CrossMark

Local law and Tracy–Widom limit for sparse random matrices

Ji Oon Lee[1] · **Kevin Schnelli[2,3]**

Abstract We consider spectral properties and the edge universality of sparse random matrices, the class of random matrices that includes the adjacency matrices of the Erdős–Rényi graph model $G(N, p)$. We prove a local law for the eigenvalue density up to the spectral edges. Under a suitable condition on the sparsity, we also prove that the rescaled extremal eigenvalues exhibit GOE Tracy–Widom fluctuations if a deterministic shift of the spectral edge due to the sparsity is included. For the adjacency matrix of the Erdős–Rényi graph this establishes the Tracy–Widom fluctuations of the second largest eigenvalue when p is much larger than $N^{-2/3}$ with a deterministic shift of order $(Np)^{-1}$.

Keywords Local law · Sparse random matrices · Erdős–Rényi graph

Mathematics Subject Classification 60B20 · 62H10

✉ Ji Oon Lee
jioon.lee@kaist.edu

Kevin Schnelli
schnelli@kth.se

[1] KAIST, Daejeon, South Korea

[2] KTH Royal Institute of Technology, Stockholm, Sweden

[3] IST Austria, Klosterneuburg, Austria

Springer

1 Introduction

We consider spectral properties of sparse random matrices. One of the most prominent examples in the class of sparse random matrices is the (centered) adjacency matrix of the Erdős–Rényi graph on N vertices, where an edge is independently included in the graph with a fixed probability $p \equiv p(N)$. Introduced in [11,12,24], the Erdős–Rényi graph model $G(N, p)$ serves as a null model in the theory of random graphs and has numerous applications in many fields including network theory. Information about a random graph can be obtained by investigating its adjacency matrix, especially the properties of its eigenvalues and eigenvectors.

The sparsity of a real symmetric N by N random matrix may be measured by a sparsity parameter $q \equiv q(N)$, with $0 \leq q \leq N^{1/2}$, such that the expected number of non-vanishing entries is asymptotically equal to Nq^2. For example, for the adjacency matrices of the Erdős–Rényi graph we have $q^2 = Np$ as the expected number of non-vanishing entries is $N(N - 1)p = (N - 1)q^2$. For the Gaussian orthogonal ensemble (GOE) we have $q = N^{1/2}$ as the expected number of non-vanishing entries is N^2. We call a random matrix sparse if q is much smaller than $N^{1/2}$.

For Wigner matrices, one of the fundamental inputs in the proof of universality results is the local semicircle law [18–20,23], which provides an estimate of the local eigenvalue density down to the optimal scale. The framework built on the local law can also help understanding the spectral properties of sparse random matrices [15]. However, in contrast to Wigner matrices, the local eigenvalue density for a sparse random matrix depends on its sparsity. For this reason, the universality of the local eigenvalue statistics for sparse random matrices was proved at first only for $q \geq N^{1/3}$ in [15,16]. Recently, bulk universality was proved in [29] under the much weaker condition $q \geq N^\phi$, for any $\phi > 0$. The main obstacle in the proof of the edge universality is that the local law obtained in [15] deteriorates at the edge of the spectrum.

Our first main result is a local law for sparse random matrices up to the edges. More precisely, we prove a local law for the eigenvalue density in the regime $q \geq N^\phi$, for arbitrarily small $\phi > 0$. The main observation is that, although the empirical spectral measure of sparse random matrices converges in the large N limit to the semicircle measure, there exists a deterministic correction term that is not negligible for large but finite N. As a result, we establish a local law that compares the empirical spectral measure not with the semicircle law but with its refinement. We also explicitly compute the leading term of the deterministic shift of the spectral edge, which equals $1/(Np)$ in case of the (centered) adjacency matrix of the Erdős–Rényi graph (N, p). (See Theorem 2.4 and Corollary 2.5 for more detail.) This information on the shift of the spectral edge is crucial in understanding the fluctuations of the largest eigenvalues.

The largest eigenvalue μ_1 of a real symmetric N by N Wigner matrix (whose entries are centered and have variance $1/N$) converges almost surely to two under the finite fourth-moment condition, and $N^{2/3}(\mu_1 - 2)$ converges in distribution to the GOE Tracy–Widom law. However, for sparse random matrices with sparsity $q \leq N^{1/3-\phi}$, $\phi > 0$, the deterministic shift of the edge is far greater than $N^{-2/3}$, the typical size of the Tracy–Widom fluctuations.

Our second main result is the edge universality that states the limiting law for the fluctuations of the rescaled largest eigenvalues of a (centered) sparse random matrices is given by the Tracy–Widom law if the shift is taken into consideration and if $q \geq N^{1/6+\phi}$, where $\phi > 0$ is arbitrarily small. We expect that the exponent one-sixth is critical, *i.e.* we expect a different limiting law for the fluctuations of the rescaled largest eigenvalues when $q \leq N^{1/6-\phi}$. (See Theorem 2.10 and the discussion below it for more detail.) For the adjacency matrices of the Erdős–Rényi graphs, the sparsity conditions corresponds to $p \geq N^{-2/3+\phi}$, for any $\phi > 0$, and our result then assures that the rescaled second largest eigenvalue has GOE Tracy–Widom fluctuations; see Corollary 2.13.

In the proof of the local law, we introduce a new method based on a recursive moment estimate for the normalized trace m of the Green function, *i.e.* we recursively control high moments of $|P(m)|$, for some specifically chosen polynomial P, by using lower moments of $|P(m)|$. This recursive moment estimate allows us to track the effects of the sparsity efficiently and it directly establishes the fluctuation averaging mechanism without fully expanding all powers of m; see Sect. 3 for details. To establish the recursive moment estimate we use a cumulant expansion of $P(m)$ replacing customary resolvent expansions of m using Schur's complement or finite-rank perturbation formulas.

Our proof of the Tracy–Widom limit of the extremal eigenvalues relies on the Green function comparison method [21,23]. However, instead of applying the conventional Lindeberg replacement approach, we use a continuous flow that interpolates between the sparse random matrix and the Gaussian orthogonal ensemble (GOE). The main advantage of using a continuous interpolation is that we may estimate the rate of change of m along the flow even if the moments of the entries in the sparse matrix are significantly different from those of the entries in the GOE matrix. The change of m over time is offset by the deterministic shift of the edge. A similar idea was used in the proof of edge universality of other random matrix models in [37,38].

This paper is organized as follows: In Sect. 2, we define the model, present the main results and outline applications to adjacency matrices of the Erdős–Rényi graph ensemble. In Sect. 3, we explain the main strategy of our proofs. In Sect. 4, we prove several properties of the deterministic refinement of Wigner's semicircle law. In Sect. 5, we prove the local law using our technical result on the recursive moment estimate, Lemma 5.1. In Sect. 6, we prove Lemma 5.1 with technical detail. In Sect. 7, we prove our second main result on the edge universality.

Notational conventions We use the symbols $O(\,\cdot\,)$ and $o(\,\cdot\,)$ for the standard big-O and little-o notation. The notations O, o, $\ll$, $\gg$, refer to the limit $N \to \infty$ unless otherwise stated. Here $a \ll b$ means $a = o(b)$. We use c and C to denote positive constants that do not depend on N, usually with the convention $c \leq C$. Their value may change from line to line. We write $a \asymp b$, if there is $C \geq 1$ such that $C^{-1}|b| \leq |a| \leq C|b|$. Throughout the paper we denote for $z \in \mathbb{C}^+$ the real part by $E = \operatorname{Re} z$ and the imaginary part by $\eta = \operatorname{Im} z$. For $a \in \mathbb{R}$, we let $(a)_+ = \max(0, a)$, and $(a)_- = -\min(a, 0)$. Finally, we use double brackets to denote index sets, *i.e.* for $n_1, n_2 \in \mathbb{R}$, $[\![n_1, n_2]\!] := [n_1, n_2] \cap \mathbb{Z}$.

2 Definitions and main results

2.1 Motivating examples

2.1.1 Adjacency matrix of Erdős–Rényi graph

One motivation for this work is the study of adjacency matrices of the *Erdős–Rényi random graph model* $G(N, p)$. The off-diagonal entries of the adjacency matrix associated with an Erdős–Rényi graph are independent, up to the symmetry constraint, Bernoulli random variables with parameter p, *i.e.* the entries are equal to 1 with probability p and 0 with probability $1 - p$. The diagonal entries are set to zero, corresponding to the choice that the graph has no self-loops. Rescaling this matrix ensemble so that the bulk eigenvalues typically lie in an order one interval we are led to the following random matrix ensemble. Let A be a real symmetric $N \times N$ matrix whose entries, A_{ij}, are independent random variables (up to the symmetry constraint $A_{ij} = A_{ji}$) with distributions

$$\mathbb{P}\Big(A_{ij} = \frac{1}{\sqrt{Np(1-p)}}\Big) = p, \qquad \mathbb{P}(A_{ij} = 0) = 1 - p,$$
$$\mathbb{P}(A_{ii} = 0) = 1, \qquad (i \neq j). \tag{2.1}$$

Note that the matrix A typically has $N(N-1)p$ non-vanishing entries. For our analysis it is convenient to extract the mean of the entries of A by considering the matrix $\widetilde{A}$ whose entries, $\widetilde{A}_{ij}$, have distribution

$$\mathbb{P}\Big(\widetilde{A}_{ij} = \frac{1-p}{\sqrt{Np(1-p)}}\Big) = p, \quad \mathbb{P}\Big(\widetilde{A}_{ij} = -\frac{p}{\sqrt{Np(1-p)}}\Big) = 1 - p,$$
$$\mathbb{P}\big(\widetilde{A}_{ii} = 0\big) = 1,$$

with $i \neq j$. A simple computation then reveals that

$$\mathbb{E}\widetilde{A}_{ij} = 0, \qquad \mathbb{E}\widetilde{A}_{ij}^2 = \frac{1}{N}, \tag{2.2}$$

and

$$\mathbb{E}\widetilde{A}_{ij}^k = \frac{(-p)^k(1-p) + (1-p)^k p}{(Np(1-p))^{k/2}} = \frac{1}{Nd^{(k-2)/2}}(1 + O(p)), \qquad (k \geq 3), \tag{2.3}$$

with $i \neq j$, where $d := pN$ denotes the *expected degree* of a vertex, which we allow to depend on N. As already suggested by (2.3), we will assume that $p \ll 1$.

2.1.2 Diluted Wigner matrices

Another motivation for this work are *diluted Wigner matrices*. Consider the matrix ensemble of real symmetric $N \times N$ matrices of the form

$$D_{ij} = B_{ij} V_{ij}, \qquad (1 \le i \le j \le N), \tag{2.4}$$

where $(B_{ij} : i \le j)$ and $(V_{ij} : i \le j)$ are two independent families of independent and identically distributed random variables. The random variables (V_{ij}) satisfies $\mathbb{E} V_{ij}^2 = 1$ and $\mathbb{E} V_{ij}^{2k} \le (Ck)^{ck}$, $k \ge 4$, for some constants c and C, and their distribution is, for simplicity, often assumed to be symmetric. The random variables (B_{ij}) are chosen to have a Bernoulli type distribution given by

$$\mathbb{P}\left(B_{ij} = \frac{1}{\sqrt{Np}}\right) = p, \qquad \mathbb{P}(B_{ij} = 0) = 1 - p, \qquad \mathbb{P}(B_{ii} = 0) = 1, \tag{2.5}$$

with $i \ne j$. We introduce the *sparsity parameter q* through

$$p = \frac{q^2}{N}, \tag{2.6}$$

with $0 < q \le N^{1/2}$. We allow q to depend on N. We refer to the random matrix $D = (D_{ij})$ as a *diluted Wigner matrix* whenever $q \lll N^{1/2}$. For $q = N^{1/2}$, we recover the usual Wigner ensemble.

2.2 Notation

In this subsection we introduce some of the notation and conventions used.

2.2.1 Probability estimates

We first introduce a suitable notion for high-probability estimates.

Definition 2.1 (*High probability event*) We say that an N-dependent event $\Xi \equiv \Xi^{(N)}$ holds with high probability if, for any (large) $D > 0$,

$$\mathbb{P}(\Xi^{(N)}) \ge 1 - N^{-D}, \tag{2.7}$$

for sufficiently large $N \ge N_0(D)$.

Definition 2.2 (*Stochastic domination*) Let $X \equiv X^{(N)}$, $Y \equiv Y^{(N)}$ be N-dependent non-negative random variables. We say that Y stochastically dominates X if, for all (small) $\epsilon > 0$ and (large) $D > 0$,

$$\mathbb{P}(X^{(N)} > N^\epsilon Y^{(N)}) \le N^{-D}, \tag{2.8}$$

for sufficiently large $N \geq N_0(\epsilon, D)$, and we write $X \prec Y$. When $X^{(N)}$ and $Y^{(N)}$ depend on a parameter $u \in U$ (typically an index label or a spectral parameter), then $X(u) \prec Y(u)$, uniformly in $u \in U$, means that the threshold $N_0(\epsilon, D)$ can be chosen independently of u. A slightly modified version of stochastic domination appeared first in [14].

In Definition 2.2 and hereinafter we implicitly choose $\epsilon > 0$ strictly smaller than $\phi/10 > 0$, where $\phi > 0$ is the fixed parameter appearing in (2.13) below.

The relation $\prec$ is transitive and it satisfies the following arithmetic rules: If $X_1 \prec Y_1$ and $X_2 \prec Y_2$ then $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1 X_2 \prec Y_1 Y_2$. Furthermore, the following property will be used on many occasions: If $\Phi(u) \geq N^{-C}$ is deterministic, if $Y(u)$ is a nonnegative random variable satisfying $\mathbb{E}[Y(u)^2] \leq N^{C'}$ for all u, and if $Y(u) \prec \Phi(u)$ uniformly in u, then, for any $\epsilon > 0$, we have $\mathbb{E}[Y(u)] \leq N^{\epsilon} \Phi(u)$ for $N \geq N_0(\epsilon)$, with a threshold independent of u. This can easily be checked since

$$\mathbb{E}[Y(u)\mathbb{1}(Y(u) > N^{\epsilon/2}\Phi)] \leq \left(\mathbb{E}[Y(u)^2]\right)^{1/2} \left(\mathbb{P}[Y(u) > N^{\epsilon/2}\Phi]\right)^{1/2} \leq N^{-D},$$

for any (large) $D > 0$, and $\mathbb{E}[Y(u)\mathbb{1}(Y(u) \leq N^{\epsilon/2}\Phi(u))] \leq N^{\epsilon/2}\Phi(u)$, hence $\mathbb{E}[Y(u)] \leq N^{\epsilon}\Phi(u)$.

2.2.2 Stieltjes transform

Given a probability measure ν on $\mathbb{R}$, we define its Stieltjes transform as the analytic function $m_\nu : \mathbb{C}^+ \to \mathbb{C}^+$, with $\mathbb{C}^+ := \{z = E + i\eta : E \in \mathbb{R}, \eta > 0\}$, defined by

$$m_\nu(z) := \int_{\mathbb{R}} \frac{d\nu(x)}{x - z}. \tag{2.9}$$

Note that $\lim_{\eta \nearrow 0} i\eta\, m_\nu(i\eta) = -1$ since ν is a probability measure. Conversely, if an analytic function $m : \mathbb{C}^+ \to \mathbb{C}^+$ satisfies $\lim_{\eta \nearrow 0} i\eta\, m(i\eta) = -1$, then it is the Stieltjes transform of a probability measure.

Choosing ν to be the standard semicircle law with density $\frac{1}{2\pi}\sqrt{4 - x^2}$ on $[-2, 2]$, on easily shows that m_ν, for simplicity hereinafter denoted by m_{sc}, is explicitly given by

$$m_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}, \qquad (z \in \mathbb{C}^+), \tag{2.10}$$

where we choose the branch of the square root so that $m_{sc}(z) \in \mathbb{C}^+, z \in \mathbb{C}^+$. It directly follows that

$$1 + z m_{sc}(z) + m_{sc}(z)^2 = 0, \qquad (z \in \mathbb{C}^+). \tag{2.11}$$

2.3 Main results

In this section we present our main results. We first generalize the matrix ensembles derived from the Erdős–Rényi graph model and the diluted Wigner matrices in Section 2.1.

Assumption 2.3 Fix any small $\phi > 0$. We assume that $H = (H_{ij})$ is a real symmetric $N \times N$ matrix whose diagonal entries are almost surely zero and whose off-diagonal entries are independent, up to the symmetry constraint $H_{ij} = H_{ji}$, identically distributed random variables. We further assume that (H_{ij}) satisfy the moment conditions

$$\mathbb{E}H_{ij} = 0, \qquad \mathbb{E}(H_{ij})^2 = \frac{1 - \delta_{ij}}{N}, \qquad \mathbb{E}|H_{ij}|^k \leq \frac{(Ck)^{ck}}{Nq^{k-2}}, \qquad (k \geq 3), \quad (2.12)$$

with sparsity parameter q satisfying

$$N^\phi \leq q \leq N^{1/2}. \qquad (2.13)$$

We assume that the diagonal entries satisfy $H_{ii} = 0$ a.s., yet this condition can easily be dropped. For the choice $\phi = 1/2$ we recover the *real symmetric Wigner ensemble* (with vanishing diagonal). For the rescaled adjacency matrix of the Erdős–Rényi graph, the sparsity parameter q, the edge probability p and the expected degree of a vertex d are linked by $q^2 = pN = d$.

We denote by $\kappa^{(k)}$ the k-th *cumulant* of the i.i.d. random variables $(H_{ij} : i < j)$. Under Assumption 2.3 we have $\kappa^{(1)} = 0$, $\kappa^{(2)} = 1/N$, and

$$|\kappa^{(k)}| \leq \frac{(2Ck)^{2(c+1)k}}{Nq^{k-2}}, \qquad (k \geq 3). \qquad (2.14)$$

We further introduce the *normalized cumulants*, $s^{(k)}$, by setting

$$s^{(1)} := 0, \qquad s^{(2)} := 1, \qquad s^{(k)} := Nq^{k-2}\kappa^{(k)}, \qquad (k \geq 3). \qquad (2.15)$$

In case H is given by the centered adjacency matrix $\widetilde{A}$ introduced in Sect. 2.1.1, we have that $s^{(k)} = 1 + O(d/N)$, $k \geq 3$, as follows from (2.3).

We start with the local law for the Green function of this matrix ensemble.

2.3.1 Local law up to the edges for sparse random matrices

Given a real symmetric matrix H we define its Green function, G^H, and the normalized trace of its Green function, m^H, by setting

$$G^H(z) := \frac{1}{H - z\mathrm{I}}, \qquad m^H(z) := \frac{1}{N}\mathrm{Tr}\, G^H(z), \qquad (z \in \mathbb{C}^+). \qquad (2.16)$$

The matrix entries of $G^H(z)$ are denoted by $G^H_{ij}(z)$. In the following we often drop the explicit z-dependence from the notation for $G^H(z)$ and $m^H(z)$.

Denoting by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ the ordered eigenvalues of H, we note that m^H is the Stieltjes transform of the empirical eigenvalue distributions, μ^H, of H given by

$$\mu^H := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} . \tag{2.17}$$

We further introduce the following domain of the upper-half plane

$$\mathcal{E} := \left\{ z = E + i\eta \in \mathbb{C}^+ : |E| < 3,\ 0 < \eta \leq 3 \right\} . \tag{2.18}$$

Our first main result is the local law for m^H up to the spectral edges.

Theorem 2.4 *Let H satisfy Assumption 2.3 with $\phi > 0$. Then, there exists an algebraic function $\widetilde{m} : \mathbb{C}^+ \to \mathbb{C}^+$ and $2 < L < 3$ such that the following hold:*

(1) The function $\widetilde{m}$ is the Stieltjes transform of a deterministic symmetric probability measure $\widetilde{\rho}$, i.e. $\widetilde{m}(z) = m_{\widetilde{\rho}}(z)$. Moreover, supp $\widetilde{\rho} = [-L, L]$ and $\widetilde{\rho}$ is absolutely continuous with respect to the Lebesgue measure with a strictly positive density on $(-L, L)$.

(2) The function $\widetilde{m} \equiv \widetilde{m}(z)$ is a solution to the polynomial equation

$$P_z(\widetilde{m}) := 1 + z\widetilde{m} + \widetilde{m}^2 + \frac{s^{(4)}}{q^2}\widetilde{m}^4 = 0 , \qquad (z \in \mathbb{C}^+) . \tag{2.19}$$

(3) The normalized trace m^H of the Green function of H satisfies

$$|m^H(z) - \widetilde{m}(z)| \prec \frac{1}{q^2} + \frac{1}{N\eta} , \tag{2.20}$$

uniformly on the domain $\mathcal{E}$, $z = E + i\eta$.

Some properties of $\widetilde{\rho}$ and its Stieltjes transform $\widetilde{m}$ are collected in Lemma 4.1 below.

The local law (2.20) implies estimates on the *local density of states* of H. For $E_1 < E_2$ define

$$\mathfrak{n}(E_1, E_2) := \frac{1}{N} |\{i\ :\ E_1 < \lambda_i \leq E_2\}| , \qquad n_{\widetilde{\rho}}(E_1, E_2) := \int_{E_1}^{E_2} \widetilde{\rho}(x)\,\mathrm{d}x . \tag{2.21}$$

Corollary 2.5 *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Let $E_1, E_2 \in \mathbb{R}$, $E_1 < E_2$. Then,*

$$|\mathfrak{n}(E_1, E_2) - n_{\widetilde{\rho}}(E_1, E_2)| \prec \frac{E_2 - E_1}{q^2} + \frac{1}{N} . \tag{2.22}$$

The proof of Corollary 2.5 from Theorem 2.4 is a standard application of the Helffer-Sjöstrand calculus; see *e.g.* Section 7.1 of [17] for a similar argument.

An interesting effect of the sparsity of the entries of H is that its eigenvalues follow, for large N, the deterministic law $\widetilde{\rho}$ that depends on the sparsity parameter q. While this law approaches the standard semicircle law ρ_{sc} in the limit $N \to \infty$, its deterministic refinement to the standard semicircular law for finite N accounts for the non-optimality at the edge of results obtained in [15], *i.e.* when (2.20) is compared with (2.23) below.

Proposition 2.6 (Local semicircle law, Theorem 2.8 of [15]) *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Then, the following estimates hold uniformly for $z \in \mathcal{E}$:*

$$\left| m^H(z) - m_{sc}(z) \right| \prec \min\left\{ \frac{1}{q^2 \sqrt{\varkappa + \eta}}, \frac{1}{q} \right\} + \frac{1}{N\eta}, \tag{2.23}$$

where m_{sc} denote the Stieltjes transform of the standard semicircle law, and

$$\max_{1 \leq i,j \leq N} \left| G_{ij}^H(z) - \delta_{ij} m_{sc}(z) \right| \prec \frac{1}{q} + \sqrt{\frac{\operatorname{Im} m_{sc}(z)}{N\eta}} + \frac{1}{N\eta}, \tag{2.24}$$

where $\varkappa \equiv \varkappa(z) := |E - 2|$, $z = E + i\eta$.

We remark that the estimate (2.23) is essentially optimal as long as the spectral parameter z stays away from the spectral edges, *e.g.* for energies in the bulk $E \in [-2+\delta, 2-\delta]$, $\delta > 0$. For the individual Green function entries, G_{ij}, we believe that the estimate (2.24) is already essentially optimal (m_{sc} therein may be replaced by $\widetilde{m}$ without changing the error bound). A consequence of Proposition 2.6 is that all eigenvectors of H are completely delocalized.

Proposition 2.7 (Theorem 2.16 and Remark 2.18 in [15]) *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Denote by $(\boldsymbol{u}_i^H)$ the ℓ^2-normalized eigenvectors of H. Then,*

$$\max_{1 \leq i \leq N} \|\boldsymbol{u}_i^H\|_\infty \prec \frac{1}{\sqrt{N}}. \tag{2.25}$$

Using (2.23) as *a priori* input it was proved in [29] that the local eigenvalue statistics in the bulk agree with the local statistics of the GOE, for $\phi > 0$; see also [16] for $\phi > 1/3$. When combined with a high moment estimates of H (see Lemma 4.3 in [15]), the estimate in (2.23) implies the following bound on the operator norm of H.

Proposition 2.8 (Lemma 4.4 of [15]) *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Then,*

$$\left| \|H\| - 2 \right| \prec \frac{1}{q^2} + \frac{1}{N^{2/3}}. \tag{2.26}$$

The following estimates of the operator norm of H sharpens the estimates of Proposition 2.8 by including the deterministic refinement to the semicircle law as expressed by Theorem 2.4.

Theorem 2.9 *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Then,*

$$\left| \|H\| - L \right| \prec \frac{1}{q^4} + \frac{1}{N^{2/3}}, \tag{2.27}$$

where $\pm L$ are the endpoints of the support of the measure $\widetilde{\rho}$ given by

$$L = 2 + \frac{s^{(4)}}{q^2} + O(q^{-4}). \tag{2.28}$$

Here and above, we restricted the choice of the sparsity parameter q to the range $N^\phi \le q \le N^{1/2}$ for arbitrary small $\phi > 0$. Yet, pushing our estimates and formalism we expect also to cover the range $(\log N)^{A_0 \log \log N} \le q \le N^{1/2}$, $A_0 \ge 30$, considered in [15]. In fact, Khorunzhiy showed for diluted Wigner matrices (*cf.* Sect. 2.1.2) that $\|H\|$ converges almost surely to 2 for $q \ggg (\log N)^{1/2}$, while $\|H\|$ diverges for $q \lll (\log N)^{1/2}$; see Theorem 2.1 and Theorem 2.2 of [32] for precise statements.

As noted in Theorem 2.9, the local law allows strong statements on the locations of the extremal eigenvalues of H. We next discuss implications for the fluctuations of the rescaled extremal eigenvalues.

2.3.2 Tracy–Widom limit of the extremal eigenvalues

Let W be a real symmetric Wigner matrix and denote by λ_1^W its largest eigenvalue. The *edge universality* for Wigner matrices asserts that

$$\lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_1^W - 2) \le s \right) = F_1(s), \tag{2.29}$$

where F_1 is the Tracy–Widom distribution function [49,50] for the GOE. Statement (2.29) holds true for the smallest eigenvalue λ_N^W as well. We henceforth focus on the largest eigenvalues, the smallest eigenvalues can be dealt with in exactly the same way.

The universality of the Tracy–Widom laws for Wigner matrices was first proved in [45,46] for real symmetric and complex Hermitian ensembles with symmetric distributions. The symmetry assumption on the entries' distribution was partially removed in [42,43]. Edge universality without any symmetry assumption was proved in [48] under the condition that the distribution of the matrix elements has subexponential decay and its first three moments match those of the Gaussian distribution, *i.e.* the third moment of the entries vanish. The vanishing third moment condition was removed in [23]. A necessary and sufficient condition on the entries' distribution for the edge universality of Wigner matrices was given in [35].

Our second main result shows that the fluctuations of the rescaled largest eigenvalue of the sparse matrix ensemble are governed by the Tracy–Widom law, if the sparsity parameter q satisfies $q \ggg N^{1/6}$.

Theorem 2.10 *Suppose that H satisfies Assumption 2.3 with $\phi > 1/6$. Denote by λ_1^H the largest eigenvalue of H. Then,*

$$\lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_1^H - L) \le s\right) = F_1(s), \qquad (2.30)$$

where L denotes the upper-edge of the deterministic measure $\widetilde{\rho}$ given in (2.28).

The convergence result (2.30) was obtained in Theorem 2.7 of [16] under the assumption that the sparsity parameter q satisfies $q \ggg N^{1/3}$, *i.e.* $\phi > 1/3$ (and with 2 replacing L).

In the regime $N^{1/6} \lll q \le N^{1/3}$, the deterministic shift of the upper edge by $L - 2 = O(q^{-2})$ is essential for (2.30) to hold since then $q^{-2} \ge N^{-2/3}$, the latter being the scale of the Tracy–Widom fluctuations. In other words, to observe the Tracy–Widom fluctuations in the regime $N^{1/6} \lll q \le N^{1/3}$ corrections from the fourth moment of the matrix entries' distribution have to be accounted for. This is in accordance with high order moment computations for diluted Wigner matrices in [33].

It is expected that the order of the fluctuations of the largest eigenvalue exceeds $N^{-2/3}$ if $q \lll N^{1/6}$. The heuristic reasoning is that, in this regime, the fluctuations of the eigenvalues in the bulk of the spectrum are much larger than $N^{-2/3}$ and hence affect the fluctuations of the eigenvalues at the edges. Indeed, the linear eigenvalue statistics of sparse random matrices were studied in [6,44]. For any sufficiently smooth function φ, it was shown there that

$$\frac{q}{\sqrt{N}} \sum_{i=1}^{N} \varphi(\lambda_i) - \mathbb{E}\left[\frac{q}{\sqrt{N}} \sum_{i=1}^{N} \varphi(\lambda_i)\right]$$

converges to a centered Gaussian random variable with variance of order one. This suggests that the fluctuations of an individual eigenvalue in the bulk are of order $N^{-1/2}q^{-1}$, which is far greater than the Tracy–Widom scale $N^{-2/3}$ if $q \lll N^{1/6}$. Moreover, comparing the size of the fluctuations with the estimate in (2.23), this also suggests that, for large η, $\mathbb{E}m^H(z)$ is not well approximated by $m_{sc}(z)$.

Remark 2.11 Theorem 2.10 can be extended to correlation functions of extreme eigenvalues as follows: For any fixed k, the joint distribution function of the first k rescaled eigenvalues converges to that of the GOE, *i.e.* if we denote by $\lambda_1^{GOE} \ge \lambda_2^{GOE} \ge \cdots \ge \lambda_N^{GOE}$ the eigenvalues of a GOE matrix independent of H, then

$$\lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_1^H - L) \le s_1, N^{2/3}(\lambda_2^H - L) \le s_2, \ldots, N^{2/3}(\lambda_k^H - L) \le s_k\right)$$
$$= \lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_1^{GOE} - 2) \le s_1, N^{2/3}(\lambda_2^{GOE} - 2) \le s_2, \ldots, N^{2/3}(\lambda_k^{GOE} - 2) \le s_k\right).$$
$$(2.31)$$

We further mention that all our results also hold for complex Hermitian sparse random matrices with the GUE Tracy–Widom law describing the limiting edge fluctuations.

2.3.3 Applications to the adjacency matrix of the Erdős–Rényi graph

We briefly return to the adjacency matrix A of the Erdős–Rényi graph ensemble introduced in Sect. 2.1.1. Since the entries of A are not centered, the largest eigenvalue λ_1^A is an outlier well-separated from the other eigenvalues. Recalling the definition of the matrix $\widetilde{A}$ whose entries are centered, we notice that

$$A = \widetilde{A} + f|\boldsymbol{e}\rangle\langle\boldsymbol{e}| - a\mathrm{I}\,, \tag{2.32}$$

with $f := q(1 - q^2/N)^{-1/2}$, $a := f/N$ and $\boldsymbol{e} := N^{-1/2}(1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$. (Here, $|\boldsymbol{e}\rangle\langle\boldsymbol{e}|$ denotes the orthogonal projection onto $\boldsymbol{e}$.) The expected degree d and the sparsity parameter q are linked by

$$d = pN = q^2\,.$$

Applying a simple rank-one perturbation formula and shifting the spectrum by a we get from Theorem 2.4 the following corollary whose proof we leave aside.

Corollary 2.12 *Fix $\phi > 0$. Let A satisfies (2.2) and (2.3) with expected degree $N^{2\phi} \le d \le N^{1-2\phi}$. Then the normalized trace m^A of the Green function of A satisfies*

$$|m^A(z) - \widetilde{m}(z + a)| \prec \frac{1}{d} + \frac{1}{N\eta}\,, \tag{2.33}$$

uniformly on the domain $\mathcal{E}$, where $z = E + \mathrm{i}\eta$ and $a = \frac{q}{N}(1 - q^2/N)^{-1/2}$.

Let $\lambda_1^A \ge \lambda_2^A \ge \cdots \ge \lambda_N^A$ denote the eigenvalues of A. The behavior of the largest eigenvalue λ_1^A was fully determined in [15], where it was shown that it has Gaussian fluctuations, *i.e.*

$$\sqrt{\frac{N}{2}}(\lambda_1^A - \mathbb{E}\lambda_1^A) \longrightarrow \mathcal{N}(0, 1)\,, \tag{2.34}$$

in distribution as $N \to \infty$, with $\mathbb{E}\lambda_1^A = f - a + \frac{1}{f} + O(q^{-3})$; see Theorem 6.2 in [15].

Combining Theorem 2.10 with the reasoning of Section 6 of [16], we have the following corollary on the behavior of the second largest eigenvalue λ_2^A of the adjacency matrix A.

Corollary 2.13 *Fix $\phi > 1/6$ and $\phi' > 0$. Let A satisfies (2.2) and (2.3) with expected degree $N^{2\phi} \le d \le N^{1-2\phi'}$. Then,*

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\lambda_2^A - L - a) \le s\right) = F_1(s)\,, \tag{2.35}$$

where $L = 2 + d^{-1} + O(d^{-2})$ is the upper edge of the measure $\widetilde{\rho}$; see (2.28).

We skip the proof of Corollary 2.13 from Theorem 2.10, since it is essentially the same as the proof of Theorem 2.7 in [16], where the result was obtained for $\phi > 1/3$, with L replaced by 2. In analogy with Remark 2.11, the convergence result in (2.35) extends in an obvious way to the eigenvalues $\lambda_{k-1}^A, \ldots, \lambda_2^A$, for any fixed k. The analogous results apply to the k-smallest eigenvalues of A. We leave the details to the interested reader.

Remark 2.14 The largest eigenvalues of sparse random matrices, especially the (shifted and normalized) adjacency matrices of the Erdős–Rényi graphs, can be used to determine the number of clusters in automated community detection algorithms [7,40] in stochastic block models. Corollary 2.13 suggests that the test statistics for such algorithms should reflect the shift of the largest eigenvalues if $N^{-2/3} \lll p \lll N^{-1/3}$, or equivalently, $N^{1/3} \lll d \lll N^{2/3}$. If $p \lll N^{-2/3}$, the test based on the edge universality of random matrices may fail as we have discussed after Theorem 2.10.

In applications, the sparsity should be taken into consideration due to small N even if p is reasonably large. For example, in the Erdős–Rényi graph with $N \simeq 10^3$, the deterministic shift is noticeable if $p \simeq 0.1$, which is in the vicinity of the parameters used in numerical experiments in [7,40].

3 Strategy and outline of proofs

In this section, we outline the strategy of our proofs. We begin with the local law of Theorem 2.4.

3.1 Local laws for the Green function

We start by recalling the approach to the local law for Wigner matrices initiated in [18–20]. Using Schur's complement (or the Feshbach formula) and large deviation estimates for quadratic forms by Hanson and Wright [27], one shows that the normalized trace $m^W(z)$ approximately satisfies the equation $1 + zm^W(z) + m^W(z)^2 \simeq 0$, with high probability, for any z in some appropriate subdomain of $\mathcal{E}$. Using that m_{sc} satisfies (2.11), a local stability analysis then yields $|m^W(z) - m_{sc}(z)| \prec (N\eta)^{-1/2}$, $z \in \mathcal{E}$. In fact, the same quadratic equation is approximately satisfied by each diagonal element of the resolvent, G_{ii}^W, and not only by their average m^W. This observation and an extension of the stability analysis to vectors instead of scalars then yields the *entry-wise local law* [21–23], $|G_{ij}^W(z) - \delta_{ij}m_{sc}(z)| \prec (N\eta)^{-1/2}$, $z \in \mathcal{E}$. (See Sect. 3.2 for some details of this argument.) Taking the normalized trace of the Green function, one expects further cancellations of fluctuations to improve the bound. Exploring the *fluctuation averaging mechanism* for m^W and refining the local stability analysis, one obtains the *strong local law* up to the spectral edges [23], $|m^W(z) - m_{sc}(z)| \prec (N\eta)^{-1}$, $z \in \mathcal{E}$. The fluctuation averaging mechanism was first introduced in [22] and substantially extended in [14,17] to generalized Wigner matrices. We refer to [13,17] for reviews of this general approach. Parallel results were obtained in [47,48]. For further improvements of convergence rates we refer to [9] and [25,26]. For other recent

developments on local laws we refer *e.g.* to [2,3,10] for Wigner matrices with general variance profile and Wigner matrices with correlated entries, and to [8] for heavy-trailed Wigner matrices. Local laws for additive random matrix models were studied in [4,30,34,36,39]. Entry-wise local laws for random graphs models other than the Erdős-Rényi graph were obtained in [5] for d-regular graphs and in [1] for random graphs with given expected degree sequences.

The strategy outlined in the preceding paragraph was applied to sparse random matrices in [15]. The sparsity of the entries manifests itself in the large deviation estimate for quadratic forms, *e.g.* letting (H_{ij}) satisfy (2.12) and choosing (B_{ij}) to be any deterministic $N \times N$ matrix, Lemma 3.8 of [15] assures that

$$\Big| \sum_{k,l} H_{ik} B_{kl} H_{li} - \frac{1}{N} \sum_{k=1}^{N} B_{kk} \Big| \prec \frac{\max_{k,l} |B_{kl}|}{q} + \Big(\frac{1}{N^2} \sum_{k,l} |B_{kl}|^2 \Big)^{1/2}, \quad (3.1)$$

for all $i \in [\![1, N]\!]$. Using the above ideas the entry-wise local law in (2.24) was obtained in [15]. Exploiting the fluctuation averaging mechanism for the normalized trace of the Green function, an additional power of q^{-1} can be gained, leading to (2.23) with the deteriorating factor $(\varkappa + \eta)^{-1/2}$.

To establish the local law for the normalized trace of the Green function which does not deteriorate at the edges, we propose in this paper a novel recursive moment estimate for the Green function. When applied to the proof of the strong local law for a Wigner matrix, it is estimating $\mathbb{E}|1 + zm^W(z) + m^W(z)^2|^D$ by the lower moments $\mathbb{E}|1 + zm^W(z) + m^W(z)^2|^{D-l}$, $l \geq 1$. The use of recursive moment estimate has three main advantages over the previous fluctuation averaging arguments: (1) it is more convenient in conjunction with the cumulant expansion in Lemma 3.2, (2) it is easier to track the higher order terms involving the fourth and higher moments if needed, and (3) it does not require to fully expand the higher power terms and thus simplifies bookkeeping and combinatorics. The same strategy can also be applied to individual entries of the Green function by establishing a recursive moment estimate for $\mathbb{E}|1 + zG_{ii}^W(z) + m_{sc}(z)G_{ii}^W(z)|^D$ and $\mathbb{E}|G_{ij}|^D$ leading to the entry-wise local law. We remark that estimating $\mathbb{E}|1 + zm^W(z) + m^W(z)^2|^D$ is also a key step for the proof of the local law in [25,26], where the estimate is obtained by establishing a bound for the diagonal resolvent entries instead of directly relating it with the lower moments $\mathbb{E}|1 + zm^W(z) + m^W(z)^2|^{D-l}$.

We illustrate this approach for the simple case of the GOE next.

3.2 Local law for the GOE

Choose W to be a GOE matrix. Since $m_{sc} \equiv m_{sc}(z)$, the Stieltjes transform of the semicircle law, satisfies $1 + zm_{sc} + m_{sc}^2 = 0$, we expect that the moments of the polynomial $1 + zm(z) + m(z)^2$, with $m \equiv m^W(z)$ the normalized trace of the Green function $G \equiv G^W(z)$, are small. We introduce the subdomain $\mathcal{D}$ of $\mathcal{E}$ by setting

$$\mathcal{D} := \{z = E + i\eta \in \mathbb{C}^+ : |E| < 3, \ N^{-1} < \eta < 3\}. \quad (3.2)$$

We are going to derive the following *recursive moment estimate* for m. For any $D \geq 2$,

$$
\mathbb{E}[|1 + zm + m^2|^{2D}] \leq \mathbb{E}\left[\frac{\operatorname{Im} m}{N\eta}|1 + zm + m^2|^{2D-1}\right]
$$
$$
+ (4D - 2)\mathbb{E}\left[\frac{\operatorname{Im} m}{(N\eta)^2}|z + 2m| \cdot |1 + zm + m^2|^{2D-2}\right],
$$
$$(3.3)$$

for $z \in \mathbb{C}^+$. Fix now $z \in \mathcal{D}$. Using Young's inequality, the second order Taylor expansion of $m(z)$ around $m_{sc}(z)$ and the *a priori* estimate $|m(z) - m_{sc}(z)| \prec 1$, we conclude from (3.3) with Markov's inequality that

$$
\left|\alpha_{sc}(z)(m(z) - m_{sc}(z)) + (m(z) - m_{sc}(z))^2\right| \prec \frac{|m(z) - m_{sc}(z)|}{N\eta}
$$
$$
+ \frac{\operatorname{Im} m_{sc}(z)}{N\eta} + \frac{1}{(N\eta)^2},
$$
$$(3.4)$$

where $\alpha_{sc}(z) := z + 2m_{sc}(z)$; see Sect. 5.1 for a similar computation. An elementary computation reveals that $|\alpha_{sc}(z)| \asymp \operatorname{Im} m_{sc}(z)$. Equation (3.4) is a *self-consistent equation* for the quantity $m(z) - m_{sc}(z)$. Its local stability properties up to the edges were examined [21,22]. From these stability properties and (3.4) it follows that, for fixed $z \in \mathcal{D}$, $|m(z) - m_{sc}(z)| \prec 1$ implies $|m(z) - m_{sc}(z)| \prec \frac{1}{N\eta}$. To obtain the local law on all of $\mathcal{D}$ one applies a continuity or bootstrapping argument [18,21,22] by decreasing the imaginary part of the spectral parameter from $\eta \asymp 1$ to $\eta \geq N^{-1}$. Using the monotonicity of the Stieltjes transform, this conclusion is extended to all of $\mathcal{E}$. This establishes that

$$
|m(z) - m_{sc}(z)| \prec \frac{1}{N\eta},
$$
$$(3.5)$$

uniformly on the domain $\mathcal{E}$, which is the local law for the GOE.

Hence, to obtain the strong local law for the GOE, it suffices to establish (3.3) for fixed $z \in \mathcal{D}$. By the definition of the normalized trace, $m \equiv m(z)$, of the Green function we have

$$
\mathbb{E}[|1 + zm + m^2|^{2D}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}(1 + zG_{ii}) + m^2\right)(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)}^D\right].
$$
$$(3.6)$$

We expand the diagonal Green function entry $G_{ii} \equiv G_{ii}^W$ using the following identity:

$$
1 + zG_{ii} = \sum_{k=1}^{N} W_{ik}G_{ki},
$$
$$(3.7)$$

which follows directly from the defining relation $(W - z\mathrm{I})G = \mathrm{I}$. To some extent (3.7) replaces the conventional Schur complement formula. We then obtain from (3.6) and (3.7) that

$$\mathbb{E}[|1 + zm + m^2|^{2D}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i,k} W_{ik}G_{ki} + m^2\right)(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}\right].$$
(3.8)

Using that the matrix entries W_{ik} are Gaussian random variables, integration by parts shows that

$$\mathbb{E}_{ik}[W_{ik}F(W_{ik})] = \frac{1 + \delta_{ik}}{N}\mathbb{E}_{ik}[\partial_{ik}F(W_{ik})],$$
(3.9)

for differentiable functions $F : \mathbb{R} \to \mathbb{C}$, where $\partial_{ik} \equiv \partial/(\partial W_{ik})$. Here we used that $\mathbb{E}W_{ij} = 0$ and $\mathbb{E}W_{ij}^2 = (1 + \delta_{ij})/N$ for the GOE. Identity (3.9) is often called *Stein's lemma* in the statistics literature. Combining (3.8) and (3.9) we obtain

$$\mathbb{E}[|1 + zm + m^2|^{2D}] = \frac{1}{N^2}\mathbb{E}\left[\sum_{i,k}(1 + \delta_{ik})\partial_{ik}\left(G_{ki}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}\right)\right]$$
$$+ \mathbb{E}\left[m^2(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}\right].$$
(3.10)

We next expand and estimate the first term on the right side of (3.10). It is easy to see that

$$(1 + \delta_{ik})\partial_{ik}\left(G_{ki}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}\right)$$
$$= -G_{ii}G_{kk}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}$$
$$- G_{ki}G_{ki}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}$$
$$- \frac{2(D-1)}{N}G_{ki}(z + 2m)\sum_{j=1}^{N}G_{jk}G_{ij}(1 + zm + m^2)^{D-2}\overline{(1 + zm + m^2)^D}$$
$$- \frac{2D}{N}G_{ki}(\overline{z} + 2\overline{m})\sum_{j=1}^{N}\overline{G_{jk}G_{ij}}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^{D-1}}.$$
(3.11)

After averaging over the indices i and k, the first term on the right side of (3.11) becomes

$$-\frac{1}{N^2}\mathbb{E}\left[\sum_{i,k}G_{ii}G_{kk}(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}\right]$$
$$= -\mathbb{E}[m^2(1 + zm + m^2)^{D-1}\overline{(1 + zm + m^2)^D}],$$

which exactly cancels with the second term on the right side of (3.10). The second term on the right side of (3.11) can be estimated as

$$\left| \mathbb{E}\left[\frac{1}{N^2} \sum_{i,k} G_{ki} G_{ki} (1 + zm + m^2)^{D-1} \overline{(1 + zm + m^2)^D} \right] \right|$$

$$\leq \mathbb{E}\left[\frac{1}{N^2} \sum_{i,k} |G_{ki}|^2 |1 + zm + m^2|^{2D-1} \right] = \mathbb{E}\left[\frac{\operatorname{Im} m}{N\eta} |1 + zm + m^2|^{2D-1} \right],$$

where we used the identity

$$\frac{1}{N} \sum_{i=1}^{N} |G_{ik}(z)|^2 = \frac{\operatorname{Im} G_{kk}(z)}{N\eta}, \tag{3.12}$$

which we refer to as the *Ward identity* below. It follows from the spectral decomposition of W.

For the third term on the right side of (3.11) we have that

$$\left| \mathbb{E}\left[\frac{1}{N^3} \sum_{i,j,k} G_{ki} G_{jk} G_{ij} (z + 2m)(1 + zm + m^2)^{D-2} \overline{(1 + zm + m^2)^D} \right] \right|$$

$$\leq \mathbb{E}\left[\frac{|\operatorname{Tr} G^3|}{N^3} |z + 2m| \cdot |1 + zm + m^2|^{2D-2} \right]$$

$$\leq \mathbb{E}\left[\frac{\operatorname{Im} m}{(N\eta)^2} |z + 2m| \cdot |1 + zm + m^2|^{2D-2} \right], \tag{3.13}$$

where we used that

$$\frac{1}{N^3} |\operatorname{Tr} G^3| \leq \frac{1}{N^3} \sum_{\alpha=1}^{N} \frac{1}{|\lambda_\alpha - z|^3} \leq \frac{1}{N^2 \eta^2} \frac{1}{N} \sum_{\alpha=1}^{N} \frac{\eta}{|\lambda_\alpha - z|^2} = \frac{\operatorname{Im} m}{N^2 \eta^2}. \tag{3.14}$$

The fourth term on the right side of (3.11) can be estimated in a similar manner since

$$\frac{1}{N^3} \left| \sum_{i,j,k} G_{ki} \overline{G_{ij} G_{jk}} \right| = \frac{1}{N^3} \left| \operatorname{Tr} G \overline{G}^2 \right| \leq \frac{1}{N^3} \sum_{\alpha=1}^{N} \frac{1}{|\lambda_\alpha - z|^3} \leq \frac{\operatorname{Im} m}{N^2 \eta^2}. \tag{3.15}$$

Returning to (3.10), we hence find, for $z \in \mathbb{C}^+$, that

$$\mathbb{E}[|1 + zm + m^2|^{2D}] \leq \mathbb{E}\left[\frac{\operatorname{Im} m}{N\eta} |1 + zm + m^2|^{2D-1} \right]$$

$$+ (4D - 2)\mathbb{E}\left[\frac{\operatorname{Im} m}{(N\eta)^2} |z + 2m| \cdot |1 + zm + m^2|^{2D-2} \right],$$

which is the recursive moment estimate for the GOE stated in (3.3).

Remark 3.1 The above presented method can also be used to obtain the entry-wise local law for the Green function of the GOE. Assuming the local law for m^W has been obtained, one may establish a recursive moment estimate for $1 + zG_{ii}^W(z) + m_{sc}(z)G_{ii}^W(z)$ to derive that

$$\left| G_{ii}^W(z) - m_{sc}(z) \right| \prec \sqrt{\frac{\operatorname{Im} m_{sc}(z)}{N\eta}} + \frac{1}{N\eta}, \tag{3.16}$$

uniformly in $i \in [\![1, N]\!]$ and $z \in \mathcal{E}$. (One may also consider high moments of $1 + zG_{ii}^W(z) + m^W(z)W_{ii}(z)$ to arrive at the same conclusion.) We leave the details to the reader. Yet, for later illustrative purposes in Sects. 6 and 7, we sketch the derivation of the recursive moment estimate for the off-diagonal Green function entries $G_{ij} \equiv G_{ij}^W(z)$, $i \neq j$. Let $z \in \mathcal{D}$. Using the relation $(W - z\mathrm{I})G = \mathrm{I}$, we get

$$\mathbb{E}\left[z|G_{ij}|^{2D} \right] = z\mathbb{E}\left[G_{ij}G_{ij}^{D-1}\overline{G_{ij}^D} \right] = \sum_{k=1}^{N} \mathbb{E}\left[W_{ik}G_{kj}G_{ij}^{D-1}\overline{G_{ij}^D} \right]$$

$$= \sum_{k=1}^{N} \frac{1 + \delta_{ik}}{N} \mathbb{E}\left[\partial_{ik}\left(G_{kj}G_{ij}^{D-1}\overline{G_{ij}^D} \right) \right],$$

where we used Stein's lemma in (3.9) in the last step. Upon computing the derivative we get, for $i \neq j$,

$$\mathbb{E}\left[z|G_{ij}|^{2D} \right] = -\frac{1}{N}\sum_{k=1}^{N} \mathbb{E}\left[G_{kk}G_{ij}G_{ij}^{D-1}\overline{G_{ij}^D} \right] - \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}\left[G_{ki}G_{kj}G_{ij}^{D-1}\overline{G_{ij}^D} \right]$$

$$- \frac{D-1}{N}\sum_{k=1}^{N} \mathbb{E}\left[G_{kj}^2 G_{ii}G_{ij}^{D-2}\overline{G_{ij}^D} \right]$$

$$- \frac{D-1}{N}\sum_{k=1}^{N} \mathbb{E}\left[G_{kj}G_{ik}G_{ij}^{D-1}\overline{G_{ij}^D} \right]$$

$$- \frac{D}{N}\sum_{k=1}^{N} \mathbb{E}\left[|G_{kj}|^2\overline{G_{ii}}G_{ij}^{D-1}\overline{G_{ij}^{D-1}} \right]$$

$$- \frac{D}{N}\sum_{k=1}^{N} \mathbb{E}\left[G_{kj}\overline{G_{ik}}G_{ij}^{D-1}\overline{G_{ij}^D} \right]. \tag{3.17}$$

The first term in the right side of (3.17) equals $-\mathbb{E}\left[m|G_{ij}|^{2D} \right]$. Using the Ward identity (3.12) we have

$$\frac{1}{N}\sum_{k=1}^{N} |G_{ki}G_{kj}| \leq \frac{1}{2N}\sum_{k=1}^{N} |G_{ki}|^2 + \frac{1}{2N}\sum_{k=1}^{N} |G_{kj}|^2 = \frac{\operatorname{Im} G_{ii} + \operatorname{Im} G_{jj}}{2N\eta}.$$

Thus using (3.16), we get the bound

$$\frac{1}{N} \sum_{k=1}^{N} |G_{ki} G_{kj}| \prec \frac{\operatorname{Im} m_{\mathrm{sc}}}{N\eta} + \frac{1}{(N\eta)^2},$$

uniformly on $\mathcal{E}$. We can now easily bound the right side of (3.17), *e.g.* for any given $\epsilon > 0$,

$$\left| \frac{D}{N} \sum_{k=1}^{N} \mathbb{E} \left[|G_{kj}|^2 \overline{G_{ii}} G_{ij}^{D-1} \overline{G_{ij}^{D-1}} \right] \right| \le N^\epsilon \mathbb{E} \left[\left(\frac{\operatorname{Im} m_{\mathrm{sc}}}{N\eta} + \frac{1}{(N\eta)^2} \right) |G_{ij}|^{2D-2} \right],$$

for N sufficiently large. Thus we get from (3.17) that, for $i \ne j$,

$$\begin{aligned}
|z + m_{\mathrm{sc}}| \, \mathbb{E}\big[|G_{ij}|^{2D}\big] \le{} & \mathbb{E}\Big[|m - m_{\mathrm{sc}}||G_{ij}|^{2D}\Big] \\
& + N^\epsilon \mathbb{E} \left[\left(\frac{\operatorname{Im} m_{\mathrm{sc}}}{N\eta} + \frac{1}{(N\eta)^2} \right) |G_{ij}|^{2D-2} \right] \\
& + N^\epsilon \mathbb{E} \left[\left(\frac{\operatorname{Im} m_{\mathrm{sc}}}{N\eta} + \frac{1}{(N\eta)^2} \right) |G_{ij}|^{2D-1} \right],
\end{aligned} \qquad (3.18)$$

uniformly on $\mathcal{E}$, for N sufficiently large. Since $|z + m_{\mathrm{sc}}| > c$ on $\mathcal{E}$, for some N-independent constant $c > 0$, we find from (3.18) and (3.5) by Young's and Markov's inequality that

$$\left| G_{ij}^W(z) \right| \prec \sqrt{\frac{\operatorname{Im} m_{\mathrm{sc}}(z)}{N\eta}} + \frac{1}{N\eta}, \qquad (i \ne j), \qquad (3.19)$$

for fixed $z \in \mathcal{D}$. Using continuity and monotonicity of $G_{ij}(z)$ the bound can be made uniform on the domain $\mathcal{E}$. Together with (3.16), this shows the entry-wise local law for the GOE.

3.3 Local law for sparse matrices

When applying the strategy of Sect. 3.2 to sparse matrices we face two difficulties. First, since the matrix entries are not Gaussian random variables, the simple integration by parts formula (3.9) needs to replaced by a full-fletched cumulant expansion. Second, since the higher order cumulants are not small (in the sense that the $(\ell+2)$-nd cumulant is only $O(N^{-1}q^{-\ell})$) we need to retain higher orders in the cumulant expansion. The following result generalizes (3.9).

Lemma 3.2 (Cumulant expansion, generalized Stein lemma) *Fix* $\ell \in \mathbb{N}$ *and let* $F \in C^{\ell+1}(\mathbb{R}; \mathbb{C}^+)$. *Let* Y *be a centered random variable with finite moments to order* $\ell+2$. *Then,*

$$\mathbb{E}[YF(Y)] = \sum_{r=1}^{\ell} \frac{\kappa^{(r+1)}(Y)}{r!} \mathbb{E}\big[F^{(r)}(Y)\big] + \mathbb{E}\big[\Omega_\ell(YF(Y))\big], \qquad (3.20)$$

where $\mathbb{E}$ denotes the expectation with respect to Y, $\kappa^{(r+1)}(Y)$ denotes the $(r+1)$-th cumulant of Y and $F^{(r)}$ denotes the r-th derivative of the function F. The error term $\Omega_\ell(YF(Y))$ in (3.20) satisfies

$$\big|\mathbb{E}\big[\Omega_\ell(YF(Y))\big]\big| \le C_\ell \mathbb{E}[|Y|^{\ell+2}] \sup_{|t| \le Q} |F^{(\ell+1)}(t)|$$
$$+ C_\ell \mathbb{E}[|Y|^{\ell+2}\mathbb{1}(|Y| > Q)] \sup_{t \in \mathbb{R}} |F^{(\ell+1)}(t)|, \qquad (3.21)$$

where $Q \ge 0$ is an arbitrary fixed cutoff and C_ℓ satisfies $C_\ell \le \frac{(C\ell)^\ell}{\ell!}$ for some numerical constant C.

In case Y is a standard Gaussian we recover (3.20) and we thus sometimes refer to Lemma 3.2 as generalized Stein lemma. The cumulant expansion in combination with the Green function was successfully used to study linear eigenvalue statistics of Wigner matrices in [31,41] and more recently in [28]. For a proof of Lemma 3.2, we refer to Proposition 3.1 in [41].

Let H be a sparse matrix satisfying Assumption 2.3 with $\phi > 0$. Recall the polynomial $P \equiv P_z$ and the function $\widetilde{m} \equiv \widetilde{m}(z)$ of Theorem 2.4 which satisfy $P(\widetilde{m}) = 0$. Let $m \equiv m^H(z)$ be given by (2.16). Following the ideas of Sect. 3.2, we derive in Sect. 6 a recursive estimate for $\mathbb{E}|P(m)|^{2D}$, for large D with $z \in \mathcal{D}$; see Lemma 5.1 for the precise statement and Sect. 6 for its proof. We start with

$$\mathbb{E}[|P|^{2D}] = \mathbb{E}\Big[\Big(1 + zm + m^2 + \frac{s^{(4)}}{q_t^2}m^4\Big)P^{D-1}\overline{P^D}\Big], \qquad (3.22)$$

for $D \ge 2$, and expand zm using the identity

$$zG_{ij} = \sum_{k=1}^{N} H_{ik}G_{kj} - \delta_{ij}, \qquad (3.23)$$

which follows from the definition of the Green function. We then obtain the identity

$$\mathbb{E}[(1 + zm)P^{D-1}\overline{P^D}] = \mathbb{E}\Big[\Big(\frac{1}{N}\sum_{i \ne k} H_{ik}G_{ki}\Big)P^{D-1}\overline{P^D}\Big]. \qquad (3.24)$$

Using the generalized Stein lemma, Lemma 3.2, we get

$$\mathbb{E}[(1 + zm)P^{D-1}\overline{P^D}] = \frac{1}{N}\sum_{r=1}^{\ell} \frac{\kappa^{(r+1)}}{r!} \mathbb{E}\Big[\sum_{i \ne k} \partial_{ik}^r\Big(G_{ki}P^{D-1}\overline{P^D}\Big)\Big]$$
$$+ \mathbb{E}\Omega_\ell\Big((1 + zm)P^{D-1}\overline{P^D}\Big), \qquad (3.25)$$

where $\partial_{ik} = \partial/(\partial H_{ik})$ and $\kappa^{(k)}$ are the cumulants of H_{ij}, $i \neq j$. The detailed form of the error $\mathbb{E}\Omega_\ell(\cdot)$ is discussed in Sect. 6.1. Anticipating the outcome, we mention that we can truncate the expansion at order $\ell \geq 8D$ so that the error term becomes sufficiently small for our purposes.

In Sect. 6, we prove that the relevant terms in (3.25) cancel with the third and fourth term on the right side of (3.22). This yields the recursive moment estimate for $P(m)$, respectively m. The proof is based on the following observations:

- The leading term on the right side of (3.25) comes from the $r = 1$ term and is identified to be $\mathbb{E}[-m^2 P^{D-1}\overline{P^D}]$, which cancels with the third term on the right side of (3.22). See the first part of Lemma 6.4 and Sect. 6.3, where the estimate is obtained by following the discussion in Sect. 3.2.
- The second largest term on the right side of (3.25) comes from the $r = 3$ term and is identified to be $s^{(4)}q^{-2}\mathbb{E}[-N^{-2}(\sum_i (G_{ii})^2)^2 P^{D-1}\overline{P^D}]$. In Lemma 6.12 we prove that it cancels with the fourth term on the right side of (3.22) after removing some negligible contributions (see the beginning of Sect. 6 for quantitative statement of negligible contributions). See the second part of Lemma 6.4 and Sect. 6.7.
- The $r = 2$ term is negligible. The main ideas in this estimate are the identification of an unmatched index and the cumulant expansion of the term in the unmatched index. See Sect. 6.4.
- The negligible contributions, including the terms with $r \geq 4$, can be identified by power counting or further expansions using cumulant series. For this analysis, we rely on the entry-wise local law (2.24) and ideas inspired by the GOE computation in Remark 3.1 above.

As we will see in Sect. 4, the inclusion of the fourth moment $s^{(4)}/q^2$ in $P(m)$ enables us to compute the deterministic shift of edge which is of order q^{-2}. While it is possible to include a higher order correction term involving the sixth moment, $s^{(6)}/q^4$, this would not improve the local law in our proof since the largest among the negligible contributions originates from the $r = 3$ term in (3.25). (More precisely, it is $I_{3,2}$ of (6.6).)

In Sect. 5, we then prove Theorem 2.4 and Theorem 2.9 using the recursive moment estimate for m and a local stability analysis. The local stability analysis relies on some properties of the Stieltjes transform $\widetilde{m}$ of the deterministic distribution $\widetilde{\rho}$ obtained in Sect. 4.

3.4 Tracy–Widom limit and Green function comparison

To establish the edge universality (for $\phi > 1/6$), we first show in Sect. 7.1 that the distribution of the largest eigenvalue of H may be obtained as the expectation (of smooth functions) of the imaginary part of $m(z)$, for appropriately chosen spectral parameters z. Such a relation was the basic structure for proving the edge universality in [16,23], and the main ingredients in the argument are the local law, the square-root decay at the edge of the limiting density, and an upper bound on the largest eigenvalue, which are Theorems 2.4, 2.9 and Lemma 4.1 for the case at hand. For the sake of completeness, we redo some parts of these estimates in Sect. 7.1.

In Sect. 7.2, we then use the *Green function comparison* method [21,23] to compare the edge statistics of H with the edge statistics of a GOE matrix. Together with the argument of Sect. 7.1, this will yield the Tracy–Widom limit of the largest eigenvalue. However, the conventional discrete Lindeberg type replacement approach to the Green function comparison does not work due to the slow decaying moments of the sparse matrix. We therefore use a continuous flow that interpolates between the sparse matrix ensemble and the GOE. Such an approach has shown to be effective in proving edge universality for deformed Wigner matrices [37] and for sample covariance matrices [38].

More concretely, we consider the *Dyson matrix flow* with initial condition H_0 defined by

$$H_t := e^{-t/2} H_0 + \sqrt{1 - e^{-t}} W^{\mathrm{GOE}}, \qquad (t \geq 0), \qquad (3.26)$$

where W^{GOE} is a GOE matrix independent of H_0. In fact, since we will choose H_0 to be a sparse matrix H with vanishing diagonal entries, we assume with some abuse of terminology that $W^{\mathrm{GOE}} = (W^{\mathrm{GOE}})^*$ has vanishing diagonal, *i.e.* we assume that $W^{\mathrm{GOE}}_{ii} = 0$ and that $(W^{\mathrm{GOE}}_{ij}, i < j)$ are independent centered Gaussian random variables of variance $1/N$. It was shown in Lemma 3.5 of [35] that the local edge statistics of W^{GOE} is described by the GOE Tracy–Widom statistics.

Let $\kappa_t^{(k)}$ be the k-th cumulant of $(H_t)_{ij}, i \neq j$. Then, by the linearity of the cumulants under the addition of independent random variables, we have $\kappa_t^{(1)} = 0$, $\kappa_t^{(2)} = 1/N$ and $\kappa_t^{(k)} = e^{-kt/2} \kappa^{(k)}, k \geq 3$. In particular, we have the bound

$$\left| \kappa_t^{(k)} \right| \leq e^{-t} \frac{(Ck)^{ck}}{N q_t^{k-2}}, \qquad (k \geq 3), \qquad (3.27)$$

where we introduced the *time-dependent sparsity parameter*

$$q_t := q \, e^{t/2}. \qquad (3.28)$$

Choosing $t = 6 \log N$, a straightforward perturbation argument shows that the local statistics, at the edges and in the bulk, of H_t and W^{GOE} agree up to negligible error. It thus suffices to consider $t \in [0, 6 \log N]$.

We first establish the local law for the normalized trace of the Green function of H_t. Let

$$G_t(z) \equiv G^{H_t}(z) = \frac{1}{H_t - z\mathrm{I}}, \qquad m_t(z) \equiv m^{H_t}(z) = \frac{1}{N} \sum_{i=1}^{N} (G_t)_{ii}(z), \qquad (z \in \mathbb{C}^+). \qquad (3.29)$$

Proposition 3.3 *Let H_0 satisfy Assumption 2.3 with $\phi > 0$. Then, for any $t \geq 0$, there exists an algebraic function $\widetilde{m}_t : \mathbb{C}^+ \to \mathbb{C}^+$ and $2 \leq L_t < 3$ such that the following holds:*

(1) $\widetilde{m}_t$ is the Stieltjes transform of a deterministic symmetric probability measure $\widetilde{\rho}_t$, i.e. $\widetilde{m}_t(z) = m_{\widetilde{\rho}_t}(z)$. Moreover, supp $\widetilde{\rho}_t = [-L_t, L_t]$ and $\widetilde{\rho}_t$ is absolutely continuous with respect to the Lebesgue measure with a strictly positive density on $(-L_t, L_t)$.

(2) $\widetilde{m}_t \equiv \widetilde{m}_t(z)$ is a solution to the polynomial equation

$$\begin{aligned} P_{t,z}(\widetilde{m}_t) &:= 1 + z\widetilde{m}_t + \widetilde{m}_t^2 + \mathrm{e}^{-t}q_t^{-2}s^{(4)}\widetilde{m}_t^4 \\ &= 1 + z\widetilde{m}_t + \widetilde{m}_t^2 + \mathrm{e}^{-2t}q^{-2}s^{(4)}\widetilde{m}_t^4 = 0. \end{aligned} \tag{3.30}$$

(3) The normalized trace of the Green function satisfies

$$|m_t(z) - \widetilde{m}_t(z)| \prec \frac{1}{q_t^2} + \frac{1}{N\eta}, \tag{3.31}$$

uniformly on the domain $\mathcal{E}$ and uniformly in $t \in [0, 6\log N]$.

Note that Theorem 2.4 is a special case of Proposition 3.3. Given Proposition 3.3, Corollary 2.5 extends in the obvious way from H to H_t. Proposition 3.3 is proved in Sect. 5.1.

The endpoints $\pm L_t$ of the support of $\widetilde{\rho}_t$ are given by $L_t = 2 + \mathrm{e}^{-t}s^{(4)}q_t^{-2} + O(\mathrm{e}^{-2t}q_t^{-4})$ and satisfy

$$\dot{L}_t = -2\mathrm{e}^{-t}s^{(4)}q_t^{-2} + O(\mathrm{e}^{-2t}q_t^{-4}), \tag{3.32}$$

where $\dot{L}_t$ denotes the derivative with respect to t of L_t; cf. Remark 4.2 below.

Choose now $q \geq N^{\phi}$ with $\phi > 1/6$. In our proof of the Green function comparison theorem, Proposition 7.2, we estimate the rate of change of m_t along the Dyson matrix flow over the time interval $[0, 6\log N]$, where it undergoes a change of $o(1)$. The continuous changes in m_t can be compensated by letting evolve the spectral parameter $z \equiv z(t)$ according to (3.32). This type of cancellation argument appeared first in [37] in the context of deformed Wigner matrices. However, one cannot prove the Green function comparison theorem for sparse random matrices by directly applying the cancellation argument since the error bound for the entry-wise local law in Proposition 2.6 is not sufficiently small. Thus the proof of the Green function comparison theorem requires some non-trivial estimates on functions of Green functions as is explained in Sect. 7.2.

4 The measure $\widetilde{\rho}$ and its Stieltjes transform

In this section, we prove important properties of $\widetilde{m}_t \equiv \widetilde{m}_t(z)$ introduced in Proposition 3.3. Recall that $\widetilde{m}_t$ is a solution to the polynomial equation $P_{t,z}(\widetilde{m}_t) = 0$ in (3.30) and that we set $q_t := \mathrm{e}^{t/2}q$. Recall moreover the domain $\mathcal{E}$ defined in (2.18).

Lemma 4.1 *For any fixed $z = E + \mathrm{i}\eta \in \mathcal{E}$ and any $t \geq 0$, the polynomial equation $P_{t,z}(w_t) = 0$ has a unique solution $w_t \equiv w_t(z)$ satisfying $\mathrm{Im}\, w_t > 0$ and $|w_t| \leq 5$. Moreover, w_t has the following properties:*

(1) *There exists a probability measure $\widetilde{\rho}_t$ such that the analytic continuation of w_t coincides with the Stieltjes transform of $\widetilde{\rho}_t$.*

(2) *The probability measure $\widetilde{\rho}_t$ is supported on $[-L_t, L_t]$, for some $L_t \geq 2$, it has a strictly positive density inside its support and it vanishes as a square-root at the edges, i.e. letting*

$$\varkappa_t \equiv \varkappa_t(E) := \min\{|E + L_t|, |E - L_t|\}, \tag{4.1}$$

we have

$$\widetilde{\rho}_t(E) \asymp \varkappa_t^{1/2}(E), \qquad (E \in (-L_t, L_t)). \tag{4.2}$$

Moreover, $L_t = 2 + e^{-t}q_t^{-2}s^{(4)} + O(e^{-2t}q_t^{-4})$.

(3) *The solution w_t satisfies*

$$\begin{aligned}
\operatorname{Im} w_t(E + i\eta) &\asymp \sqrt{\varkappa_t + \eta} && \text{if } E \in [-L_t, L_t], \\
\operatorname{Im} w_t(E + i\eta) &\asymp \frac{\eta}{\sqrt{\varkappa_t + \eta}} && \text{if } E \notin [-L_t, L_t].
\end{aligned} \tag{4.3}$$

Proof For simplicity, we abbreviate $P \equiv P_{t,z}$. Let

$$Q(w) \equiv Q_{t,z}(w_t) := -\frac{1}{w_t} - w_t - e^{-t}q_t^{-2}s^{(4)}w_t^3. \tag{4.4}$$

By definition, $P(w) = 0$ if and only if $z = Q(w)$. It is easily checked that the derivative

$$Q'(w) = \frac{1}{w^2} - 1 - 3e^{-t}q_t^{-2}s^{(4)}w^2 \tag{4.5}$$

is monotone increasing on $(-\infty, 0)$. Furthermore, we have

$$\begin{aligned}
Q'(-1) &= -3e^{-t}q_t^{-2}s^{(4)} < 0, \quad Q'(-1 + 2q_t^{-2}s^{(4)}) \\
&= (4 - 3e^{-t})q_t^{-2}s^{(4)} + O(q_t^{-4}) > 0.
\end{aligned}$$

Hence, $Q'(w) = 0$ has a unique solution on $(-1, -1 + 2q_t^{-2}s^{(4)})$, which we will denote by τ_t, and $Q(w) \equiv Q_t(w_t)$ attains its minimum on $(-\infty, 0)$ at $w_t = \tau_t$. We let $L_t := Q_t(\tau_t)$, or equivalently, $w_t = \tau_t$ if $z = L_t$. A direct calculation shows that

$$\tau_t = -1 + \frac{3e^{-t}}{2}q_t^{-2}s^{(4)} + O(e^{-2t}q_t^{-4}), \qquad L_t = 2 + e^{-t}q_t^{-2}s^{(4)} + O(e^{-2t}q_t^{-4}). \tag{4.6}$$

For simplicity we let $L \equiv L_t$ and $\tau \equiv \tau_t$. Choosing now $w = Q^{-1}(z)$ we have the expansion

$$z = Q(\tau) + Q'(\tau)(w - \tau) + \frac{Q''(\tau)}{2}(w - \tau)^2 + O(|w - \tau|^3)$$
$$= L + \frac{Q''(\tau)}{2}(w - \tau)^2 + O(|w - \tau|^3), \qquad (4.7)$$

in a $q_t^{-1/2}$-neighborhood of τ_t. We hence find that

$$w = \tau + \left(\frac{2}{Q''(\tau)}\right)^{1/2} \sqrt{z - L} + O(|z - L|), \qquad (4.8)$$

in that neighborhood. In particular, choosing the branch of the square root so that $\sqrt{z - L} \in \mathbb{C}^+$, we find that $\operatorname{Im} w > 0$ since $Q''(\tau) > 0$. We note that there exists another solution with negative imaginary part, which corresponds to the different branch of the square root.

We can apply the same argument with a solution of the equation $Q'(w) = 0$ on $(1 - 2q_t^{-2}s^{(4)}, 1)$ leading us to the relation

$$w = -\tau + \left(\frac{2}{Q''(\tau)}\right)^{1/2} \sqrt{z + L} + O(|z + L|), \qquad (4.9)$$

in a $q_t^{1/2}$-neighborhood of $-\tau$.

For uniqueness, we consider the disk $B_5 = \{w \in \mathbb{C} : |w| < 5\}$. On its boundary ∂B_5

$$|w^2 + zw + 1| \geq |w|^2 - |z| \cdot |w| - 1 > 1 > \left|q_t^{-2}w^4 s^{(4)}\right|, \qquad (4.10)$$

for $z \in \mathcal{E}$. Hence, by Rouché's theorem, the equation $P(w) = 0$ has the same number of roots as the quadratic equation $w^2 + zw + 1 = 0$ in B_5. Since $w^2 + zw + 1 = 0$ has two solutions on B_5, we find that $P(w) = 0$ has two solutions on it. For $z = L + iq_t^{-1}$, we can easily check that one solution of $P(w) = 0$ has positive imaginary part (from choosing the branch of the square root as in (4.8)) and the other solution has negative imaginary part. If both solutions of $P(w) = 0$ are in $\mathbb{C}^+ \cup \mathbb{R}$ (or in $\mathbb{C}^- \cup \mathbb{R}$) for some $z = \tilde{z} \in \mathbb{C}^+$, then by continuity, there exists z' on the line segment joining $L + iq_t^{-1}$ and $\tilde{z}$ such that $P_{t,z'}(w') = 0$ for some $w' \in \mathbb{R}$. By the definition of P this cannot happen, hence one solution of $P(w) = 0$ is in $\mathbb{C}^+$ and the other in $\mathbb{C}^-$, for any $z \in \mathcal{E}$. This shows the uniqueness statement of the lemma.

Next, we extend $w \equiv w(z)$ to cover $z \notin \mathcal{E}$. (With slight abuse of notation, the extension of w will also be denoted by $w \equiv w(z)$.) Repeating the argument of the previous paragraph, we find that $P(w) = 0$ has two solutions for $z \in (-L, L)$. Furthermore, we can also check that exactly one of them is in $\mathbb{C}^+$ by considering $z = \pm L \mp q_t^{-1}$ and using continuity. Thus, $w(z)$ forms a curve on $\mathbb{C}^+$, joining $-\tau$

and τ, which we will denote by Γ. We remark that, by the inverse function theorem, $w(z)$ is analytic for $z \in (-L, L)$ since $Q'(w) \neq 0$ for such z.

By symmetry, Γ intersects the imaginary axis at $w(0)$. On the imaginary axis, we find that

$$Q'(w) = \frac{1}{w^2} - 1 - 3e^{-t}q_t^{-2}s^{(4)}w^2 < 0 \qquad \text{if } |w| < 5. \qquad (4.11)$$

Thus, we get from $Q(w) = z$ that

$$\frac{dw}{d\eta} = \frac{1}{Q'(w)}\frac{dz}{d\eta} = \frac{i}{Q'(w)}, \qquad (4.12)$$

which shows in particular that $w(i)$ is purely imaginary and $\operatorname{Im} w(i) < \operatorname{Im} w(0)$. By continuity, this shows that the analytic continuation of $w(z)$ for $z \in \mathbb{C}^+$ is contained in the domain D_Γ enclosed by Γ and the interval $[-L, L]$. We also find that $|w(z)| < 5$, for all $z \in \mathbb{C}^+$.

To prove that $w(z)$ is analytic in $\mathbb{C}^+$, it suffices to show that $Q'(w) \neq 0$ for $w \in D_\Gamma$. If $Q'(w) = 0$ for $w \in D_\Gamma$, we have

$$w^2 Q'(w) = 1 - w^2 - 3e^{-t}q_t^{-2}s^{(4)}w^4 = 0. \qquad (4.13)$$

On the circle $\{w \in \mathbb{C} : |w| = 5\}$,

$$|1 - w^2| \geq 24 > \left|3e^{-t}q_t^{-2}w^4 s^{(4)}\right|. \qquad (4.14)$$

Hence, again by Rouché's theorem, $w^2 Q'(w) = 0$ has two solutions in the disk $\{w \in \mathbb{C} : |w| < 5\}$. We already know that those two solutions are $\pm\tau$. Thus, $Q'(w) \neq 0$ for $w \in D_\Gamma$ and $w(z)$ is analytic.

Let $\widetilde{\rho}$ be the measure obtained by the Stieltjes inversion of $w \equiv w(z)$. To show that $\widetilde{\rho}$ is a probability measure, it suffices to show that $\lim_{y\to\infty} iy\, w(iy) = -1$. Since w is bounded, one can easily check from the definition of w that $|w| \to 0$ as $|z| \to \infty$. Thus, by considering $z = iy$ and letting $y \to \infty$,

$$0 = \lim_{y\to\infty} P(w(iy)) = \lim_{y\to\infty}(1 + iy\, w(iy)), \qquad (4.15)$$

which implies that $\lim_{y\to\infty} iy\, w(iy) = -1$. This proves the first property of $\widetilde{\rho}$. Other properties can be easily proved from the first property and Eqs. (4.8) and (4.9). $\qquad \square$

Remark 4.2 Recall that $q_t = e^{t/2}q$. As we have seen in (4.6),

$$L_t = 2 + e^{-t}q_t^{-2}s^{(4)} + O(e^{-2t}q_t^{-4}) = 2 + e^{-2t}q^{-2}s^{(4)} + O(e^{-4t}q^{-4}).$$

Moreover, the time derivative of L_t satisfies

$$\dot{L}_t = \frac{d}{dt}Q(\tau) = \frac{\partial Q}{\partial t}(\tau) + Q'(\tau)\dot{\tau} = \frac{\partial Q}{\partial t}(\tau) = 2e^{-2t}q^{-2}s^{(4)}\tau^3,$$

hence, referring once more to (4.6),

$$\dot{L}_t = -2\mathrm{e}^{-t}q_t^{-2}s^{(4)} + O(\mathrm{e}^{-2t}q_t^{-4}) = -2\mathrm{e}^{-2t}q^{-2}s^{(4)} + O(\mathrm{e}^{-4t}q^{-4}).$$

Remark 4.3 It can be easily seen from the definition of $P_{t,z}$ that $w_t \to m_{\mathrm{sc}}$ as $N \to \infty$ or $t \to \infty$. For $z \in \mathcal{E}$, we can also check the *stability condition* $|z + w_t| > 1/6$ since

$$|z + w_t| = \frac{|1 + \mathrm{e}^{-2t}q^{-2}s^{(4)}|w_t|^4|}{|w_t|} \tag{4.16}$$

and $|w_t| < 5$, as we have seen in the proof of Lemma 4.1.

5 Proof of Proposition 3.3 and Theorem 2.9

5.1 Proof of Proposition 3.3

In this section, we prove Proposition 3.3. The main ingredient of the proof is the recursive moment estimate for $P(m_t)$. Recall the subdomain $\mathcal{D}$ of $\mathcal{E}$ defined in (3.2) and the matrix H_t, $t \geq 0$, defined in (3.26). We have the following result.

Lemma 5.1 (Recursive moment estimate) *Fix $\phi > 0$ and suppose that H_0 satisfies Assumption 2.3. Fix any $t \geq 0$. Recall the definition of the polynomial $P \equiv P_{t,z}$ in (3.30). Then, for any $D > 10$ and (small) $\epsilon > 0$, the normalized trace of the Green function, $m_t \equiv m_t(z)$, of the matrix H_t satisfies*

$$
\begin{aligned}
\mathbb{E}\,|P(m_t)|^{2D} \leq\ & N^\epsilon\,\mathbb{E}\!\left[\left(\frac{1}{q_t^4} + \frac{\mathrm{Im}\,m_t}{N\eta}\right)|P(m_t)|^{2D-1}\right] \\
& + N^{-\epsilon/8}q_t^{-1}\,\mathbb{E}\!\left[|m_t - \widetilde{m}_t|^2|P(m_t)|^{2D-1}\right] \\
& + N^\epsilon q_t^{-1}\sum_{s=2}^{2D}\sum_{s'=0}^{s-2}\mathbb{E}\!\left[\left(\frac{\mathrm{Im}\,m_t}{N\eta}\right)^{2s-s'-2}|P'(m_t)|^{s'}|P(m_t)|^{2D-s}\right] \\
& + N^\epsilon q_t^{-8D} + N^\epsilon\sum_{s=2}^{2D}\mathbb{E}\!\left[\left(\frac{1}{N\eta} + \frac{1}{q_t}\left(\frac{\mathrm{Im}\,m_t}{N\eta}\right)^{1/2} + \frac{1}{q_t^2}\right)\left(\frac{\mathrm{Im}\,m_t}{N\eta}\right)^{s-1}\right. \\
& \hspace{2cm} \left. \times\,|P'(m_t)|^{s-1}|P(m_t)|^{2D-s}\right],
\end{aligned}
\tag{5.1}
$$

uniformly on the domain $\mathcal{D}$, for N sufficiently large.

The proof of Lemma 5.1 is postponed to Sect. 6. We are now ready to prove Proposition 3.3.

Proof of Proposition 3.3 and Theorem 2.4 Let $\widetilde{m}_t$ be the solution w_t in Lemma 4.1. Parts (1) and (2) of Proposition 3.3 and Theorem 2.4 were already proved in

Lemma 4.1, so it suffices to prove the statements (3). For simplicity we omit the z-dependence from the notation. Fix $t \in [0, 6 \log N]$. Let

$$\Lambda_t := |m_t - \widetilde{m}_t|. \tag{5.2}$$

We remark that from the local law in Lemma 2.6, we have $\Lambda_t \prec 1$, $z \in \mathcal{D}$. We also define the following z-dependent deterministic parameters

$$\alpha_1 := \operatorname{Im} \widetilde{m}_t, \qquad \alpha_2 := P'(\widetilde{m}_t), \qquad \beta := \frac{1}{q_t^2} + \frac{1}{N\eta}, \tag{5.3}$$

with $z = E + i\eta$. We note that

$$|\alpha_2| \geq \operatorname{Im} P'(\widetilde{m}_t) = \operatorname{Im} z + 2 \operatorname{Im} m_t + 4\mathrm{e}^{-t} s^{(4)} q_t^{-2} \left(3 (\operatorname{Re} \widetilde{m}_t)^2 \operatorname{Im} \widetilde{m}_t - (\operatorname{Im} \widetilde{m}_t)^3 \right)$$

$$\geq \operatorname{Im} \widetilde{m}_t = \alpha_1,$$

since $|\widetilde{m}_t| \leq 5$ as proved in Lemma 4.1. Recall that $\widetilde{m}_t(L) = \tau$ in the proof of Lemma 4.1. Recalling the definition of $\varkappa_t \equiv \varkappa_t(E)$ in (4.1) and using (4.8) we have

$$|\widetilde{m}_t - \tau| \asymp \sqrt{z - L} \asymp \sqrt{\varkappa_t + \eta}.$$

By the definitions of τ and L in the proof of Lemma 4.1, we also have that

$$L + 2\tau + 4\mathrm{e}^{-t} q_t^{-2} s^{(4)} \tau^3 = -\frac{1}{\tau} - \tau - \mathrm{e}^{-t} q_t^{-2} \tau^3 s^{(4)} + 2\tau + 4\mathrm{e}^{-t} q_t^{-2} \tau^3 s^{(4)}$$

$$= -\tau \left(\frac{1}{\tau^2} - 1 - 3\mathrm{e}^{-t} q_t^{-2} \tau^2 s^{(4)} \right) = 0,$$

hence

$$P'(\widetilde{m}_t) = z + 2\widetilde{m}_t + 4\mathrm{e}^{-t} q_t^{-2} \widetilde{m}_t^3 s^{(4)} = (z - L) + 2(\widetilde{m}_t - \tau)$$

$$+ 4\mathrm{e}^{-t} q_t^{-2} s^{(4)} (\widetilde{m}_t^3 - \tau^3), \tag{5.4}$$

and we find from (4.8) that

$$|\alpha_2| = |P'(\widetilde{m}_t)| \asymp \sqrt{\varkappa_t + \eta}.$$

We remark that the parameter α_1 is needed only for the proof of Theorem 2.9; the proof of Proposition 3.3 can be done simply by substituting every α_1 below with $|\alpha_2|$.

Recall that, for any $a, b \geq 0$ and $\mathsf{p}, \mathsf{q} > 1$ with $\mathsf{p}^{-1} + \mathsf{q}^{-1} = 1$, Young's inequality states

$$ab \leq \frac{a^{\mathsf{p}}}{\mathsf{p}} + \frac{b^{\mathsf{q}}}{\mathsf{q}}. \tag{5.5}$$

Let $D \geq 10$ and choose any (small) $\epsilon > 0$. All estimates below hold for N sufficiently large (depending on D and ϵ). For brevity, N is henceforth implicitly assumed to be sufficiently large. Using first that $\operatorname{Im} m_t \leq \operatorname{Im} \widetilde{m}_t + |m_t - \widetilde{m}_t| = \alpha_1 + \Lambda_t$ and then applying (5.5) with $\mathsf{p} = 2D$ and $\mathsf{q} = 2D/(2D-1)$, we get for the first term on the right side of (5.1) that

$$
\begin{aligned}
N^\epsilon \left(\frac{\operatorname{Im} m_t}{N\eta} + q_t^{-4} \right) |P(m_t)|^{2D-1} &\leq N^\epsilon \frac{\alpha_1 + \Lambda_t}{N\eta} |P(m_t)|^{2D-1} + N^\epsilon q_t^{-4} |P(m_t)|^{2D-1} \\
&\leq \frac{N^{(2D+1)\epsilon}}{2D} \left(\frac{\alpha_1 + \Lambda_t}{N\eta} \right)^{2D} + \frac{N^{(2D+1)\epsilon}}{2D} q_t^{-8D} + \frac{2(2D-1)}{2D} \cdot N^{-\frac{\epsilon}{2D-1}} |P(m_t)|^{2D} .
\end{aligned}
$$
(5.6)

Similarly, for the second term on the right side of (5.1), we have

$$
N^{-\epsilon/8} q_t^{-1} \Lambda_t^2 |P(m_t)|^{2D-1} \leq \frac{N^{-(D/4-1)\epsilon}}{2D} q_t^{-2D} \Lambda_t^{4D} + \frac{2D-1}{2D} N^{-\frac{\epsilon}{2D-1}} |P(m_t)|^{2D} .
$$
(5.7)

From the Taylor expansion of $P'(m_t)$ around $\widetilde{m}_t$, we have

$$
|P'(m_t) - P'(\widetilde{m}_t) - P''(\widetilde{m}_t)(m_t - \widetilde{m}_t)| \leq C q_t^{-2} \Lambda_t^2 ,
$$
(5.8)

and $|P'(m_t)| \leq |\alpha_2| + 3\Lambda_t$, for all $z \in \mathcal{D}$, with high probability since $P''(\widetilde{m}_t) = 2 + O(q_t^{-2})$ and $\Lambda_t \prec 1$ by assumption. We note that, for any fixed $s \geq 2$,

$$
\begin{aligned}
(\alpha_1 + \Lambda_t)^{2s-s'-2} (|\alpha_2| + 3\Lambda_t)^{s'} &\leq N^{\epsilon/2} (\alpha_1 + \Lambda_t)^{s-1} (|\alpha_2| + 3\Lambda_t)^{s-1} \\
&\leq N^\epsilon (\alpha_1 + \Lambda_t)^{s/2} (|\alpha_2| + 3\Lambda_t)^{s/2}
\end{aligned}
$$

with high probability, uniformly on $\mathcal{D}$, since $\alpha_1 \leq |\alpha_2| \leq C$ and $\Lambda_t \prec 1$. In the third term on the right side of (5.1), note that $2s - s' - 2 \geq s$ since $s' \leq s - 2$. Hence, for $2 \leq s \leq 2D$,

$$
\begin{aligned}
N^\epsilon q_t^{-1} &\left(\frac{\operatorname{Im} m_t}{N\eta} \right)^{2s-s'-2} |P'(m_t)|^{s'} |P(m_t)|^{2D-s} \\
&\leq N^\epsilon q_t^{-1} \beta^s (\alpha_1 + \Lambda_t)^{2s-s'-2} (|\alpha_2| + 3\Lambda_t)^{s'} |P(m_t)|^{2D-s} \\
&\leq N^{2\epsilon} q_t^{-1} \beta^s (\alpha_1 + \Lambda_t)^{s/2} (|\alpha_2| + 3\Lambda_t)^{s/2} |P(m_t)|^{2D-s} \\
&\leq N^{2\epsilon} q_t^{-1} \frac{s}{2D} \beta^{2D} (\alpha_1 + \Lambda_t)^D (|\alpha_2| + 3\Lambda_t)^D + N^{2\epsilon} q_t^{-1} \frac{2D-s}{2D} |P(m_t)|^{2D}
\end{aligned}
$$
(5.9)

uniformly on $\mathcal{D}$ with high probability. For the last term on the right side of (5.1), we note that

$$
\frac{1}{N\eta} + q_t^{-1} \left(\frac{\operatorname{Im} m_t}{N\eta} \right)^{1/2} + q_t^{-2} \prec \beta ,
$$
(5.10)

uniformly on $\mathcal{D}$. Thus, similar to (5.9) we find that, for $2 \leq s \leq 2D$,

$$
N^\epsilon \left(\frac{1}{N\eta} + q_t^{-1} \left(\frac{\operatorname{Im} m_t}{N\eta} \right)^{1/2} + q_t^{-2} \right) \left(\frac{\operatorname{Im} m_t}{N\eta} \right)^{s-1} \left| P'(m_t) \right|^{s-1} |P(m_t)|^{2D-s}
$$

$$
\leq N^{2\epsilon} \beta \cdot \beta^{s-1} (\alpha_1 + \Lambda_t)^{s/2} (|\alpha_2| + 3\Lambda_t)^{s/2} |P(m_t)|^{2D-s}
$$

$$
\leq \frac{s}{2D} \left(N^{2\epsilon} N^{\frac{(2D-s)\epsilon}{4D^2}} \right)^{\frac{2D}{s}} \beta^{2D} (\alpha_1 + \Lambda_t)^D (|\alpha_2| + 3\Lambda_t)^D
$$

$$
+ \frac{2D-s}{2D} \left(N^{-\frac{(2D-s)\epsilon}{4D^2}} \right)^{\frac{2D}{2D-s}} |P(m_t)|^{2D}
$$

$$
\leq N^{(2D+1)\epsilon} \beta^{2D} (\alpha_1 + \Lambda_t)^D (|\alpha_2| + 3\Lambda_t)^D + N^{-\frac{\epsilon}{2D}} |P(m_t)|^{2D} , \qquad (5.11)
$$

for all $z \in \mathcal{D}$, with high probability. We hence have from (5.1), (5.6), (5.7), (5.9) and (5.11) that

$$
\mathbb{E}\left[|P(m_t)|^{2D} \right] \leq N^{(2D+1)\epsilon} \mathbb{E}\left[\beta^{2D} (\alpha_1 + \Lambda_t)^D (|\alpha_2| + 3\Lambda_t)^D \right] + \frac{N^{(2D+1)\epsilon}}{2D} q_t^{-8D}
$$

$$
+ \frac{N^{-(D/4-1)\epsilon}}{2D} q_t^{-2D} \mathbb{E}\left[\Lambda_t^{4D} \right] + C N^{-\frac{\epsilon}{2D}} \mathbb{E}\left[|P(m_t)|^{2D} \right] , \qquad (5.12)
$$

for all $z \in \mathcal{D}$. Note that the last term on the right side can be absorbed into the left side. Thence

$$
\mathbb{E}\left[|P(m_t)|^{2D} \right]
$$

$$
\leq C N^{(2D+1)\epsilon} \mathbb{E}\left[\beta^{2D} (\alpha_1 + \Lambda_t)^D (|\alpha_2| + 3\Lambda_t)^D \right] + \frac{N^{(2D+1)\epsilon}}{D} q_t^{-8D}
$$

$$
+ \frac{N^{-(D/4-1)\epsilon}}{D} q_t^{-2D} \mathbb{E}\left[\Lambda_t^{4D} \right]
$$

$$
\leq N^{3D\epsilon} \beta^{2D} |\alpha_2|^{2D} + N^{3D\epsilon} \beta^{2D} \mathbb{E}\left[\Lambda_t^{2D} \right] + N^{3D\epsilon} q_t^{-8D} + N^{-D\epsilon/8} q_t^{-2D} \mathbb{E}\left[\Lambda_t^{4D} \right] , \qquad (5.13)
$$

uniformly on $\mathcal{D}$, where we used that $D > 10$ and the inequality

$$
(a+b)^{\mathsf{p}} \leq 2^{\mathsf{p}-1} (a^{\mathsf{p}} + b^{\mathsf{p}}) , \qquad (5.14)
$$

for any $a, b \geq 0$ and $\mathsf{p} \geq 1$, to get the second line.

Next, from the third order Taylor expansion of $P(m_t)$ around $\widetilde{m}_t$, we have

$$
\left| P(m_t) - \alpha_2 (m_t - \widetilde{m}_t) - \frac{P''(\widetilde{m}_t)}{2} (m_t - \widetilde{m}_t)^2 \right| \leq C q_t^{-2} \Lambda_t^3 \qquad (5.15)
$$

since $P(\widetilde{m}_t) = 0$ and $P'''(\widetilde{m}_t) = 4! e^{-t} q_t^{-2} s^{(4)} \widetilde{m}_t$. Thus, using $\Lambda_t \prec 1$ and $P''(\widetilde{m}_t) = 2 + O(q_t^{-2})$ we get

$$\Lambda_t^2 \prec 2|\alpha_2|\Lambda_t + 2|P(m_t)|, \qquad (z \in \mathcal{D}). \qquad (5.16)$$

Taking the $2D$-power of the inequality and using once more (5.14), we get after taking the expectation

$$\mathbb{E}\left[\Lambda_t^{4D}\right] \leq 4^{2D} N^{\epsilon/2} |\alpha_2|^{2D} \mathbb{E}\left[\Lambda_t^{2D}\right] + 4^{2D} N^{\epsilon/2} \mathbb{E}\left[|P(m_t)|^{2D}\right], \qquad (z \in \mathcal{D}). \qquad (5.17)$$

Replacing from (5.13) for $\mathbb{E}[|P(m_t)|^{2D}]$ we obtain, using that $4^{2D} \leq N^{\epsilon/2}$, for N sufficiently large,

$$\begin{aligned} \mathbb{E}[\Lambda_t^{4D}] \leq\ & N^{\epsilon} |\alpha_2|^{2D} \mathbb{E}[\Lambda_t^{2D}] + N^{(3D+1)\epsilon} \beta^{2D} |\alpha_2|^{2D} \\ & + N^{(3D+1)\epsilon} \beta^{2D} \mathbb{E}\left[\Lambda_t^{2D}\right] + N^{(3D+1)\epsilon} q_t^{-8D} \\ & + N^{-D\epsilon/8+\epsilon} q_t^{-2D} \mathbb{E}\left[\Lambda_t^{4D}\right], \end{aligned} \qquad (5.18)$$

uniformly on $\mathcal{D}$. Applying the Schwarz inequality to the first term and the third term on the right, absorbing the terms $o(1)\mathbb{E}[\Lambda_t^{4D}]$ into the left side and using $q_t^{-2} \leq \beta$ in the fourth term, we get

$$\mathbb{E}[\Lambda_t^{4D}] \leq N^{2\epsilon} |\alpha_2|^{4D} + N^{(3D+2)\epsilon} \beta^{2D} |\alpha_2|^{2D} + N^{(3D+2)\epsilon} \beta^{4D}, \qquad (5.19)$$

uniformly on $\mathcal{D}$. Feeding (5.19) back into (5.13) we get, for any $D \geq 10$ and (small) $\epsilon > 0$,

$$\begin{aligned} \mathbb{E}\left[|P(m_t)|^{2D}\right] \leq\ & N^{3D\epsilon} \beta^{2D} |\alpha_2|^{2D} + N^{3D\epsilon} \beta^{2D} \mathbb{E}\left[\Lambda_t^{2D}\right] \\ & + N^{(3D+1)\epsilon} \beta^{4D} + q_t^{-2D} |\alpha_2|^{4D} \\ \leq\ & N^{5D\epsilon} \beta^{2D} |\alpha_2|^{2D} + N^{5D\epsilon} \beta^{4D} + q_t^{-2D} |\alpha_2|^{4D}, \end{aligned} \qquad (5.20)$$

uniformly on $\mathcal{D}$, for N sufficiently large, where we used the Schwarz inequality and once more (5.19) to get the second line.

By Markov's inequality, we therefore obtain from (5.20) that for fixed $z \in \mathcal{D}$, $|P(m_t)| \prec |\alpha_2|\beta + \beta^2 + q_t^{-1}|\alpha_2|^2$. It then follows from the Taylor expansion of $P(m_t)$ around $\widetilde{m}_t$ in (5.15) that

$$|\alpha_2(m_t - \widetilde{m}_t) + (m_t - \widetilde{m}_t)^2| \prec \beta\Lambda_t^2 + |\alpha_2|\beta + \beta^2 + q_t^{-1}|\alpha_2|^2, \qquad (5.21)$$

for each fixed $z \in \mathcal{D}$, where we used that $q_t^{-2} \leq \beta$. To get a uniform bound on $\mathcal{D}$, we choose $18N^8$ lattice points $z_1, z_2, \ldots, z_{18N^8}$ in $\mathcal{D}$ such that, for any $\widetilde{z} \in \mathcal{D}$, there exists z_n satisfying $|\widetilde{z} - z_n| \leq N^{-4}$. Since

$$|m_t(\widetilde{z}) - m_t(z_n)| \le |\widetilde{z} - z_n| \sup_{z \in \mathcal{D}} \left| \frac{\partial m_t(z)}{\partial z} \right| \le |\widetilde{z} - z_n| \sup_{z \in \mathcal{D}} \frac{1}{(\operatorname{Im} z)^2} \le N^{-2}$$

and since a similar estimate holds for $|\widetilde{m}_t(\widetilde{z}) - \widetilde{m}_t(z)|$, a union bound yields that (5.21) holds uniformly on $\mathcal{D}$ with high probability. In particular, for any (small) $\epsilon > 0$ and (large) D there is an event $\widetilde{\Xi}$ with $\mathbb{P}(\widetilde{\Xi}) \ge 1 - N^D$ such that, for all $z \in \mathcal{D}$,

$$|\alpha_2(m_t - \widetilde{m}_t) + (m - m_t)^2| \le N^\epsilon \beta \Lambda_t^2 + N^\epsilon |\alpha_2| \beta + N^\epsilon \beta^2 + N^\epsilon q_t^{-1} |\alpha_2|^2, \tag{5.22}$$

on $\widetilde{\Xi}$, for N sufficiently large.

Recall next that there is a constant $C_0 > 1$ such that $C_0^{-1} \sqrt{\varkappa_t(E) + \eta} \le |\alpha_2| \le C_0 \sqrt{\varkappa_t(E) + \eta}$, where we can choose C_0 uniform in $z \in \mathcal{E}$. Note further that $\beta = \beta(E + i\eta)$ is for fixed E a decreasing function of η while $\sqrt{\kappa_t(E) + \eta}$ is increasing. Thus, there exists $\widetilde{\eta}_0 \equiv \widetilde{\eta}_0(E)$ such that $\sqrt{\varkappa(E) + \widetilde{\eta}_0} = C_0 q_t \beta(E + i\widetilde{\eta}_0)$. We then consider the subdomain $\widetilde{\mathcal{D}} \subset \mathcal{D}$ defined by

$$\widetilde{\mathcal{D}} := \{ z = E + i\eta \in \mathcal{D} : \eta > \widetilde{\eta}_0(E) \}. \tag{5.23}$$

On $\widetilde{\mathcal{D}}$, $\beta \le q_t^{-1} |\alpha_2|$, hence we obtain from the estimate (5.22) that

$$|\alpha_2(m_t - \widetilde{m}_t) + (m - m_t)^2| \le N^\epsilon \beta \Lambda_t^2 + 3N^\epsilon q_t^{-1} |\alpha_2|^2$$

and thus

$$|\alpha_2| \Lambda_t \le (1 + N^\epsilon \beta) \Lambda_t^2 + 3N^\epsilon q_t^{-1} |\alpha_2|^2,$$

uniformly on $\widetilde{\mathcal{D}}$ on $\widetilde{\Xi}$. Hence, we get on $\widetilde{\Xi}$ that either

$$|\alpha_2| \le 4\Lambda_t \qquad \text{or} \qquad \Lambda_t \le 6N^\epsilon q_t^{-1} |\alpha_2|, \qquad (z \in \widetilde{\mathcal{D}}). \tag{5.24}$$

Note that any $z \in \mathcal{E}$ with $\eta = \operatorname{Im} z = 3$ is in $\widetilde{\mathcal{D}}$. When $\eta = 3$, we easily see that

$$|\alpha_2| \ge |z + 2\widetilde{m}_t| - Cq_t^{-2} \ge \eta = 3 \ggg 6N^\epsilon q_t^{-1} |\alpha_2|,$$

for sufficiently large N. In particular we have that either $3/4 \le \Lambda_t$ or $\Lambda_t \le 6N^\epsilon q_t^{-1} |\alpha_2|$ on $\widetilde{\Xi}$ for $\eta = 3$. Moreover, since m_t and $\widetilde{m}_t$ are Stieltjes transforms, we have

$$\Lambda_t \le \frac{2}{\eta} = \frac{2}{3}.$$

We conclude that, for $\eta = 3$, the second possibility, $\Lambda_t \le 6N^\epsilon q_t^{-1} |\alpha_2|$ holds on $\widetilde{\Xi}$. Since $6N^\epsilon q_t^{-1} \ll 1$ on $\widetilde{\mathcal{D}}$, in particular $6N^\epsilon q_t^{-1} |\alpha_2| < |\alpha_2|/8$, we find from (5.24) by continuity that

$$\Lambda_t \le 6N^\epsilon q_t^{-1}|\alpha_2|, \qquad (z \in \widetilde{\mathcal{D}}), \qquad (5.25)$$

holds on the event $\widetilde{\Xi}$. Putting the estimate (5.25) back into (5.13), we find that

$$\mathbb{E}\left[|P(m_t)|^{2D}\right] \le N^{4D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{3D\epsilon}q_t^{-8D} + q_t^{-6D}|\alpha_2|^{4D}$$
$$\le N^{6D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{6D\epsilon}\beta^{4D}, \qquad (5.26)$$

for any (small) $\epsilon > 0$ and (large) D, uniformly on $\widetilde{\mathcal{D}}$. Note that, for $z \in \mathcal{D}\backslash\widetilde{\mathcal{D}}$, the estimate $\mathbb{E}[|P(m_t)|^{2D}] \le N^{6D\epsilon}\beta^{2D}|\alpha_2|^{2D} + N^{6D\epsilon}\beta^{4D}$ can be directly checked from (5.20). Considering lattice points $\{z_i\} \subset \mathcal{D}$ again, a union bound yields for any (small) $\epsilon > 0$ and (large) D there is an event Ξ with $\mathbb{P}(\Xi) \ge 1 - N^D$ such that

$$|\alpha_2(m_t - \widetilde{m}_t) + (m - m_t)^2| \le N^\epsilon\beta\Lambda_t^2 + N^\epsilon|\alpha_2|\beta + N^\epsilon\beta^2, \qquad (5.27)$$

on Ξ, uniformly on $\mathcal{D}$ for N sufficiently large.

Next, recall that $\beta = \beta(E + i\eta)$ is for fixed E a decreasing function of η while $\sqrt{\varkappa_t(E) + \eta}$ is increasing. Thus there is $\eta_0 \equiv \eta_0(E)$ such that $\sqrt{\varkappa(E) + \eta_0} = 10C_0N^\epsilon\beta(E + i\eta_0)$. Further notice that $\eta_0(E)$ is a continuous function. We consider the three subdomains of $\mathcal{E}$ defined by

$$\mathcal{E}_1 := \left\{z = E + i\eta \in \mathcal{E} : \eta \le \eta_0(E), 10N^\epsilon \le N\eta\right\},$$
$$\mathcal{E}_2 := \left\{z = E + i\eta \in \mathcal{E} : \eta > \eta_0(E), 10N^\epsilon \le N\eta\right\},$$
$$\mathcal{E}_3 := \left\{z = E + i\eta \in \mathcal{E} : 10N^\epsilon > N\eta\right\}.$$

Note that $\mathcal{E}_1 \cup \mathcal{E}_2 \subset \mathcal{D}$. We split the stability analysis of (5.27) according to whether $z \in \mathcal{E}_1, \mathcal{E}_2$ or $\mathcal{E}_3$. *Case 1* If $z \in \mathcal{E}_1$, we note that $|\alpha_2| \le C_0\sqrt{\varkappa(E) + \eta} \le 10C_0^2N^\epsilon\beta(E + i\eta)$. We then obtain from (5.27) that

$$\Lambda_t^2 \le |\alpha_2|\Lambda_t + N^\epsilon\beta\Lambda_t^2 + N^\epsilon|\alpha_2|\beta + N^\epsilon\beta^2$$
$$\le 10C_0^2N^\epsilon\beta\Lambda_t + N^\epsilon\beta\Lambda_t^2 + (10C_0^2N^\epsilon + 1)N^\epsilon\beta^2,$$

on Ξ. Thus,

$$\Lambda_t \le CN^\epsilon\beta, \qquad (z \in \mathcal{E}_1), \qquad (5.28)$$

on Ξ, for some finite constant C.

Case 2 If $z \in \mathcal{E}_2$, we obtain from (5.27) that

$$|\alpha_2|\Lambda_t \le (1 + N^\epsilon\beta)\Lambda_t^2 + |\alpha_2|N^\epsilon\beta + N^\epsilon\beta^2, \qquad (5.29)$$

on Ξ. We then note that $C_0|\alpha_2| \ge \sqrt{\varkappa_t(E) + \eta} \ge 10C_0N^\epsilon\beta$, i.e. $N^\epsilon\beta \le |\alpha_2|/10$, so that

$$|\alpha_2|\Lambda_t \le 2\Lambda_t^2 + (1 + N^{-\epsilon})|\alpha_2|N^\epsilon\beta, \qquad (5.30)$$

on Ξ, where we used that $N^\epsilon \beta \leq 1$. Hence, we get on Ξ that either

$$|\alpha_2| \leq 4\Lambda_t \quad \text{or} \quad \Lambda_t \leq 3N^\epsilon \beta, \quad (z \in \mathcal{E}_2). \tag{5.31}$$

We follow the dichotomy argument and the continuity argument that were used to obtain (5.25). Since $3N^\epsilon \beta \leq |\alpha_2|/8$ on $\mathcal{E}_2$, we find by continuity that

$$\Lambda_t \leq 3N^\epsilon \beta, \quad (z \in \mathcal{E}_2), \tag{5.32}$$

holds on the event Ξ. *Case 3* For $z \in \mathcal{E}_3 = \mathcal{E} \backslash (\mathcal{E}_1 \cup \mathcal{E}_2)$ we use that $|m_t'(z)| \leq \text{Im}\, m_t(z)/\text{Im}\, z, z \in \mathbb{C}^+$, since m_t is a Stieltjes transform. Set now $\widetilde{\eta} := 10N^{-1+\epsilon}$. By the fundamental theorem of calculus we can estimate

$$|m_t(E + i\eta)| \leq \int_\eta^{\widetilde{\eta}} \frac{\text{Im}\, m_t(E + is)}{s} ds + |m_t(E + i\widetilde{\eta})|$$

$$\leq \int_\eta^{\widetilde{\eta}} \frac{s\, \text{Im}\, m_t(E + is)}{s^2} ds + \Lambda_t(E + i\widetilde{\eta}) + |\widetilde{m}_t(E + i\widetilde{\eta})|.$$

Using that $s \to s\, \text{Im}\, m_t(E+is)$ is a monotone increasing function as is easily checked from the definition of the Stieltjes transform, we find that

$$
\begin{aligned}
|m_t(E + i\eta)| &\leq \frac{2\widetilde{\eta}}{\eta} \text{Im}\, m_t(E + i\widetilde{\eta}) + \Lambda_t(E + i\widetilde{\eta}) + |\widetilde{m}_t(E + i\widetilde{\eta})| \\
&\leq C \frac{N^\epsilon}{N\eta} \big(\text{Im}\, \widetilde{m}_t(E + i\widetilde{\eta}) + \Lambda_t(E + i\widetilde{\eta}) \big) + |\widetilde{m}_t(E + i\widetilde{\eta})|,
\end{aligned}
\tag{5.33}
$$

for some C where we used $\widetilde{\eta} := 10N^{-1+\epsilon}$ to get the second line. Noticing that $z = E + i\widetilde{\eta} \in \mathcal{E}_1 \cup \mathcal{E}_2$, we have on the event Ξ that $\Lambda_t(E + i\widetilde{\eta}) \leq CN^\epsilon \beta(E + i\widetilde{\eta}) \leq C$ as follows from (5.28) and (5.32). Using moreover that $\widetilde{m}_t$ is uniformly bounded by a constant on $\mathcal{E}$, we then get that, on the event Ξ,

$$\Lambda_t \leq CN^\epsilon \beta, \quad (z \in \mathcal{E}_3). \tag{5.34}$$

Combining (5.28), (5.32) and (5.34), and recalling the definition of the event Ξ, we get $\Lambda_t \prec \beta$, uniformly on $\mathcal{E}$ for fixed $t \in [0, 6 \log N]$. Choosing $t = 0$, we have completed the proof of Theorem 2.4. To extend this bound to all $t \in [0, 6 \log N]$, we use the continuity of the Dyson matrix flow. Choosing a lattice $\mathcal{L} \subset [0, 6 \log N]$ with spacings of order N^{-3}, we get $\Lambda_t \prec \beta$, uniformly on $\mathcal{E}$ and on $\mathcal{L}$, by a union bound. By continuity we extend the conclusion to all of $[0, 6 \log N]$. This proves Proposition 3.3.

5.2 Proof of Theorem 2.9

We start with an upper bound on the largest eigenvalue $\lambda_1^{H_t}$ of H_t.

Lemma 5.2 *Let H_0 satisfy Assumption 2.3 with $\phi > 0$. Let L_t be deterministic number defined in Lemma 4.1. Then,*

$$\lambda_1^{H_t} - L_t \prec \frac{1}{q_t^4} + \frac{1}{N^{2/3}}, \tag{5.35}$$

uniformly in $t \in [0, 6 \log N]$.

Proof To prove Lemma 5.2 we follow the strategy of the proof of Lemma 4.4 in [15]. Fix $t \in [0, 6 \log N]$. Recall first from (5.3) the deterministic z-dependent parameters

$$\alpha_1 := \operatorname{Im} \widetilde{m}_t, \qquad \alpha_2 := P'(\widetilde{m}_t), \qquad \beta := \frac{1}{q_t^2} + \frac{1}{N\eta}. \tag{5.36}$$

We mostly drop the z-dependence for brevity. We further introduce the z-independent quantity

$$\widetilde{\beta} := \left(\frac{1}{q_t^4} + \frac{1}{N^{2/3}} \right)^{1/2}. \tag{5.37}$$

Fix a small $\epsilon > 0$ and define the domain $\mathcal{D}_\epsilon$ by

$$\mathcal{D}_\epsilon := \left\{ z = E + i\eta : N^{4\epsilon} \widetilde{\beta}^2 \le \varkappa_t \le q_t^{-1/3}, \ \eta = \frac{N^\epsilon}{N\sqrt{\varkappa_t}} \right\}, \tag{5.38}$$

where $\varkappa_t \equiv \varkappa_t(E) = E - L_t$. Note that on $\mathcal{D}_\epsilon$,

$$N^{-1+\epsilon} \lll \eta \le \frac{N^{-\epsilon}}{N\widetilde{\beta}}, \qquad \varkappa \ge N^{5\epsilon} \eta.$$

In particular we have $N^\epsilon \widetilde{\beta} \le (N\eta)^{-1}$, hence $N^\epsilon q_t^{-2} \le C(N\eta)^{-1}$ so that q_t^{-2} is negligible when compared to $(N\eta)^{-1}$ and β on $\mathcal{D}_\epsilon$. Note moreover that

$$|\alpha_2| \asymp \sqrt{\varkappa_t + \eta} \asymp \sqrt{\varkappa_t} = \frac{N^\epsilon}{N\eta} \asymp N^\epsilon \beta,$$

$$\alpha_1 = \operatorname{Im} \widetilde{m}_t \asymp \frac{\eta}{\sqrt{\varkappa_t + \eta}} \asymp \frac{\eta}{\sqrt{\varkappa_t}} \le N^{-5\epsilon} \sqrt{\varkappa_t} \asymp N^{-5\epsilon} |\alpha_2| \asymp N^{-4\epsilon} \beta. \tag{5.39}$$

In particular we have $\alpha_1 \lll |\alpha_2|$ on $\mathcal{D}_\epsilon$.

We next claim that

$$\Lambda_t := |m_t - \widetilde{m}_t| \lll \frac{1}{N\eta}$$

with high probability on the domain $\mathcal{D}_\epsilon$.

Since $\mathcal{D}_\epsilon \subset \mathcal{E}$, we find from Proposition 3.3 that $\Lambda_t \leq N^{\epsilon'}\beta$ for any $\epsilon' > 0$ with high probability. Fix $0 < \epsilon' < \epsilon/7$. From (5.13), we get

$$
\mathbb{E}\left[|P(m_t)|^{2D}\right] \leq CN^{(4D-1)\epsilon'}\mathbb{E}\left[\beta^{2D}(\alpha_1 + \Lambda_t)^D(|\alpha_2| + 3\Lambda_t)^D\right]
$$
$$
+ \frac{N^{(2D+1)\epsilon'}}{D}q_t^{-8D} + \frac{N^{-(D/4-1)\epsilon'}}{D}q_t^{-2D}\mathbb{E}\left[\Lambda_t^{4D}\right]
$$
$$
\leq C^{2D}N^{6D\epsilon'}\beta^{4D} + \frac{N^{(2D+1)\epsilon'}}{D}q_t^{-8D} + \frac{N^{4D\epsilon'}}{D}q_t^{-2D}\beta^{4D}
$$
$$
\leq C^{2D}N^{6D\epsilon'}\beta^{4D},
$$

for N sufficiently large, where we used that $\Lambda_t \leq N^{\epsilon'}\beta \lll N^\epsilon\beta$ with high probability and, by (5.39), $\alpha_1 \lll |\alpha_2|$, $|\alpha_2| \leq CN^\epsilon\beta$ on $\mathcal{D}_\epsilon$. Applying $(2D)$-th order Markov inequality and a simple lattice argument combined with a union bound, we get $|P(m_t)| \leq CN^{3\epsilon'}\beta^2$ uniformly on $\mathcal{D}_\epsilon$ with high probability. From the Taylor expansion of $P(m_t)$ around $\widetilde{m}_t$ in (5.15), we then get that

$$
|\alpha_2|\Lambda_t \leq 2\Lambda_t^2 + CN^{3\epsilon'}\beta^2, \tag{5.40}
$$

uniformly on $\mathcal{D}_\epsilon$ with high probability, where we also used that $\Lambda_t \lll 1$ on $\mathcal{D}_\epsilon$ with high probability.

Since $\Lambda_t \leq N^{\epsilon'}\beta \leq CN^{\epsilon'-\epsilon}|\alpha_2|$ with high probability on $\mathcal{D}_\epsilon$, we have $|\alpha_2|\Lambda_t \geq CN^{\epsilon-\epsilon'}\Lambda_t^2 \ggg 2\Lambda_t^2$. Thus the first term on the right side of (5.40) can be absorbed into the left side and we conclude that

$$
\Lambda_t \leq CN^{3\epsilon'}\frac{\beta}{|\alpha_2|}\beta \leq CN^{3\epsilon'-\epsilon}\beta,
$$

must hold with high probability on $\mathcal{D}_\epsilon$. Hence, using that $0 < \epsilon' < \epsilon/7$, we obtain that

$$
\Lambda_t \leq N^{-\epsilon/2}\beta \leq 2\frac{N^{-\epsilon/2}}{N\eta},
$$

with high probability on $\mathcal{D}_\epsilon$. This proves the claim that $\Lambda_t \lll (N\eta)^{-1}$ on $\mathcal{D}_\epsilon$ with high probability. Moreover, this also shows that

$$
\operatorname{Im} m_t \leq \operatorname{Im}\widetilde{m}_t + \Lambda_t = \alpha_1 + \Lambda_t \lll \frac{1}{N\eta}, \tag{5.41}
$$

on $\mathcal{D}_\epsilon$ with high probability, where we used (5.39).

Now we prove the estimate (5.35). If $\lambda_1^{H_t} \in [E - \eta, E + \eta]$ for some $E \in [L_t + N^\epsilon(q_t^{-4} + N^{-2/3}), L_t + q^{-1/3}]$ with $z = E + \mathrm{i}\eta \in \mathcal{D}_\epsilon$,

$$
\operatorname{Im} m_t(z) \geq \frac{1}{N}\operatorname{Im}\frac{1}{\lambda_1^{H_t} - E - \mathrm{i}\eta} = \frac{1}{N}\frac{\eta}{(\lambda_1^{H_t} - E)^2 + \eta^2} \geq \frac{1}{5N\eta}, \tag{5.42}
$$

which contradicts the high probability bound $\operatorname{Im} m_t \lll (N\eta)^{-1}$ in (5.41). The size of each interval $[E - \eta, E + \eta]$ is at least $N^{-1+\epsilon} q_t^{1/6}$. Thus, considering $O(N)$ such intervals, we can conclude that $\lambda_1 \notin [L + N^\epsilon(q_t^{-4} + N^{-2/3}), L + q_t^{-1/3}]$ with high probability. From Proposition 2.8, we find that $\lambda_1^{H_t} - L_t \prec q_t^{-1/3}$ with high probability, hence we conclude that (5.35) holds, for fixed $t \in [0, 6\log N]$. Using a lattice argument and the continuity of the Dyson matrix flow, we easily obtain (5.35) uniformly in $t \in [0, 6\log N]$. □

We are now well-prepared for the proof of Theorem 2.9. It follows immediately from the next result.

Lemma 5.3 *Let H_0 satisfy Assumption (2.3) with $\phi > 0$. Then, uniformly in $t \in [0, 6\log N]$,*

$$\left| \|H_t\| - L_t \right| \prec \frac{1}{q_t^4} + \frac{1}{N^{2/3}} \, . \tag{5.43}$$

Proof of Lemma 5.3 and Theorem 2.9 Fix $t \in [0, \log N]$. Consider the largest eigenvalue $\lambda_1^{H_t}$. In Proposition 5.2, we already showed that $(L_t - \lambda_1^{H_t})_- \prec q_t^{-4} + N^{-2/3}$. It thus suffices to consider $(L_t - \lambda_1^{H_t})_+$. By Lemma 4.1 there is $c > 0$ such that $c(L_t - \lambda_1^{H_t})_+^{3/2} \le n_{\widetilde{\rho}_t}(\lambda_1^{H_t}, L_t)$, with $n_{\widetilde{\rho}_t}$ as in (2.21). Hence, by Corollary 2.5 (and its obvious generalization to H_t), we have the estimate

$$\left(L_t - \lambda_1^{H_t}\right)_+^{3/2} \prec \frac{\left(L_t - \lambda_1^{H_t}\right)_+}{q_t^2} + \frac{1}{N} \, , \tag{5.44}$$

so that $(L_t - \lambda_1^{H_t})_+ \prec q_t^{-4} + N^{-2/3}$. Thus $|\lambda_1^{H_t} - L_t| \prec q_t^{-4} + N^{-2/3}$. Similarly, one shows the estimate $|\lambda_N^{H_t} + L_t| \prec q_t^{-4} + N^{-2/3}$ for the smallest eigenvalue $\lambda_N^{H_t}$. This proves (5.43) for fixed $t \in [0, 6\log N]$. Uniformity follows easily from the continuity of the Dyson matrix flow. □

6 Recursive moment estimate: Proof of Lemma 5.1

In this section, we prove Lemma 5.1. Recall the definitions of the Green functions G_t and m_t in (3.29). We fix $t \in [0, 6\log N]$ throughout this section, and we will omit t from the notation in the matrix H_t, its matrix elements and its Green functions. Given a (small) $\epsilon > 0$, we introduce the z-dependent control parameter $\Phi_\epsilon \equiv \Phi_\epsilon(z)$ by setting

$$\Phi_\epsilon(z) := N^\epsilon \, \mathbb{E}\left[\left(\frac{1}{q_t^4} + \frac{\operatorname{Im} m_t}{N\eta}\right)|P(m_t)|^{2D-1}\right]$$
$$+ N^{-\epsilon/4} q_t^{-1} \, \mathbb{E}\left[|m_t - \widetilde{m}_t|^2 |P(m_t)|^{2D-1}\right]$$

$$+ N^\epsilon q_t^{-1} \sum_{s=2}^{2D} \sum_{s'=0}^{s-2} \mathbb{E}\left[\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{2s-s'-2} |P'(m_t)|^{s'} |P(m_t)|^{2D-s}\right]$$

$$+ N^\epsilon q_t^{-8D}$$

$$+ N^\epsilon \sum_{s=2}^{2D} \mathbb{E}\left[\left(\frac{1}{N\eta} + \frac{1}{q_t}\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{1/2} + \frac{1}{q_t^2}\right)\left(\frac{\operatorname{Im} m_t}{N\eta}\right)^{s-1}\right.$$

$$\left. \times |P'(m_t)|^{s-1} |P(m_t)|^{2D-s}\right]. \tag{6.1}$$

Recall the domain $\mathcal{D}$ defined in (3.2). Lemma 5.1 states that, for any (small) $\epsilon > 0$,

$$\mathbb{E}[|P|^{2D}(z)] \leq \Phi_\epsilon(z), \qquad (z \in \mathcal{D}), \tag{6.2}$$

for N sufficiently large. We say that a *random variable Z is negligible* if $|\mathbb{E}[Z]| \leq C\Phi_\epsilon$ for some N-independent constant C.

To prove the recursive moment estimate of Lemma 5.1, we return to (3.25) which reads

$$\mathbb{E}[(1 + zm)P^{D-1}\overline{P^D}] = \frac{1}{N} \sum_{r=1}^{\ell} \frac{\kappa_t^{(r+1)}}{r!} \mathbb{E}\left[\sum_{i \neq k} \partial_{ik}^r \left(G_{ki} P^{D-1}\overline{P^D}\right)\right]$$
$$+ \mathbb{E}\Omega_\ell\left((1 + zm)P^{D-1}\overline{P^D}\right), \tag{6.3}$$

where $\partial_{ik} = \partial/(\partial H_{ik})$ and $\kappa_t^{(k)}$ are the cumulants of $(H_t)_{ij}$, $i \neq j$. The detailed form of the error $\mathbb{E}\Omega_\ell(\cdot)$ is discussed in Sect. 6.1.

It is convenient to condense the notation a bit. Abbreviate

$$I \equiv I(z, m, D) := (1 + zm)P(m)^{D-1}\overline{P(m)^D}. \tag{6.4}$$

We rewrite the cumulant expansion (6.3) as

$$\mathbb{E}I = \sum_{r=1}^{\ell} \sum_{s=0}^{r} w_{I_{r,s}} \mathbb{E}I_{r,s} + \mathbb{E}\Omega_\ell(I), \tag{6.5}$$

where we set

$$I_{r,s} := N\kappa_t^{(r+1)} \frac{1}{N^2} \sum_{i \neq k} \left(\partial_{ik}^{r-s} G_{ki}\right)\left(\partial_{ik}^s (P^{D-1}\overline{P^D})\right). \tag{6.6}$$

(By convention $\partial_{ik}^0 G_{ki} = G_{ki}$.) The weights $w_{I_{r,s}}$ are combinatoric coefficient given by

$$w_{I_{r,s}} := \frac{1}{r!}\binom{r}{s} = \frac{1}{(r-s)!s!}. \tag{6.7}$$

Returning to (3.22), we have in this condensed form the expansion

$$\mathbb{E}[|P|^{2D}] = \sum_{r=1}^{\ell} \sum_{s=0}^{r} w_{I_{r,s}} \mathbb{E} I_{r,s} + \mathbb{E}\left[\left(m^2 + \frac{s^{(4)}}{q_t^2} m^4\right) P^{D-1}\overline{P^D}\right] + \mathbb{E}\Omega_\ell(I).$$

(6.8)

6.1 Truncation of the cumulant expansion

In this subsection, we bound the error term $\mathbb{E}\Omega_\ell(I)$ in (6.8) for large ℓ. We need some more notation. Let $E^{[ik]}$ denote the $N \times N$ matrix determined by

$$(E^{[ik]})_{ab} = \begin{cases} \delta_{ia}\delta_{kb} + \delta_{ib}\delta_{ka} & \text{if } i \neq k, \\ \delta_{ia}\delta_{ib} & \text{if } i = k, \end{cases} \qquad (i, k, a, b \in [\![1, N]\!]).$$

(6.9)

For each pair of indices (i, k), we define the matrix $H^{(ik)}$ from H through the decomposition

$$H = H^{(ik)} + H_{ik} E^{[ik]}.$$

(6.10)

With this notation we have the following estimate.

Lemma 6.1 *Suppose that H satisfies Assumption 2.3 with $\phi > 0$. Let $i, k \in [\![1, N]\!]$, $D \in \mathbb{N}$ and $z \in \mathcal{D}$. Define the function F_{ki} by*

$$F_{ki}(H) := G_{ki} P^{D-1}\overline{P}^D,$$

(6.11)

where $G \equiv G^H(z)$ and $P \equiv P(m^H(z))$. Choose an arbitrary $\ell \in \mathbb{N}$. Then, for any (small) $\epsilon > 0$,

$$\mathbb{E}\left[\sup_{x \in \mathbb{R},\, |x| \leq q_t^{-1/2}} |\partial_{ik}^\ell F_{ki}(H^{(ik)} + x E^{[ik]})|\right] \leq N^\epsilon,$$

(6.12)

uniformly $z \in \mathcal{D}$, for N sufficiently large. Here ∂_{ik}^ℓ denotes the partial derivative $\frac{\partial^\ell}{\partial H_{ik}^\ell}$.

Proof Fix two pairs of indices (a, b) and (i, k). From the definition of the Green function and (6.10) we easily get

$$G_{ab}^{H^{(ik)}} = G_{ab}^H + H_{ik}\left(G^{H^{(ik)}} E^{[ik]} G^H\right)_{ab} = G_{ab}^H + H_{ik} G_{ai}^{H^{(ik)}} G_{kb}^H + H_{ik} G_{ak}^{H^{(ik)}} G_{ib}^H,$$

where we omit the z-dependence. Letting $\Lambda_o^{H^{(ik)}} := \max_{a,b}|G_{ab}^{H^{(ik)}}|$ and $\Lambda_o^H := \max_{a,b}|G_{ab}^H|$, we get

$$\Lambda_o^{H^{(ik)}} \prec \Lambda_o^H + \frac{1}{q_t} \Lambda_o^H \Lambda_o^{H^{(ik)}},$$

since $|H_{ik}| \prec q_t^{-1}$ by (2.12). Moreover by (2.24) we have $\Lambda_o^H \prec 1$, uniformly in $z \in \mathcal{D}$. It thus follows that $\Lambda_o^{H^{(ik)}} \prec \Lambda_o^H \prec 1$, uniformly in $z \in \mathcal{D}$, where we used (2.24). Similarly, for $x \in \mathbb{R}$, we have

$$G_{ab}^{H^{(ik)}+xE^{[ik]}} = G_{ab}^{H^{(ik)}} - x\big(G^{H^{(ik)}} E^{[ik]} G^{H^{(ik)}+xE^{[ik]}}\big)_{ab},$$

and we get

$$\sup_{|x| \leq q_t^{-1/2}} \max_{a,b} |G_{ab}^{H^{(ik)}+xE^{[ik]}}| \prec \Lambda_o^{H^{(ik)}} \prec 1, \tag{6.13}$$

uniformly in $z \in \mathcal{D}$, where we used once more (2.24).

Recall that P is a polynomial of degree 4 in m. Then F_{ki} is a multivariate polynomial of degree $4(2D-1)+1$ in the Green function entries and the normalized trace m whose number of member terms is bounded by 4^{2D-1}. Hence $\partial_{ik}^\ell F_{ki}$ is a multivariate polynomial of degree $4(2D-1)+1+\ell$ whose number of member terms is roughly bounded by $4^{2D-1} \times (4(2D-1)+1+2l)^l$. Next, to control the individual monomials in $\partial_{ik}^\ell F_{ki}$, we apply (6.13) to each factor of Green function entries (at most $4(2D-1)+1+\ell$ times). Thus, altogether we obtain

$$\mathbb{E}\left[\sup_{|x| \leq q_t^{-1/2}} |(\partial_{ik}^\ell F_{ki})(H^{(ik)} + xE^{[ik]})|\right] \leq 4^{2D}(8D+\ell)N^{(8D+\ell)\epsilon'}, \tag{6.14}$$

for any small $\epsilon' > 0$ and sufficiently large N. Choosing $\epsilon' = \epsilon/(2(8D+\ell))$ with get (6.12). $\square$

Recall that we set $I = (1+zm)P(m)^{D-1}\overline{P(m)^D}$ in (6.4). To control the error term $\mathbb{E}\Omega_\ell(I)$ in (6.8), we use the following result.

Corollary 6.2 *Let $\mathbb{E}\Omega_\ell(I)$ be as in (6.8). With the assumptions and notation of Lemma 6.1, we have, for any (small) $\epsilon > 0$,*

$$|\mathbb{E}\Omega_\ell(I)| \leq N^\epsilon \left(\frac{1}{q_t}\right)^\ell, \tag{6.15}$$

uniformly in $z \in \mathcal{D}$, for N sufficiently large. In particular, the error $\mathbb{E}\Omega_\ell(I)$ is negligible for $\ell \geq 8D$.

Proof First, fix a pair of indices (k, i), $k \neq i$. Recall the definition of F_{ik} in (6.11). Denoting $\mathbb{E}_{ik}$ the partial expectation with respect to H_{ik}, we have from Lemma 3.2, with $Q = q_t^{-1/2}$, that

$$\left|\mathbb{E}_{ik}\Omega_\ell(H_{ik}F_{ki})\right| \leq C_\ell \mathbb{E}_{ik}\left[|H_{ik}|^{\ell+2}\right]\sup_{|x|\leq q_t^{-1/2}}\left|\partial_{ik}^{\ell+1}F_{ki}\left(H^{(ik)}+xE^{[ik]}\right)\right|$$
$$+ C_\ell \mathbb{E}_{ik}\left[|H_{ik}|^{\ell+2}\mathbb{1}\left(|H_{ik}|>q_t^{-1/2}\right)\right]$$
$$\times \sup_{x\in\mathbb{R}}\left|\partial_{ik}^{\ell+1}F_{ki}\left(H^{(ik)}+xE^{[ik]}\right)\right|, \qquad (6.16)$$

with $C_\ell \leq (C\ell)^\ell/\ell!$, for some numeral constant C. To control the full expectation of the first term on the right side, we use the moment assumption (2.12) and Lemma 6.1 to conclude that, for any $\epsilon > 0$,

$$C_\ell \mathbb{E}\left[\mathbb{E}_{ik}\left[|H_{ki}|^{\ell+2}\right]\sup_{|x|\leq q_t^{-1/2}}\left|\partial_{ik}^{\ell+1}F_{ki}\left(H^{(ik)}+xE^{[ik]}\right)\right|\right] \leq C_\ell \frac{(C(\ell+2))^{c(\ell+2)}}{Nq_t^\ell}N^\epsilon$$
$$\leq \frac{N^{2\epsilon}}{Nq_t^\ell},$$

for N sufficiently large. To control the second term on the right side of (6.16), we use the operator norm of $G(z)$ is deterministically bounded by η^{-1}, *i.e.* $\|G(z)\| \leq \eta^{-1}$, to conclude that

$$\sup_{x\in\mathbb{R}}\left|\partial_{ik}^{\ell+1}F_{ki}\left(H^{(ik)}+xE^{[ik]}\right)\right| \leq 4^{2D}(8D+\ell)\left(\frac{C}{\eta}\right)^{(8D+\ell)}, \quad (z\in\mathbb{C}^+);$$

cf. the paragraph above (6.14). On the other hand, we have from Hölder's inequality and the moment assumptions in (2.12) that, for any $D'\in\mathbb{N}$,

$$\mathbb{E}_{ik}\left[|H_{ik}|^{\ell+2}\mathbb{1}\left(|H_{ik}|>q_t^{-1/2}\right)\right] \leq \left(\frac{C}{q_t}\right)^{D'},$$

for N sufficiently large. Using that $q_t \geq N^\phi$ by (2.3), we hence obtain, for any $D'\in\mathbb{N}$,

$$C_\ell \mathbb{E}_{ik}\left[|H_{ik}|^{\ell+2}\mathbb{1}\left(|H_{ik}|>q_t^{-1/2}\right)\right]\sup_{x\in\mathbb{R}}\left|\partial_{ik}^{\ell+1}F_{ki}\left(H^{(ik)}+xE^{[ik]}\right)\right| \leq \left(\frac{C}{q_t}\right)^{D'},$$
$$(6.17)$$

uniformly on $\mathbb{C}^+$, for N sufficiently large.

Next, summing over i, k and choosing $D' \geq \ell$ sufficiently large in (6.17) we obtain, for any $\epsilon > 0$

$$\left|\mathbb{E}\left[\Omega_\ell\left((1+zm)P^{D-1}\overline{P^D}\right)\right]\right| = \left|\mathbb{E}\left[\Omega_\ell\left(\frac{1}{N}\sum_{i\neq k}H_{ik}F_{ki}\right)\right]\right| \leq \frac{N^\epsilon}{q_t^\ell},$$

uniformly on $\mathcal{D}$, for N sufficiently large. This proves (6.15). □

Remark 6.3 We will also consider slight generalizations of the cumulant expansion in (6.3). Let i, j, $k \in [\![1, N]\!]$. Let $n \in \mathbb{N}_0$ and choose indices $a_1, \ldots, a_n, b_1, \ldots, b_n \in [\![1, N]\!]$. Let $D \in \mathbb{N}$ and choose $s_1, s_2, s_3, s_4 \in [\![0, D]\!]$. Fix $z \in \mathcal{D}$. Define the function F_{ki} by setting

$$F_{ki}(H) := G_{ki} \prod_{l=1}^{n} G_{a_l b_l} P^{D-s_1} \overline{P^{D-s_2}} \left(P'\right)^{s_3} \left(\overline{P'}\right)^{s_4}. \qquad (6.18)$$

It is then straightforward to check that we have the cumulant expansion

$$\mathbb{E}\left[\frac{1}{N}\sum_{i \neq k} H_{ik} F_{ki}\right] = \sum_{r=1}^{\ell} \frac{\kappa_t^{(r+1)}}{r!} \mathbb{E}\left[\frac{1}{N}\sum_{i \neq k} \partial_{ik}^r F_{ki}\right] + \mathbb{E}\Omega_\ell\left(\frac{1}{N}\sum_{i \neq k} H_{ik} F_{ki}\right),$$
$$(6.19)$$

where the error $\mathbb{E}\Omega_\ell(\cdot)$ satisfies the same bound as in (6.15). This follows easily by extending Lemma 6.1 and Corollary 6.2.

6.2 Truncated cumulant expansion

Armed with the estimates on $\mathbb{E}\Omega_\ell(\cdot)$ of the previous subsection, we now turn to the main terms on the right side of (6.8). In the remainder of this section we derive the following result from which Lemma 5.1 follows directly. Recall the definition of Φ_ϵ in (6.1).

Lemma 6.4 *Fix $D \geq 2$ and $\ell \geq 8D$. Let $I_{r,s}$ be given by (6.6). Then we have, for any (small) $\epsilon > 0$,*

$$w_{I_{1,0}}\mathbb{E}[I_{1,0}] = -\mathbb{E}\big[m^2 P(m)^{D-1}\overline{P(m)^D}\big] + O(\Phi_\epsilon), \qquad w_{I_{2,0}}\mathbb{E}[I_{2,0}] = O(\Phi_\epsilon),$$
$$w_{I_{3,0}}\mathbb{E}[I_{3,0}] = -\frac{s^{(4)}}{q_t^2}\mathbb{E}\big[m^4 P(m)^{D-1}\overline{P(m)^D}\big] + O(\Phi_\epsilon), \quad w_{I_{r,0}}\mathbb{E}[I_{r,0}] = O(\Phi_\epsilon), \quad (4 \leq r \leq \ell),$$
$$(6.20)$$

uniformly in $z \in \mathcal{D}$, for N sufficiently large. Moreover, we have, for any (small) $\epsilon > 0$,

$$w_{I_{r,s}}|\mathbb{E}[I_{r,s}]| \leq \Phi_\epsilon, \qquad (1 \leq s \leq r \leq \ell), \qquad (6.21)$$

uniformly in $z \in \mathcal{D}$, for N sufficiently large.

Proof of Lemma 5.1 By the definition of Φ_ϵ in (6.1) (with a sufficiently large (small) $\epsilon > 0$), it suffices to show that $\mathbb{E}[|P|^{2D}(z)] \leq \Phi_\epsilon(z)$, for all $z \in \mathcal{D}$, for N sufficiently large. Choosing $\ell \geq 8D$, Corollary 6.2 asserts that $\mathbb{E}\Omega_\ell(I)$ in (6.8) is negligible. By Lemma 6.4 the only non-negligible terms in the expansion of the first term on the right side of (6.8) are $w_{I_{1,0}}\mathbb{E}I_{1,0}$ and $w_{I_{3,0}}\mathbb{E}I_{3,0}$, yet these two terms cancel with the

middle term on the right side of (6.8), up to negligible terms. Thus the whole right side of (6.8) is negligible. This proves Lemma 5.1. □

We now choose an initial (small) $\epsilon > 0$. Below we use the factor N^ϵ to absorb numerical constants in the estimates by allowing ϵ to increase by a tiny amount from line to line. We often drop z from the notation; it is always understood that $z \in \mathcal{D}$ and all estimates are uniform on $\mathcal{D}$. The proof of Lemma 6.4 is done in the remaining Sects. 6.3–6.7 where $(\mathbb{E}I_{r,s})$ are controlled.

6.3 Estimate on $I_{1,s}$

Starting from the definition of $I_{1,0}$ in (6.5), a direct computation yields

$$
\begin{aligned}
\mathbb{E}I_{1,0} &= \frac{\kappa_t^{(2)}}{N} \mathbb{E}\left[\sum_{i_1 \neq i_2} (\partial_{i_1 i_2} G_{i_2 i_1}) P^{D-1} \overline{P^D} \right] \\
&= -\mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} (G_{i_2 i_2} G_{i_1 i_1} + G_{i_1 i_2} G_{i_2 i_1}) P^{D-1} \overline{P^D} \right] \\
&= -\mathbb{E}\left[m^2 P^{D-1} \overline{P^D} \right] + \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 = 1}^{N} (G_{i_1 i_1})^2 P^{D-1} \overline{P^D} \right] \\
&\quad - \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} (G_{i_2 i_1})^2 P^{D-1} \overline{P^D} \right].
\end{aligned}
\tag{6.22}
$$

The middle term on the last line is negligible since

$$
\left| \frac{1}{N^2} \mathbb{E}\left[\sum_{i_1=1}^{N} (G_{i_1 i_1})^2 P^{D-1} \overline{P^D} \right] \right| \leq \frac{N^\epsilon}{N} \mathbb{E}\left[P^{D-1} \overline{P^D} \right],
$$

where we used $|G_{i_1 i_1}| \prec 1$, and so is the third term since

$$
\frac{1}{N^2} \left| \mathbb{E}\left[\sum_{i_1 \neq i_2} (G_{i_2 i_1})^2 P^{D-1} \overline{P^D} \right] \right| \leq N^\epsilon \mathbb{E}\left[\frac{\operatorname{Im} m}{N\eta} |P|^{2D-1} \right],
$$

where we used Lemma 6.5. We thus obtain from (6.22) that

$$
\left| I_{1,0} + \mathbb{E}\left[m^2 P^{D-1} \overline{P^D} \right] \right| \leq \Phi_\epsilon,
\tag{6.23}
$$

for N sufficiently large. This proves the first estimate in (6.20).

Consider next $I_{1,1}$. Similar to (3.11), we have

$$
\begin{aligned}
\mathbb{E}I_{1,1} &= \frac{1}{N^2} \sum_{i_1 \neq i_2} \mathbb{E}\left[G_{i_2 i_1} \partial_{i_1 i_2} (P^{D-1}\overline{P^D}) \right] \\
&= -\frac{2(D-1)}{N^3}\mathbb{E}\left[\sum_{i_1 \neq i_2} \sum_{i_3=1}^{N} G_{i_2 i_1} G_{i_2 i_3} G_{i_3 i_2} P'(m) P^{D-2}\overline{P^D} \right] \\
&\quad - \frac{2D}{N^3} \sum_{i_1 \neq i_2} \sum_{i_3=1}^{N} \mathbb{E}\left[G_{i_2 i_1} \overline{G_{i_3 i_2} G_{i_2 i_3} P'(m)} P^{D-1}\overline{P^{D-1}} \right].
\end{aligned}
$$

Here the *fresh summation index* i_3 originated from $\partial_{i_1 i_2} P(m) = P'(m)\frac{1}{N}\sum_{i_3=1}^{N}\partial_{i_1 i_2} G_{i_3 i_3}$. Note that we can add the terms with $i_1 = i_2$ at the expense of a negligible error, so that

$$
\begin{aligned}
\mathbb{E}I_{1,1} &= -2(D-1)\mathbb{E}\left[\frac{1}{N^3}\operatorname{Tr} G^3 P'(m) P^{D-2}\overline{P^D} \right] \\
&\quad - 2D\,\mathbb{E}\left[\frac{1}{N^3}G(G^*)^2\overline{P'(m)} P^{D-1}\overline{P^{D-1}} \right] + O(\Phi_\epsilon). \tag{6.24}
\end{aligned}
$$

In the remainder of this section, we will freely include or exclude negligible terms with coinciding indices. Using (3.14) and (3.15) we obtain from (6.24) the estimate

$$
\begin{aligned}
|\mathbb{E}I_{1,1}| &= \left| \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} G_{i_2 i_1} \partial_{i_1 i_2}(P^{D-1}\overline{P^D}) \right] \right| \\
&\leq (4D-2)\mathbb{E}\left[\frac{\operatorname{Im} m}{(N\eta)^2}|P'||P|^{2D-2} \right] + \Phi_\epsilon, \tag{6.25}
\end{aligned}
$$

for N sufficiently large. This proves (6.21) for $r = s = 1$.

6.4 Estimate on $I_{2,0}$

We start with a lemma that is used in the power counting arguments below.

Lemma 6.5 *For any $i, k \in [\![1, N]\!]$,*

$$
\frac{1}{N}\sum_{j=1}^{N}|G_{ij}(z)G_{jk}(z)| \prec \frac{\operatorname{Im} m(z)}{N\eta},
$$

$$
\frac{1}{N}\sum_{j=1}^{N}|G_{ij}(z)| \prec \left(\frac{\operatorname{Im} m(z)}{N\eta} \right)^{1/2}, \qquad (z \in \mathbb{C}^+). \tag{6.26}
$$

Moreover, for fixed $n \in \mathbb{N}$,

$$\frac{1}{N^n} \sum_{j_1, j_2, \ldots, j_n = 1}^{N} |G_{ij_1}(z) G_{j_1 j_2}(z) G_{j_2 j_3}(z) \cdots G_{j_n k}(z)| \prec \left(\frac{\operatorname{Im} m(z)}{N \eta} \right)^{n/2}, \quad (z \in \mathbb{C}^+).$$

(6.27)

Proof Let $\lambda_1^{H_t} \geq \lambda_2^{H_t} \geq \cdots \geq \lambda_N^{H_t}$ be the eigenvalues of H_t, and let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N, \boldsymbol{u}_\alpha \equiv \boldsymbol{u}_\alpha^{H_t}$, denoted the associated normalized eigenvectors. Then, by spectral decomposition, we get

$$\sum_{j=1}^{N} |G_{ij}|^2 = \sum_{j=1}^{N} \sum_{\alpha, \beta} \frac{\boldsymbol{u}_\alpha(i) \overline{\boldsymbol{u}_\alpha(j)}}{\lambda_\alpha - z} \frac{\overline{\boldsymbol{u}_\beta(i)} \boldsymbol{u}_\beta(j)}{\lambda_\beta - \overline{z}} = \sum_{\alpha, \beta} \frac{\boldsymbol{u}_\alpha(i) \langle \boldsymbol{u}_\alpha, \boldsymbol{u}_\beta \rangle \overline{\boldsymbol{u}_\beta(i)}}{(\lambda_\alpha - z)(\lambda_\beta - \overline{z})}$$

$$= \sum_{\alpha=1}^{N} \frac{|\boldsymbol{u}_\alpha(i)|^2}{|\lambda_\alpha - z|^2}.$$

Since the eigenvectors are delocalized by Proposition 2.7, we find that

$$\frac{1}{N} \sum_{j=1}^{N} |G_{ij}|^2 \prec \frac{1}{N^2} \sum_{\alpha=1}^{N} \frac{1}{|\lambda_\alpha - z|^2} = \frac{\operatorname{Im} m}{N \eta}.$$

This proves the first inequality in (6.26) for $i = k$. The inequality for $i \neq k$, the second inequality in (6.26) and (6.27) then follow directly from the Schwarz inequality. □

Recalling the definition of $I_{r,s}$ in (6.6) we have

$$I_{2,0} := N \kappa_t^{(3)} \frac{1}{N^2} \sum_{i_1 \neq i_2} \left(\partial_{i_1 i_2}^2 G_{i_2 i_1} \right) P^{D-1} \overline{P^D}.$$

We then notice that $I_{2,0}$ contains terms with one or three off-diagonal Green function entries $G_{i_1 i_2}$. We split accordingly

$$w_{I_{2,0}} I_{2,0} = w_{I_{2,0}^{(1)}} I_{2,0}^{(1)} + w_{I_{2,0}^{(3)}} I_{2,0}^{(3)},$$

(6.28)

where $I_{2,0}^{(1)}$ contains all terms with one off-diagonal Green function entries (and, necessarily, two diagonal Green function entries) and where $I_{2,0}^{(3)}$ contains all terms with three off-diagonal Green function entries (and zero diagonal Green function entries), and $w_{I_{2,0}^{(1)}}, w_{I_{2,0}^{(3)}}$ denote the respective weights. Explicitly,

$$\mathbb{E}I_{2,0}^{(1)} = N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2} G_{i_2i_1}G_{i_2i_2}G_{i_1i_1} P^{D-1}\overline{P^D}\right],$$

$$\mathbb{E}I_{2,0}^{(3)} = N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2} (G_{i_2i_1})^3 P^{D-1}\overline{P^D}\right], \tag{6.29}$$

and $w_{I_{2,0}} = 1$, $w_{I_{2,0}^{(1)}} = 3$, $w_{I_{2,0}^{(3)}} = 1$.

We first note that $I_{2,0}^{(3)}$ satisfies, for N sufficiently large,

$$|\mathbb{E}I_{2,0}^{(3)}| \leq \frac{N^\epsilon s^{(3)}}{q_t}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}|G_{i_1i_2}|^2|P|^{2D-1}\right] \leq \frac{N^\epsilon}{q_t}\mathbb{E}\left[\frac{\operatorname{Im}m}{N\eta}|P|^{2D-1}\right] \leq \Phi_\epsilon. \tag{6.30}$$

Remark 6.6 (Power counting I) Consider the terms $I_{r,0}$, $r \geq 1$. For $n \geq 1$, we then split

$$w_{I_{2n}}I_{2n,0} = \sum_{l=0}^{n} w_{I_{2n,0}^{(2l+1)}} I_{2n,0}^{(2l+1)},$$

$$w_{I_{2n-1}}I_{2n-1,0} = \sum_{l=0}^{n} w_{I_{2n-1,0}^{(2l)}} I_{2n-1,0}^{(2l)}, \tag{6.31}$$

according to the parity of r. For example, for $r = 1$, $I_{1,0} = I_{1,0}^{(0)} + I_{1,0}^{(2)}$, with

$$\mathbb{E}I_{1,0}^{(0)} = -\mathbb{E}\left[\frac{1}{N^2}\sum_{i \neq k} G_{kk}G_{ii} P^{D-1}\overline{P^D}\right],$$

$$\mathbb{E}I_{1,0}^{(2)} = -\mathbb{E}\left[\frac{1}{N^2}\sum_{i \neq k} G_{ik}G_{ki} P^{D-1}\overline{P^D}\right];$$

cf. (6.22). Now, using a simple power counting, we bound the summands in (6.31) as follows. First, we note that each term in $I_{r,0}$ contains a factor of $q_t^{(r-2)+}$. Second, for $\mathbb{E}I_{2n,0}^{(2l+1)}$ and $\mathbb{E}I_{2n-1,0}^{(2l)}$, with $n \geq 1$, $l \geq 1$, we can by Lemma 6.5 extract one factor of $\frac{\operatorname{Im}m}{N\eta}$ (other Green function entries are bounded using $|G_{ik}| \prec 1$). Thus, for $n \geq 1$, $l \geq 1$,

$$\left|\mathbb{E}I_{2n,0}^{(2l+1)}\right| \leq \frac{N^\epsilon}{q_t^{2n-2}}\mathbb{E}\left[\left(\frac{\operatorname{Im}m}{N\eta}\right)|P|^{2D-1}\right],$$

$$\left|\mathbb{E}I_{2n-1,0}^{(2l)}\right| \leq \frac{N^\epsilon}{q_t^{(2n-3)+}}\mathbb{E}\left[\left(\frac{\operatorname{Im}m}{N\eta}\right)|P|^{2D-1}\right], \tag{6.32}$$

for N sufficiently large, and we conclude that all these terms are negligible.

We next consider $\mathbb{E}I_{2,0}^{(1)}$ which is not covered by (6.32). Using $|G_{ii}| \prec 1$ and Lemma 6.5 we get

$$\left|\mathbb{E}I_{2,0}^{(1)}\right| \leq \frac{N^\epsilon s^{(3)}}{q_t} \mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}|G_{i_1 i_2}||P|^{2D-1}\right] \leq \frac{N^\epsilon s^{(3)}}{q_t}\mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{1/2}|P|^{2D-1}\right],$$

(6.33)

for N sufficiently large. Yet, this bound is not negligible. We need to gain an additional factor of q_t^{-1} with which it will become negligible. We have the following result.

Lemma 6.7 *For any (small) $\epsilon > 0$, we have, for all $z \in \mathcal{D}$,*

$$\left|\mathbb{E}I_{2,0}^{(1)}\right| \leq \frac{N^\epsilon}{q_t^2}\mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{1/2}|P|^{2D-1}\right] + \Phi_\epsilon$$

$$\leq N^\epsilon \mathbb{E}\left[\left(\frac{1}{q_t^4} + \frac{\mathrm{Im}\, m}{N\eta}\right)|P(m)|^{2D-1}\right] + \Phi_\epsilon,$$

(6.34)

for N sufficiently large. In particular, the term $\mathbb{E}I_{2,0}$ is negligible.

Proof Fix a (small) $\epsilon > 0$. Recalling (6.29), we have

$$\mathbb{E}I_{2,0}^{(1)} = N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2} G_{i_2 i_1}G_{i_1 i_1}G_{i_2 i_2}P^{D-1}\overline{P^D}\right].$$

(6.35)

The key feature here is that the Green function entries are $G_{i_2 i_1}G_{i_1 i_1}G_{i_2 i_2}$, where at least one index, say i_2, appears an odd number of times. We say that the index i_2 is *unmatched*. Using the resolvent formula (3.23) we expand in the unmatched index i_2 to get

$$z\mathbb{E}I_{2,0}^{(1)} = N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2\neq i_3} H_{i_2 i_3}G_{i_3 i_1}G_{i_2 i_2}G_{i_1 i_1}P^{D-1}\overline{P^D}\right].$$

(6.36)

We now proceed in a similar way as in Remark 3.1 where we estimated $|G_{i_1 i_2}^W|$, $i_1 \neq i_2$, for the GOE. Applying the cumulant expansion to the right side of (6.36), we will show that the leading term is $-\mathbb{E}[m I_{2,0}^{(1)}]$. Then, upon substituting $m(z)$ by the deterministic quantity $\widetilde{m}(z)$ and showing that all other terms in the cumulant expansion of the right side of (6.36) are negligible, we will get that

$$|z + \widetilde{m}(z)|\left|\mathbb{E}I_{2,0}^{(1)}\right| \leq \frac{N^\epsilon}{q_t^2}\mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{1/2}|P|^{2D-1}\right] + \Phi_\epsilon \leq 2\Phi_\epsilon,$$

(6.37)

for N sufficiently large. Since $|z + \widetilde{m}(z)| \geq 1/6$ uniformly on $\mathcal{E} \supset \mathcal{D}$, as shown in Remark 4.3, the lemma directly follows. The main efforts in the proof go into showing that the sub-leading terms in the cumulant expansion of the right side of (6.36) are indeed negligible.

For simplicity we abbreviate $\widehat{I} \equiv I_{2,0}^{(1)}$. Then using Lemma 3.2 and Remark 6.3, we have, for arbitrary $\ell' \in \mathbb{N}$, the cumulant expansion

$$z\mathbb{E}I_{2,0}^{(1)} = z\mathbb{E}\widehat{I} = \sum_{r'=1}^{\ell'}\sum_{s'=0}^{r'} w_{\widehat{I}_{r',s'}}\mathbb{E}\widehat{I}_{r',s'} + O\left(\frac{N^\epsilon}{q_t^{\ell'}}\right), \tag{6.38}$$

with

$$\widehat{I}_{r',s'} := N\kappa_t^{(r'+1)}N\kappa_t^{(3)}\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}\left(\partial_{i_2 i_3}^{r'-s'}(G_{i_3 i_1}G_{i_2 i_2}G_{i_1 i_1})\right)\left(\partial_{i_2 i_3}^{s'}(P^{D-1}\overline{P^D})\right) \tag{6.39}$$

and $w_{\widehat{I}_{r',s'}} = \frac{1}{r'!}\binom{r'}{s'}$. Here, we used Corollary 6.2 to truncate the series in (6.38) at order ℓ'. Choosing $\ell' \geq 8D$ the remainder is indeed negligible.

We first focus on $\widehat{I}_{r',0}$. For $r' = 1$, we compute

$$\begin{aligned}
\mathbb{E}\widehat{I}_{1,0} = &-\frac{s^{(3)}}{q_t}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}G_{i_2 i_1}G_{i_3 i_3}G_{i_2 i_2}G_{i_1 i_1}P^{D-1}\overline{P^D}\right]\\
&-3\frac{s^{(3)}}{q_t}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}G_{i_2 i_3}G_{i_3 i_1}G_{i_2 i_2}G_{i_1 i_1}P^{D-1}\overline{P^D}\right]\\
&-2\frac{s^{(3)}}{q_t}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}G_{i_1 i_2}G_{i_2 i_3}G_{i_3 i_1}G_{i_2 i_2}P^{D-1}\overline{P^D}\right]\\
=: &\,\mathbb{E}\widehat{I}_{1,0}^{(1)} + 3\mathbb{E}\widehat{I}_{1,0}^{(2)} + 2\mathbb{E}\widehat{I}_{1,0}^{(3)}, \tag{6.40}
\end{aligned}$$

where we organized the terms according to the off-diagonal Green functions entries. By Lemma 6.5,

$$\begin{aligned}
\left|\mathbb{E}\widehat{I}_{1,0}^{(2)}\right| &\leq \frac{N^\epsilon}{q_t}\mathbb{E}\left[\frac{\text{Im}\,m}{N\eta}|P|^{2D-1}\right] \leq \Phi_\epsilon,\\
\left|\mathbb{E}\widehat{I}_{1,0}^{(3)}\right| &\leq \frac{N^\epsilon}{q_t}\mathbb{E}\left[\left(\frac{\text{Im}\,m}{N\eta}\right)^{3/2}|P|^{2D-1}\right] \leq \Phi_\epsilon. \tag{6.41}
\end{aligned}$$

Recall $\widetilde{m} \equiv \widetilde{m}_t(z)$ defined in Proposition 3.3. We rewrite $\widehat{I}_{1,0}^{(1)}$ with $\widetilde{m}$ as

$$\begin{aligned}
\mathbb{E}\widehat{I}_{1,0}^{(1)} = &-\mathbb{E}\frac{s^{(3)}}{q_t}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}\widetilde{m}G_{i_2 i_1}G_{i_2 i_2}G_{i_1 i_1}P^{D-1}\overline{P^D}\right]\\
&-\mathbb{E}\frac{s^{(3)}}{q_t}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}(m-\widetilde{m})G_{i_2 i_1}G_{i_2 i_2}G_{i_1 i_1}P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon). \tag{6.42}
\end{aligned}$$

By the Schwarz inequality and the high probability bounds $|G_{kk}|, |G_{ii}| \leq N^{\epsilon/8}$, for N sufficiently large, the second term in (6.42) is bounded as

$$\frac{|s^{(3)}|}{q_t} \left| \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} (m - \tilde{m}) G_{i_2 i_1} G_{i_2 i_2} G_{i_1 i_1} P^{D-1} \overline{P^D} \right] \right|$$

$$\leq N^{\epsilon/4} \frac{|s^{(3)}|}{q_t} \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} |m - \tilde{m}| |G_{i_1 i_2}| |P|^{2D-1} \right]$$

$$\leq N^{-\epsilon/4} \frac{|s^{(3)}|}{q_t} \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} |m - \tilde{m}|^2 |P|^{2D-1} \right]$$

$$+ N^{3\epsilon/4} \frac{|s^{(3)}|}{q_t} \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} |G_{i_1 i_2}|^2 |P|^{2D-1} \right]$$

$$= N^{-\epsilon/4} \frac{|s^{(3)}|}{q_t} \mathbb{E}\left[|m - \tilde{m}|^2 |P|^{2D-1} \right] + N^{3\epsilon/4} \frac{|s^{(3)}|}{q_t} \mathbb{E}\left[\frac{\operatorname{Im} m}{N\eta} |P|^{2D-1} \right].$$
$$(6.43)$$

We thus get from (6.40), (6.41), (6.42) and (6.43) that

$$\mathbb{E}\widehat{I}_{1,0} = -\tilde{m}\mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} G_{i_2 i_1} G_{i_2 i_2} G_{i_1 i_1} P^{D-1} \overline{P^D} \right] + O(\Phi_\epsilon) = -\mathbb{E}\tilde{m} I_{2,0}^{(1)} + O(\Phi_\epsilon),$$
$$(6.44)$$

where we used (6.35). We remark that in the expansion of $\mathbb{E}\widehat{I} = \mathbb{E}I_{2,0}^{(1)}$ the only term with one off-diagonal entry is $\mathbb{E}\widehat{I}_{2,0}^{(1)}$. All the other terms contain at least two off-diagonal entries.

Remark 6.8 (Power counting II) Comparing (6.5) and (6.38), we have $\widehat{I}_{r',s'} = (I_{2,0}^{(1)})_{r',s'}$. Consider now the terms with $s' = 0$. As in (6.31) we organize the terms according to the number of off-diagonal Green function entries. For $r' \geq 2$,

$$w_{\widehat{I}_{r',0}} \widehat{I}_{r',0} = \sum_{l=0}^{n} w_{\widehat{I}_{r',0}^{(l+1)}} \widehat{I}_{r',0}^{(l+1)} = \sum_{l=0}^{n} w_{\widehat{I}_{r',0}^{(l+1)}} (I_{2,0}^{(1)})_{r',0}^{(l+1)}.$$
$$(6.45)$$

A simple power counting as in Remark 6.6 then directly yields

$$\left| \mathbb{E}\widehat{I}_{r',0}^{(1)} \right| \leq \frac{N^\epsilon}{q_t^{r'}} \mathbb{E}\left[\left(\frac{\operatorname{Im} m}{N\eta} \right)^{1/2} |P|^{2D-1} \right],$$

$$\left| \mathbb{E}\widehat{I}_{r',0}^{(l+1)} \right| \leq \frac{N^\epsilon}{q_t^{r'}} \mathbb{E}\left[\left(\frac{\operatorname{Im} m}{N\eta} \right) |P|^{2D-1} \right], \qquad (l \geq 1),$$
$$(6.46)$$

for N sufficiently large. Here, we used that each term contains a factor $\kappa_t^{(3)}\kappa_t^{(r'+1)} \leq CN^{-2}q_t^{-r'}$. We conclude that all terms in (6.46) with $r' \geq 2$ are negligible. Yet we remark that $|\mathbb{E}\widehat{I}_{2,0}^{(1)}|$ is the leading error term in $|\mathbb{E}I_{2,0}^{(1)}|$, which is explicitly listed on the right side of (6.34).

Remark 6.9 (Power counting III) Consider the terms $\widehat{I}_{r',s'}$, with $1 \leq s' \leq r'$. For $s' = 1$, note that $\partial_{i_2i_3}(P^{D-1}\overline{P^D})$ contains two off-diagonal Green function entries. Explicitly,

$$
\begin{aligned}
\widehat{I}_{r',1} = & -2(D-1)\frac{N\kappa_t^{(r'+1)}N\kappa_t^{(3)}}{N^3} \\
& \times \sum_{i_1 \neq i_2 \neq i_3} \left(\partial_{i_2i_3}^{r'-1}(G_{i_3i_1}G_{i_2i_2}G_{i_1i_1})\right)\left(\frac{1}{N}\sum_{i_4=1}^{N}G_{i_4i_2}G_{i_3i_4}\right)P'P^{D-2}\overline{P^D} \\
& -2D\frac{N\kappa_t^{(r'+1)}N\kappa_t^{(3)}}{N^3} \\
& \times \sum_{i_1 \neq i_2 \neq i_3} \left(\partial_{i_2i_3}^{r'-1}(G_{i_3i_1}G_{i_2i_2}G_{i_1i_1})\right)\left(\frac{1}{N}\sum_{i_4=1}^{N}\overline{G_{i_4i_2}G_{i_3i_4}}\right)\overline{P'}P^{D-1}\overline{P^{D-1}},
\end{aligned}
$$

where the fresh summation index i_4 is generated from $\partial_{i_2i_3}P$. Using Lemma 6.5 we get, for $r' \geq 1$,

$$
|\mathbb{E}\widehat{I}_{r',1}| \leq \frac{N^\epsilon}{q_t^{r'}}\mathbb{E}\left[\left(\frac{\operatorname{Im} m}{N\eta}\right)^{3/2}|P'||P|^{2D-2} + \left(\frac{\operatorname{Im} m}{N\eta}\right)^{3/2}|P'||P|^{2D-2}\right] \leq 2\Phi_\epsilon,
$$
(6.47)

for N sufficiently large, where we used that $\partial_{i_2i_3}^{r'-1}(G_{i_3i_1}G_{i_2i_2}G_{i_1i_1})$, $r' \geq 1$, contains at least one off-diagonal Green function entry.

For $2 \leq s' \leq r'$, we first note that, for N sufficiently large,

$$
\begin{aligned}
|\mathbb{E}\widehat{I}_{r',s'}| &\leq \frac{N^\epsilon}{q_t^{r'}}\left|\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1 \neq i_2 \neq i_3}\left(\partial_{i_2i_3}^{r'-s'}G_{i_3i_1}G_{i_2i_2}G_{i_1i_1}\right)\left(\partial_{i_2i_3}^{s'}(P^{D-1}\overline{P^D})\right)\right]\right| \\
&\leq \frac{N^\epsilon}{q_t^{r'}}\mathbb{E}\left[\left(\frac{\operatorname{Im} m}{N\eta}\right)^{1/2}\frac{1}{N^2}\sum_{i_2 \neq i_3}|\partial_{i_2i_3}^{s'}(P^{D-1}\overline{P^D})|\right].
\end{aligned}
$$
(6.48)

Next, since $s' \geq 2$, the partial derivative in $\partial_{i_2i_3}^{s'}(P^{D-1}\overline{P^D})$ acts on P and $\overline{P}$ (and on their derivatives) more than once. For example, for $s' = 2$,

$$\partial_{i_1 i_2}^2 P^{D-1} = \frac{4(D-1)(D-2)}{N^2} \left(\sum_{i_3=1}^N G_{i_3 i_2} G_{i_1 i_3} \right)^2 (P')^2 P^{D-3}$$

$$+ \frac{2(D-1)}{N^2} \left(\sum_{i_3=1}^N G_{i_2 i_3} G_{i_3 i_1} \right)^2 P'' P^{D-2}$$

$$- \frac{2(D-1)}{N} \sum_{i_3=1}^N \partial_{i_1 i_2} \left(G_{i_3 i_2} G_{i_1 i_3} \right) P' P^{D-2},$$

where $\partial_{i_1 i_2}$ acted twice on P, respectively P', to produce the first two terms. More generally, for $s' \geq 2$, consider a resulting term containing

$$P^{D-s_1'} \overline{P^{D-s_2'}} \left(P' \right)^{s_3'} \left(\overline{P'} \right)^{s_4'} \left(P'' \right)^{s_5'} \left(\overline{P''} \right)^{s_6'} \left(P''' \right)^{s_7'} \left(\overline{P'''} \right)^{s_8'}, \tag{6.49}$$

with $1 \leq s_1' \leq D$, $0 \leq s_2' \leq D$ and $\sum_{n=1}^8 s_n' \leq s'$. Since $P^{(4)}$ is constant we did not list it. We see that such a term above was generated from $P^{D-1} \overline{P^D}$ by letting the partial derivative $\partial_{i_2 i_3}$ act $(s_1' - 1)$-times on P and s_2'-times on $\overline{P}$, which implies that $s_1' - 1 \geq s_3'$ and $s_2' \geq s_4'$. If $s_1' - 1 > s_3'$, then $\partial_{i_2 i_3}$ acted on the derivatives of $P, \overline{P}$ directly $(s_1' - 1 - s_3')$-times, and a similar argument holds for $\overline{P'}$. Whenever $\partial_{i_2 i_3}$ acted on $P, \overline{P}$ or their derivatives, it generated a term $2N^{-1} \sum_{i_l} G_{i_2 i_l} G_{i_l i_3}$, with $i_l, l \geq 3$, a fresh summation index. For each fresh summation index we apply Lemma 6.5 to gain a factor $\frac{\operatorname{Im} m}{N \eta}$. The total number of fresh summation indices in a term corresponding to (6.49) is

$$s_1' - 1 + s_2' + (s_1' - 1 - s_3') + (s_2' - s_4') = 2s_1' + 2s_2' - s_3' - s_4' - 2 = 2\tilde{s}_0 - \tilde{s} - 2,$$

with $\tilde{s}_0 := s_1' + s_2'$ and $\tilde{s} := s_3 + s_4$, and we note this number does not decrease when $\partial_{i_2 i_3}$ acts on off-diagonal Green functions entries later. Thus, from (6.48) we conclude, upon using $|G_{i_1 i_2}|$, $|P''(m)|$, $|P'''(m)|$, $|P^{(4)}(m)| \prec 1$ that, for $2 \leq s' \leq r'$,

$$|\mathbb{E} \widehat{I}_{r', s'}| \leq \frac{N^\epsilon}{q_t^{r'}} \mathbb{E} \left[\left(\frac{\operatorname{Im} m}{N \eta} \right)^{1/2} \frac{1}{N^2} \sum_{i_2 \neq i_3} \left| \partial_{i_2 i_3}^{s'} \left(P^{D-1} \overline{P^D} \right) \right| \right]$$

$$\leq \frac{N^{2\epsilon}}{q_t^{r'}} \sum_{\tilde{s}_0=2}^{2D} \sum_{\tilde{s}=1}^{\tilde{s}_0-2} \mathbb{E} \left[\left(\frac{\operatorname{Im} m}{N \eta} \right)^{1/2 + 2\tilde{s}_0 - \tilde{s} - 2} |P'|^{\tilde{s}} |P|^{2D - \tilde{s}_0} \right]$$

$$+ \frac{N^{2\epsilon}}{q_t^{r'}} \sum_{\tilde{s}_0=2}^{2D} \mathbb{E} \left[\left(\frac{\operatorname{Im} m}{N \eta} \right)^{1/2 + 2\tilde{s}_0 - 1} |P'|^{\tilde{s}_0 - 1} |P|^{2D - \tilde{s}_0} \right], \tag{6.50}$$

for N sufficiently large. Here the last term on the right corresponds to $\tilde{s} = \tilde{s}_0 - 1$. Thus, we conclude from (6.50) and the definition of Φ_ϵ in (6.1) that $\mathbb{E}[\widehat{I}_{r', s'}]$, $2 \leq s' \leq r'$, is negligible.

To sum up and conclude this remark, we established that all $\mathbb{E}[\widetilde{I}_{r',s'}]$ with $1 \leq s' \leq r'$ are negligible.

From (6.35), (6.38), (6.44), (6.46), (6.47) and (6.50) we find that

$$|z + \widetilde{m}| \left|\mathbb{E}I_{2,0}^{(1)}\right| \leq \frac{N^\epsilon}{q_t^2}\mathbb{E}\left[\left(\frac{\mathrm{Im}\,m}{N\eta}\right)^{1/2}|P|^{2D-1}\right] + \Phi_\epsilon \,,$$

for N sufficiently large. Since $(z+\widetilde{m})$ is deterministic and $|z+\widetilde{m}| > 1/6$, as we showed in Remark 4.3, we obtain that $|\mathbb{E}I_{2,0}^{(1)}| \leq \Phi_\epsilon$. This concludes the proof of (6.34). $\square$

Combining (6.28), (6.30) and (6.34) we showed that

$$|\mathbb{E}I_{2,0}| \leq \Phi_\epsilon \,, \tag{6.51}$$

for N sufficiently large, *i.e.* all terms in $\mathbb{E}I_{2,0}$ are negligible and the second estimate in (6.20) is proved.

6.5 Estimate on $I_{r,0}, r \geq 4$

For $r \geq 5$ we use the bounds $|G_{i_1i_1}|, |G_{i_1i_2}| \prec 1$ to get

$$|\mathbb{E}I_{r,0}| = \left|N\kappa_t^{(r+1)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}\left(\partial_{i_1i_2}^r G_{i_2i_1}\right)P^{D-1}\overline{P^D}\right]\right|$$
$$\leq \frac{N^\epsilon}{q_t^4}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}|P|^{2D-1}\right] \leq \frac{N^\epsilon}{q_t^4}\mathbb{E}[|P|^{2D-1}] \leq \Phi_\epsilon \,, \tag{6.52}$$

for N sufficiently large. For $r = 4$, $\partial_{i_1i_2}^r G_{i_2i_1}$ contains at least one off-diagonal term $G_{i_1i_2}$. Thus

$$\left|N\kappa_t^{(5)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}\left(\partial_{i_1i_2}^4 G_{i_2i_1}\right)P^{D-1}\overline{P^D}\right]\right| \leq \frac{N^\epsilon}{q_t^3}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}|G_{i_2i_1}||P|^{2D-1}\right]$$
$$\leq \frac{N^\epsilon}{q_t^3}\mathbb{E}\left[\left(\frac{\mathrm{Im}\,m}{N\eta}\right)^{1/2}|P|^{2D-1}\right] \leq \Phi_\epsilon \,, \tag{6.53}$$

for N sufficiently large, where we used Lemma 6.5 to get the last line. We conclude that all terms $\mathbb{E}I_{r,0}$ with $r \geq 4$ are negligible. This proves the fourth estimate in (6.20).

6.6 Estimate on $I_{r,s}$, $r \geq 2$, $s \geq 1$

For $r \geq 2$ and $s = 1$, we have

$$\mathbb{E}I_{r,1} = N\kappa_t^{(r+1)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2}(\partial_{i_1i_2}^{r-1}G_{i_2i_1})\partial_{i_1i_2}(P^{D-1}\overline{P^D})\right].$$

Note that each term in $\mathbb{E}I_{r,1}$, $r \geq 2$, contains at least two off-diagonal Green function entries. For the terms with at least three off-diagonal Green function entries, we use the bound $|G_{i_1i_2}|, |G_{i_1i_1}| \prec 1$ and

$$N\kappa_t^{(r+1)}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1,i_2,i_3}|G_{i_2i_1}G_{i_1i_3}G_{i_3i_2}||P'||P|^{2D-2}\right]$$

$$\leq N^\epsilon\frac{s^{(r+1)}}{q_t}\mathbb{E}\left[\left(\frac{\operatorname{Im}m}{N\eta}\right)^{3/2}|P'||P|^{2D-2}\right]$$

$$\leq N^\epsilon s^{(r+1)}\mathbb{E}\left[\sqrt{\operatorname{Im}m}\left(\frac{\operatorname{Im}m}{N\eta}\right)\left(\frac{1}{N\eta}+q_t^{-2}\right)|P'||P|^{2D-2}\right], \tag{6.54}$$

for N sufficiently large, where we used Lemma 6.5. Note that the right side is negligible since $\operatorname{Im}m \prec 1$.

Denoting the terms with two off-diagonal Green function entries in $\mathbb{E}I_{r,1}$ by $\mathbb{E}I_{r,1}^{(2)}$, we have

$$\mathbb{E}I_{r,1}^{(2)} = N\kappa^{(r+1)}\mathbb{E}\left[\frac{2(D-1)}{N^2}\sum_{i_1 \neq i_2}G_{i_2i_2}^{r/2}G_{i_1i_1}^{r/2}\left(\frac{1}{N}\sum_{i_3=1}^N G_{i_2i_3}G_{i_3i_1}\right)P'P^{D-2}\overline{P^D}\right]$$

$$+ N\kappa^{(r+1)}\mathbb{E}\left[\frac{2D}{N^2}\sum_{i_1 \neq i_2}G_{i_2i_2}^{r/2}G_{i_1i_1}^{r/2}\left(\frac{1}{N}\sum_{i_3=1}^N \overline{G_{i_2i_3}G_{i_3i_1}}\right)\overline{P'}P^{D-1}\overline{P^{D-1}}\right], \tag{6.55}$$

where i_3 is a fresh summation index and where we noted that r is necessarily even in this case. Lemma 6.5 then give us the upper bound

$$|\mathbb{E}I_{r,1}^{(2)}| \leq \frac{N^\epsilon}{q_t^{r-1}}\mathbb{E}\left[\frac{\operatorname{Im}m}{N\eta}|P'||P|^{2D-2}\right],$$

which is negligible for $r > 2$. However, for $r = 2$, we need to gain an additional factor q_t^{-1}. This can be done as in the proof of Lemma 6.7 by considering the off-diagonal entries $G_{i_2i_3}G_{i_3i_1}$, generated from $\partial_{i_1i_2}P(m)$, since the index i_2 appears an odd number of times.

Lemma 6.10 *For any (small) $\epsilon > 0$, we have*

$$\left|\mathbb{E}I_{2,1}^{(2)}\right| \leq \frac{N^\epsilon}{q_t^2}\mathbb{E}\left[\frac{\operatorname{Im}m}{N\eta}|P'||P|^{2D-2}\right] + \Phi_\epsilon, \tag{6.56}$$

uniformly on $\mathcal{D}$, for N sufficiently large. In particular, the term $\mathbb{E}I_{2,1}$ is negligible.

Proof We start with the first term on the right side of (6.55). Using (3.23), we write

$$zN\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}G_{i_2i_3}G_{i_2i_2}G_{i_1i_1}G_{i_1i_3}P'P^{D-2}\overline{P^D}\right]$$

$$= N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3\neq i_4}H_{i_2i_4}G_{i_4i_3}G_{i_2i_2}G_{i_1i_1}G_{i_1i_3}P'P^{D-2}\overline{P^D}\right].$$

As in the proof of Lemma 6.7, we now apply the cumulant expansion to the right side. The leading terms of the expansion is

$$N\kappa_t^{(3)}\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}mG_{i_2i_3}G_{i_2i_2}G_{i_1i_1}G_{i_1i_3}P'P^{D-2}\overline{P^D}\right], \qquad (6.57)$$

and, thanks to the additional factor of q_t^{-1} from the cumulant $\kappa_t^{(3)}$, all other terms in the cumulant expansion are negligible, as can be checked by power counting as in the proof of Lemma 6.7. Replacing in (6.57) m by $\widetilde{m}$, we then get

$$|z+\widetilde{m}|\left|N\kappa_t^{(3)}\right|\left|\mathbb{E}\left[\frac{1}{N^3}\sum_{i_1\neq i_2\neq i_3}G_{i_2i_3}G_{i_2i_2}G_{i_1i_1}G_{i_1i_3}P'P^{D-2}\overline{P^D}\right]\right| \leq C\Phi_\epsilon,$$

for N sufficiently large; *cf.* (6.37). Since $|z+\widetilde{m}(z)| \geq 1/6$, $z \in \mathcal{D}$ by Remark 4.3, we conclude that the first term on the right side of (6.55) is negligible. In the same way one shows that the second term is negligible, too. We leave the details to the reader. $\square$

We hence conclude from (6.54) and (6.56) that $\mathbb{E}I_{r,1}$ is negligible for all $r \geq 2$.

Consider next the terms

$$\mathbb{E}I_{r,s} = N\kappa_t^{(r+1)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}(\partial_{i_1i_2}^{r-s}G_{i_2i_1})\partial_{i_1i_2}^s(P^{D-1}\overline{P^D})\right],$$

with $2 \leq s \leq r$. We proceed in a similar way as in Remark 6.8. We note that each term in $\partial_{i_1i_2}^{r-s}G_{i_2i_1}$ contains at least one off-diagonal Green function when $r - s$ is even, yet when $r - s$ is odd there is a term with no off-diagonal Green function entries. Since $s \geq 2$, the partial derivative $\partial_{i_1i_2}^s$ acts on P or $\overline{P}$ (or their derivatives) more than once in total; *cf.* Remark 6.8. Consider such a term with

$$P^{D-s_1}\overline{P^{D-s_2}}\left(P'\right)^{s_3}\left(\overline{P'}\right)^{s_4},$$

for $1 \leq s_1 \leq D$ and $0 \leq s_2 \leq D$. Since $P''(m), P'''(m), P^{(4)}(m) \prec 1$ and $P^{(5)} = 0$, we do not include derivatives of order two and higher here. We see that such a term was generated from $P^{D-1}\overline{P^D}$ by letting the partial derivative $\partial_{i_1i_2}$ act $(s_1 - 1)$-times

on P and s_2-times on $\overline{P}$, which implies that $s_3 \leq s_1 - 1$ and $s_4 \leq s_2$. If $s_3 < s_1 - 1$, then $\partial_{i_1 i_2}$ acted on P' as well $[(s_1 - 1) - s_3]$-times, and a similar argument holds for $\overline{P'}$. Whenever $\partial_{i_1 i_2}$ acts on P or $\overline{P}$ (or their derivatives), it generates a fresh summation index i_l, $l \geq 3$, with a term $2N^{-1} \sum_{i_l} G_{i_2 i_l} G_{i_l i_1}$. The total number of fresh summation indices in this case is

$$(s_1 - 1) + s_2 + [(s_1 - 1) - s_3] + [s_2 - s_4] = 2s_1 + 2s_2 - s_3 - s_4 - 2 \,.$$

Assume first that $r = s$ so that $\partial_{i_1 i_2}^{r-s} G_{i_2 i_1} = G_{i_2 i_1}$. Then applying Lemma 6.5 $(2s_1 + 2s_2 - s_3 - s_4 - 2)$-times and letting $s_0 := s_1 + s_2$ and $s' := s_3 + s_4$, we obtain the upper bound, $r = s \geq 2$,

$$|\mathbb{E}I_{r,r}| \leq \frac{N^\epsilon}{q_t^{r-1}} \sum_{s_0=2}^{2D} \sum_{s'=1}^{s_0-1} \mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{1/2} \left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{2s_0-s'-2} |P'|^{s'} |P|^{2D-s_0}\right] \leq \Phi_\epsilon \,, \tag{6.58}$$

for N sufficiently large, i.e. $\mathbb{E}I_{r,r}$, $r \geq 2$, is negligible.

Second, assume that $2 \leq s < r$. Then applying Lemma 6.5 $(2s_1 + 2s_2 - s_3 - s_4 - 2)$-times, we get

$$|\mathbb{E}I_{r,s}| \leq \frac{N^\epsilon}{q_t^{r-1}} \sum_{s_0=2}^{2D} \sum_{s'=1}^{s_0-2} \mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{2s_0-s'-2} |P'|^{s'} |P|^{2D-s_0}\right]$$
$$+ \frac{N^\epsilon}{q_t^{r-1}} \sum_{s_0=2}^{2D} \mathbb{E}\left[\left(\frac{\mathrm{Im}\, m}{N\eta}\right)^{s_0-1} |P'|^{s_0-1} |P|^{2D-s_0}\right], \tag{6.59}$$

for N sufficiently large with $2 \leq s < r$. In particular, $|\mathbb{E}I_{r,s}| \leq \Phi_\epsilon$, $2 \leq s < r$. In (6.59) the second term bounds the terms corresponding to $s_0 - 1 = s'$ obtained by acting on $\partial_{i_1 i_2}$ exactly $(s_1 - 1)$-times on P and s_2-times on $\overline{P}$ but never on their derivatives.

To sum up, we showed that $\mathbb{E}I_{r,s}$ is negligible, for $1 \leq s < r$. This proves (6.21) for $1 \leq s < r$.

6.7 Estimate on $I_{3,0}$

We first notice that $I_{3,0}$ contains terms with zero, two or four off-diagonal Green function entries and we split accordingly

$$w_{I_{3,0}} I_{3,0} = w_{I_{3,0}^{(0)}} I_{3,0}^{(0)} + w_{I_{3,0}^{(2)}} I_{3,0}^{(2)} + w_{I_{3,0}^{(4)}} I_{3,0}^{(4)} \,.$$

When there are two off-diagonal entries, we can use Lemma 6.5 to get the bound

$$\left| \mathbb{E} I_{3,0}^{(2)} \right| = \left| N \kappa_t^{(4)} \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} G_{i_2 i_2} G_{i_1 i_1} (G_{i_2 i_1})^2 P^{D-1} \overline{P^D} \right] \right|$$

$$\leq \frac{N^\epsilon}{q_t^2} \mathbb{E}\left[\frac{\mathrm{Im}\, m}{N \eta} |P|^{2D-1} \right] \leq \Phi_\epsilon , \tag{6.60}$$

for N sufficiently large. A similar estimate holds for $|\mathbb{E} I_{3,0}^{(2)}|$. The only non-negligible term is $I_{3,0}^{(0)}$.

For $n \in \mathbb{N}$, set

$$S_n \equiv S_n(z) := \frac{1}{N} \sum_{i=1}^{N} (G_{ii}(z))^n, \quad (z \in \mathbb{C}^+). \tag{6.61}$$

By definition $S_1 = m$. We remark that $|S_n| \prec 1$ on $\mathcal{D}$, for any fixed n, by Proposition 2.6.

Lemma 6.11 *We have*

$$w_{I_{3,0}^{(0)}} \mathbb{E} I_{3,0}^{(0)} = -N \kappa_t^{(4)} \mathbb{E}\left[S_2^2 P^{D-1} \overline{P^D} \right]. \tag{6.62}$$

Proof Recalling the definition of $I_{r,s}$ in (6.5), we have

$$w_{I_{3,0}} I_{3,0} = \frac{N \kappa_t^{(4)}}{3!} \mathbb{E}\left[\frac{1}{N^2} \sum_{i_1 \neq i_2} \left(\partial_{i_1 i_2}^3 G_{i_2 i_1} \right) P^{D-1} \overline{P^D} \right].$$

We then easily see that the terms with no off-diagonal entries in $\partial_{i_1 i_2}^3 G_{i_2 i_1}$ are of the form

$$-G_{i_2 i_2} G_{i_1 i_1} G_{i_2 i_2} G_{i_1 i_1} .$$

We only need to determine the weight $w_{I_{3,0}^{(0)}}$. With regard to the indices, taking the third derivative corresponds to putting the indices $i_2 i_1$ or $i_1 i_2$ three times. In that sense, the very first i_2 and the very last i_1 are from the original $G_{i_2 i_1}$. The choice of $i_2 i_1$ or $i_1 i_2$ must be exact in the sense that the connected indices in the following diagram must have been put at the same time:

$$i_2 \underbrace{i_2 \quad i_1}_{} \underbrace{i_1 \quad i_2}_{} \underbrace{i_2 \quad i_1}_{} i_1 .$$

Thus, the only combinatorial factor we have to count is the order of putting the indices. In this case, we have three connected indices, so the number of terms must be $3! = 6$. Thus, $w_{I_{3,0}^{(0)}} = 1$ and (6.62) indeed holds. $\qquad \square$

Lemma 6.12 *For any (small) $\epsilon > 0$, we have, for all $z \in \mathcal{D}$,*

$$N\kappa_t^{(4)}\mathbb{E}\left[S_2^2 P^{D-1}\overline{P^D}\right] = N\kappa_t^{(4)}\mathbb{E}\left[m^4 P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon). \qquad (6.63)$$

Proof Fix $\epsilon > 0$. We first claim that

$$N\kappa_t^{(4)}\mathbb{E}\left[S_2^2 P^{D-1}\overline{P^D}\right] = N\kappa_t^{(4)}\mathbb{E}\left[m^2 S_2 P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon). \qquad (6.64)$$

The idea is to expand the term $\mathbb{E}[zmS_2^2 P^{D-1}\overline{P^D}]$ in two different ways and compare the results. Using the resolvent identity (3.23) and Lemma 3.2, we get

$$
\begin{aligned}
\mathbb{E}\left[zmS_2^2 P^{D-1}\overline{P^D}\right] = {}&- \mathbb{E}\left[S_2^2 P^{D-1}\overline{P^D}\right] + \mathbb{E}\left[\frac{1}{N}\sum_{i_1 \neq i_2} H_{i_1 i_2} G_{i_2 i_1} S_2^2 P^{D-1}\overline{P^D}\right] \\
= {}&- \mathbb{E}\left[S_2^2 P^{D-1}\overline{P^D}\right] \\
&+ \sum_{r=1}^{\ell'} \frac{N\kappa_t^{(r+1)}}{r!}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2} \partial_{i_1 i_2}^r\left(G_{i_2 i_1} S_2^2 P^{D-1}\overline{P^D}\right)\right] \\
&+ \mathbb{E}\left[\Omega_{\ell'}\left(\frac{1}{N}\sum_{i_1 \neq i_2} H_{i_1 i_2} G_{i_2 i_1} S_2^2 P^{D-1}\overline{P^D}\right)\right],
\end{aligned}
$$

(6.65)

for arbitrary $\ell' \in \mathbb{N}$. Using the resolvent identity (3.23) once more, we write

$$zS_2 = \frac{1}{N}\sum_{i_1=1}^{N} zG_{i_1 i_1}G_{i_1 i_1} = -\frac{1}{N}\sum_{i_1=1}^{N} G_{i_1 i_1} + \frac{1}{N}\sum_{i_1 \neq i_2} H_{i_1 i_2} G_{i_2 i_1} G_{i_1 i_1}.$$

Thus, using Lemma 3.2, we also have

$$
\begin{aligned}
\mathbb{E}\left[zmS_2^2 P^{D-1}\overline{P^D}\right] = {}&- \mathbb{E}\left[m^2 S_2 P^{D-1}\overline{P^D}\right] \\
&+ \mathbb{E}\left[\frac{1}{N}\sum_{i_1 \neq i_2} H_{i_1 i_2} G_{i_2 i_1} G_{i_1 i_1} m S_2 P^{D-1}\overline{P^D}\right] \\
= {}&- \mathbb{E}[m^2 S_2 P^{D-1}\overline{P^D}] \\
&+ \sum_{r=1}^{\ell'} \frac{N\kappa_t^{(r+1)}}{r!}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1 \neq i_2} \partial_{i_1 i_2}^r\left(G_{i_2 i_1} G_{i_1 i_1} m S_2 P^{D-1}\overline{P^D}\right)\right] \\
&+ \mathbb{E}\left[\Omega_{\ell'}\left(\frac{1}{N}\sum_{i_1 \neq i_2} H_{i_1 i_2} G_{i_2 i_1} G_{i_1 i_1} m S_2 P^{D-1}\overline{P^D}\right)\right], \quad (6.66)
\end{aligned}
$$

for arbitrary $\ell' \in \mathbb{N}$. By Corollary 6.2 and Remark 6.3, the two error terms $\mathbb{E}[\Omega_{\ell'}(\cdot)]$ in (6.65) and (6.66) are negligible for $\ell' \geq 8D$.

With the extra factor $N\kappa_t^{(4)}$, we then write

$$N\kappa_t^{(4)}\mathbb{E}\left[zmS_2^2P^{D-1}\overline{P^D}\right] = -N\kappa_t^{(4)}\mathbb{E}\left[S_2^2P^{D-1}\overline{P^D}\right] + \sum_{r=1}^{\ell'}\sum_{s=0}^{r}w_{\widetilde{I}_{r,s}}\mathbb{E}\widetilde{I}_{r,s} + O(\Phi_\epsilon),$$

$$N\kappa_t^{(4)}\mathbb{E}\left[zmS_2^2P^{D-1}\overline{P^D}\right] = -N\kappa_t^{(4)}\mathbb{E}\left[m^2S_2P^{D-1}\overline{P^D}\right] + \sum_{r=1}^{\ell'}\sum_{s=0}^{r}w_{\widetilde{\widetilde{I}}_{r,s}}\mathbb{E}\widetilde{\widetilde{I}}_{r,s} + O(\Phi_\epsilon),$$

$$(6.67)$$

with

$$\widetilde{I}_{r,s} := N\kappa_t^{(4)}N\kappa_t^{(r+1)}\frac{1}{N^2}\sum_{i_1\neq i_2}\left(\partial_{i_1i_2}^{r-s}\left(G_{i_2i_1}S_2^2\right)\right)\left(\partial_{i_1i_2}^s\left(P^{D-1}\overline{P^D}\right)\right),$$

$$\widetilde{\widetilde{I}}_{r,s} := N\kappa_t^{(4)}N\kappa_t^{(r+1)}\frac{1}{N^2}\sum_{i_1\neq i_2}\left(\partial_{i_1i_2}^{r-s}\left(G_{i_2i_1}G_{i_1i_1}mS_2\right)\right)\left(\partial_{i_1i_2}^s\left(P^{D-1}\overline{P^D}\right)\right)$$

$$(6.68)$$

and $w_{\widetilde{I}_{r,s}} = w_{\widetilde{\widetilde{I}}_{r,s}} = (r!)^{-1}((r-s)!)^{-1}$.

For $r = 1$, $s = 0$, we find that

$$\mathbb{E}\widetilde{I}_{1,0} = -N\kappa_t^{(4)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}G_{i_1i_1}G_{i_2i_2}S_2^2P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon)$$

$$= -N\kappa_t^{(4)}\mathbb{E}\left[m^2S_2^2P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon), \qquad (6.69)$$

and similarly

$$\mathbb{E}\widetilde{\widetilde{I}}_{1,0} = -N\kappa_t^{(4)}\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}G_{i_1i_1}^2G_{i_2i_2}mS_2P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon)$$

$$= -N\kappa_t^{(4)}\mathbb{E}\left[m^2S_2^2P^{D-1}\overline{P^D}\right] + O(\Phi_\epsilon), \qquad (6.70)$$

where we used (6.61). We hence conclude that $\mathbb{E}\widetilde{I}_{1,0}$ equals $\mathbb{E}\widetilde{\widetilde{I}}_{1,0}$ up to negligible error.

Following the ideas in Sect. 6.4, we can bound

$$\left|\mathbb{E}\widetilde{I}_{1,1}\right| \leq \frac{N^\epsilon}{q_t^2}\left|\mathbb{E}\left[\frac{1}{N^2}\sum_{i_1\neq i_2}G_{i_1i_2}S_2^2(\partial_{i_1i_2}P^{D-1}\overline{P^D})\right]\right|$$

$$\leq \frac{N^\epsilon}{q_t^2}\mathbb{E}\left[\left(\frac{\mathrm{Im}\,m}{N\eta}\right)^{3/2}|P'||P|^{2D-2}\right] \leq \Phi_\epsilon,$$

and similarly $|\mathbb{E}\widetilde{\widetilde{I}}_{1,1}| \leq \Phi_\epsilon$, for N sufficiently large. In fact, for $r \geq 2$, $s \geq 0$ we can use, with small notational modifications the power counting outlined in Remark 6.8

and Remark 6.9 to conclude that

$$\mathbb{E}\widetilde{I}_{r,s} = O(\Phi_\epsilon), \qquad \mathbb{E}\widetilde{\widetilde{I}}_{r,s} = O(\Phi_\epsilon), \qquad (r \geq 2, s \geq 0).$$

Therefore the only non-negligible terms on the right hand side of (6.67) are $N\kappa_t^{(4)}\mathbb{E}[S_2^2 P^{D-1}\overline{PD}]$, $N\kappa_t^{(4)}\mathbb{E}[m^2 S_2 P^{D-1}\overline{PD}]$ as well as $\widetilde{I}_{1,0}$ and $\widetilde{\widetilde{I}}_{1,0}$. Since, by (6.69) and (6.70), the latter agree up to negligible error terms, we conclude that the former two must be equal up do negligible error terms. Thus (6.64) holds.

Next, expanding the term $\mathbb{E}[zm^3 S_2 P^{D-1}\overline{PD}]$ in two different ways similar to above, we further get

$$N\kappa_t^{(4)}\mathbb{E}\Big[m^2 S_2 P^{D-1}\overline{PD}\Big] = N\kappa_t^{(4)}\mathbb{E}\Big[m^4 P^{D-1}\overline{PD}\Big] + O(\Phi_\epsilon). \qquad (6.71)$$

Together with (6.64) this shows (6.63) and concludes the proof of the lemma. $\qquad\square$

Finally, from Lemmas 6.11 and 6.12, we conclude that

$$w_{I_{3,0}}\mathbb{E}[I_{3,0}] = -\frac{s_t^{(4)}}{q_t^2}\mathbb{E}\Big[m^4 P^{D-1}\overline{PD}\Big] + O(\Phi_\epsilon). \qquad (6.72)$$

This proves the third estimate in (6.20).

Proof of Lemma 6.4 The estimates in (6.20) were obtained in (6.23), (6.51), (6.52), (6.53) and (6.72). Estimate (6.21) follows from (6.25), (6.54), (6.58) and (6.59). $\quad\square$

7 Tracy–Widom limit: Proof of Theorem 2.10

In this section, we prove the Tracy–Widom limit of the extremal eigenvalues, Theorem 2.10, by following the strategy outlined in Sect. 3.4. In the first step, Proposition 7.1, we introduce a smooth cutoff to express the distribution of the largest eigenvalue of H as a functional of the imaginary part of the normalized trace of the Green function m of H. Then in Proposition 7.2, we compare the expectations of such functionals with respect to H and GOE, respectively.

To make the first step more precise in Proposition 7.1 below, we need some more notation. For $\eta > 0$, we set

$$\theta_\eta(y) := \frac{\eta}{\pi(y^2 + \eta^2)}, \qquad (y \in \mathbb{R}). \qquad (7.1)$$

Using the functional calculus and the definition of the Green function, we then get

$$\mathrm{Im}\, m(E + i\eta) = \frac{\pi}{N}\mathrm{Tr}\, \theta_\eta(H - E), \qquad (z = E + i\eta \in \mathbb{C}^+). \qquad (7.2)$$

We have the following proposition, which corresponds to Corollary 6.2 in [23], respectively to Lemma 6.5 in [16]. Recall the definition of L in (2.28).

Proposition 7.1 *Suppose that H satisfies Assumption 2.3 with $\phi > 1/6$. Denote by λ_1^H the largest eigenvalue of H. Fix $\epsilon > 0$. Let $E \in \mathbb{R}$ be such that $|E - L| \leq N^{-2/3+\epsilon}$. Let $E_+ := L + 2N^{-2/3+\epsilon}$ and define $\chi_E := \mathbb{1}_{[E,E_+]}$. Set $\eta_1 := N^{-2/3-3\epsilon}$ and $\eta_2 := N^{-2/3-9\epsilon}$. Let moreover $K : \mathbb{R} \to [0, \infty)$ be a smooth function satisfying*

$$K(x) = \begin{cases} 1 & \text{if } |x| < 1/3 \\ 0 & \text{if } |x| > 2/3 \end{cases}, \tag{7.3}$$

which is monotone decreasing on $[0, \infty)$. Then, for any $D > 0$,

$$\mathbb{E}\left[K\left(\text{Tr}(\chi_E * \theta_{\eta_2})(H) \right) \right] > \mathbb{P}(\lambda_1^H \leq E - \eta_1) - N^{-D} \tag{7.4}$$

and

$$\mathbb{E}\left[K\left(\text{Tr}(\chi_E * \theta_{\eta_2})(H) \right) \right] < \mathbb{P}(\lambda_1^H \leq E + \eta_1) + N^{-D}, \tag{7.5}$$

for N sufficiently large, with θ_{η_2} as in (7.1).

We postpone the proof of Proposition 7.1 to Sect. 7.1 and move on to the Green function comparison theorem. Let W^{GOE} be a GOE matrix independent of H with vanishing diagonal entries as introduced in Sect. 3.4 and denote by $m^{\text{GOE}} \equiv m^{W^{\text{GOE}}}$ the normalized trace of its Green function.

Proposition 7.2 (Green function comparison) *Let $\epsilon > 0$ and set $\eta_0 := N^{-2/3-\epsilon}$. Let $E_1, E_2 \in \mathbb{R}$ satisfy $|E_1|, |E_2| \leq N^{-2/3+\epsilon}$. Let $F : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying*

$$\max_{x \in \mathbb{R}} |F^{(l)}(x)|(|x| + 1)^{-C} \leq C, \qquad (l \in [\![1, 11]\!]). \tag{7.6}$$

Then, for any sufficiently small $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left| \mathbb{E}F\left(N \int_{E_1}^{E_2} \text{Im } m(x + L + i\eta_0) \, dx \right) - \mathbb{E}F\left(N \int_{E_1}^{E_2} \text{Im } m^{\text{GOE}}(x + 2 + i\eta_0) \, dx \right) \right| \leq N^{-\delta}, \tag{7.7}$$

for sufficiently large N, where L is given in (2.28).

Proposition 7.2 is proved in Sect. 7.2. We are ready to prove Theorem 2.10.

Proof of Theorem 2.10 Fix $\epsilon > 0$ and set $\eta_1 := N^{-2/3-3\epsilon}$ and $\eta_2 := N^{-2/3-9\epsilon}$. Consider $E = L + sN^{-2/3}$ with $s \in (-N^{-2/3+\epsilon}, N^{-2/3+\epsilon})$. For any $D > 0$, we find from Proposition 7.1 that

$$\mathbb{P}(\lambda_1^H \leq E) < \mathbb{E}K\left(\text{Tr}(\chi_{E+\eta_1} * \theta_{\eta_2})(H) \right) + N^{-D},$$

for N sufficiently large. Applying Proposition 7.2 with 9ϵ instead of ϵ and setting $E_1 := E - L + \eta_1$, $E_2 := E_+ - L$, we find that there is $\delta > 0$ such that

$$\mathbb{E}K\left(\mathrm{Tr}(\chi_{E+\eta_1} * \theta_{\eta_2})(H)\right) = \mathbb{E}K\left(\frac{N}{\pi}\int_{E_1}^{E_2} \mathrm{Im}\, m(x + L + i\eta_2)\,\mathrm{d}x\right)$$

$$\leq \mathbb{E}K\left(\frac{N}{\pi}\int_{E_1}^{E_2} \mathrm{Im}\, m^{\mathrm{GOE}}(x + 2 + i\eta_2)\,\mathrm{d}x\right) + N^{-\delta}$$

$$= \mathbb{E}K\left(\mathrm{Tr}(\chi_{[E_1+2, E_2+2]} * \theta_{\eta_2})(W^{\mathrm{GOE}})\right) + N^{-\delta},$$

for N sufficiently large. Hence, applying Proposition 7.1 to the matrix W^{GOE}, we get

$$\mathbb{P}\left(N^{2/3}(\lambda_1^H - L) \leq s\right) = \mathbb{P}\left(\lambda_1^H \leq E\right)$$

$$< \mathbb{P}\left(\lambda_1^{\mathrm{GOE}} \leq E_1 + 2 + \eta_1\right) + N^{-D} + N^{-\delta}$$

$$= \mathbb{P}\left(N^{2/3}(\lambda_1^{\mathrm{GOE}} - 2) \leq s + 2N^{-3\epsilon}\right) + N^{-D} + N^{-\delta}. \qquad (7.8)$$

Similarly, we can also check that

$$\mathbb{P}\left(N^{2/3}(\lambda_1^H - L) \leq s\right) > \mathbb{P}\left(N^{2/3}(\lambda_1^{\mathrm{GOE}} - 2) \leq s - 2N^{-3\epsilon}\right) - N^{-D} - N^{-\delta}. \qquad (7.9)$$

Since the right sides of Eqs. (7.8) and (7.9) converge both to $F_1(s)$, the GOE Tracy–Widom distribution, as N tends to infinity we conclude that

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\lambda_1^H - L) \leq s\right) = F_1(s). \qquad (7.10)$$

This proves Theorem 2.10.

In the rest of this section, we prove Propositions 7.1 and 7.2

7.1 Proof of Proposition 7.1

In principle, we could adopt the strategy in the proof of Corollary 6.2 of [23] after proving the optimal rigidity estimate at the edge with the assumption $q \ggg N^{1/6}$ and checking that such an optimal bound is required only for the eigenvalues at the edge. However, we introduce a slightly different approach that directly compares $\mathrm{Tr}(\chi_E * \theta_{\eta_2})(H)$ and $\mathrm{Tr}\, \chi_{E-\eta_1}(H)$ by using the local law.

Proof of Proposition 7.1 For an interval $I \subset \mathbb{R}$, let $\mathcal{N}_I$ be the number of the eigenvalues in I, *i.e.*

$$\mathcal{N}_I := |\{\lambda_i : \lambda_i \in I\}|. \qquad (7.11)$$

We compare $\mathrm{Tr}(\chi_E * \theta_{\eta_2})(H)$ and $\mathrm{Tr}\,\chi_{E-\eta_1}(H)$ by considering the following four cases:

Case 1 If $x \in [E + \eta_1, E_+ - \eta_1)$, then $\chi_{E-\eta_1}(x) = 1$ and

$$
\begin{aligned}
(\chi_E * \theta_{\eta_2})(x) - 1 &= \frac{1}{\pi} \int_E^{E_+} \frac{\eta_2}{(x-y)^2 + \eta_2^2} \, dy - 1 \\
&= \frac{1}{\pi} \left(\tan^{-1} \frac{E_+ - x}{\eta_2} - \tan^{-1} \frac{E - x}{\eta_2} \right) - 1 \\
&= -\frac{1}{\pi} \left(\tan^{-1} \frac{\eta_2}{E_+ - x} + \tan^{-1} \frac{\eta_2}{x - E} \right) = O\left(\frac{\eta_2}{\eta_1} \right) = O(N^{-6\epsilon}).
\end{aligned}
\tag{7.12}
$$

For any $E' \in [E + \eta_1, E_+ - \eta_1)$, with the local law, Proposition 3.3, we can easily see that

$$
\frac{\mathcal{N}_{[E'-\eta_1, E'+\eta_1)}}{5N\eta_1} \leq \frac{1}{N} \sum_{i=1}^N \frac{\eta_1}{|\lambda_i - E'|^2 + \eta_1^2} = \mathrm{Im}\, m(E' + i\eta_1) \leq \frac{N^{\epsilon/2}}{N\eta_1} + \frac{N^{\epsilon/2}}{q^2},
\tag{7.13}
$$

with high probability, where we used $\mathrm{Im}\,\widetilde{m}(E' + i\eta_1) \leq C\sqrt{\varkappa(E') + \eta_1} \ll N^{\epsilon/2}/(N\eta_1)$. Thus, considering at most $[E_+ - (L - N^{-2/3+\epsilon})]/\eta_1 = 3N^{4\epsilon}$ intervals, we find that

$$
\mathcal{N}_{[E+\eta_1, E_+-\eta_1)} \leq CN^{9\epsilon/2}
\tag{7.14}
$$

and

$$
\sum_{i:E+\eta_1 < \lambda_i < E_+-\eta_1} \left((\chi_E * \theta_{\eta_2})(\lambda_i) - \chi_{E-\eta_1}(\lambda_i) \right) \leq N^{-\epsilon},
\tag{7.15}
$$

with high probability.

Case 2 For $x < E - \eta_1$, choose $k \geq 0$ such that $3^k \eta_1 \leq E - x < 3^{k+1}\eta_1$. Then, $\chi_{E-\eta_1}(x) = 0$ and

$$
\begin{aligned}
(\chi_E * \theta_{\eta_2})(x) &= \frac{1}{\pi} \left(\tan^{-1} \frac{E_+ - x}{\eta_2} - \tan^{-1} \frac{E - x}{\eta_2} \right) \\
&= \frac{1}{\pi} \left(\tan^{-1} \frac{\eta_2}{E - x} - \tan^{-1} \frac{\eta_2}{E_+ - x} \right) \\
&< \frac{1}{2} \left(\frac{\eta_2}{E - x} - \frac{\eta_2}{E_+ - x} \right) \\
&= \frac{1}{2} \frac{\eta_2(E_+ - E)}{(E - x)(E_+ - x)} < 2N^{-4/3-8\epsilon} \cdot 3^{-2k}\eta_1^{-2}.
\end{aligned}
\tag{7.16}
$$

Abbreviate $\mathcal{N}_k = \mathcal{N}_{(E-3^{k+1}\eta_1, E-3^k\eta_1]}$. Consider

$$\operatorname{Im} m(E - 2 \cdot 3^k\eta_1 + \mathrm{i} \cdot 3^k\eta_1) = \frac{1}{N}\sum_{i=1}^{N} \frac{3^k\eta_1}{|\lambda_i - (E - 2 \cdot 3^k\eta_1)|^2 + (3^k\eta_1)^2} > \frac{1}{N}\frac{\mathcal{N}_k}{2 \cdot 3^k\eta_1}.$$

(7.17)

With the local law, Proposition 3.3, and the estimate $\operatorname{Im}\widetilde{m}(x + \mathrm{i}y) \asymp \sqrt{|x - L| + y}$, we find that

$$\operatorname{Im} m(E - 2 \cdot 3^k\eta_1 + \mathrm{i} \cdot 3^k\eta_1) \leq C\sqrt{3^k\eta_1} + \frac{N^{\epsilon/2}}{N \cdot 3^k\eta_1} \leq N^{5\epsilon}\sqrt{3^k\eta_1}$$

(7.18)

and hence

$$\frac{1}{N}\frac{\mathcal{N}_k}{2 \cdot 3^k\eta_1} < N^{5\epsilon}\sqrt{3^k\eta_1},$$

(7.19)

with high probability. Thus, with high probability,

$$\sum_{i:\lambda_i < E-\eta_1} \left((\chi_E * \theta_{\eta_2})(\lambda_i) - \chi_{E-\eta_1}(\lambda_i)\right) \leq 2\sum_{k=0}^{2\log N} N^{-4/3-8\epsilon} \cdot 3^{-2k}\eta_1^{-2}\mathcal{N}_k$$

$$\leq 4N^{-1/3-3\epsilon}\eta_1^{-1/2}\sum_{k=0}^{\infty} 3^{-k/2} \leq 10N^{-3\epsilon/2}.$$

(7.20)

Case 3 By Proposition 2.9 there are with high probability no eigenvalues in $[E_+ - \eta_1, \infty)$.

Case 4 For $x \in [E - \eta_1, E + \eta_1)$, we use the trivial estimate

$$(\chi_E * \theta_{\eta_2})(x) < 1 = \chi_{E-\eta_1}(x).$$

(7.21)

Combining the four cases above, we find that

$$\operatorname{Tr}(\chi_E * \theta_{\eta_2})(H) \leq \operatorname{Tr}\chi_{E-\eta_1}(H) + N^{-\epsilon},$$

(7.22)

with high probability. From the definition of the cutoff K and the fact that $\operatorname{Tr}\chi_{E-\eta_1}(H)$ is an integer,

$$K(\operatorname{Tr}\chi_{E-\eta_1}(H) + N^{-\epsilon}) = K(\mathcal{N}_{[E-\eta_1, E_+]}).$$

Thus, since K is monotone decreasing on $[0, \infty)$, (7.22) implies that

$$K(\operatorname{Tr}(\chi_E * \theta_{\eta_2})(H)) \geq K(\operatorname{Tr}\chi_{E-\eta_1}(H)),$$

with high probability. Taking the expectation, we get for any fixed $D > 0$ that

$$\mathbb{E}\left[K\left(\mathrm{Tr}(\chi_E * \theta_{\eta_2})(H)\right)\right] > \mathbb{P}(\lambda_1 \leq E - \eta_1) - N^{-D},$$

for N sufficiently large. This proves the first part of Proposition 7.1. The second part can also be proved in a similar manner by showing that

$$\mathrm{Tr}(\chi_E * \theta_{\eta_2})(H) \geq \mathrm{Tr}\, \chi_{E+\eta_1}(H) - N^{-\epsilon},\tag{7.23}$$

with high probability, applying the cutoff K and taking the expectation. In this argument (7.21) gets replaced by

$$\chi_{E+\eta_1}(x) = 0 < (\chi_E * \theta_{\eta_2})(x),\tag{7.24}$$

for $x \in [E - \eta_1, E + \eta_1)$. This proves Proposition 7.1. $\qquad\qquad\square$

7.2 Green function comparison: Proof of Proposition 7.2

Recall the definition of the matrix H_t in (3.26). We first state a lemma that has the analogue role of Lemma 3.2 in calculations involving $H \equiv H_t$ and $\dot{H} \equiv \mathrm{d}H_t/\mathrm{d}t$.

Lemma 7.3 *Fix $\ell \in \mathbb{N}$ and let $F \in C^{\ell+1}(\mathbb{R}; \mathbb{C}^+)$. Let $Y \equiv Y_0$ be a random variable with finite moments to order $\ell + 2$ and let W be a Gaussian random variable independent of Y. Assume that $\mathbb{E}[Y] = \mathbb{E}[W] = 0$ and $\mathbb{E}[Y^2] = \mathbb{E}[W^2]$. Define*

$$Y_t := \mathrm{e}^{-t/2}Y_0 + \sqrt{1 - \mathrm{e}^{-t}}\,W,\tag{7.25}$$

and abbreviate $\dot{Y}_t \equiv \frac{\mathrm{d}Y_t}{\mathrm{d}t}$. Then,

$$\mathbb{E}\left[\dot{Y}_t F(Y_t)\right] = -\frac{1}{2}\sum_{r=2}^{\ell} \frac{\kappa^{(r+1)}(Y_0)}{r!}\mathrm{e}^{-\frac{(r+1)t}{2}}\mathbb{E}\left[F^{(r)}(Y_t)\right] + \mathbb{E}\left[\Omega_\ell(\dot{Y}_t F(Y_t))\right],\tag{7.26}$$

where $\mathbb{E}$ denotes the expectation with respect to Y and W, $\kappa^{(r+1)}(Y)$ denotes the $(r + 1)$-th cumulant of Y and $F^{(r)}$ denotes the r-th derivative of the function F. The error term Ω_ℓ in (7.26) satisfies

$$\left|\mathbb{E}\left[\Omega_\ell(\dot{Y}F(Y_t))\right]\right| \leq C_\ell \mathbb{E}[|Y_t|^{\ell+2}]\sup_{|x|\leq Q}|F^{(\ell+1)}(x)|$$
$$+ C_\ell \mathbb{E}\left[|Y_t|^{\ell+2}\mathbb{1}(|Y_t| > Q)\right]\sup_{x\in\mathbb{R}}|F^{(\ell+1)}(x)|,\tag{7.27}$$

where $Q \geq 0$ is an arbitrary fixed cutoff and C_ℓ satisfies $C_\ell \leq \frac{(C\ell)^\ell}{\ell!}$ for some numerical constant C.

Proof We follow the proof of Corollary 3.1 in [41]. First, note that

$$\dot{Y}_t = -\frac{e^{-t/2}}{2} Y + \frac{e^{-t}}{2\sqrt{1 - e^{-t}}} W \,. \tag{7.28}$$

Thus,

$$\mathbb{E}\left[\dot{Y}_t F(Y_t) \right] = -\frac{e^{-t/2}}{2} \mathbb{E}\left[Y F \left(e^{-t/2} Y + \sqrt{1 - e^{-t}} W \right) \right]$$
$$+ \frac{e^{-t}}{2\sqrt{1 - e^{-t}}} \mathbb{E}\left[W F \left(e^{-t/2} Y + \sqrt{1 - e^{-t}} W \right) \right].$$

Applying Lemma 3.2 and (3.9), we get (7.26) since the first two moments of W and Y agree. $\qquad \square$

Proof of Proposition 7.2 Fix a (small) $\epsilon > 0$. Consider $x \in [E_1, E_2]$. Recall the definition of H_t in (3.26). For simplicity, let

$$G \equiv G_t(x + L_t + i\eta_0), \qquad m \equiv m_t(x + L_t + i\eta_0), \tag{7.29}$$

with $\eta_0 = N^{-2/3 - \epsilon}$, and define

$$X \equiv X_t := N \int_{E_1}^{E_2} \operatorname{Im} m(x + L_t + i\eta_0) \, dx \,. \tag{7.30}$$

Note that $X \prec N^\epsilon$ and $|F^{(l)}(X)| \prec N^{C\epsilon}$, for $l \in [\![1, 11]\!]$. Recall from (3.32) that

$$L = 2 + e^{-t} q_t^{-2} s^{(4)} + O(e^{-2t} q_t^{-4}), \qquad \dot{L} = -2e^{-t} q_t^{-2} s^{(4)} + O(e^{-2t} q_t^{-4}), \tag{7.31}$$

where $q_t = e^{t/2} q_0$. Let $z := x + L + i\eta_0$ and $G \equiv G(z)$. Differentiating $F(X)$ with respect to t, we get

$$\frac{d}{dt} \mathbb{E} F(X) = \mathbb{E}\left[F'(X) \frac{dX}{dt} \right] = \mathbb{E}\left[F'(X) \operatorname{Im} \int_{E_1}^{E_2} \sum_{i_1=1}^{N} \frac{dG_{i_1 i_1}}{dt} \, dx \right]$$
$$= \mathbb{E}\left[F'(X) \operatorname{Im} \int_{E_1}^{E_2} \left(\sum_{i_1=1}^{N} \sum_{i_2 \le i_3} \dot{H}_{i_2 i_3} \frac{\partial G_{i_1 i_1}}{\partial H_{i_2 i_3}} + \dot{L} \sum_{i_1, i_2} G_{i_1 i_2} G_{i_2 i_1} \right) dx \right], \tag{7.32}$$

where by definition

$$\dot{H}_{i_2 i_3} \equiv (\dot{H}_t)_{i_2 i_3} = -\frac{1}{2} e^{-t/2} (H_0)_{i_2 i_3} + \frac{e^{-t}}{2\sqrt{1 - e^{-t}}} W_{i_2 i_3}^{\mathrm{GOE}} \,. \tag{7.33}$$

Thus, from Lemma 7.3, we find that

$$
\sum_{i_1=1}^{N} \sum_{i_2 \leq i_3} \mathbb{E}\left[\dot{H}_{i_2 i_3} F'(X) \frac{\partial G_{i_1 i_1}}{\partial H_{i_2 i_3}}\right] = -\sum_{i_1, i_2, i_3} \mathbb{E}\left[\dot{H}_{i_2 i_3} F'(X) G_{i_1 i_2} G_{i_3 i_1}\right]
$$

$$
= \frac{e^{-t}}{2N} \sum_{r=2}^{\ell} \frac{q_t^{-(r-1)} s^{(r+1)}}{r!} \sum_{i_1, i_2, i_3} \mathbb{E}\left[\partial_{i_2 i_3}^r \left(F'(X) G_{i_1 i_2} G_{i_3 i_1}\right)\right] + O(N^{1/3 + C\epsilon}),
\tag{7.34}
$$

for $\ell = 10$, where we abbreviate $\partial_{i_2 i_3} \equiv \partial/(\partial H_{i_2 i_3})$. Note that the $O(N^{1/3 + C\epsilon})$ error term in (7.34) originates from Ω_ℓ in (7.26), which is $O(N^{C\epsilon} N^2 q_t^{-10})$ for the choice $Y = H_{i_2 i_3}$.

We claim the following lemma.

Lemma 7.4 *Let $z := x + L + i\eta_0$ and $G \equiv G(z)$. For any integer $r \geq 2$, set*

$$
J_r := \frac{e^{-t}}{2N} \frac{q_t^{-(r-1)} s^{(r+1)}}{r!} \sum_{i_1, i_2, i_3} \partial_{i_2 i_3}^r \left(F'(X) G_{i_1 i_2} G_{i_3 i_1}\right).
\tag{7.35}
$$

Then,

$$
\mathbb{E} J_3 = 2e^{-t} s^{(4)} q_t^{-2} \sum_{i_1, i_2} \mathbb{E}\left[F'(X) G_{i_1 i_2} G_{i_2 i_1}\right] + O(N^{2/3 - \epsilon'})
\tag{7.36}
$$

and, for $r = 2$ and for $r \geq 4$,

$$
\mathbb{E} J_r = O(N^{2/3 - \epsilon'}).
\tag{7.37}
$$

Assuming that Lemma 7.4 holds, we obtain that there exists $\epsilon' > 2\epsilon$ such that, for all $t \in [0, 6 \log N]$,

$$
\sum_{i_1} \sum_{i_2 \leq i_3} \mathbb{E}\left[\dot{H}_{i_2 i_3} F'(X) \frac{\partial G_{i_1 i_1}}{\partial H_{i_2 i_3}}\right] = -\dot{L} \sum_{i_1, i_2} \mathbb{E}\left[G_{i_1 i_2} G_{i_2 i_1} F'(X)\right] + O(N^{2/3 - \epsilon'}),
\tag{7.38}
$$

which implies that the right side of (7.32) is $O(N^{-\epsilon'/2})$. Integrating from $t = 0$ to $t = 6 \log N$, we get

$$
\left| \mathbb{E} F\left(N \int_{E_1}^{E_2} \operatorname{Im} m(x + L + i\eta_0)\, dx\right)_{t=0} \right.
$$
$$
\left. - \mathbb{E} F\left(N \int_{E_1}^{E_2} \operatorname{Im} m(x + L + i\eta_0)\, dx\right)_{t=6 \log N} \right| \leq N^{-\epsilon'/4}.
$$

Comparing $\operatorname{Im} m|_{t=6\log N}$ and $\operatorname{Im} m^{\mathrm{GOE}}$ is trivial; if we let $\lambda_k(6\log N)$ be the k-th largest eigenvalue of $H_{6\log N}$ and λ_k^{GOE} the k-th largest eigenvalue of W^{GOE}, then $|\lambda_k(6\log N) - \lambda_k^{\mathrm{GOE}}| \prec N^{-3}$, hence

$$\left|\operatorname{Im} m|_{t=6\log N} - \operatorname{Im} m^{\mathrm{GOE}}\right| \prec N^{-5/3} . \tag{7.39}$$

This proves Proposition 7.2. □

It remains to prove Lemma 7.4. The proof uses quite similar ideas as used in Sect. 6. We thus sometimes omit some details and refer to the corresponding paragraph in Sect. 6.

Proof of Lemma 7.4 First, we note that in the definition of J_r we may freely include or exclude terms with $i_1 = i_2$ or $i_1 = i_3$ in the summation $\sum_{i_1,i_2,i_3}$, since they contain at least one off-diagonal Green function entry, $G_{i_1 i_3}$ or $G_{i_1 i_2}$, and the sizes of these terms are, in expectation, at most of order

$$q_t^{-1} q_t^{-1} N^2 N^{-1} \lll N^{2/3-\epsilon'} \mathrm{e}^{-t} ,$$

for any sufficiently small $\epsilon' > 0$. (When $i_2 = i_3$ the summand is zero by construction.)

For $r \geq 5$, it is easy to see that $\mathbb{E} J_r = O(N^{2/3-\epsilon'})$, since it contains at least two off-diagonal entries in $\partial_{i_2 i_3}^r \left(F'(X) G_{i_1 i_2} G_{i_3 i_1} \right)$ and $|\mathbb{E} J_r|$ is bounded by

$$q_t^{-4} N^3 N^{-1} N^{-2/3+2\epsilon} \lll N^{2/3-\epsilon'} \mathrm{e}^{-2t} ,$$

which can be checked using Lemma 6.5 and a simple power counting.

Therefore, we only need to consider the cases $r = 2, 3, 4$. In the following subsections, we check each case and complete the proof of Lemma 7.4. See Eqs. (7.48), (7.52) and (7.53) below.

7.2.1 Proof of Lemma 7.4 for $r = 2$

We proceed as in Lemma 6.7 of Sect. 6.4 and apply the idea of an unmatched index. Observe that

$$\partial_{i_2 i_3}^2 \left(F'(X) G_{i_1 i_2} G_{i_2 i_1} \right) = F'(X) \partial_{i_2 i_3}^2 (G_{i_1 i_2} G_{i_3 i_1}) + 2\partial_{i_2 i_3} F'(X) \partial_{i_2 i_3} (G_{i_1 i_2} G_{i_3 i_1})$$
$$+ (\partial_{i_2 i_3}^2 F'(X)) G_{i_1 i_2} G_{i_3 i_1} . \tag{7.40}$$

We first consider the expansion of $F'(X) \partial_{i_2 i_3}^2 (G_{i_1 i_2} G_{i_3 i_1})$. We can easily estimate the terms with four off-diagonal Green function entries, since, for example,

$$\sum_{i_1,i_2,i_3} \left| \mathbb{E} \left[F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_2} G_{i_2 i_1} \right] \right| \leq N^{C\epsilon} \sum_{i_1,i_2,i_3} \mathbb{E} |G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_2} G_{i_3 i_1}|$$
$$\leq N^{C\epsilon} N^3 \left(\frac{\operatorname{Im} m}{N\eta_0} \right)^2 \leq N^{3-4/3+C\epsilon} ,$$

where we used Lemma 6.5. Thus, for sufficiently small ϵ and ϵ',

$$\frac{e^{-t}q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \left| \mathbb{E}\left[F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_2} G_{i_3 i_1} \right] \right| \lll N^{2/3-\epsilon'}. \qquad (7.41)$$

For the terms with three off-diagonal Green function entries, the bound we get from Lemma 6.5 is

$$q_t^{-1} N^{-1} N^3 N^{C\epsilon} \left(\frac{\operatorname{Im} m}{N \eta_0} \right)^{3/2} \asymp q_t^{-1} N^{1+C\epsilon},$$

which is not sufficiently small. To gain an additional factor of q_t^{-1}, making the above bound $q_t^{-2} N^{1+C\epsilon} \lll N^{2/3-\epsilon'}$, we use Lemma 3.2 to expand in an unmatched index. For example, such a term is of the form

$$G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1}$$

and we focus on the unmatched index i_3 in $G_{i_3 i_2}$. Then, multiplying by z and expanding, we get

$$\frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[z F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$

$$= \frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3,i_4} \mathbb{E}\left[F'(X) G_{i_1 i_2} H_{i_3 i_4} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$

$$= \frac{q_t^{-1}}{N} \sum_{r'=1}^{\ell} \frac{\kappa^{(r'+1)}}{r'!} \sum_{i_1,i_2,i_3,i_4} \mathbb{E}\left[\partial_{i_3 i_4}^{r'} \left(F'(X) G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1} \right) \right] + O(N^{2/3-\epsilon'}),$$

for $\ell = 10$.

For $r' = 1$, we need to consider $\partial_{i_3 i_4}(F'(X) G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1})$. When $\partial_{i_3 i_4}$ acts on $F'(X)$ it creates a fresh summation index i_5 and we get a term

$$\frac{q_t^{-1}}{N^2} \sum_{i_1,i_2,i_3,i_4} \mathbb{E}\left[\left(\partial_{i_3 i_4} F'(X) \right) G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$

$$= -\frac{2q_t^{-1}}{N^2} \int_{E_1}^{E_2} \sum_{i_1,i_2,i_3,i_4,i_5} \mathbb{E}\left[G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1} F''(X) \operatorname{Im} \left(\widetilde{G}_{i_3 i_5} \widetilde{G}_{i_5 i_4} \right) \right] \mathrm{d}y,$$

$$(7.42)$$

where we abbreviate $\widetilde{G} \equiv G(y + L + i\eta_0)$. Applying Lemma 6.5 to the index a and $\widetilde{G}$, we get,

$$\frac{1}{N} \sum_{i_5=1}^{N} \left| \widetilde{G}_{i_3 i_5} \widetilde{G}_{i_5 i_4} \right| \prec N^{-2/3+2\epsilon},$$

for any i_3 and i_4, which also shows that

$$|\partial_{i_3 i_4} F'(X)| \prec N^{-1/3+C\epsilon} . \tag{7.43}$$

Applying Lemma 6.5 to the remaining off-diagonal Green function entries, we obtain from (7.42) that

$$\frac{q_t^{-1}}{N^2} \sum_{i_1,i_2,i_3,i_4} \left| \mathbb{E}[(\partial_{i_3 i_4} F'(X)) G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1}] \right|$$
$$\leq q_t^{-1} N^2 N^{-1/3+C\epsilon} N^{-1+3\epsilon} = q_t^{-1} N^{2/3+C\epsilon} . \tag{7.44}$$

When $\partial_{i_3 i_4}$ acts on $G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1}$, then we always get four or more off-diagonal Green function entries with the only exception being

$$-G_{i_1 i_2} G_{i_4 i_4} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} .$$

To the terms with four or more off-diagonal Green function entries, we apply Lemma 6.5 and obtain a bound similar to (7.44) by power counting. For the term of the exception, we rewrite it as

$$-\frac{q_t^{-1}}{N^2} \sum_{i_1,i_2,i_3,i_4} \mathbb{E}\left[F'(X) G_{i_1 i_2} G_{i_4 i_4} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$
$$= -\frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[m F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$
$$= -\widetilde{m} \frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$
$$+ \frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[(\widetilde{m} - m) F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right] .$$

Here, the last term is again bounded by $q_t^{-1} N^{2/3+C\epsilon}$ as we can easily check with Proposition 3.3 and Lemma 6.5. We thus arrive at

$$\frac{q_t^{-1}}{N} (z + \widetilde{m}) \sum_{i_1,i_2,i_3} \mathbb{E}\left[F'(X) G_{i_1 i_2} G_{i_3 i_2} G_{i_3 i_3} G_{i_2 i_1} \right]$$
$$= \frac{q_t^{-1}}{N} \sum_{r'=2}^{\ell} \frac{\kappa^{(r'+1)}}{r'!} \sum_{i_1,i_2,i_3,i_4} \mathbb{E}\left[\partial_{i_3 i_4}^{r'} \left(F'(X) G_{i_1 i_2} G_{i_4 i_2} G_{i_3 i_3} G_{i_2 i_1} \right) \right] + O(N^{2/3-\epsilon}) . \tag{7.45}$$

On the right side, the summation starts from $r' = 2$, hence we have gained a factor $(N q_t)^{-1}$ from $\kappa^{(r'+1)}$ and added the fresh summation index i_4, so the net gain is q_t^{-1}. Since $|z + \widetilde{m}| \asymp 1$, this shows that

$$\frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[F'(X)G_{i_1i_2}G_{i_3i_2}G_{i_3i_3}G_{i_2i_1}\right] = O(N^{2/3-\epsilon'}). \qquad (7.46)$$

Together with (7.41), this takes care of the first term on the right side of (7.40).

For the second term on the right side of (7.40), we focus on

$$\partial_{i_2i_3} F'(X) = -\int_{E_1}^{E_2} \sum_{i_4=1}^{N} \left[F''(X)\,\mathrm{Im}\left(\widetilde{G}_{i_2i_4}\widetilde{G}_{i_4i_3}\right)\right] dy \qquad (7.47)$$

and apply the same argument to the unmatched index i_3 in $\widetilde{G}_{i_3i_4}$.

For the third term, we focus on the factor $G_{i_1i_2}G_{i_3i_1}$ and apply the same argument once more with the unmatched index i_3 in $G_{i_3i_1}$. We omit the detail.

Estimating all terms in this way, we eventually get the desired estimate

$$\frac{q_t^{-1}}{N} \sum_{i_1,i_2,i_3} \left|\mathbb{E}\left[\partial_{i_2i_3}^2\left(F'(X)\partial_{i_2i_3}G_{i_1i_1}\right)\right]\right| = O(N^{2/3-\epsilon'}). \qquad (7.48)$$

7.2.2 Proof of Lemma 7.4 for $r = 3$

We proceed as in Sect. 6.7. Note that there will be no unmatched indices in the Green function entries for this case. If $\partial_{i_2i_3}$ acts on $F'(X)$ at least once, then that term is bounded by

$$q_t^{-2}N^{\epsilon}N^{-1}N^3N^{-1/3+C\epsilon}N^{-2/3+2\epsilon} = q_t^{-2}N^{1+C\epsilon} \lll N^{2/3-\epsilon'},$$

where we used (7.43) and the fact that $G_{i_1i_2}G_{i_3i_1}$ or $\partial_{i_2i_3}(G_{i_1i_2}G_{i_3i_1})$ contains at least two off-diagonal entries. Moreover, in the expansion of $\partial_{i_2i_3}^2(G_{i_1i_2}G_{i_3i_1})$, the terms with three or more off-diagonal Green function entries entries can be bounded by

$$q_t^{-2}N^{\epsilon}N^{-1}N^3N^{C\epsilon}N^{-1+3\epsilon} = q_t^{-2}N^{1+C\epsilon} \lll N^{2/3-\epsilon'}.$$

Thus,

$$\begin{aligned}
\mathbb{E}J_3 &= \frac{\mathrm{e}^{-t}}{2N}\frac{q_t^{-2}s^{(4)}}{3!} \sum_{i_1,i_2,i_3} \mathbb{E}\left[\partial_{i_2i_3}^3\left(F'(X)G_{i_1i_2}G_{i_3i_1}\right)\right] \\
&= -\frac{4!}{2}\frac{\mathrm{e}^{-t}q_t^{-2}s^{(4)}}{3!} \sum_{i_1,i_2} \mathbb{E}\left[F'(X)G_{i_1i_2}G_{i_2i_2}G_{i_2i_1}S_2\right] + O(N^{2/3-\epsilon'}),
\end{aligned} \qquad (7.49)$$

where the combinatorial factor 4! is computed as in Lemma 6.11 and S_2 is as in (6.61). As in Lemma 6.12 of Sect. 6.7, the first term on right side of (7.49) is computed by expanding

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[zm\, S_2 F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right]$$

in two different ways. We then obtain that

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} S_2 \right] = q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[m^2 F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right] + O(N^{2/3-\epsilon'})$$

$$= q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right] + O(N^{2/3-\epsilon'}),$$

$$(7.50)$$

where we used that $m \equiv m(z) = -1 + O(N^{-1/3+\epsilon})$ with high probability. Indeed, since $\widetilde{m}(L)$, which was denoted by τ in the proof of Lemma 4.1, satisfies $\widetilde{m}(L) = -1 + O(e^{-t}q_t^{-2})$ by (4.6), and since $|\widetilde{m}(z) - \widetilde{m}(L)| \asymp \sqrt{\varkappa + \eta_0} \le N^{-1/3+\epsilon}$ by (4.8), we have from Proposition 3.3 that $m(z) = -1 + O(N^{-1/3+\epsilon})$ with high probability.

Consider next the first term on the very right of (7.50). Since $|z - 2| \le CN^{-1/3}$, we get

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[zF'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right] = 2q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right] + O(N^{2/3-\epsilon'}),$$

$$(7.51)$$

where we used (7.43). Expanding on left hand side the factor $G_{i_2i_2}$ using (3.23), we obtain

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}[zF'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1}] = -q_t^{-2} \sum_{i_1,i_2} \mathbb{E}[F'(X) G_{i_1i_2} G_{i_2i_1}]$$

$$+ q_t^{-2} \sum_{i_1,i_2,i_3} \mathbb{E}[F'(X) H_{i_2i_3} G_{i_3i_2} G_{i_1i_2} G_{i_2i_1}].$$

Applying Lemma 3.2 to the second term on the right side, we find that most of the terms are $O(N^{2/3-\epsilon'})$ either due to three (or more) off-diagonal entries, the partial derivative $\partial_{i_2i_3}$ acting on $F'(X)$, or higher cumulants. The only term that does not fall into one these categories is

$$-\frac{q_t^{-2}}{N} \sum_{i_1,i_2,i_3} \mathbb{E}\left[F'(X) G_{i_1i_2} G_{i_3i_3} G_{i_2i_2} G_{i_2i_1} \right],$$

which is generated when $\partial_{i_2i_3}$ acts on $G_{i_3i_2}$. From this argument, we find that

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[zF'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right]$$

$$= -q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X) G_{i_1i_2} G_{i_2i_1} \right] - q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[m F'(X) G_{i_1i_2} G_{i_2i_2} G_{i_2i_1} \right] + O(N^{2/3-\epsilon'}).$$

Hence, combining it with (7.51) and the fact that $m = -1 + O(N^{-1/3+\epsilon})$ with high probability, we get

$$q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X)G_{i_1i_2}G_{i_2i_2}G_{i_2i_1}\right] = -q_t^{-2} \sum_{i_1,i_2} \mathbb{E}\left[F'(X)G_{i_1i_2}G_{i_2i_1}\right] + O(N^{2/3-\epsilon'}).$$

In combination with (7.49) and (7.50), we conclude that

$$\begin{aligned} \mathbb{E}J_3 &= \frac{\mathrm{e}^{-t}}{2N}\frac{q_t^{-2}s^{(4)}}{3!} \sum_{i_1,i_2,i_3} \mathbb{E}\left[\partial_{i_2i_3}^3\left(F'(X)G_{i_1i_2}G_{i_3i_1}\right)\right] \\ &= 2\mathrm{e}^{-t}q_t^{-2}s^{(4)} \sum_{i_1,i_2} \mathbb{E}\left[F'(X)G_{i_1i_2}G_{i_2i_1}\right] + O(N^{2/3-\epsilon'}). \end{aligned} \tag{7.52}$$

7.2.3 Proof of Lemma 7.4 for r = 4

In this case, we estimate the term as in the case $r = 2$ and get

$$\frac{q_t^{-3}}{N} \sum_{i_1,i_2,i_3} \left|\mathbb{E}\left[\partial_{i_2i_3}^4\left(F'(X)G_{i_1i_2}G_{i_3i_1}\right)\right]\right| = O(N^{2/3-\epsilon'}). \tag{7.53}$$

We leave the details to the interested reader.

References

1. Adlam, B., Che, Z.: Spectral statistics of sparse random graphs with a general degree distribution. arXiv:1509.03368 (2015)
2. Ajanki, O., Erdős, L., Krüger, T.: Universality for general Wigner-type matrices. arXiv:1506.05098 (2015)
3. Ajanki, O., Erdős, L., Krüger, T.: Local eigenvalue statistics for random matrices with general short range correlations. arXiv:1604.08188 (2016)
4. Bao, Z.G., Erdős, L., Schnelli, K.: Local law of addition of random matrices on optimal scale. Commun. Math. Phys. **349**(3), 947–990 (2017)
5. Bauerschmidt, R., Knowles, A., Yau, H.-T.: Local semicircle law for random regular graphs. arXiv:1503.08702 (2015)
6. Benaych-Georges, F., Guionnet, A., Male, C.: Central limit theorems for linear statistics of heavy tailed random matrices. Commun. Math. Phys. **329**(2), 641–686 (2014)
7. Bickel, P.J., Sarkar, P.: Hypothesis testing for automated community detection in networks. J. R. Stat. Soc. B **78**, 253–273 (2016)
8. Bordenave, C., Guionnet, A.: Localization and delocalization of eigenvectors for heavy-tailed random matrices. Probab. Theory Relat. Fields **157**(3–4), 885–953 (2013)
9. Cacciapuoti, C., Maltsev, A., Schlein, B.: Bounds for the Stieltjes transform and the density of states of wigner matrices. Probab. Theory Relat. Fields **163**(1–2), 1–59 (2015)
10. Che, Z.: Universality of random matrices with correlated entries. Electron. J. Probab. **22**, 1–38 (2017)
11. Erdős, P., Rényi, A.: On random graphs I. Publ. Math. **6**, 290–297 (1959)

12. Erdős, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5**, 17–61 (1960)
13. Erdős, L.: Universality of Wigner random matrices: a survey of recent results. Russ. Math. Surv. **66**(3), 507 (2011)
14. Erdős, L., Knowles, A., Yau, H.-T.: Averaging fluctuations in resolvents of random band matrices. Ann. Henri Poincaré **14**, 1837–1926 (2013)
15. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős-Rényi graphs I: local semicircle law. Ann. Probab. **41**, 2279–2375 (2013)
16. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős-Rnyi graphs II: Eigenvalue spacing and the extreme eigenvalues. Commun. Math. Phys. **314**(3), 587–640 (2012)
17. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: The local semicircle law for a general class of random matrices. Electron. J. Probab. **18**(59), 1–58 (2013)
18. Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for wigner random matrices. Ann. Probab. **37**, 815–852 (2009)
19. Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for wigner random matrices. Commun. Math. Phys. **287**, 641–655 (2009)
20. Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. Int. Math. Res. Not. **2010**(3), 436–479 (2010)
21. Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized wigner matrices. Probab. Theory Relat. Fields **154**(1–2), 341–407 (2012)
22. Erdős, L., Yau, H.-T., Yin, J.: Universality for generalized Wigner matrices with Bernoulli distribution. J. Comb. **2**, 15–82 (2012)
23. Erdős, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized wigner matrices. Adv. Math. **229**, 1435–1515 (2012)
24. Gilbert, E.N.: Random graphs. Ann. Math. Stat. **30**(4), 1141–1144 (1959)
25. Götze, F., Naumov, A., Tikhomirov, A.: Local semicircle law under moment conditions. Part I: the stieltjes transform. arXiv:1510.07350 (2015)
26. Götze, F., Naumov, A., Timushev, D., Tikhomirov, A.: On the local semicircular law for Wigner ensembles. arXiv:1602.03073 (2016)
27. Hanson, D.L., Wright, E.T.: A bound on tail probabilities for quadratic forms in independent random variables. Ann. Math. Stat. **42**, 1079–1083 (1971)
28. He, Y., Knowles, A.: Mesoscopic eigenvalue statistics of wigner matrices. arXiv:1603.01499 (2016)
29. Huang, J., Landon, B., Yau, H.-T.: Bulk universality of sparse random matrices. J. Math. Phys. **56**(12), 123301 (2015)
30. Kargin, V.: Subordination for the sum of two random matrices. Ann. Probab. **43**(4), 2119–2150 (2015)
31. Khorunzhy, A., Khoruzhenko, B., Pastur, L.: Asymptotic properties of large random matrices with independent entries. J. Math. Phys. **37**(10), 5033–5060 (1996)
32. Khorunzhy, A.: Sparse random matrices: spectral edge and statistics of rooted trees. Adv. Appl. Probab. **33**(1), 124–140 (2001)
33. Khorunzhiy, O.: On high moments and the spectral norm of large dilute Wigner random matrices. Zh. Mat. Fiz. Anal. Geom. **10**(1), 64–125 (2014)
34. Knowles, A., Yin, J.: Anisotropic local laws for random matrices. arXiv:1410.3516 (2014)
35. Lee, J.O., Yin, J.: A necessary and sufficient condition for edge universality of Wigner matrices. Duke Math. J. **163**(1), 117–173 (2014)
36. Lee, J.O., Schnelli, K.: Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. J. Math. Phys. **54**, 103504 (2013)
37. Lee, J.O., Schnelli, K.: Edge universality for deformed Wigner matrices. Rev. Math. Phys. **27**(8), 1550018 (2015)
38. Lee, J.O., Schnelli, K.: Tracy-widom distribution for the largest eigenvalue of real sample covariance matrices with general population. Ann. Appl. Probab. **26**(6), 3786–3839 (2016)
39. Lee, J.O., Schnelli, K., Stetler, B., Yau, H.-T.: Bulk universality for deformed Wigner matrices. Ann. Probab. **44**(3), 2349–2425 (2016)
40. Lei, J.: A goodness-of-fit test for stochastic block models. Ann. Stat. **44**, 401–424 (2016)
41. Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Ann. Probab. **37**, 1778–1840 (2009)
42. Péché, S., Soshnikov, A.: On the Lower bound of the spectral norm of symmetric random matrices with independent entries. Electron. Commun. Probab. **13**, 280290 (2008)

43. Péché, S., Soshnikov, A.: Wigner random matrices with non-symmetrically distributed entries. J. Stat. Phys. **129**, 857884 (2007)
44. Shcherbina, M., Tirozzi, B.: Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs: diluted regime. J. Math. Phys. **53**, 043501 (2012)
45. Sinai, Y., Soshnikov, A.: A refinement of wigners semicircle law in a neighborhood of the spectrum edge. Funct. Anal. Appl. **32**, 114131 (1998)
46. Soshnikov, A.: Universality at the edge of the spectrum in wigner random matrices. Commun. Math. Phys. **207**, 697–733 (1999)
47. Tao, T., Vu, V.: Random matrices: universality of the local eigenvalue statistics. Acta Math. **206**, 127–204 (2011)
48. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics up to the edge. Commun. Math. Phys. **298**, 549–572 (2010)
49. Tracy, C., Widom, H.: Level-spacing distributions and the airy kernel. Commun. Math. Phys. **159**, 151–174 (1994)
50. Tracy, C., Widom, H.: On orthogonal and symplectic matrix ensembles. Commun. Math. Phys. **177**, 727–754 (1996)