}$, the rescaled extreme eigenvalues of H and $H(t)$ have the same distribution.

THEOREM 4.1. *Let H be as in Definition* 1.1 *with* $q \gg N^{1/9}$. *Let $H(t)$ be as in* (3.1), *with eigenvalues denoted by* $\lambda_1(t), \lambda_2(t), \ldots, \lambda_N(t)$, *and* $t = N^{-1/3+\mathfrak{d}}$ *with* $\mathfrak{d} > 0$ *sufficiently small. Fix* $k \geq 1$ *and numbers* $s_1, s_2, \ldots, s_k$, *there is a universal constant* $\mathfrak{c} > 0$ *so that*

$$(4.1) \quad \mathbb{P}_H\big(N^{2/3}(\lambda_i(0) - L - \mathcal{X}) \geq s_i, 1 \leq i \leq k\big)$$

$$= \mathbb{P}_{H(t)}\big(N^{2/3}(\lambda_i(t) - L_t - \mathcal{X}_t) \geq s_i, 1 \leq i \leq k\big) + \mathcal{O}(N^{-\mathfrak{c}}),$$

where L_t, $\mathcal{X}_t$ *are as defined in* (3.2). *In the special case,* $q = CN^{1/6}$, *we have*

$$(4.2) \quad \mathbb{P}_H\big(N^{2/3}(\lambda_i(0) - L) \geq s_i, 1 \leq i \leq k\big)$$

$$= \mathbb{P}_{H(t)}\big(N^{2/3}(\lambda_i(t) - L_t) \geq s_i, 1 \leq i \leq k\big) + \mathcal{O}(N^{-\mathfrak{c}}).$$

The analogous statement holds for the smallest eigenvalues.

Theorem 1.6 follows from combining Theorem 3.1 and Theorem 4.1.

PROOF OF THEOREM 1.6.   The result (1.14) follows from combining (3.5) and (4.1). The result (1.15) follows from combining (3.6) and (4.2).   □

Before proving Theorem 4.1, we need to introduce some notation. For any E, we define

$$(4.3) \qquad \mathcal{N}_t(E) := \big|\{i : \lambda_i(t) \geq L_t + \mathcal{X} + E\}\big|,$$

and we write $\mathcal{N}_0(E)$ as $\mathcal{N}(E)$. We fix $\mathfrak{c} > 0$, and take $\ell = N^{-2/3-\mathfrak{c}/3}$ and $\eta = N^{-2/3-\mathfrak{c}}$. Then with overwhelming probability, from (1.11), we know that $\lambda_1(t) \leq L_t + \mathcal{X}_t + N^{-2/3+\mathfrak{c}}$. We define:

$$\chi_E(x) = \mathbf{1}_{[E, N^{-2/3+\mathfrak{c}}]}(x - L_t - \mathcal{X}_t),$$

$$(4.4)$$

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2 + \eta^2)} = \frac{1}{\pi} \operatorname{Im} \frac{1}{x + i\eta}.$$

From the same argument as in [40], Lemma 2.7, we get that

$$(4.5) \qquad \operatorname{Tr}(\chi_{E+\ell} * \theta_\eta)(H(t)) - N^{-\mathfrak{c}/9} \leq \mathcal{N}_t(E) \leq \operatorname{Tr}(\chi_{E-\ell} * \theta_\eta)(H(t)) + N^{-\mathfrak{c}/9},$$

hold with overwhelming probability. Let $K_i : \mathbb{R} \mapsto [0, 1]$ be a monotonic smooth function satisfying,

$$(4.6) \qquad K_i(x) = \begin{cases} 0 & x \leq i - 2/3, \\ 1 & x \geq i - 1/3. \end{cases}$$

We have that $\mathbf{1}_{\mathcal{N}_t(E)\geq i} = K_i(\mathcal{N}_t(E))$, and since K_i is monotonically increasing, and so

$$K_i\big(\operatorname{Tr}(\chi_{E+\ell} * \theta_\eta)(H(t))\big) + \mathcal{O}(N^{-\mathfrak{c}/9})$$

$$(4.7) \qquad \leq \mathbf{1}_{\mathcal{N}_t(E)\geq i}$$

$$\leq K_i\big(\operatorname{Tr}(\chi_{E-\ell} * \theta_\eta)(H(t))\big) + \mathcal{O}(N^{-\mathfrak{c}/9}).$$

In this way, we can express the locations of eigenvalues in terms of the integrals of the Stieltjes transform of the empirical eigenvalue densities. We have

$$\mathbb{E}_{H(t)}\left[\prod_{i=1}^k K_i\left(\operatorname{Im}\left[\frac{N}{\pi}\int_{s_i N^{-2/3}+\ell}^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t + y + i\eta)\right]dy\right)\right] + \mathcal{O}(N^{-\mathfrak{c}/9})$$

$$(4.8) \qquad \leq \mathbb{P}_{H(t)}\big(N^{2/3}(\lambda_i(t) - L_t - \mathcal{X}) \geq s_i, 1 \leq i \leq k\big) = \mathbb{E}\left[\prod_{i=1}^k \mathbf{1}_{\mathcal{N}_t(s_i N^{-2/3})\geq i}\right]$$

$$\leq \mathbb{E}_{H(t)}\left[\prod_{i=1}^k K_i\left(\operatorname{Im}\left[\frac{N}{\pi}\int_{s_i N^{-2/3}-\ell}^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t + y + i\eta)\right]dy\right)\right] + \mathcal{O}(N^{-\mathfrak{c}/9}).$$

For the product of the functions of Stieltjes transform, we have the following comparison theorem.

PROPOSITION 4.2.   *Let H be as in Definition* 1.1, *with $q \gg N^{1/9}$. We fix $\mathfrak{c} > 0$, $E_1, E_2, \ldots, E_k = \mathcal{O}(N^{-2/3})$, $\eta = N^{-2/3+\mathfrak{c}}$ and $F : \mathbb{R}^k \mapsto \mathbb{R}$ a bounded test function with*

bounded derivatives. For $t \ll 1$, we have

$$\mathbb{E}_H\bigg[F\bigg(\bigg\{\operatorname{Im}\bigg[N\int_{E_i}^{N^{-2/3+\mathfrak{c}}} m(\tilde{L}+y+\mathrm{i}\eta)\bigg]\bigg\}_{i=1}^k\bigg)\bigg]$$

(4.9)
$$= \mathbb{E}_{H(t)}\bigg[\bigg\{F\bigg(\operatorname{Im}\bigg[N\int_{E_i}^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t+y+\mathrm{i}\eta)\bigg]\bigg\}_{i=1}^k\bigg)\bigg]$$

$$+ \mathcal{O}\bigg(N^{6\mathfrak{c}}\bigg(\frac{N^{2/3}t}{q^3}+\frac{N^{1/3}t}{q}+\frac{t^{1/2}N^{1/6}}{q}\bigg)\bigg).$$

In the special case, $q = CN^{1/6}$, we have

$$\mathbb{E}_H\bigg[F\bigg(\bigg\{\operatorname{Im}\bigg[N\int_{E_i}^{N^{-2/3+\mathfrak{c}}} m(L+y+\mathrm{i}\eta)\bigg]\bigg\}_{i=1}^k\bigg)\bigg]$$

(4.10)
$$= \mathbb{E}_{H(t)}\bigg[\bigg\{F\bigg(\operatorname{Im}\bigg[N\int_{E_i}^{N^{-2/3+\mathfrak{c}}} m_t(L_t+y+\mathrm{i}\eta)\bigg]\bigg\}_{i=1}^k\bigg)\bigg]$$

$$+ \mathcal{O}\bigg(N^{6\mathfrak{c}}\bigg(\frac{N^{2/3}t}{q^3}+\frac{N^{1/3}t}{q}+\frac{t^{1/2}N^{1/6}}{q}\bigg)\bigg).$$

PROOF OF THEOREM 4.1.   Since $q \gg N^{1/9}$ and $t = N^{-1/3+\mathfrak{d}}$, we can take $\mathfrak{c}$ and $\mathfrak{d}$ small, and then the error terms in (4.9) and (4.10) are of order $\mathcal{O}(N^{-\mathfrak{c}})$. By combining (4.8) and (4.9), we get

$$\mathbb{P}_{H(t)}\big(N^{2/3}\big(\lambda_i(t)-L_t-\mathcal{X}\big)\geq s_i+2N^{2/3}\ell, 1\leq i\leq k\big)+\mathcal{O}(N^{-\mathfrak{c}/9})$$

$$\leq \mathbb{E}_{H(t)}\bigg[\prod_{i=1}^k K_i\bigg(\operatorname{Im}\bigg[\frac{N}{\pi}\int_{s_i N^{-2/3}+\ell}^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t+y+\mathrm{i}\eta)\bigg]\mathrm{d}y\bigg)\bigg]+\mathcal{O}(N^{-\mathfrak{c}/9})$$

(4.11)
$$\leq \mathbb{P}_H\big(N^{2/3}\big(\lambda_i(0)-L-\mathcal{X}\big)\geq s_i, 1\leq i\leq k\big)$$

$$\leq \mathbb{E}_{H(t)}\bigg[\prod_{i=1}^k K_i\bigg(\operatorname{Im}\bigg[\frac{N}{\pi}\int_{s_i N^{-2/3}-\ell}^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t+y+\mathrm{i}\eta)\bigg]\mathrm{d}y\bigg)\bigg]+\mathcal{O}(N^{-\mathfrak{c}/9})$$

$$\leq \mathbb{P}_{H(t)}\big(N^{2/3}\big(\lambda_i(t)-L_t-\mathcal{X}\big)\geq s_i-2N^{2/3}\ell, 1\leq i\leq k\big)+\mathcal{O}(N^{-\mathfrak{c}/9}).$$

Since $N^{2/3}\ell = N^{-\mathfrak{c}/3} \ll 1$, (4.1) follows. For (4.2), analogous to (4.8), we have

$$\mathbb{E}_{H(t)}\bigg[\prod_{i=1}^k K_i\bigg(\operatorname{Im}\bigg[\frac{N}{\pi}\int_{s_i N^{-2/3}+\ell}^{N^{-2/3+\mathfrak{c}}} m_t(L_t+y+\mathrm{i}\eta)\bigg]\mathrm{d}y\bigg)\bigg]+\mathcal{O}(N^{-\mathfrak{c}/9})$$

(4.12)
$$\leq \mathbb{P}_{H(t)}\big(N^{2/3}\big(\lambda_i(t)-L_t\big)\geq s_i, 1\leq i\leq k\big)$$

$$\leq \mathbb{E}_{H(t)}\bigg[\prod_{i=1}^k K_i\bigg(\operatorname{Im}\bigg[\frac{N}{\pi}\int_{s_i N^{-2/3}-\ell}^{N^{-2/3+\mathfrak{c}}} m_t(L_t+y+\mathrm{i}\eta)\bigg]\mathrm{d}y\bigg)\bigg]+\mathcal{O}(N^{-\mathfrak{c}/9}).$$

Hence (4.2) follows from combining (4.12) and (4.10).   □

PROOF OF PROPOSITION 4.2.   For simplicity of notation, we only prove the case $k = 1$. The general case can be proved in the same way. Let

(4.13)
$$X_t = \operatorname{Im}\bigg[N\int_E^{N^{-2/3+\mathfrak{c}}} m_t(\tilde{L}_t+y+\mathrm{i}\eta)\,\mathrm{d}y\bigg].$$

We prove that

$$(4.14) \qquad |\mathbb{E}[F(X_t)] - \mathbb{E}[F(X_0)]| \lesssim N^{6\mathfrak{c}}\left(\frac{N^{2/3}t}{q^3} + \frac{N^{1/3}t}{q} + \frac{t^{1/2}N^{1/6}}{q}\right).$$

Similar to [47], Proposition 7.2,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}&\mathbb{E}[F(X_t)]\\
(4.15) \qquad &= \mathbb{E}\left[F'(X_t)\frac{\mathrm{d}X_t}{\mathrm{d}t}\right]\\
&= \mathbb{E}\left[F'(X_t)\operatorname{Im}\int_E^{N^{-2/3+\mathfrak{c}}}\left(\sum_{ijk}\dot{h}_{jk}(t)\frac{\partial G_{ii}}{\partial H_{jk}} + (\dot{L}_t + \dot{\mathcal{X}}_t)\sum_{ij}G_{ij}^2\right)\mathrm{d}y\right],
\end{aligned}
$$

where by definition

$$
\begin{aligned}
(4.16) \qquad \dot{h}_{jk}(t) &= -\frac{1}{2}e^{-t/2}h_{jk} + \frac{e^{-t}}{2\sqrt{1-e^{-t}}}w_{jk},\\
\dot{\mathcal{X}}(t) &= \frac{2}{N}\sum_{ij}h_{ij}(t)\dot{h}_{ij}(t).
\end{aligned}
$$

By a large deviation estimate, we have $|\dot{\mathcal{X}}_t| \prec 1/(t^{1/2}N^{1/2}q)$ and

$$(4.17) \qquad \mathbb{E}\left[F'(X_t)\operatorname{Im}\int_E^{N^{-2/3+\mathfrak{c}}}\dot{\mathcal{X}}_t\sum_{ij}G_{ij}^2\,\mathrm{d}x\right] \prec \frac{N^{1/6}}{t^{1/2}q}.$$

By the cumulant expansion formula, we get

$$(4.18) \qquad -\sum_{ijk}\mathbb{E}[\dot{h}_{jk}(t)F'(X_t)G_{ij}G_{ki}] = \sum_{p\geq 2}\frac{e^{-(p+1)t}C_p}{2Nq^{p-1}}\mathbb{E}[\partial_{jk}^p(F'(X_t)G_{ij}G_{ki})].$$

We notice that

$$(4.19) \qquad \dot{L}_t = \mp\frac{12C_4}{q^2} + \mathcal{O}\left(\frac{1}{q^4}\right).$$

Thus, (4.14) follows from the following proposition. The proof of (4.10) follows from the same argument as above, with replacing $\tilde{L}_t$ by L_t. $\square$

PROPOSITION 4.3. *Under the assumptions of Proposition* 4.2, *for any* $p \geq 2$, *let*

$$(4.20) \qquad J_p := \frac{e^{-(p+1)t}C_{p+1}}{2Nq^{p-1}}\mathbb{E}[\partial_{jk}^p(F'(X_t)G_{ij}G_{ki})].$$

Then, with overwhelming probability $J_2 = \mathcal{O}(N^{1+5\mathfrak{c}}/q)$,

$$(4.21) \qquad J_3 = \pm\frac{12C_4}{q^2}\sum_{ij}\mathbb{E}[F'(X_t)G_{ij}^2] + \mathcal{O}\left(\frac{N^{4/3+2\mathfrak{c}}}{q^3} + \frac{N^{1+5\mathfrak{c}}}{q^2}\right)$$

and for any $p \geq 4$, $J_p = \mathcal{O}(N^{4/3+2\mathfrak{c}}/q^3)$.

PROOF. Let $\tilde{z}_t = \tilde{L}_t + y + i\eta$. We have

$$
(4.22) \qquad |\partial_{ij}^p \tilde{z}_t| \lesssim \frac{1}{N},
$$

$$
|\partial_{ij}^p \tilde{m}_t(\tilde{z}_t)| \lesssim \frac{\text{Im}[\tilde{m}_t(\tilde{z}_t)]}{N\eta} \lesssim \frac{\sqrt{y+\eta}}{N\eta} \lesssim N^{-2/3+3\mathfrak{c}/2}, \quad p \geq 1
$$

and

$$
(4.23) \qquad \frac{1}{N^3} \sum_{ijk} |\partial_{jk}^p (G_{ij}(\tilde{z}_t) G_{ki}(\tilde{z}_t))| \lesssim \frac{\text{Im}[\tilde{m}_t(\tilde{z}_t)]}{N\eta} \lesssim N^{-2/3+3\mathfrak{c}/2}, \quad p \geq 0,
$$

with overwhelming probability. Therefore, we have that

$$
(4.24) \qquad \begin{aligned}
|\partial_{jk}^p X_t| &= \left| \text{Im}\left[N \int_E^{N^{-2/3+\mathfrak{c}}} \partial_{jk}^p \tilde{m}_t(\tilde{L}_t + y + i\eta)\, dy \right] \right| \\
&\lesssim N^{-1/3+5\mathfrak{c}/2}, \quad p \geq 1
\end{aligned}
$$

with overwhelming probability. For any monomial of Green's function with at least three off-diagonal terms, for example,

$$
(4.25) \qquad \begin{aligned}
&\frac{1}{N}\sum_{ijk} |G_{ij}^2 G_{jk} G_{ki}|, \qquad \frac{1}{N}\sum_{ijk} |G_{ij}^2 G_{jk}|, \qquad \frac{1}{N}\sum_{ijk} |G_{ij} G_{jk} G_{ki}|, \\
&\frac{1}{N}\sum_{ijk} |G_{ij}^2 G_{jk}^2|, \qquad \frac{1}{N}\sum_{ijk} |G_{ij}^2 G_{jk}^3|, \qquad \frac{1}{N}\sum_{ijk} |G_{ij} G_{ik} G_{jk}^3|
\end{aligned}
$$

are of order $\mathcal{O}(N^{1+3\mathfrak{c}})$ with overwhelming probability.

For any $p \geq 4$, using (4.23) and (4.24), we have the following bound:

$$
(4.26) \qquad J_p = \frac{\mathcal{O}(1)}{Nq^3} \sum_{ijk} \mathbb{E}[\partial_{jk}^p (F'(X_t) G_{ij} G_{ki})] \lesssim \frac{N^{4/3+2\mathfrak{c}}}{q^3},
$$

with overwhelming probability.

For $p = 2$, with overwhelming probability, we have

$$
\begin{aligned}
J_2 &= \frac{\mathcal{O}(1)}{Nq} \sum_{ijk} \mathbb{E}[\partial_{jk}^2 (F'(X_t) G_{ij} G_{ki})] \\
&= \frac{\mathcal{O}(1)}{Nq} \mathbb{E}[F'(X_t)\partial_{jk}^2 (G_{ij} G_{ki})] + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q}\right) \\
(4.27) \qquad &= \frac{\mathcal{O}(1)}{Nq} \sum_{ijk} \mathbb{E}[F'(X_t) G_{ij} G_{ki} G_{jj} G_{kk}] + \frac{1}{q}\{\text{terms in (4.25)}\} \\
&\quad + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q}\right) \\
&= \frac{\mathcal{O}(1)}{Nq} \sum_{ijk} \mathbb{E}[F'(X_t) G_{ij} G_{ki} G_{jj} G_{kk}] + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q}\right).
\end{aligned}
$$

For the first term, we use the identity $G_{ij} = \sum_{\ell \neq j} G_{jj} h_{j\ell} G_{\ell i}^{(j)}$, and the cumulant expansion

$$\frac{\mathcal{O}(1)}{Nq} \sum_{ijk} \mathbb{E}[F'(X_t) G_{ij} G_{ki} G_{jj} G_{kk}]$$

$$= \frac{\mathcal{O}(1)}{Nq} \sum_{ijk, \ell \neq j} \mathbb{E}[F'(X_t) G_{jj} h_{j\ell} G_{\ell i}^{(j)} G_{ki} G_{jj} G_{kk}]$$

(4.28)

$$= \sum_{p \geq 1} \frac{\mathcal{O}(1)}{N^2 q^p} \sum_{ijk, \ell \neq i} \mathbb{E}[G_{\ell i}^{(j)} \partial_{j\ell}^p (F'(X_t) G_{jj} G_{ki} G_{jj} G_{kk})]$$

$$= \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q}\right),$$

where in the last estimate we used (4.23), (4.24) and (4.25).

For $p = 3$, we have

$$J_3 = \frac{e^{-4t} C_4}{2Nq^2} \sum_{ijk} \mathbb{E}[F'(X_t) \partial_{jk}^3 (G_{ij} G_{ki})] + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q^2}\right)$$

$$= -\frac{12 e^{-4t} C_4}{Nq^2} \sum_{ijk} \mathbb{E}[F'(X_t) G_{ik} G_{jj} G_{kk} G_{jj} G_{ki})]$$

$$+ \frac{1}{q^2} \{\text{terms in (4.25)}\} + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q^2}\right)$$

(4.29)

$$= -\frac{12 e^{-4t} C_4}{Nq^2} \sum_{ijk} \mathbb{E}[F'(X_t) G_{ik} G_{jj} G_{kk} G_{jj} G_{ki})] + \mathcal{O}\left(\frac{N^{1+5\mathfrak{c}}}{q^2}\right)$$

$$= \frac{12 e^{-4t} C_4}{q^2} \sum_{ik} \mathbb{E}[F'(X_t) G_{ik} G_{ki})] + \mathcal{O}\left(\frac{N^{4/3+2\mathfrak{c}}}{q^3} + \frac{N^{1+5\mathfrak{c}}}{q^2}\right)$$

$$= \frac{12 C_4}{q^2} \sum_{ik} \mathbb{E}[F'(X_t) G_{ik} G_{ki})] + \mathcal{O}\left(\frac{N^{4/3+2\mathfrak{c}}}{q^3} + \frac{N^{1+5\mathfrak{c}}}{q^2}\right),$$

where in the second to last line we used that $G_{jj} = \tilde{m}_t(\tilde{z}) + \mathcal{O}_\prec(1/q + 1/N\eta) = -1 + \mathcal{O}_\prec(1/q)$, and (4.23). This completes the proof of Proposition 4.3. $\square$

APPENDIX A: WARD IDENTITY

Let H be a symmetric matrix, with Green's function G, then for any $z = E + i\eta$ with $\eta > 0$, we have the following Ward identity:

(A.1)
$$\sum_j |G_{ij}(z)|^2 = \frac{\text{Im}[G_{ii}(z)]}{\eta}.$$

The following follows from local semicircle estimates for sparse random matrices and the eigenvector delocalization.

PROPOSITION A.1. *Let H be as in Definition* 1.1 *and fix a large constant $\mathfrak{b} > 0$. Then uniformly for any $z = E + i\eta$ such that $-\mathfrak{b} \leq E \leq \mathfrak{b}$ and $1/N \ll \eta \leq \mathfrak{b}$, we have*

(A.2)
$$\max_{ij} |G_{ij}| = \mathcal{O}(1),$$

with overwhelming probability and

$$\text{(A.3)} \qquad \max_i \text{Im}[G_{ii}(z)] \prec \text{Im}[m_N(z)],$$

where G is the Green's function of the matrix H.

APPENDIX B: PROPERTIES OF THE EQUILIBRIUM MEASURES

The proof of Propositions 2.5 and 2.6 follow similar to [47], Lemma 4.1. We first prove Proposition 2.5.

PROOF OF PROPOSITION 2.5.   We introduce the following domains of the complex plane:

$$\text{(B.1)} \qquad \begin{aligned} \mathcal{D}_z &:= \{E + i\eta : -2 \le E \le 2, 0 \le \eta \le 2\}, \\ \mathcal{D}_w &:= \{w \in \mathbb{C} : |w| < 3\}. \end{aligned}$$

Let

$$\text{(B.2)} \qquad R(w) := -\frac{1}{w} - w - \frac{1}{q^2}(a_2 w^3 + a_3 w^5 + \cdots).$$

By definition, $P_0(z, w) = 0$ if and only if $z = R(w)$. We first check that $R(w)$ has exactly two critical points on $\mathcal{D}_w$. The derivatives of $R(w)$ are given by

$$\text{(B.3)} \qquad \begin{aligned} R'(w) &= \frac{1}{w^2} - 1 - \frac{1}{q^2}(3a_2 w^2 + 5a_3 w^4 + \cdots), \\ R''(w) &= -\frac{2}{w^3} - \frac{1}{q^2}(6a_2 w + 20a_3 w^3 + \cdots). \end{aligned}$$

For q large enough, $R''(w) < 0$ on $(0, 3)$ and $R''(w) > 0$ on $(-3, 0)$. Therefore, $R'(w)$ is monotonically decreasing on $(0, 3)$, and monotonically increasing on $(-3, 0)$. Furthermore, for C large enough, we have $R'(1 - C/q^2) > 0$ and $R'(1 + C/q^2) < 0$. $R'(w)$ is an even function and so $R'(w) = 0$ has two solutions on $(-3, 3)$, which we denote by $\pm\tau$. Note that $\tau \in (1 - C/q^2, 1 + C/q^2)$. We let $L = R(-\tau)$, and so $-L = R(\tau)$. We have that $L = 2 + \mathcal{O}(1/q^2)$. Next we show that $R(w)$ does not have other critical points on $\mathcal{D}_w$. The equation $R'(w) = 0$ is equivalent to

$$\text{(B.4)} \qquad 1 - w^2 - \frac{1}{q^2}(3a_2 w^4 + 5a_3 w^6 + \cdots) = 0.$$

On the boundary of $\mathcal{D}_w$, we have

$$\text{(B.5)} \qquad |1 - w^2| \ge 8 \gg \frac{1}{q^2}|3a_2 w^4 + 5a_3 w^6 + \cdots|.$$

Hence, by Rouché's theorem, the equation (B.4) has the same number of roots as the quadratic equation $1 - w^2 = 0$ on $\mathcal{D}_w$. Since $1 - w^2 = 0$ has two solutions on $\mathcal{D}_w$, we find that $R(w)$ has exactly two critical points, that is, $\pm\tau$ on $\mathcal{D}_w$.

We next show that for any $z \in \mathcal{D}_z$, $P_0(z, w)$ has exactly two (counting multiplicity) solutions on $\mathcal{D}_w$. On the boundary of $\mathcal{D}_w$, we have

$$\text{(B.6)} \qquad |w^2 + zw + 1| \ge |w|^2 - |z||w| - 1 \ge 2 \gg \frac{1}{q^2}|a_2 w^4 + a_3 w^6 + \cdots|.$$

Hence, by Rouché's theorem, the equation $P_0(z, w) = 0$ has the same number of roots as the quadratic equation $1 + zw + w^2 = 0$ on $\mathcal{D}_w$. Since $1 + zw + w^2 = 0$ has two solutions on $\mathcal{D}_w$,
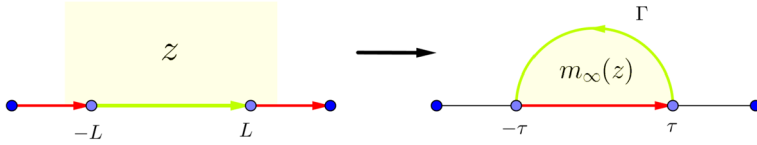
FIG. 1. m_∞ is a biholomorphic map from the upper half-plane to the region bounded by the curve Γ and the interval $[-\tau, \tau]$, where the boundary $[-L, L]$ is mapped to the curve Γ, $[-\infty, -L]$ is mapped to $[0, \tau]$, and $[L, \infty]$ is mapped to $[-\tau, 0]$.

we find that $P_0(z, w) = 0$ has exactly solutions (counting multiplicity) on $\mathcal{D}_w$. In particular, for any $z \in (-L, L)$, $P(z, w) = 0$ has two solutions $w(z) \in \mathbb{C}_+$ and $\bar{w}(z) \in \mathbb{C}_-$, and $w(z)$ forms a curve on $\mathbb{C}_+$, joining $-\tau$ and τ, which we denote by Γ.

Then we show that $R(w)$ is biholomorphic from the region $\mathcal{D}_\Gamma$, bounded by the curve Γ and the interval $[-L, L]$ to the upper half-plane, as in Figure 1. By definition (B.2), $R(w)$ is holomorphic on $\mathcal{D}_\Gamma$, and continuously extends to the boundary except at $w = 0$. In a small neighborhood of 0, the imaginary part of $R(w)$ satisfies

(B.7)
$$\mathrm{Im}[R(w)] = \mathrm{Im}[w]\left(\frac{1}{|w|^2} - 1\right) + \mathcal{O}\left(\frac{\mathrm{Im}[w]}{q^2}\right) > 0.$$

Hence, by maximum principle, we have $\mathrm{Im}[R(w)] > 0$ on $w \in \mathcal{D}_\Gamma$, that is, $R(w)$ maps $\mathcal{D}_\Gamma$ to the upper half-plane. If we view $R(w)$ as a map to the compactified complex plane, that is, $\mathbb{C} \cup \{\infty\}$, then $R(0) = \infty$. Moreover, it extends to an orientation preserving bijection between the boundary of $\mathcal{D}_\Gamma$ and the boundary of $\mathcal{C}_+$, where the boundary curve Γ is mapped to $[-L, L]$, $[0, \tau]$ is mapped to $[-\infty, -L]$, and $[-\tau, 0]$ is mapped to $[L, \infty]$. For any $z \in \mathbb{C}_+$, the number of solutions to $z = R(w)$ is given by

(B.8)
$$\frac{1}{2\pi i} \int_{\partial \mathcal{D}_\Gamma} \frac{R'(w)}{R(w) - z} \, dw = \frac{1}{2\pi i} \int_{\partial \mathbb{C}_+ \cup \{\infty\}} \frac{1}{\xi - z} \, d\xi = 1.$$

It follows $R(w)$ is a biholomorphic map from $\mathcal{D}_\Gamma$ to the upper half-plane. We take m_∞ to be the inverse of $R(w)$.

Finally, we check the properties of $m_\infty(z)$. The function $m_\infty(z)$ is a holomorphic map from the upper half-plane $\mathbb{C}_+$ to $\mathcal{D}_\Gamma$, satisfying $P_0(z, m_\infty(z)) = 0$. Let ρ_∞ be the measure obtained by Stieltjes inversion of $m_\infty(z)$. To show that ρ_∞ is a probability measure, it suffices to check that $\lim_{y \to \infty} iy m_\infty(iy) = -1$. Since $m_\infty(z)$ is bounded, one can check that $\lim_{z \to \infty} m_\infty(z) = 0$. Thus,

(B.9)
$$0 = \lim_{y \to \infty} P_0(z, m_\infty(iy)) = \lim_{y \to \infty} (1 + iy m_\infty(iy)),$$

which implies that $\lim_{y \to \infty} iy m_\infty(iy) = -1$, and ρ_∞ is a probability measure. Moreover,

(B.10)
$$\lim_{\eta \to 0} \mathrm{Im}[m_\infty(E + i\eta)] = 0,$$

for $E \notin [-L, L]$. Hence, ρ_∞ is a measure supported on $[-L, L]$.

In the following, we study the behaviors of m_∞ and ρ_∞ at the edges $\pm L$. Let $z = R(w)$. In a small neighborhood of the critical point $-\tau$, we have the expansion,

(B.11)
$$z = R(-\tau) + R'(-\tau)(w + \tau) + \frac{R''(-\tau)}{2}(w + \tau)^2(1 + \mathcal{E}(w))$$
$$= L + \frac{R''(-\tau)}{2}(w + \tau)^2(1 + \mathcal{E}(w)),$$

where the error term satisfies $\mathrm{Re}[\mathcal{E}(w)] = \mathcal{O}(|w + \tau|)$ and $\mathrm{Im}[\mathcal{E}[w]] = \mathcal{O}(\mathrm{Im}[w])$. Since $R''(-\tau) > 0$, we find that in a small neighborhood of $-\tau$,

$$(B.12) \qquad \left(\frac{2}{R''(-\tau)}\right)^{1/2} \sqrt{z - L} = (w + \tau)(1 + \mathcal{E}(w))^{1/2},$$

where we choose the branch of the square root so that $\sqrt{z - L} \in \mathbb{C}^+$. By taking imaginary part on both sides of (B.12), and noting that $\mathrm{Im}[(1 + \mathcal{E}(w))^{1/2}] = \mathcal{O}(\mathrm{Im}[w])$, we get

$$(B.13) \qquad \mathrm{Im}[\sqrt{z - L}] \asymp \mathrm{Im}[w](1 + \mathcal{O}(w + \tau)).$$

In a small neighborhood of $-\tau$, or equivalently, in a small neighborhood of L, we get

$$(B.14) \qquad \mathrm{Im}[m_\infty(z)] = \mathrm{Im}[w] \asymp \mathrm{Im}[\sqrt{z - L}] \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } \mathrm{Re}[z] \leq L, \\ \eta/\sqrt{\kappa + \eta} & \text{if } \mathrm{Re}[z] \geq L, \end{cases}$$

where $\kappa = |\mathrm{Re}[z] - z|$. In particular, by the Stieltjes inversion formula, ρ_∞ has square root behavior at edge L. For the derivative of $P_0(z, m_\infty(z))$, by the chain rule, we have

$$(B.15) \qquad 0 = \partial_z P_0(z, m_\infty(z)) = m_\infty(z) + \partial_z m_\infty(z) \partial_2 P_0(z, m_\infty(z)).$$

Since $m_\infty(z)$ has square root behavior at L, we have $|\partial_z m_\infty(z)| \asymp |z - L|^{-1/2}$, and

$$(B.16) \qquad |\partial_2 P_0(z, m_\infty(z))| = \frac{|m_\infty(z)|}{|\partial_z m_\infty(z)|} \asymp \sqrt{|z - L|}.$$

We can apply the same argument and get the same estimates of the behaviors of m_∞ and ρ_∞ in a small neighborhood of $-L$. This completes the proof of Proposition 2.5. $\quad\square$

PROOF OF PROPOSITION 2.6. We only prove the estimate (2.20). The rest of Proposition 2.6 follows from the same argument as in Proposition 2.5. Similar to the proof of Proposition 2.5, we define $\tilde{R}(w)$, $\tilde{\tau}$ and $\tilde{L} = \tilde{R}(-\tilde{\tau})$. Moreover, we have $\tilde{R}(w) = R(w) - \mathcal{X}w$. By a perturbation argument, we get

$$(B.17) \qquad \tilde{\tau} = \tau + \frac{\mathcal{X}}{R''(\tau)} + \mathcal{O}_\prec\left(\frac{1}{Nq^2}\right), \qquad \tilde{L} = L + \mathcal{X}\tau + \mathcal{O}_\prec\left(\frac{1}{Nq^2}\right).$$

Moreover, since $\tau = 1 + \mathcal{O}(1/q^2)$. The claim follows. $\quad\square$

APPENDIX C: FREE CONVOLUTION CALCULATIONS

In this appendix, we consider H as in Definition 1.1, with $q \gg N^{1/9}$. We will calculate various free convolution quantities. We let

$$(C.1) \qquad t := \frac{N^0}{N^{1/3}}.$$

We recall the construction of L, $\tilde{L}$, $\tilde{\rho}_\infty(x)$ and $\tilde{m}_\infty(z)$ from Propositions 2.5 and 2.6. We use the definitions of Section 3.

LEMMA C.1. *For all $E \asymp t^2$ and $k \geq 1$, we have*

$$(C.2) \qquad \left| \int \frac{1}{(x - (\tilde{L} + E))^k} \, d\mu_H(x) - \int \frac{1}{(x - (\tilde{L} + E))^k} \tilde{\rho}_\infty(x) \, dx \right|$$
$$\prec \frac{1}{t^{2(k-1)}} \left(\frac{1}{Nt^2} + \frac{1}{N^{1/2} t q^{3/2}} \right).$$

PROOF. Let $\chi(y)$ be a smooth cut-off function such that $\chi(y) = 1$ for $|y| \leq t^2$ and $\chi(y) = 0$ for $|y| \geq 2t^2$, such that $|\chi'(y)| \leq Ct^{-2}$. Let $f(x)$ be a cut-off function so that $f(x) = 1$ for $x \leq \tilde{L} + E/3$ and $f(x) = 0$ for $x \geq \tilde{L} + 2E/3$. We choose $f(x)$ so that the derivatives obey $|f^{(k)}(x)| \leq Ct^{-2k}$. Since by (1.11) with overwhelming probability all the eigenvalues of H are less than $\tilde{L} + N^{-2/3+\varepsilon}$ for any $\varepsilon > 0$, we can consider instead the test function

$$\text{(C.3)} \qquad g(x) := f(x)\frac{1}{(x - (\tilde{L} + E))^k}.$$

Let

$$\text{(C.4)} \qquad S(x + iy) := m_N(x + iy) - \tilde{m}_\infty(x + iy).$$

Define $\eta_* := N^{-2/3+\varepsilon}$ for $\varepsilon > 0$, with $\varepsilon < \mathfrak{d}/4$, so that $\eta_* < t^2$. Note that for $y \geq \eta_*$ we have by Theorem 2.1 that

$$\text{(C.5)} \qquad |S(x + iy)| \prec \frac{1}{Ny} + \frac{1}{(Ny)^{1/2}q^{3/2}} =: \Phi(y),$$

for $q \geq N^{1/9}$ regardless of whether $x \geq \tilde{L}$ or $x \leq \tilde{L}$. By the Helffer–Sjöstrand formula,

$$
\text{(C.6)} \qquad
\begin{aligned}
&\left| \int g(x)\, d\mu_{H_0}(x) - \int g(x)\tilde{\rho}_\infty(x)\, dx \right| \\
&\qquad \leq \left| \int g''(x)\chi(y)y\, \mathrm{Im}[S(x + iy)]\, dx\, dy \right| \\
&\qquad + \int |g(x)\chi'(y)\, \mathrm{Im}[S(x + iy)]|\, dx\, dy \\
&\qquad + \int |g'(x)\chi'(y)y\, \mathrm{Re}[S(x + iy)]|\, dx\, dy.
\end{aligned}
$$

For the second two terms, note that

$$
\text{(C.7)} \qquad
\begin{aligned}
\int |g(x)|\, dx &\leq \frac{C\log(N)}{t^{2k-2}}, \qquad \int |g'(x)|\, dx \leq \frac{C}{t^{2k}}, \\
\int |g''(x)|\, dx &\leq \frac{C}{t^{2k+2}},
\end{aligned}
$$

and that the y integral is only over $t^2 < |y| < 2t^2$. So,

$$
\text{(C.8)} \qquad
\begin{aligned}
&\int |g(x)\chi'(y)\, \mathrm{Im}[S(x + iy)]|\, dx\, dy \\
&\qquad + \int |g'(x)\chi'(y)y\, \mathrm{Re}[S(x + iy)]|\, dx\, dy \prec \frac{\Phi(t^2)}{t^{2k-2}}.
\end{aligned}
$$

For the last term, we first use the fact that $y \to y\, \mathrm{Im}[m_N]$ is increasing to bound the $|y| \leq \eta_*$ integral by

$$
\text{(C.9)} \qquad
\begin{aligned}
&\left| \int_{|y| \leq \eta_*} g''(x)\chi(y)y\, \mathrm{Im}[S(x + iy)]\, dx\, dy \right| \\
&\qquad \leq 2\int_{y \leq \eta_*} |g''(x)|\eta_*\, \mathrm{Im}[\tilde{m}_\infty(x + i\eta_*)]\, dx\, dy \\
&\qquad + \int_{y \leq \eta_*} |g''(x)|\eta_*|S(x + i\eta_*)|\, dx\, dy.
\end{aligned}
$$

The second term is bounded by

$$
\text{(C.10)} \qquad \int_{y \leq \eta_*} |g''(x)| \eta_* |S(x + i\eta_*)| \, dx \, dy \prec \frac{\Phi(\eta_*)\eta_*^2}{t^{2k+2}} \lesssim \frac{\Phi(t^2)}{t^{2k-2}},
$$

where we used $\eta_* \leq t^2$. We now estimate the first term of (C.9). Using (2.21), we find

$$
\int_{y \leq \eta_*} |g''(x)| \eta_* \operatorname{Im}[m(x + i\eta_*)] \, dx \, dy
$$

$$
\text{(C.11)} \qquad \lesssim \int_{x<0} |g''(x + \tilde{L})| \eta_*^2 \sqrt{\eta_* + |x|} \, dx
$$

$$
+ \int_{x \geq 0} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{\eta_* + x}} \, dx.
$$

We note that $|g''(x + \tilde{L})| \lesssim C|x - E|^{-(k+2)} \mathbf{1}_{x \leq 2E/3}$. The second integral is bounded by

$$
\int_{x \geq 0} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{\eta_* + x}} \, dx \leq \int_{0 < x < \eta_*} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{\eta_* + x}} \, dx
$$

$$
\text{(C.12)} \qquad\qquad + \int_{x \geq \eta_*} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{x}} \, dx
$$

$$
\lesssim \frac{\eta_*^{3.5}}{t^{2k+4}} + \int_{x \geq \eta_*} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{x}} \, dx.
$$

The last term is bounded by

$$
\int_{x \geq \eta_*} |g''(x + \tilde{L})| \frac{\eta_*^3}{\sqrt{x}} \, dx \lesssim \int_{\eta_* \leq x < E/3} \frac{1}{|x - E|^{k+2}} \frac{\eta_*^3}{\sqrt{x}} \, dx
$$

$$
\text{(C.13)} \qquad\qquad + \int_{E/3 < x < 2E/3} \frac{1}{t^4 |x - E|^k} \frac{\eta_*^3}{\sqrt{x}} \, dx
$$

$$
\lesssim \frac{\eta_*^3}{t^{2k+3}}.
$$

For the other term on the right-hand side of (C.11), similarly we get

$$
\text{(C.14)} \qquad \int_{x<0} |g''(x + \tilde{L})| \eta_*^2 \sqrt{\eta_* + |x|} \, dx \lesssim \frac{\eta_*^{2.5}}{t^{2k+2}} + \frac{\eta_*^2}{t^{2k+1}}.
$$

In summary, we have the estimate

$$
\text{(C.15)} \qquad \left| \int_{y \leq \eta_*} g''(x) \chi(y) y \operatorname{Im}[S(x + iy)] \, dx \, dy \right| \prec \frac{\Phi(t^2)}{t^{2k-2}} + \frac{1}{t^{2k-2}} \frac{\eta_*^2}{t^3}
$$

$$
\prec \frac{\Phi(t^2)}{t^{2k-2}},
$$

where we used that $\eta_*^2/t^3 = N^{2\varepsilon}/(N^{4/3} t^3) \leq 1/(Nt^2)$ by our choice of ε above. Finally,

$$
\int_{y > \eta_*} |g''(x)| y \chi(y) \operatorname{Im}[S(x + iy)] \, dx \, dy
$$

$$
\text{(C.16)} \qquad\qquad \lesssim \frac{1}{t^{2k+2}} \int_{|y| \leq 2t^2} y \Phi(y) \, dy \lesssim \frac{\Phi(t^2)}{t^{2k-2}},
$$

where we used the fact that the support of χ is in $|y| \le 2t^2$. Putting together our estimates yields the claim. $\square$

PROOF OF LEMMA 3.2.   From the square root behavior of $\tilde{m}_\infty$ we easily conclude that $e^{t/2}\tilde{\xi}_+ - \tilde{L} \asymp t^2$. Then, using (C.2) we conclude the same for $\hat{\xi}_+$. By taking difference of the defining equations (3.12) for $\tilde{\xi}_+$ and $\hat{\xi}_+$, we get

(C.17)
$$
\int \frac{d\mu_H(x)}{(x - e^{t/2}\hat{\xi}_+)^2} - \int \frac{\tilde{\rho}_\infty(x)\,dx}{(x - e^{t/2}\hat{\xi}_+)^2}
$$
$$
= \int \left(\frac{1}{(x - e^{t/2}\tilde{\xi}_+)^2} - \frac{1}{(x - e^{t/2}\hat{\xi}_+)^2} \right) \tilde{\rho}_\infty(x)\,dx.
$$

We calculate the right-hand side

(C.18)
$$
\int \left(\frac{1}{(x - e^{t/2}\tilde{\xi}_+)^2} - \frac{1}{(x - e^{t/2}\hat{\xi}_+)^2} \right) \tilde{\rho}_\infty(x)\,dx
$$
$$
= \int \frac{e^{t/2}(\tilde{\xi}_+ - \hat{\xi}_+))(2x - e^{t/2}\tilde{\xi}_+ - e^{t/2}\hat{\xi}_+)}{(x - e^{t/2}\tilde{\xi}_+)^2(x - e^{t/2}\hat{\xi}_+)^2} \tilde{\rho}_\infty(x)\,dx
$$
$$
= e^{t/2}(\tilde{\xi}_+ - \hat{\xi}_+) \int \frac{1}{(x - e^{t/2}\hat{\xi}_+)(x - e^{t/2}\tilde{\xi}_+)^2}
$$
$$
+ \frac{1}{(x - e^{t/2}\tilde{\xi}_+)(x - e^{t/2}\hat{\xi}_+)^2} \tilde{\rho}_\infty(x)\,dx.
$$

Since both $e^{t/2}\hat{\xi}_+ - \tilde{L} \asymp e^{t/2}\tilde{\xi}_+ - \tilde{L} \asymp t^2$, we see that the integrand is negative, and moreover, by the square root behavior of $\tilde{\rho}_\infty$,

(C.19)
$$
\left| \int \frac{1}{(x - e^{t/2}\hat{\xi}_+)(x - e^{t/2}\tilde{\xi}_+)^2} + \frac{1}{(x - e^{t/2}\tilde{\xi}_+)^2(x - e^{t/2}\hat{\xi}_+)} \tilde{\rho}_\infty(x)\,dx \right|
$$
$$
\asymp \frac{1}{t^3}.
$$

We immediately conclude (3.16) from this and the estimate (C.2) for the left-hand side of (C.17). Next, we estimate

(C.20) $|\hat{E}_+ - \tilde{E}_+| \le |\hat{\xi}_+ - \tilde{\xi}_+| + Ct|\tilde{m}_\infty(e^{t/2}\tilde{\xi}_+) - m_N(e^{t/2}\hat{\xi}_+)|.$

The latter quantity we estimate by

(C.21)
$$
|\tilde{m}_\infty(e^{t/2}\tilde{\xi}_+) - m_N(e^{t/2}\hat{\xi}_+)| \le |\tilde{m}_\infty(e^{t/2}\tilde{\xi}_+) - \tilde{m}_\infty(e^{t/2}\hat{\xi}_+)|
$$
$$
+ |\tilde{m}_\infty(e^{t/2}\hat{\xi}_+) - m_N(e^{t/2}\hat{\xi}_+)|
$$
$$
\prec t \left(\frac{1}{N^{1/2}t^2q^{3/2}} + \frac{1}{Nt^3} \right),
$$

where we used (C.2). Hence, we have proven (3.17). Finally, the estimate (3.18) is an easy consequence of (3.16) and (C.2). $\square$

LEMMA C.2.   *Let the polynomial of a self-consistent equation be given by*

(C.22) $0 = 1 + zm + m^2 + \dfrac{a}{q^2}m^4 + \dfrac{b}{q^4}m^6 + \mathcal{X}m^2 + P_1(m),$

where $P_1(m)$ is a polynomial in m of size $\mathcal{O}(q^{-6})$, and constants $a, b = \mathcal{O}(1)$, and m be the Stieltjes transform of a probability measure, supported on $[-\tilde{L}, \tilde{L}]$, with

(C.23) $$\tilde{L} = 2 + \frac{a}{q^2} + \frac{b}{q^4} - \frac{9a^2}{4q^4} + \mathcal{X} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}).$$

For any z in a small neighborhood of $\tilde{L}$, $m(z)$ is given by

(C.24) $$m(z) = -\tau + \left(\frac{2}{F''(-\tau)}\right)^{1/2} (z - \tilde{L})^{1/2} + \mathcal{O}(|z - \tilde{L}|),$$

where

(C.25) $$\tau = 1 - \frac{3a}{q^2} - \frac{5b}{q^4} + \frac{18a^2}{q^4} - \mathcal{X} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}).$$

PROOF. We have the polynomial

(C.26) $$0 = 1 + zm + (1 + \mathcal{X})m^2 + \frac{a}{q^2}m^4 + \frac{b}{q^4}m^6 + P_1(m).$$

Solving for z,

(C.27) $$z = -\frac{1}{m} - m - \frac{a}{q^2}m^3 - \frac{b}{q^4}m^5 - \mathcal{X}m - P_2(m) := F(m),$$

where $P_2(m) = P_1(m)/m$ is another polynomial of size $\mathcal{O}(q^{-6})$. There is a unique positive solution to $F'(-\tau) = 0$. We record the expressions

(C.28) $$F'(m) = \frac{1}{m^2} - 1 - \frac{3a}{q^2}m^2 - \frac{5b}{q^4}m^4 - \mathcal{X} - P_2'(m)$$

and

(C.29) $$F''(m) = -\frac{2}{m^3} - \frac{6a}{q^2}m - \frac{20b}{q^4}m^3 - P_2''(m).$$

We first find

(C.30) $$\tau^2 = 1 - \mathcal{X} - \frac{3a}{q^2} - \frac{5b}{q^4} + \frac{18a^2}{q^4} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}),$$

and subsequently, we can expand to obtain

(C.31) $$\tau = 1 - \frac{3a}{2q^2} - \frac{5b}{2q^4} + \frac{63a^2}{8q^4} - \frac{\mathcal{X}}{2} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}).$$

Substituting this in $F(-\tau)$, we get

(C.32) $$\tilde{L} = 2 + \frac{a}{q^2} + \frac{b}{q^4} - \frac{9a^2}{4q^4} + \mathcal{X} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}).$$

We see that

(C.33) $$F''(-\tau) = 2 + \mathcal{O}(q^{-2}).$$

We expand

(C.34) $$z = \tilde{L} + \frac{F''(-\tau)}{2}(m - \tau)^2 + \mathcal{O}(|m - \tau|^3)$$

and invert

(C.35)
$$m(z) = -\tau + \left(\frac{2}{F''(-\tau)}\right)^{1/2}(z - \tilde{L})^{1/2} + \mathcal{O}(|z - \tilde{L}|).$$
□

PROOF OF LEMMA 3.3.   In the following proof, we write $m = \tilde{m}_\infty$, $\xi = \tilde{\xi}_+$ and $E_+ = \tilde{E}_+$. Recall from (2.16), m satisfies a self-consistent equation,

(C.36)
$$0 = 1 + zm + m^2 + \frac{6\mathcal{C}_4}{q^2}m^4 + \frac{120\mathcal{C}_6}{q^4}m^6 + \mathcal{X}_t m^2 + \mathcal{O}(q^{-6}).$$

Then by Lemma C.2, we have

(C.37)
$$\tilde{L} = 2 + \frac{6\mathcal{C}_4}{q^2} + \frac{120\mathcal{C}_6}{q^4} - \frac{81\mathcal{C}_4^2}{q^4} + \mathcal{X} + \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6}).$$

In the following, we first compute $E_+ = \xi - (1 - e^{-t})e^{t/2}m(e^{t/2}\xi)$. For ξ (as defined in (3.12)), by Lemma C.2 and (C.33), we have

(C.38)
$$\frac{e^{-t}}{1 - e^{-t}} = m'(e^{t/2}\xi) = \frac{1}{(4 + \mathcal{O}(q^{-2}))^{1/2}}\frac{1}{\sqrt{e^{t/2}\xi - \tilde{L}}} + \mathcal{O}(1).$$

We can solve for ξ and get

(C.39)
$$e^{t/2}\xi = \tilde{L} + \frac{1}{4}(e^t - 1)^2 + \mathcal{O}(t^3 + t^2 q^{-2}).$$

Using Lemma 2.16, we compute

(C.40)
$$m(e^{t/2}\xi) = -1 + \frac{18\mathcal{C}_4}{q^2} + (1 + \mathcal{O}(q^{-2}))^{1/2}(e^{t/2}\xi - \tilde{L})^{1/2} + \mathcal{O}(t^2)$$
$$= -1 + \frac{9\mathcal{C}_4}{q^2} + \frac{e^t - 1}{2} + \mathcal{O}(t^2 + q^{-4}).$$

Hence,

(C.41)
$$E_+ = e^{-t/2}\tilde{L} + \frac{e^{-t/2}(e^t - 1)^2}{4} - (1 - e^{-t})e^{t/2}\left(-1 + \frac{9\mathcal{C}_4}{q^2} + \frac{e^t - 1}{2}\right)$$
$$+ \mathcal{O}(t^3 + tq^{-4} + t^2 q^{-2})$$
$$= 2 + \frac{6(1 - 2t)\mathcal{C}_4}{q^2} + \frac{120\mathcal{C}_6}{q^2} - \frac{81\mathcal{C}_4^2}{q^4} + \mathcal{X}$$
$$+ \mathcal{O}(\mathcal{X}q^{-2} + \mathcal{X}^2 + q^{-6} + t\mathcal{X} + t^3 + tq^{-4} + t^2 q^{-2})$$
$$= 2 + \frac{6(1 - 2t)\mathcal{C}_4}{q^2} + \frac{120\mathcal{C}_6}{q^2} - \frac{81\mathcal{C}_4^2}{q^4} + \mathcal{X}$$
$$+ \mathcal{O}_\prec(t^2 q^{-2} + tq^{-4} + q^{-6} + tN^{-1/2}q^{-1} + N^{-1/2}q^{-3} + t^{1/2}N^{-1}),$$

where we used that $\mathcal{X} \prec N^{-1/2}q^{-1}$.

The kth cumulant of the entries of $H(t)$ (as defined in (3.1)) is $1/N$ when $k = 2$, and

$$(C.42) \qquad \frac{(k-1)! e^{-tn/2} \mathcal{C}_k}{N q^{k-2}},$$

for $k \geq 3$. The self-consistent equation for $H(t)$ is given by

$$(C.43) \qquad 0 = 1 + z m_t + m_t^2 + \frac{6 e^{-2t} \mathcal{C}_4}{q^2} m_t^4 + \frac{120 e^{-3t} \mathcal{C}_6}{q^4} m_t^6 + \mathcal{X}_t m_t^2 + \mathcal{O}(q^{-6}).$$

By Lemma C.2, we have

$$(C.44) \qquad \begin{aligned} \tilde{L}_t &= 2 + \frac{6 e^{-2t} \mathcal{C}_t}{q^2} + \frac{120 e^{-3t} \mathcal{C}_6}{q^4} - \frac{81 e^{-4t} \mathcal{C}_4}{q^4} + \mathcal{X}_t \\ &\quad + \mathcal{O}(\mathcal{X}_t q^{-2} + \mathcal{X}_t^2 + q^{-6}). \end{aligned}$$

We note that

$$(C.45) \qquad \mathcal{X}_t = \mathcal{X} + \mathcal{O}_{\prec}(t N^{-1/2} q^{-1} + t^{1/2} N^{-1}) = \mathcal{O}_{\prec}(N^{-1/2} q^{-1}),$$

and rewrite (C.44) as

$$(C.46) \qquad \begin{aligned} \tilde{L}_t &= 2 + \frac{6 \mathcal{C}_4}{q^2}(1 - 2t) + \frac{120 \mathcal{C}_6}{q^4} - \frac{81 \mathcal{C}_4^2}{q^4} + \mathcal{X} \\ &\quad + \mathcal{O}_{\prec}(t^2 q^{-2} + t q^{-4} + q^{-6} + t N^{-1/2} q^{-1} \\ &\quad + N^{-1/2} q^{-3} + t^{1/2} N^{-1}). \end{aligned}$$

Lemma 3.3 follows from comparing (C.41) and (C.46), and recalling $q \gg N^{1/9}$. $\quad\square$

REFERENCES

[1] AJANKI, O., ERDŐS, L. and KRÜGER, T. (2017). Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Comm. Pure Appl. Math.* **70** 1672–1705. MR3684307 https://doi.org/10.1002/cpa.21639

[2] AJANKI, O. H., ERDŐS, L. and KRÜGER, T. (2016). Local spectral statistics of Gaussian matrices with correlated entries. *J. Stat. Phys.* **163** 280–302. MR3478311 https://doi.org/10.1007/s10955-016-1479-y

[3] AJANKI, O. H., ERDŐS, L. and KRÜGER, T. (2017). Universality for general Wigner-type matrices. *Probab. Theory Related Fields* **169** 667–727. MR3719056 https://doi.org/10.1007/s00440-016-0740-2

[4] BAUERSCHMIDT, R., HUANG, J., KNOWLES, A. and YAU, H.-T. (2019). Edge rigidity and universality of random regular graphs of intermediate degree. Preprint. Available at arXiv:1910.10121.

[5] BELINSCHI, S. T. and CAPITAINE, M. (2017). Spectral properties of polynomials in independent Wigner and deterministic matrices. *J. Funct. Anal.* **273** 3901–3963. MR3711884 https://doi.org/10.1016/j.jfa.2017.07.010

[6] BENAYCH-GEORGES, F., BORDENAVE, C. and KNOWLES, A. (2017). Spectral radii of sparse random matrices. Preprint. Available at arXiv:1704.02945.

[7] BENAYCH-GEORGES, F., BORDENAVE, C. and KNOWLES, A. (2019). Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *Ann. Probab.* **47** 1653–1676. MR3945756 https://doi.org/10.1214/18-AOP1293

[8] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** no. 33, 53. MR3183577 https://doi.org/10.1214/ejp.v19-3054

[9] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966 https://doi.org/10.1017/CBO9780511814068

[10] BOURGADE, P., ERDŐS, L., YAU, H.-T. and YIN, J. (2016). Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.* **69** 1815–1881. MR3541852 https://doi.org/10.1002/cpa.21624

[11] BROUWER, A. E. and HAEMERS, W. H. (2012). *Spectra of Graphs. Universitext.* Springer, New York. MR2882891 https://doi.org/10.1007/978-1-4614-1939-6

[12] CHE, Z. (2017). Universality of random matrices with correlated entries. *Electron. J. Probab.* **22** Paper No. 30, 38. MR3629874 https://doi.org/10.1214/17-EJP46

[13] CHUNG, F. R. K. (1997). *Spectral Graph Theory. CBMS Regional Conference Series in Mathematics* **92**. Amer. Math. Soc., Providence, RI. MR1421568

[14] COIFMAN, R. R., LAFON, S., LEE, A. B., MAGGIONI, M., NADLER, B., WARNER, F. and ZUCKER, S. W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods. *Proc. Natl. Acad. Sci. USA* **102** 7432–7437.

[15] COIFMAN, R. R., LAFON, S., LEE, A. B., MAGGIONI, M., NADLER, B., WARNER, F. and ZUCKER, S. W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. USA* **102** 7426–7431.

[16] ERDŐS, L., KNOWLES, A. and YAU, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y

[17] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770 https://doi.org/10.1007/s00220-012-1527-7

[18] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Delocalization and diffusion profile for random band matrices. *Comm. Math. Phys.* **323** 367–416. MR3085669 https://doi.org/10.1007/s00220-013-1773-3

[19] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** no. 59, 58. MR3068390 https://doi.org/10.1214/EJP.v18-2473

[20] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 https://doi.org/10.1214/11-AOP734

[21] ERDŐS, L., KRÜGER, T. and SCHRÖDER, D. (2019). Random matrices with slow correlation decay. *Forum Math. Sigma* **7** e8, 89. MR3941370 https://doi.org/10.1017/fms.2019.2

[22] ERDŐS, L., RAMÍREZ, J. A., SCHLEIN, B. and YAU, H.-T. (2010). Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electron. J. Probab.* **15** 526–603. MR2639734 https://doi.org/10.1214/EJP.v15-768

[23] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.* **287** 641–655. MR2481753 https://doi.org/10.1007/s00220-008-0636-9

[24] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** 815–852. MR2537522 https://doi.org/10.1214/08-AOP421

[25] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2011). Universality of random matrices and local relaxation flow. *Invent. Math.* **185** 75–119. MR2810797 https://doi.org/10.1007/s00222-010-0302-7

[26] ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1–46. MR2919197 https://doi.org/10.1214/10-AIHP388

[27] ERDŐS, L. and SCHNELLI, K. (2017). Universality for random matrix flows with time-dependent density. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1606–1656. MR3729630 https://doi.org/10.1214/16-AIHP765

[28] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427 https://doi.org/10.1007/s00440-011-0390-3

[29] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 https://doi.org/10.1016/j.aim.2011.12.010

[30] ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

[31] EVANGELOU, S. N. and ECONOMOU, E. N. (1992). Spectral density singularities, level statistics, and localization in a sparse random matrix ensemble. *Phys. Rev. Lett.* **68** 361–364.

[32] FELDHEIM, O. N. and SODIN, S. (2010). A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20** 88–123. MR2647136 https://doi.org/10.1007/s00039-010-0055-x

[33] GILBERT, E. N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098

[34] HAAGERUP, U. and THORBJØRNSEN, S. (2005). A new application of random matrices: Ext($C^*_{red}(F_2)$) is not a group. *Ann. of Math.* (2) **162** 711–775. MR2183281 https://doi.org/10.4007/annals.2005.162.711

[35] HE, Y. and KNOWLES, A. (2017). Mesoscopic eigenvalue statistics of Wigner matrices. *Ann. Appl. Probab.* **27** 1510–1550. MR3678478 https://doi.org/10.1214/16-AAP1237

[36] HE, Y., KNOWLES, A. and ROSENTHAL, R. (2018). Isotropic self-consistent equations for mean-field random matrices. *Probab. Theory Related Fields* **171** 203–249. MR3800833 https://doi.org/10.1007/s00440-017-0776-y

[37] HOORY, S., LINIAL, N. and WIGDERSON, A. (2006). Expander graphs and their applications. *Bull. Amer. Math. Soc.* (*N.S.*) **43** 439–561. MR2247919 https://doi.org/10.1090/S0273-0979-06-01126-8

[38] HUANG, J., LANDON, B. and YAU, H.-T. (2015). Bulk universality of sparse random matrices. *J. Math. Phys.* **56** 123301, 19. MR3429490 https://doi.org/10.1063/1.4936139

[39] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization.* Wiley Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

[40] KNOWLES, A. and YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields* **155** 543–582. MR3034787 https://doi.org/10.1007/s00440-011-0407-y

[41] LANDON, B. and CHE, Z. (2017). Local spectral statistics of the addition of random matrices. Preprint. Available at arXiv:1701.00513.

[42] LANDON, B., SOSOE, P. and YAU, H.-T. (2019). Fixed energy universality of Dyson Brownian motion. *Adv. Math.* **346** 1137–1332. MR3914908 https://doi.org/10.1016/j.aim.2019.02.010

[43] LANDON, B. and YAU, H.-T. (2017). Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.* **355** 949–1000. MR3687212 https://doi.org/10.1007/s00220-017-2955-1

[44] LANDON, B. and YAU, H.-T. (2017). Edge statistics of Dyson Brownian motion. Preprint. Available at arXiv:1712.03881.

[45] LEE, J. O. and SCHNELLI, K. (2015). Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** 1550018, 94. MR3405746 https://doi.org/10.1142/S0129055X1550018X

[46] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818 https://doi.org/10.1214/16-AAP1193

[47] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 https://doi.org/10.1007/s00440-017-0787-8

[48] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** 507–536. MR0208649

[49] MOHAR, B. (1997). Some applications of Laplace eigenvalues of graphs. In *Graph Symmetry* (*Montreal, PQ*, 1996). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **497** 225–275. Kluwer Academic, Dordrecht. MR1468791

[50] MOHAR, B. and POLJAK, S. (1993). Eigenvalues in combinatorial optimization. In *Combinatorial and Graph-Theoretical Problems in Linear Algebra* (*Minneapolis, MN*, 1991). *IMA Vol. Math. Appl.* **50** 107–151. Springer, New York. MR1240959 https://doi.org/10.1007/978-1-4613-8354-3_5

[51] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670 https://doi.org/10.1007/s00440-007-0133-7

[52] POTHEN, A., SIMON, H. D. and LIOU, K.-P. (1990). Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.* **11** 430–452. MR1054210 https://doi.org/10.1137/0611030

[53] SHCHERBINA, M. and TIROZZI, B. (2012). Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs: Diluted regime. *J. Math. Phys.* **53** 043501, 18. MR2953145 https://doi.org/10.1063/1.3698291

[54] SHI, J. and MALIK, J. (1997). Normalized cuts and image segmentation. In *Computer Vision and Pattern Recognition*, 1997. *Proceedings*, 1997 *IEEE Computer Society Conference on* 731–737.

[55] SODIN, S. (2010). The spectral edge of some random band matrices. *Ann. of Math.* (2) **172** 2223–2251. MR2726110 https://doi.org/10.4007/annals.2010.172.2223

[56] SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** 697–733. MR1727234 https://doi.org/10.1007/s002200050743

[57] SPIELMAN, D. (2012). Spectral graph theory. In *Combinatorial Scientific Computing* 495–524. CRC Press, Boca Raton, FL. MR2952760 https://doi.org/10.1201/b11644-19

[58] SPIELMAN, D. A. and TENG, S.-H. (1996). Spectral partitioning works: Planar graphs and finite element meshes. In 37*th Annual Symposium on Foundations of Computer Science* (*Burlington, VT*, 1996) 96–105. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1450607 https://doi.org/10.1109/SFCS.1996.548468

[59] SPIELMAN, D. A. and TENG, S.-H. (2007). Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra Appl.* **421** 284–305. MR2294342 https://doi.org/10.1016/j.laa.2006.07.020

[60] TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449 https://doi.org/10.1007/s00220-010-1044-5

[61] TAO, T. and VU, V. (2011). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665 https://doi.org/10.1007/s11511-011-0061-3

[62] VU, V. H. (2007). Spectral norm of random matrices. *Combinatorica* **27** 721–736. MR2384414 https://doi.org/10.1007/s00493-007-2190-z