Commun. Math. Phys. 393, 839–907 (2022)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-022-04377-y

Communications in
Mathematical Physics

Convergence Rate to the Tracy–Widom Laws for the Largest Eigenvalue of Wigner Matrices

Kevin Schnelli[1], **Yuanyuan Xu**[2]

[1] KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: schnelli@kth.se
[2] Institute of Science and Technology Austria, Klosterneuburg, Austria. E-mail: yuanyuan.xu@ist.ac.at

Received: 19 May 2021 / Accepted: 9 March 2022
Published online: 15 April 2022 – © The Author(s) 2022

Abstract: We show that the fluctuations of the largest eigenvalue of a real symmetric or complex Hermitian Wigner matrix of size N converge to the Tracy–Widom laws at a rate $O(N^{-1/3+\omega})$, as N tends to infinity. For Wigner matrices this improves the previous rate $O(N^{-2/9+\omega})$ obtained by Bourgade (J Eur Math Soc, 2021) for generalized Wigner matrices. Our result follows from a Green function comparison theorem, originally introduced by Erdős et al. (Adv Math 229(3):1435–1515, 2012) to prove edge universality, on a finer spectral parameter scale with improved error estimates. The proof relies on the continuous Green function flow induced by a matrix-valued Ornstein–Uhlenbeck process. Precise estimates on leading contributions from the third and fourth order moments of the matrix entries are obtained using iterative cumulant expansions and recursive comparisons for correlation functions, along with uniform convergence estimates for correlation kernels of the Gaussian invariant ensembles.

1. Introduction and Main Results

In this paper we study a quantitative version of the edge universality for Wigner random matrices. Let H_N be a real symmetric or complex Hermitian Wigner matrix of size N. Then the edge universality asserts that the largest eigenvalue, λ_N, of H_N satisfies

$$\lim_{N\to\infty} \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r\Big) = \mathrm{TW}_\beta(r), \quad r \in \mathbb{R}, \tag{1.1}$$

where TW_β are the cumulative distribution functions of the Tracy–Widom laws [44,45] and $\beta = 1, 2$ indicates the symmetry class ($\beta = 1$ for real symmetric and $\beta = 2$ for complex Hermitian Wigner matrices). The universality of the Tracy–Widom laws was first proved in [40,41] for Wigner matrices whose entries have symmetric distributions.

Kevin Schnelli is supported in parts by the Swedish Research Council Grant VR-2017-05195, and the Knut and Alice Wallenberg Foundation. Yuanyuan Xu is supported by the Swedish Research Council Grant VR-2017-05195, and the ERC Advanced Grant "RMTBeyond" No. 101020331.

This symmetry assumption was partially removed in [35,36]. Edge universality for Wigner matrices whose entries have vanishing third moments was proved in [43]. Edge universality without moment matching was proved in [19] for Wigner matrices and in [2,6] for generalized Wigner matrices. A necessary and sufficient condition on the entries' distributions for the edge universality to hold was given in [31].

The main result of this paper is an estimate on the rate of convergence in (1.1) for Wigner matrices. Theorem 1.3 below states that, for any fixed $r_0 \in \mathbb{R}$ and small $\omega > 0$,

$$\sup_{r>r_0} \left| \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r \Big) - \mathrm{TW}_\beta(r) \right| \leq N^{-1/3+\omega}, \tag{1.2}$$

for N sufficiently large. For the Gaussian unitary ensemble (GUE, $\beta = 2$) and Gaussian orthogonal ensemble (GOE, $\beta = 1$) it was established in [25] that the convergence rate for the largest eigenvalue on a proper scaling is of order $O(N^{-2/3})$; see Theorem 1.2 below. The first rate of convergence for non-invariant ensembles was recently given by Bourgade in [5] where the upper bound $O(N^{-2/9+\omega})$ for the convergence rate was obtained for generalized Wigner matrices.

The proof of the estimate in (1.2) is based on the Green function comparison method for the edge universality by Erdős et al. [19]. Our main technical result given in Theorem 1.4 compares the expectation of a suitably chosen function of the Green function of the Wigner matrix H_N with the corresponding quantity for the Gaussian invariant ensembles. Instead of the traditional Lindeberg type swapping strategy [8,19,43], we use the continuous Green function flow induced by a matrix-valued Ornstein–Uhlenbeck process in combination with cumulant expansions [29,30] for the comparison. To achieve the convergence rate $O(N^{-1/3})$ in (1.2) the comparison is required on a much finer spectral scale than the typical $O(N^{-2/3})$ edge scaling. This requires in turn precise estimates on the contributions to the Green function flow from third and fourth order moments of the matrix entries.

Contributions from third moments can be estimated using the idea of unmatched indices [19], however due to the finer spectral scale, we require expansions to arbitrary order in terms of the control parameter of the strong local law for the Green function [19] to implement this idea. This step relies on applying cumulant expansions iteratively to Green functions and observing a cancellation to leading order [22,23,30]. The usefulness of cumulant expansions in random matrix theory was recognized in [27] and has widely been used since, *e.g.,* [7,16,21,32].

Contributions from fourth moments are controlled by first showing that they can be reduced to trace-like correlation functions of products of Green functions. This first step is motivated by the Weingarten calculus [10] to compute Haar integrals of products of eigenvector components for the invariant Gaussian ensembles. The actual reduction for non-invariant ensembles relies on applying cumulant expansions iteratively. In a second step we compare the resulting trace-like correlation functions between Wigner matrices and the invariant ensembles using again the interpolating flow. This leads to a hierarchy of correlation functions which, after expansion to arbitrary order, can be recursively estimated by the local law for the Green function. Finally, we need to control the trace-like correlation functions for the invariant ensembles. This is accomplished by using the uniform asymptotics [13] for correlation kernels of the invariant ensembles in the edge scaling.

Edge universality can also be studied through the dynamical approach of Erdős, Schlein and Yau. The local relaxation time of Dyson's Brownian motion (DBM) at the edges is known [1,5,28] to be of order $O(N^{-1/3})$. Combining his quantitative local

relaxation estimates for the DBM with a Green function comparison for short times, Bourgade obtained in [5] the convergence rate $O(N^{-2/9})$ to the Tracy–Widom laws for generalized Wigner matrices. In view of the local relaxation time of the DBM at the spectral edges, the convergence rate estimate in (1.2) may be optimal for Wigner matrices in general, though numerical simulations in [20] indicate that certain Wigner matrices exhibit faster convergence rates after a scaling and centering of the largest eigenvalue. We suspect that such a centering would crucially depend on the fourth moments of the entries and the symmetry type of the matrices.

The methods presented in this paper are rather robust and can be applied to other random matrix models. Of interest in statistics are in particular convergence rate estimates for sample covariance matrices. For the white Wishart ensemble the convergence rate $O(N^{-2/3})$ after a proper scaling were obtained in [14,33]. Edge universality for sample covariance matrices was established in [37] and a first quantitative version appeared recently in [46]. In the accompanying article [38] we establish the results corresponding to (1.2) for sample covariance matrices. In this paper we focus on estimating the contributions from third and fourth order moments of the matrix entries through assuming that the variances are uniform as for the invariant ensembles. Studying generalized Wigner matrices requires in addition new techniques to implement a variance profile and is thus postponed to our upcoming work [39].

1.1. Setup and main results. Let $H \equiv H_N$ be an $N \times N$ Wigner matrix satisfying the following.

Assumption 1.1. For a real symmetric ($\beta = 1$) Wigner matrix, we assume the following.

1. The matrix entries $\{H_{ij} \,|\, i \le j\}$ are independent real-valued centered random variables.
2. For $i \neq j$, $\mathbb{E}[(\sqrt{N}H_{ij})^2] = 1$, and $\mathbb{E}[(\sqrt{N}H_{ii})^2]$ are uniformly bounded.
3. All moments of the entries of $\sqrt{N}H_N$ are uniformly bounded, *i.e.,* for any $k \ge 3$, there exists C_k independent of N such that, for all $1 \le i, j \le N$,

$$\mathbb{E}[|\sqrt{N}H_{ij}|^k] \le C_k\,. \tag{1.3}$$

For a complex Hermitian ($\beta = 2$) Wigner matrix, we assume the following.

a. The matrix entries $\{H_{ij} \,|\, i \le j\}$ are independent complex-valued centered random variables.
b. For $i \neq j$, $\mathbb{E}[|\sqrt{N}H_{ij}|^2] = 1$, $\mathbb{E}[(H_{ij})^2] = 0$, and $\mathbb{E}[(\sqrt{N}H_{ii})^2]$ are uniformly bounded.
c. The bound (1.3) holds true.

The Gaussian ensembles, which we denote by GβE for short, are Wigner matrices with Gaussian entries: For the Gaussian unitary ensemble (GUE, $\beta = 2$) the off-diagonal matrix entries are standard complex-valued Gaussians (*i.e.,* $\sqrt{N}H_{ij} \overset{\mathrm{d}}{=} \mathcal{N}(0, \frac{1}{2}) + \mathrm{i}\mathcal{N}(0, \frac{1}{2})$) and the diagonal entries are standard real-valued Gaussians (*i.e.,* $\sqrt{N}H_{ii} \overset{\mathrm{d}}{=} \mathcal{N}(0, 1)$). Similarly, for the Gaussian orthogonal ensemble (GOE, $\beta = 1$) the matrix entries are real-valued Gaussians with $\sqrt{N}H_{ij} \overset{\mathrm{d}}{=} \mathcal{N}(0, 1)$ ($i \neq j$) and $\sqrt{N}H_{ii} \overset{\mathrm{d}}{=} \mathcal{N}(0, 2)$.

Let $(\lambda_j)_{j=1}^N$ be the eigenvalues of H_N arranged in a non-decreasing order. It is well known that the largest eigenvalue λ_N converges to the spectral edge 2 in probability.

The typical spacing of the top eigenvalues near 2 is of order $O(N^{-2/3})$, due to the square-root behavior at the end points of the limiting spectral density and eigenvalue rigidity. The limiting distribution of $N^{2/3}(\lambda_N - 2)$ for the Gaussian ensembles was found by Tracy and Widom in [44,45]. The corresponding convergence rate was quantized by Johnstone and Ma [25] in the following theorem.

Theorem 1.2 (Convergence rate for the Gaussian ensembles). *Let H_N be the GUE. For any fixed $r_0 \in \mathbb{R}$, there exists a constant $C = C(r_0)$ such that*

$$\sup_{r>r_0} \left| \mathbb{P}^{\mathrm{GUE}}\Big(N^{2/3}(\lambda_N - 2) < r \Big) - \mathrm{TW}_2(r) \right| \le C N^{-2/3}. \tag{1.4}$$

Moreover, considering the GOE with N even, we have

$$\sup_{r>r_0} \left| \mathbb{P}^{\mathrm{GOE}}\Big((N-1)^{1/6}\sqrt{N}\Big(\lambda_N - (4 - \frac{2}{N})^{1/2}\Big) < r \Big) - \mathrm{TW}_1(r) \right| \le C N^{-2/3}. \tag{1.5}$$

The first quantitative convergence rate $O(N^{-2/9+\omega})$ for generalized Wigner matrices was obtained by Bourgade [5] using optimal local relaxation estimates for the Dyson Brownian motion and a quantitative Green function comparison theorem for short times.

The main result of this paper is an improved bound for the convergence rate of the distribution of $N^{2/3}(\lambda_N - 2)$ for arbitrary Wigner matrices to the Tracy–Widom laws.

Theorem 1.3 (Convergence rate for Wigner matrices). *Let H_N be a real or complex Wigner matrix satisfying Assumption 1.1. For any fixed $r_0 \in \mathbb{R}$ and small $\omega > 0$,*

$$\sup_{r>r_0} \left| \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r \Big) - \mathrm{TW}_\beta(r) \right| \le N^{-\frac{1}{3}+\omega}, \tag{1.6}$$

for sufficiently large $N \ge N_0(r_0, \omega)$. The corresponding statement holds for the smallest eigenvalue λ_1.

The proof of Theorem 1.3 relies on the Green function comparison method [18,19]. Let

$$G(z) := \frac{1}{H_N - z}, \quad m_N(z) := \frac{1}{N}\mathrm{Tr} G(z), \quad z \in \mathbb{C}^+, \tag{1.7}$$

denote the resolvent or Green function of the Wigner matrix H_N and m_N its normalized trace. The distribution of the rescaled largest eigenvalue can be linked to the expectation (of smooth functions) of the imaginary part of $m_N(z)$ for appropriately chosen spectral parameters z; see Sect. 2.3. The main technical result of this paper is the following comparison theorem at the spectral edges.

Theorem 1.4 (Green function comparison theorem). *Let F be a smooth function with uniformly bounded derivatives. For any small $\epsilon > 0$, let $N^{-1+\epsilon} \le \eta \le N^{-2/3+\epsilon}$ and $|\kappa_1|, |\kappa_2| \le C_0 N^{-2/3+\epsilon}$ for some $C_0 > 0$. Then there exists some $c_0 > 0$ that does not depend on ϵ, such that*

$$\left| \Big(\mathbb{E} - \mathbb{E}^{\mathrm{G}\beta\mathrm{E}} \Big) \Big[F\Big(N \int_{\kappa_1}^{\kappa_2} \mathrm{Im}\, m_N(2 + x + \mathrm{i}\eta) \mathrm{d}x \Big) \Big] \right| \le N^{-1/3+c_0\epsilon}, \tag{1.8}$$

for sufficiently large $N \ge N_0(\epsilon, C_0)$.

Remark 1.1. A first Green function comparison theorem at the spectral edges was obtained in [19] for spectral parameters η of size $O(N^{-2/3-\epsilon})$ and with an error estimate of size $O(N^{-1/6+c_0\epsilon})$.

The constant c_0 in the upper bound in (1.8) can be chosen as any number bigger than one. An inspection of our proof in fact yields that the upper bound in (1.8) can be written as

$$\max\{K_4, |M_2 - 1|\}N^{-\frac{1}{3}+c_0\epsilon} + O(N^{-1/2+\epsilon}),$$

where $M_2 = \max_i \left|\mathbb{E}[(\sqrt{N}h_{ii})^2]\right|$; $K_4 = \max_{i\neq j}\left|c^{(4)}(\sqrt{N}h_{ij})\right|$, for $\beta = 1$, and $K_4 = \max_{i\neq j}\left|c^{(2,2)}(\sqrt{N}h_{ij})\right|$, for $\beta = 2$, with $c^{(4)}(\sqrt{N}h_{ij})$ the fourth cumulant of $\sqrt{N}h_{ij}$ given in (2.27) and $c^{(2,2)}$ the corresponding (2, 2)-cumulant defined in (2.24).

Remark 1.2. The proof of the Green function comparison is based on a continuous interpolation given by a matrix-valued Ornstein–Uhlenbeck process; see (3.8). On the level of the eigenvalues this evolution corresponds to Dyson's Brownian motion (DBM). Bourgade's proof of the convergence rate $O(N^{-2/9+\epsilon})$ consists of two parts: (1) the local relaxation estimate for the DBM for $t \gg N^{-1/3}$; (2) a quantitative version of the Green function comparison theorem for small times $t \ll 1$, which is not sharp. Optimizing the errors from these two parts, the error $N^{-2/9}$ is obtained at $t = N^{-1/9}$. In our proof, we improve the Green function comparison even for long times $t \sim \log N$ and then use standard perturbation theory to bridge to the Gaussian ensembles.

1.2. Organization of the paper and outline of proofs. The paper is organized as follows. In Sect. 2, we provide the preliminaries for the proofs, *e.g.,* local law for the Green function and cumulant expansions; and recall some properties of the invariant ensembles. In Sect. 3, following the approach of [19], we first reduce the proof of the main result Theorem 1.3 to the Green function comparison in Theorem 1.4. We then prove Theorem 1.4 using the interpolating Green function flow and the key estimates on the resulting drift term stated in Proposition 3.1 below.

In Sect. 4, before we give the proof of Proposition 3.1 for arbitrary functions F, we prove the corresponding Green function comparison theorem in the simplest case, $F(x) = x$; see Proposition 4.1. To make the statements easier, we first consider complex Hermitian Wigner matrices. The proof of Proposition 4.1 is carried out in Sects. 4, 5 and 6. We sketch the proof in the following.

1. We first set up the interpolation between a given Wigner matrix and the GUE using the matrix Ornstein–Uhlenbeck process in (3.7). Using Ito's formula, we derive the stochastic evolution for the time-dependent normalized trace of the Green function $m_N(t, z)$ in (4.4). It then suffices to estimate the drift term given in (4.6). Using the cumulant expansions of Lemma 2.4, we expand the expectation of the drift term up to the fourth order. We observe a precise cancellation of the second order terms in the cumulant expansions (4.5) for the off-diagonal entries. The cancellation of these second order terms is due to Assumption 1.1 (b.), namely that the variances of our Wigner matrices coincide with the invariant ensembles. It then suffices to estimate the third and fourth order terms in (4.5) as well as the remaining second order terms for the diagonal entries, which are averaged products of Green function entries.
2. All the third order terms, as well as the fourth order terms excluding the ones corresponding to the (2,2)-cumulants of the off-diagonal entries are unmatched; see Definition 4.1. The contributions from these unmatched terms are negligible, as stated

in Proposition 4.2 which is proved in Sect. 6. For GUE matrices, corresponding estimates can be established using the Weingarten calculus as discussed in Sect. 6.1. In Sect. 6.2, we study an example of an unmatched term and introduce the expansion mechanism used to prove Proposition 4.2 for general Wigner matrices. The key observation is that each time we perform the cumulant expansion on an unmatched term, we gain an additional off-diagonal Green function entry which slightly improves the estimate by the entrywise local law in (3.10). In Sect. 6.3, we give the proof of Proposition 4.2 for any unmatched term using the above expansion mechanism iteratively by counting the number of off-diagonal Green function entries.

3. The fourth order terms corresponding to the $(2, 2)$-cumulants of the off-diagonal entries and the second order terms stemming from the diagonal entries are given in terms of matched terms with a certain structure; see Definition 4.2. Motivated by the GUE computations based on the Weingarten calculus in Sect. 5.1, we show that such terms can be expanded into trace-like correlation functions of Green functions referred to as type-0 terms in Definition 4.2, as stated in Proposition 4.3. The proof of Proposition 4.3 is presented in Sect. 5.2 using cumulant expansions iteratively. The resulting type-0 terms are then estimated in Lemma 4.1 which is proved using recursive comparisons and iterative expansions in Sect. 5.3. The key observation is that, after deriving the stochastic evolution in (5.27) under the Ornstein–Uhlenbeck flow for any type-0 term containing d_1 off-diagonal Green function entries, we can expand the corresponding drift term to arbitrary order using Propositions 4.2 and 4.3, and end up with finitely many type-0 terms containing at least $d_1 + 1$ off-diagonal Green function entries as in (5.32). By recursive comparison, Lemma 4.1 follows from the local law in (3.10) for the Green function and the estimates of type-0 terms for the GUE in Lemma 5.2. The last Sect. 5.4 is devoted to the proof of Lemma 5.2 using the determinantal structure of the GUE and convergence properties of its correlation kernel in the edge scaling.

In Sect. 7, we extend the above ideas to general functions F, and use the estimate (4.3) from Proposition 4.1 as an input to prove Proposition 3.1. We then conclude with the Green function comparison in Theorem 1.4 and hence our main result Theorem 1.3. In the last Sect. 8, the real symmetric case is proved with the required modifications.

Notation: We will use the following definition on high-probability estimates from [15].

Definition 1.1. Let $\mathcal{X} \equiv \mathcal{X}^{(N)}$ and $\mathcal{Y} \equiv \mathcal{Y}^{(N)}$ be two sequences of nonnegative random variables. We say $\mathcal{Y}$ stochastically dominates $\mathcal{X}$ if, for all (small) $\tau > 0$ and (large) $\Gamma > 0$,

$$\mathbb{P}\big(\mathcal{X}^{(N)} > N^{\tau}\mathcal{Y}^{(N)}\big) \leq N^{-\Gamma}, \tag{1.9}$$

for sufficiently large $N \geq N_0(\tau, \Gamma)$, and we write $\mathcal{X} \prec \mathcal{Y}$ or $\mathcal{X} = O_{\prec}(\mathcal{Y})$.

We often use the notation $\prec$ also for deterministic quantities, then (1.9) holds with probability one. Properties of stochastic domination can be found in the following lemma.

Lemma 1.1 (Proposition 6.5 in [17]).

1. *$X \prec Y$ and $Y \prec Z$ imply $X \prec Z$;*
2. *If $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1X_2 \prec Y_1Y_2$;*
3. *If $X \prec Y$, $\mathbb{E}Y \geq N^{-c_1}$ and $|X| \leq N^{c_2}$ almost surely with some fixed exponents c_1, $c_2 > 0$, then we have $\mathbb{E}X \prec \mathbb{E}Y$.*

For any vector $\mathbf{v} \in \mathbb{C}^N$, let $\mathbf{v}(j)$ be the j-th entry of the vector. For any matrix $A \in \mathbb{C}^{N\times N}$, the matrix norm induced by the Euclidean vector norm is given by $\|A\|_2 := \sigma_{\max}(A)$, where $\sigma_{\max}(A)$ denotes the largest singular value of A. We denote the sup norm of the matrix by $\|A\|_{\max} := \max_{i,j} |A_{ij}|$. We use the notation $\underline{A} := \frac{1}{N}\mathrm{Tr}A$ for the normalized trace.

Throughout the paper, we use c and C to denote strictly positive constants that are independent of N. Their values may change from line to line. We use the standard Big-O and little-o notations for large N. For $X, Y \in \mathbb{R}$, we write $X \ll Y$ if there exists a small $c > 0$ such that $|X| \leq N^{-c}|Y|$ for large N. Moreover, we write $X \sim Y$ if there exist constants $c, C > 0$ such that $c|Y| \leq |X| \leq C|Y|$ for large N. Finally, we denote the upper half-plane by $\mathbb{C}^+ := \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$, and the non-negative real numbers by $\mathbb{R}^+ := \{x \in \mathbb{R} : x \geq 0\}$.

2. Preliminaries

In the section, we collect some basic notations, tools and results required in the subsequent sections, in particular we introduce the local law for the Green function of Wigner matrices and eigenvalue rigidity estimates; relate the distribution function of the largest eigenvalues to the normalized trace of the Green function; introduce the cumulant expansion formalism and finally recall properties of the GUE and the Airy kernel.

2.1. Local law for Wigner matrices. For a probability measure ν on $\mathbb{R}$ denote by m_ν its Stieltjes transform, *i.e.,*

$$m_\nu(z) := \int_{\mathbb{R}} \frac{\mathrm{d}\nu(x)}{x - z}, \quad z \in \mathbb{C}^+. \tag{2.1}$$

We refer to z as spectral parameter and often write $z = E + \mathrm{i}\eta$, $E \in \mathbb{R}$, $\eta > 0$. Note that $m_\nu : \mathbb{C}^+ \to \mathbb{C}^+$ is analytic and can be analytically continued to the real line outside the support of ν. Moreover, m_ν satisfies $\lim_{\eta\nearrow\infty} \mathrm{i}\eta m_\mu(\mathrm{i}\eta) = -1$. The Stieltjes transform of the semicircle distribution $\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{(4 - x^2)_+}$ is denoted by $m_{sc}(z)$. It is well know that $m_{sc}(z)$ is the unique solution to

$$1 + zm_{sc}(z) + m_{sc}^2(z) = 0, \tag{2.2}$$

satisfying $\mathrm{Im}\, m_{sc}(z) > 0$, for $\mathrm{Im}\, z > 0$. The Stieltjes transform of the empirical eigenvalue measure of a Wigner matrix H_N, $\mu_N := \frac{1}{N}\sum_{j=1}^N \delta_{\lambda_j}$, is then given by the normalized trace of its Green function defined in (1.7).

Let $\kappa = \kappa(E)$ be the distance from $E \in \mathbb{R}$ to the closest edge point of the semicircle law, *i.e.,*

$$\kappa := \min\{|E - 2|, |E + 2|\}. \tag{2.3}$$

Define the domain of the spectral parameter z,

$$S_0 := \{z = E + \mathrm{i}\eta : |E| \leq 5, 0 < \eta \leq 10\}. \tag{2.4}$$

The Stieltjes transform m_{sc} has the following quantitative properties, for a reference, see *e.g.,* [17].

Lemma 2.1. *The Stieltjes transform of the semicircular law has the following properties:*

1. *The imaginary part of m_{sc} satisfies*

$$|\mathrm{Im}\, m_{sc}(z)| \sim \begin{cases} \sqrt{\kappa+\eta}, & \text{if } E \in [-2,2], \\ \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{otherwise}, \end{cases} \tag{2.5}$$

uniformly in $z \in S_0$.

2. *There exists a strictly positive constant c, such that*

$$c \le |m_{sc}(z)| \le 1 - c\eta, \tag{2.6}$$

hold for all $z \in S_0$.

For any arbitrary small $\epsilon > 0$, introduce the following subdomain of S_0,

$$S \equiv S(\epsilon) := \left\{ z = E + \mathrm{i}\eta : |E| \le 5, N^{-1+\epsilon} \le \eta \le 10 \right\}. \tag{2.7}$$

We also define the deterministic control parameter

$$\Psi \equiv \Psi(z) := \sqrt{\frac{\mathrm{Im}\, m_{sc}(z)}{N\eta}} + \frac{1}{N\eta}, \quad z = E + \mathrm{i}\eta. \tag{2.8}$$

In particular, from (2.5), for any $z \in S(\epsilon)$, we have

$$\frac{C}{\sqrt{N}} \le \Psi(z) \le C' N^{-\epsilon}. \tag{2.9}$$

With these notations, we are now ready to state the following local law for the Green function of a Wigner matrix.

Theorem 2.1 (Local law for Wigner matrices [19]). *Let H be a symmetric or Hermitian N by N matrix satisfying Assumption 1.1 and recall the Green function of H and its normalized trace in (1.7). Then we have*

$$\max_{1\le i,j\le N} |G_{ij}(z) - \delta_{ij} m_{sc}(z)| \prec \Psi(z), \quad |m_N(z) - m_{sc}(z)| \prec \frac{1}{N\eta}, \tag{2.10}$$

uniformly in $z \in S$.

2.2. Rigidity of eigenvalues. The local law for the Green function in Theorem 2.1 implies the following rigidity estimates for the eigenvalues of H. Recall that the eigenvalues of H are denoted as $(\lambda_j)_{j=1}^N$ arranged in a non-decreasing order. For $E_1 < E_2$ ($E_1, E_2 \in \mathbb{R} \cup \{\pm\infty\}$) denote the eigenvalue counting function by

$$\mathcal{N}(E_1, E_2) := \#\{j : E_1 \le \lambda_j \le E_2\}. \tag{2.11}$$

We also define the classical location γ_j of the j-th eigenvalue λ_j by

$$\frac{j}{N} = \int_{-\infty}^{\gamma_j} \rho_{sc}(x)\mathrm{d}x. \tag{2.12}$$

Theorem 2.2 (Eigenvalue rigidity [19]). *For any $E_1 < E_2$, we have*

$$\left|\mathcal{N}(E_1, E_2) - N\int_{E_1}^{E_2}\rho_{sc}(x)\mathrm{d}x\right| \prec 1\,. \tag{2.13}$$

In addition, for any $1 \le j \le N$, we have

$$|\lambda_j - \gamma_j| \prec N^{-2/3}\Big(\min\{j, N-j+1\}\Big)^{-1/3}. \tag{2.14}$$

In particular, fix any C_1 and C_2, then for any small $\epsilon > 0$ and large $\Gamma > 0$ we have

$$|\lambda_N - 2| \le N^{-2/3+\epsilon}, \quad \mathcal{N}(2 - C_1N^{-2/3+\epsilon}, 2 + C_2N^{-2/3+\epsilon}) \le N^{2\epsilon}, \tag{2.15}$$

with probability bigger than $1 - N^{\Gamma}$, for N sufficiently large.

2.3. Relating the distribution of the largest eigenvalue to the Green function. Fix a small $\epsilon > 0$ and set

$$E_L := 2 + 4N^{-2/3+\epsilon}. \tag{2.16}$$

For any $E \le E_L$, we define

$$\chi_E := \mathbb{1}_{[E,E_L]}, \tag{2.17}$$

and note that $\mathcal{N}(E, E_L) = \mathrm{Tr}\chi_E(H)$. For $\eta > 0$, we define the mollifier θ_η by setting

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi}\mathrm{Im}\,\frac{1}{x - \mathrm{i}\eta}. \tag{2.18}$$

We can relate $\mathrm{Tr}\chi_E \star \theta_\eta(H)$ to the normalized trace of the Green function by the following identity,

$$\mathrm{Tr}\chi_E \star \theta_\eta(H) = \frac{N}{\pi}\int \chi_E(y)\mathrm{Im}\,m_N(y+\mathrm{i}\eta)\mathrm{d}y = \frac{N}{\pi}\int_E^{E_L}\mathrm{Im}\,m_N(y+\mathrm{i}\eta)\mathrm{d}y\,. \tag{2.19}$$

The following lemma assures that $\mathrm{Tr}\chi_E(H)$ can be sufficiently well approximated by $\mathrm{Tr}\chi_E \star \theta_\eta(H)$ for $\eta \ll N^{-2/3}$. Relying on this approximation, the lemma after, Lemma 2.3, then yields the desired link between the distribution function of the rescaled largest eigenvalue of H and the normalized trace of the Green function using a cleverly chosen observable. This line of arguments was used first in [19] to prove the edge universality of Wigner matrices, where η is chosen slightly smaller than the typical edge eigenvalue spacing $N^{-2/3}$. In order to obtain a quantitative convergence rate, we aim to choose here η much smaller with $\eta \gg N^{-1}$. A similar argument was used in [5]. The proofs of Lemmas 2.2 and 2.3 are modifications of [19] in order to accommodate the small η regime, and are postponed to Appendix.

Lemma 2.2. *Let E, η and l_1 be scale parameters satisfying $N^{-1} \ll \eta \ll l_1 \ll E_L - E \le CN^{-2/3+\epsilon}$. Then, for any $\Gamma > 0$,*

$$\left|\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_E \star \theta_\eta(H)\right| \le C\Big(\mathcal{N}(E-l_1, E+l_1) + \frac{\eta}{l_1}N^{2\epsilon}\Big), \tag{2.20}$$

holds with probability bigger than $1 - N^{-\Gamma}$, for N sufficiently large.

Let $F : \mathbb{R} \longrightarrow \mathbb{R}$ be a smooth cut-off function such that

$$F(x) = 1, \quad \text{if} \quad |x| \le 1/9; \quad F(x) = 0, \quad \text{if} \quad |x| \ge 2/9, \tag{2.21}$$

and we assume that $F(x)$ is non-increasing for $x \ge 0$. Then one obtains from Lemma 2.2 the following result.

Lemma 2.3. *Set $l_1 = N^{3\epsilon}\eta$ and $l = N^{3\epsilon} l_1$ such that $N^{-1} \ll \eta \ll l_1 \ll l \ll E_L - E \le CN^{-2/3+\epsilon}$. Then for any $\Gamma > 0$, we have*

$$\mathrm{Tr}\chi_{E+l} \star \theta_\eta(H) - N^{-\epsilon} \le \mathcal{N}(E, \infty) \le \mathrm{Tr}\chi_{E-l} \star \theta_\eta(H) + N^{-\epsilon}, \tag{2.22}$$

with probability bigger than $1 - N^{-\Gamma}$, for N sufficiently large. Furthermore, we have

$$\mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E-l}\star\theta_\eta(H)\Big)\Big] - N^{-\Gamma} \le \mathbb{P}\Big(\mathcal{N}(E,\infty) = 0\Big) \le \mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E+l}\star\theta_\eta(H)\Big)\Big] + N^{-\Gamma}, \tag{2.23}$$

where $F(x)$ is the cut-off function given in (2.21).

Hence, recalling (2.19), we have established the desired link to the normalized trace of the Green function.

2.4. Cumulant expansion formulas. A key tool of this paper are the following cumulant expansion identities. For reference, we refer to Lemma 3.1 in [21].

Lemma 2.4. *Let h be a complex-valued random variable with finite moments. Define the (p, q)-cumulant of h to be*

$$c^{(p,q)} := (-\mathrm{i})^{p+q}\Big(\frac{\partial^{p+q}}{\partial s^p \partial t^q} \log \mathbb{E}\mathrm{e}^{\mathrm{i}sh + \mathrm{i}t\bar{h}}\Big)\Big|_{s,t=0}. \tag{2.24}$$

Let $f : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ be a smooth function and denote its derivatives by

$$f^{(p,q)}(z_1, z_2) := \frac{\partial^{p+q}}{\partial z_1^p \partial z_2^q} f(z_1, z_2).$$

Then for any fixed $l \in \mathbb{N}$, we have

$$\mathbb{E}\big[\bar{h} f(h, \bar{h})\big] = \sum_{p+q+1=1}^{l} \frac{1}{p!q!} c^{(p,q+1)} \mathbb{E}\big[f^{(p,q)}(h, \bar{h})\big] + R_{l+1}, \tag{2.25}$$

where the error term R_{l+1} can be bounded as

$$\begin{aligned} |R_{l+1}| \le{}& C_l \mathbb{E}|h|^{l+1} \max_{p+q=l} \Big\{ \sup_{|z|\le M} |f^{(p,q)}(z, \bar{z})| \Big\} \\ &+ C_l \mathbb{E}\Big[|h|^{l+1} 1_{|h|>M}\Big] \max_{p+q=l} \|f^{(p,q)}(z, \bar{z})\|_\infty, \end{aligned} \tag{2.26}$$

and $M > 0$ is an arbitrary fixed cutoff.

Moreover, we have the analogous cumulant expansion formula for a real-valued random variable h with finite moments. Define the k-th cumulant of h to be

$$c^{(k)} := (-\mathrm{i})^k \Big(\frac{\mathrm{d}^k}{\mathrm{d}t^k} \log \mathbb{E}\mathrm{e}^{\mathrm{i}th}\Big)\Big|_{t=0}. \tag{2.27}$$

Let $f : \mathbb{R} \longrightarrow \mathbb{C}$ be a smooth function and denote by $f^{(k)}$ its k-th derivative. Then for any fixed $l \in \mathbb{N}$, we have

$$\mathbb{E}\big[hf(h)\big] = \sum_{k+1=1}^{l} \frac{1}{k!} c^{(k+1)} \mathbb{E}[f^{(k)}(h)] + R_{l+1}, \tag{2.28}$$

where the error term satisfies

$$|R_{l+1}| \leq C_l \mathbb{E}|h|^{l+1} \sup_{|x|\leq M} |f^{(l)}(x)| + C_l \mathbb{E}\Big[|h|^{l+1} 1_{|h|>M}\Big] \|f^{(l)}\|_\infty,$$

and $M > 0$ is an arbitrary fixed cutoff.

2.5. GUE and the Airy kernel. Let $H \equiv H_N$ belong to the GUE and denote the eigenvalues of the rescaled matrix $\sqrt{N}H$ by $(\mu_j)_{j=1}^N$ in non-decreasing order. The joint eigenvalue density is explicitly given by

$$p(\mu_1, \ldots, \mu_N) = \frac{1}{Z_{N,\beta}} \prod_{i<j} |\mu_i - \mu_j|^\beta \mathrm{e}^{-\frac{\beta}{4}\sum_{i=1}^N \mu_i^2}, \quad \beta = 2,$$

with $Z_{N,\beta}$ be the normalization constant.

The process of the eigenvalues is well known to be a determinantal point process [24,42]. The n-point correlation function of the eigenvalue process is given by

$$p_n(\mu_1, \ldots, \mu_n) = \det[K_N(\mu_i, \mu_j)]_{1\leq i,j\leq n}, \tag{2.29}$$

with the reproducing kernel given by

$$K_N(x, y) := \sum_{k=0}^{N-1} q_k(x) q_k(y) \mathrm{e}^{-\frac{x^2+y^2}{4}},$$

where q_k is the k-th Hermite polynomial given by

$$q_k(x) := (-1)^k \mathrm{e}^{\frac{x^2}{2}} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \mathrm{e}^{-\frac{x^2}{2}}.$$

The Hermite polynomials are orthogonal with respect to the weight $\mathrm{e}^{-\frac{x^2}{2}}$ over $\mathbb{R}$. We further define the k-th Hermite function by

$$\phi_k(x) := \frac{1}{\sqrt{\sqrt{2\pi} k!}} \mathrm{e}^{-\frac{x^2}{4}} q_k(x), \tag{2.30}$$

which is a solution to the differential equation

$$\phi_k''(x) + (k + \frac{1}{2} - \frac{x^2}{4})\phi_k(x) = 0. \tag{2.31}$$

One then checks that $\{\phi_k\}$ form an orthonormal basis of $L^2(\mathbb{R})$. The Christoffel–Darboux formula then states that

$$K_N(x,y)=\sum_{k=0}^{N-1}\phi_k(x)\phi_k(y)=\sqrt{N}\frac{\phi_N(x)\phi_{N-1}(y)-\phi_{N-1}(x)\phi_N(y)}{x-y},\quad x\neq y, \tag{2.32}$$

as well as

$$K_N(x,x)=\sqrt{N}\Big(\phi_N'(x)\phi_{N-1}(x)-\phi_{N-1}'(x)\phi_N(x)\Big). \tag{2.33}$$

We also have the trace identity for the kernel

$$\int_{\mathbb{R}}K_N(x,x)\mathrm{d}x=N, \tag{2.34}$$

and the reproducing formula

$$K_N(x,y)=\int_{\mathbb{R}}K_N(x,z)K_N(z,y)\mathrm{d}z. \tag{2.35}$$

More details can be found in [3,12].

Recall that the eigenvalues $(\lambda_j)_{j=1}^N$ of the GUE are given by $\lambda_j=\frac{\mu_j}{\sqrt{N}}$. Then the corresponding kernel for the eigenvalue process (λ_j) is given by

$$\widetilde{K}_N(x,y)=\sqrt{N}K_N(\sqrt{N}x,\sqrt{N}y). \tag{2.36}$$

In the edge regime, we rescale the eigenvalues as $\lambda_j=2+\frac{l_j}{N^{2/3}}$ and the corresponding kernel is then given by

$$\begin{aligned}K_N^{\mathrm{edge}}(x,y)&:=\frac{1}{N^{2/3}}\widetilde{K}_N\Big(2+\frac{x}{N^{2/3}},2+\frac{y}{N^{2/3}}\Big)\\&=\frac{1}{N^{1/6}}K_N\Big(2\sqrt{N}+\frac{x}{N^{1/6}},2\sqrt{N}+\frac{y}{N^{1/6}}\Big).\end{aligned} \tag{2.37}$$

Next, recall that the Airy kernel is defined by

$$K_{\mathrm{airy}}(x,y):=\frac{\mathrm{Ai}(x)\mathrm{Ai}'(y)-\mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}, \tag{2.38}$$

with Ai be the Airy function of first kind, which is the solution of

$$\mathrm{Ai}''(x)-x\mathrm{Ai}(x)=0,\quad x\in\mathbb{R}, \tag{2.39}$$

satisfying the boundary condition $\mathrm{Ai}(x)\to 0$ as $x\to\infty$. As $x\to y$, the Airy kernel reduces to

$$K_{\mathrm{airy}}(x,x):=(\mathrm{Ai}'(x))^2-\mathrm{Ai}''(x)\mathrm{Ai}(x)=(\mathrm{Ai}'(x))^2-x(\mathrm{Ai}(x))^2. \tag{2.40}$$

Lemma 2.5 (Lemma 3.9.33 in [3]). *For fixed $L_0\in\mathbb{R}$, there exists a constant C, such that one has uniformly in $x,y\in[L_0,+\infty)$ that*

$$\Big|\partial_x^a\partial_y^bK_{\mathrm{airy}}(x,y)\Big|\leq C,\quad a,b\in\{0,1\}. \tag{2.41}$$

Furthermore, we have the asymptotics

$$K_{\mathrm{airy}}(x,x)\sim_{x\to\infty}\frac{\mathrm{e}^{-\frac{4}{3}x^{\frac{3}{2}}}}{x};\quad K_{\mathrm{airy}}(x,x)\sim_{x\to-\infty}\sqrt{|x|}. \tag{2.42}$$

The following result of Deift and Gioev [13] quantizes the convergence rate of the edge kernel in (2.37) to the limiting Airy kernel in (2.38).

Theorem 2.3 (Theorem 1.1 in [13]). *For fixed $L_0 \in \mathbb{R}$, there exists constants $C, c > 0$ depending on L_0, such that one has uniformly for $x, y \in [L_0, +\infty)$,*

$$\left|\partial_x^a \partial_y^b \left[K_N^{\text{edge}}(x, y) - K_{\text{airy}}(x, y)\right]\right| \leq CN^{-2/3} \mathrm{e}^{-cx} \mathrm{e}^{-cy}, \quad a, b \in \{0, 1\}. \tag{2.43}$$

3. Proof of Theorem 1.3

In this section we give the proof of Theorem 1.3 from the main technical result, the Green function comparison theorem, Theorem 1.4.

Proof of Theorem 1.3. Because of the rigidity of the eigenvalues in (2.15), one easily verifies that, for any $\epsilon > 0$ and $\Gamma > 2/3$,

$$\sup_{|r| \geq N^{\epsilon}} \left|\mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r\Big) - \mathbb{P}^{\mathrm{G}\beta\mathrm{E}}\Big(N^{2/3}(\lambda_N - 2) < r\Big)\right| \leq N^{-\Gamma}, \tag{3.1}$$

for sufficiently large N. Hence in order to prove Theorem 1.3, it suffices to focus on $r_0 < r < N^{\epsilon}$ with r_0 as in Theorem 1.2 and Theorem 1.3.

Set as in (2.16)

$$E := 2 + N^{-2/3} r, \quad \text{and} \quad E_L := 2 + 4N^{-2/3+\epsilon}.$$

Fix $\eta = N^{-1+\epsilon}$ and $l = N^{-1+7\epsilon}$ as in Lemma 2.3. Here we choose $\epsilon > 0$ sufficiently small such that $l \ll N^{-2/3}$. From (2.19) and (2.23), we can relate the distribution of the largest eigenvalue to the normalized trace of the Green function as follows,

$$\begin{aligned} &\mathbb{E}\Big[F\Big(N \int_{N^{-2/3}r-l}^{4N^{-2/3+\epsilon}} \operatorname{Im} m_N(2 + x + \mathrm{i}\eta)\mathrm{d}x\Big)\Big] - N^{-\Gamma} \leq \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r\Big) = \mathbb{P}\Big(\mathcal{N}(E, \infty) = 0\Big) \\ &\leq \mathbb{E}\Big[F\Big(N \int_{N^{-2/3}r+l}^{4N^{-2/3+\epsilon}} \operatorname{Im} m_N(2 + x + \mathrm{i}\eta)\mathrm{d}x\Big)\Big] + N^{-\Gamma}. \end{aligned} \tag{3.2}$$

By shifting the value of r in the second inequality of (3.2) and combining with the first inequality of (3.2), we obtain

$$\begin{aligned} \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r - 2N^{2/3}l\Big) - N^{-\Gamma} &\leq \mathbb{E}\Big[F\Big(N \int_{N^{-2/3}r-l}^{4N^{-2/3+\epsilon}} \operatorname{Im} m_N(2 + x + \mathrm{i}\eta)\mathrm{d}x\Big)\Big] \\ &\leq \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r\Big) + N^{-\Gamma}. \end{aligned} \tag{3.3}$$

Note that the above inequalities hold true for $\beta = 1, 2$ and any Wigner matrices, including the Gaussian ensembles. From the known convergence rates for the Gaussian ensembles in Theorem 1.2 (for the GUE, and GOE with N even), and the convergence rate $N^{-1/3}$ obtained in Theorem 1.2 of [9] for the GOE with N odd, we find

$$\begin{aligned} \mathrm{TW}_\beta\Big(r - 2N^{2/3}l\Big) - CN^{-1/3} &\leq \mathbb{E}^{\mathrm{G}\beta\mathrm{E}}\Big[F\Big(N \int_{N^{-2/3}r-l}^{4N^{-2/3+\epsilon}} \operatorname{Im} m_N(2 + x + \mathrm{i}\eta)\mathrm{d}x\Big)\Big] \\ &\leq \mathrm{TW}_\beta(r) + CN^{-1/3}. \end{aligned}$$

A similar upper and lower bound can be obtained in the same way when we consider $+l$ in the integral domain instead of $-l$. Since the Tracy–Widom distributions have smooth and uniformly bounded densities and $l = N^{-1+7\epsilon}$, we have

$$\sup_{r_0<r<N^\epsilon} \left| \mathbb{E}^{\mathrm{G}\beta\mathrm{E}} \Big[F\Big(N \int_{N^{-2/3}r\pm l}^{4N^{-2/3+\epsilon}} \mathrm{Im}\, m_N(2+x+\mathrm{i}\eta)\mathrm{d}x \Big) \Big] - \mathrm{TW}_\beta(r) \right| = O(N^{-1/3+7\epsilon}). \tag{3.4}$$

Using the Green function comparison theorem, Theorem 1.4, there exists some $c_0 > 0$ independent of ϵ such that

$$\sup_{r_0<r<N^\epsilon} \left| \Big(\mathbb{E} - \mathbb{E}^{\mathrm{G}\beta\mathrm{E}} \Big) \Big[F\Big(N \int_{N^{-2/3}r\pm l}^{4N^{-2/3+\epsilon}} \mathrm{Im}\, m_N(2+x+\mathrm{i}\eta)\mathrm{d}x \Big) \Big] \right| \le N^{-1/3+c_0\epsilon}, \tag{3.5}$$

for sufficiently large N. In combination with (3.2) and (3.4), we choose $\epsilon < \frac{\omega}{\max\{c_0,7\}}$ in the setting of Theorem 1.3 and obtain

$$\sup_{r_0<r<N^\epsilon} \left| \mathbb{P}\Big(N^{2/3}(\lambda_N - 2) < r \Big) - \mathrm{TW}_\beta(r) \right| \le N^{-1/3+\omega}. \tag{3.6}$$

Together with (3.1), we have hence completed the proof of Theorem 1.3.

We now move on to the proof of the Green function comparison theorem, Theorem 1.4. In the following, we first consider complex Hermitian Wigner matrices, as the complex Hermitian case is slightly easier than the real symmetric case. The proof of the Green function comparison theorem in the real symmetric setup is presented in Sect. 8.

Proof of Theorem 1.4. Consider the matrix Ornstein–Uhlenbeck process $\big(h_{ab}(t)\big)_{a,b=1}^N$:

$$\mathrm{d}h_{ab}(t) = \frac{1}{\sqrt{N}}\mathrm{d}\beta_{ab}(t) - \frac{1}{2}h_{ab}(t)\mathrm{d}t, \quad h_{ab}(0) = (H_N)_{ab}, \tag{3.7}$$

where $\big(\beta_{ab}(t)\big)_{a<b}^N$ are independent complex standard Brownian motions, $\big(\beta_{aa}(t)\big)_{a=1}^N$ are independent real standard Brownian motions, $\big(\beta_{ab}(t)\big)_{a<b}$ are independent from $\big(\beta_{aa}(t)\big)_{a=1}^N$, and $\beta_{ba}(t) = \overline{\beta_{ab}(t)}$. The initial condition H_N is a complex Hermitian Wigner matrix satisfying Assumption 1.1. In distribution the above is equivalent to writing

$$H(t) \overset{\mathrm{d}}{=} \mathrm{e}^{-\frac{t}{2}} H_N + \sqrt{1-\mathrm{e}^{-t}}\mathrm{GUE}_N, \quad t \in \mathbb{R}^+, \tag{3.8}$$

where GUE_N belongs to the GUE. For any $t \in \mathbb{R}^+$, $z \in \mathbb{C} \setminus \mathbb{R}$, we define

$$G(t,z) := \frac{1}{H(t)-z}; \quad m_N(t,z) := \frac{1}{N}\mathrm{Tr}G(t,z). \tag{3.9}$$

Recalling the local law Theorem 2.1 and Lemma 2.1, we obtain that the local law for $G(t,z)$,

$$\max_{i,j} |G_{ij}(t,z) - \delta_{ij} m_{sc}(z)| \prec \Psi(z); \quad |m_N(t,z) - m_{sc}(z)| \prec \frac{1}{N\eta}, \tag{3.10}$$

holds uniformly in $z \in S$ given in (2.7) and $t \ge 0$. Indeed, we choose a mesh of the interval $0 \le t \le T := 8\log N$ of size N^{10}, and obtain that (3.10) holds uniformly in

$z \in S, t \in [0, 8 \log N]$ from the continuity of the process (3.8) in time. Moreover, (3.10) also holds uniformly in $t \geq 8 \log N$ from (3.34) below.

In the following, we often ignore the parameters and write for short

$$H \equiv H(t), \quad h_{ab} \equiv h_{ab}(t), \quad G \equiv G(t, z), \quad m_N \equiv m_N(t, z), \quad t \in \mathbb{R}^+, \ z \in \mathbb{C} \setminus \mathbb{R}.$$

For a fixed small $\epsilon > 0$ and some $C_0 > 0$, let

$$N^{-1+\epsilon} \leq \eta \leq N^{-2/3+\epsilon}, \quad |\kappa_1|, |\kappa_2| \leq C_0 N^{-2/3+\epsilon}, \tag{3.11}$$

with $\kappa_1 < \kappa_2$. In view of (2.19) and (2.23), we are interested in the quantity

$$\mathcal{X} \equiv \mathcal{X}(t) := N \int_{\kappa_1}^{\kappa_2} \operatorname{Im} m_N(t, 2 + x + \mathrm{i}\eta) \mathrm{d}x, \quad t \in \mathbb{R}^+. \tag{3.12}$$

Hence $\mathcal{X}$ is a function of t, η as well as κ_1 and κ_2.

Let $F : \mathbb{R} \to \mathbb{R}$ be an arbitrary smooth function with uniformly bounded derivatives. The next lemma determines the evolution of the observable $F\big(\mathcal{X}(t)\big)$ under the Ornstein–Uhlenbeck flow in (3.7). To alleviate the notation, we introduce the following abbreviations. Let $P : \mathbb{R}^+ \times \mathbb{C} \setminus \mathbb{R} \longrightarrow \mathbb{C}$ be an arbitrary function, then we introduce

$$\widetilde{\operatorname{Im}}\, P \equiv \widetilde{\operatorname{Im}}\, P(t, z) := \frac{1}{2\mathrm{i}}(P(t, z) - P(t, \bar{z})), \quad t \in \mathbb{R}^+, \quad z \in \mathbb{C} \setminus \mathbb{R}. \tag{3.13}$$

For example, for complex Wigner matrices, $\widetilde{\operatorname{Im}}\, G_{ij}(t, z) \neq \operatorname{Im} G_{ij}(t, z)$, unless $i = j$. Further, we abbreviate, for $t \in \mathbb{R}^+$, and $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$,

$$\begin{aligned} \Delta\widetilde{\operatorname{Im}}\, P &\equiv (\Delta\widetilde{\operatorname{Im}}\, P)(t, z_1, z_2) \\ &:= \frac{1}{2\mathrm{i}}\Big(P(t, z_2) - P(t, \overline{z_2})\Big) - \frac{1}{2\mathrm{i}}\Big(P(t, z_1) - P(t, \overline{z_1})\Big), \end{aligned} \tag{3.14}$$

where the spectral parameters are given as

$$z_1 = 2 + \kappa_1 + \mathrm{i}\eta, \quad z_2 = 2 + \kappa_2 + \mathrm{i}\eta, \tag{3.15}$$

with κ_1, κ_2, and η from (3.11). In particular, we have $z_1, z_2 \in S_{\text{edge}}$ defined in (4.1) below.

Returning to $F(\mathcal{X})$, Ito's lemma yields the following result.

Lemma 3.1. *The observable $F(\mathcal{X})$ satisfies the following stochastic differential equation:*

$$\mathrm{d}F(\mathcal{X}) = \mathrm{d}M + \Theta \mathrm{d}t, \tag{3.16}$$

with the diffusion term

$$\mathrm{d}M = -\frac{1}{\sqrt{N}} \sum_{a,b=1}^{N} \Big(F'(\mathcal{X}) \Delta\widetilde{\operatorname{Im}}\, G_{ba}\Big) \mathrm{d}\beta_{ab}, \tag{3.17}$$

and the drift term $\Theta \equiv \Theta(t, z_1, z_2)$ is explicitly given in (3.25) below. Moreover, $\mathbb{E}[\Theta]$ can be written as

$$\mathbb{E}[\Theta] = \sum_{\substack{p+q+1=3 \\ p,q \in \mathbb{N}}}^{4} K_{p,q+1} + E_2 + O_{\prec}(N^{-1/2}), \tag{3.18}$$

for N sufficiently large, with

$$K_{p,q+1} := \frac{1}{2p!q!N^{\frac{p+q+1}{2}}} \sum_{\substack{a,b=1 \\ a\neq b}}^{N} s_{ab}^{(p,q+1)} \mathbb{E}\Big[\frac{\partial^{p+q} F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}}\, G_{ba}}{\partial h_{ba}^p \partial h_{ab}^q}\Big]; \tag{3.19}$$

$$E_2 := \frac{1}{2N} \sum_{a=1}^{N} (s_{aa}^{(2)} - 1)\mathbb{E}\Big[\frac{\partial F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}}\, G_{aa}}{\partial h_{aa}}\Big], \tag{3.20}$$

where $s_{ab}^{(p,q+1)} \equiv s_{ab}^{(p,q+1)}(t)$ and $s_{aa}^{(2)} \equiv s_{aa}^{(2)}(t)$ are the cumulants of the rescaled time dependent entries $\sqrt{N}h_{ab}$ defined in (2.24) and (2.27).

Remark 3.1. The diffusion term $\mathrm{d}M$ in (3.17) yields a martingale $M(t)$ upon integration. Note that the operator norm of the Green function has the deterministic bound $\|G(z)\|_2 \leq \frac{1}{\eta} \leq N^{1-\epsilon}$, given $z = E + \mathrm{i}\eta$ with $\eta \geq N^{-1+\epsilon}$. Since F has bounded derivatives, $|F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\, G_{ba}| = O(N^{1-\epsilon})$. Thus $M(t)$ is a true martingale with vanishing expectation.

Remark 3.2. In (3.18), only cumulants of order three and higher appear, i.e. $p+q+1 \geq 3$. This is a consequence of our assumption that the second moments of the off-diagonal matrix entries match with the Gaussian ensembles; see item b.) in Assumption 1.1.

Proof of Lemma 3.1. Recall the dynamics of the Orstein–Uhlenbeck process in (3.7) and that G is a function of the matrix entries h_{ab}. Using the first Ito's lemma and then the relation

$$\frac{\partial G_{ij}}{\partial h_{ab}} = -G_{ia}G_{bj}, \tag{3.21}$$

we compute

$$\begin{aligned}
\mathrm{d}G_{ij}(t,z) &= \frac{\partial G_{ij}}{\partial t}\mathrm{d}t + \sum_a \frac{\partial G_{ij}}{\partial h_{aa}}\mathrm{d}h_{aa} + \frac{1}{2}\sum_a \frac{\partial^2 G_{ij}}{\partial h_{aa}\partial h_{aa}}\mathrm{d}h_{aa}\mathrm{d}h_{aa} \\
&\quad + \sum_{a<b} \frac{\partial G_{ij}}{\partial h_{ab}}\mathrm{d}h_{ab} + \sum_{a<b} \frac{\partial G_{ij}}{\partial \overline{h_{ab}}}\mathrm{d}\overline{h_{ab}} + \sum_{a<b} \frac{\partial^2 G_{ij}}{\partial h_{ab}\partial \overline{h_{ab}}}\mathrm{d}h_{ab}\mathrm{d}\overline{h_{ab}} \\
&= -\frac{1}{\sqrt{N}} \sum_{a,b=1}^{N} G_{ia}G_{bj}\mathrm{d}\beta_{ab} \\
&\quad + \frac{1}{2}\sum_{a,b=1}^{N} \Big(h_{ab}G_{ia}G_{bj} + \frac{1}{N}G_{ib}G_{bj}G_{aa} + \frac{1}{N}G_{ia}G_{aj}G_{bb}\Big)\mathrm{d}t.
\end{aligned} \tag{3.22}$$

In view of $\mathcal{X}$ from (3.12), we take the normalized trace of the Green function and the imaginary part. Using the symmetry of H and

$$G_{ij}(z) = \overline{G_{ji}(\bar{z})},$$

we obtain the following stochastic differential equation:

$$\mathrm{d}\Big(\mathrm{Im}\, m_N(t,z)\Big) = -\frac{1}{2\mathrm{i}N^{3/2}} \sum_{i,a,b=1}^{N} \Big(G_{ia}G_{bi}(z) - G_{ia}G_{bi}(\bar{z})\Big)\mathrm{d}\beta_{ab}$$

$$
\begin{aligned}
&+\frac{1}{4N\mathrm{i}}\sum_{i,a,b=1}^{N}\Big[h_{ab}\Big(G_{ia}G_{bi}(z)-G_{ia}G_{bi}(\bar{z})\Big)\\
&+\frac{1}{N}\Big(G_{ib}G_{bi}G_{aa}(z)-G_{ib}G_{bi}G_{aa}(\bar{z})\Big)\\
&+\frac{1}{N}\Big(G_{ia}G_{ai}G_{bb}(z)-G_{ia}G_{ai}G_{bb}(\bar{z})\Big)\Big]\mathrm{d}t\\
=&-\frac{1}{N^{3/2}}\sum_{i,a,b=1}^{N}\widetilde{\mathrm{Im}}\,(G_{ia}G_{bi})\mathrm{d}\beta_{ab}+\frac{1}{2N}\sum_{i,a,b=1}^{N}\Big[h_{ab}\widetilde{\mathrm{Im}}\,(G_{ia}G_{bi})\\
&+\frac{1}{N}\widetilde{\mathrm{Im}}\,\big(G_{ib}G_{bi}G_{aa}+G_{ia}G_{ai}G_{bb}\big)\Big]\mathrm{d}t,
\end{aligned}
$$

where we use the notation from (3.13).

Using Ito's formula similarly on $F(\mathcal{X})$ and combining with (3.22), we obtain the stochastic differential Eq. (3.16), with the diffusion term given by

$$
\mathrm{d}M=-F'(\mathcal{X})\Big(\int_{\kappa_1}^{\kappa_2}\frac{1}{\sqrt{N}}\sum_{i,a,b=1}^{N}\widetilde{\mathrm{Im}}\,\Big(G_{ia}G_{bi}(t,2+x+\mathrm{i}\eta)\Big)\mathrm{d}x\Big)\mathrm{d}\beta_{ab},\tag{3.23}
$$

and the drift term given by (we omit the parameters t and $z=2+x+\mathrm{i}\eta$ of the Green functions)

$$
\begin{aligned}
\Theta=&\frac{1}{2}\sum_{i,a,b=1}^{N}h_{ab}\Big(F'(\mathcal{X})\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,(G_{ia}G_{bi})\mathrm{d}x\Big)\\
&+\frac{1}{2N}\sum_{i,a,b=1}^{N}F'(\mathcal{X})\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,\Big(G_{ib}G_{bi}G_{aa}+G_{ia}G_{ai}G_{bb}\Big)\mathrm{d}x\\
&+\frac{1}{2}F''(\mathcal{X})\frac{1}{N}\sum_{i,j=1}^{N}\sum_{a,b=1}^{N}\Big(\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,(G_{ia}G_{bi})\mathrm{d}x\Big)\Big(\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,(G_{jb}G_{aj})\mathrm{d}x\Big).
\end{aligned}\tag{3.24}
$$

Using $G^2(z)=\frac{\mathrm{d}}{\mathrm{d}z}G(z)$ and the definition of $\widetilde{\mathrm{Im}}$ in (3.13), we write

$$
\begin{aligned}
&\sum_{i=1}^{N}\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,\Big((G_{ia}G_{bi})(t,2+x+\mathrm{i}\eta)\Big)\mathrm{d}x\\
&\quad=\int_{\kappa_1}^{\kappa_2}\widetilde{\mathrm{Im}}\,\Big(\frac{\mathrm{d}G_{ba}}{\mathrm{d}x}(t,2+x+\mathrm{i}\eta)\Big)\mathrm{d}x=(\Delta\widetilde{\mathrm{Im}}\,G_{ba})(t,z_1,z_2),
\end{aligned}
$$

with $\Delta\widetilde{\mathrm{Im}}$ defined in (3.14) and z_1,z_2 given in (3.15). Applied to the martingale term (3.23) we find (3.17). Applied to the drift term (3.24), we find

$$
\begin{aligned}
\Theta=&\frac{1}{2}\sum_{a,b=1}^{N}h_{ab}\Big(F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\,G_{ba}\Big)\\
&+\frac{1}{2N}\sum_{a,b=1}^{N}\Big(F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\,(G_{aa}G_{bb})+F''(\mathcal{X})(\Delta\widetilde{\mathrm{Im}}\,G_{ab})(\Delta\widetilde{\mathrm{Im}}\,G_{ba})\Big).
\end{aligned}\tag{3.25}
$$

Next, we take the expectation of Θ and apply the cumulant expansions in Lemma 2.4 with respect to the independent entries h_{ab} in the first term on the right of (3.25). Using the relation (3.21), we compute

$$\begin{aligned}\frac{\partial F'(\mathcal{X})}{\partial h_{ba}} &= F''(\mathcal{X})\sum_{i=1}^{N}\int_{\kappa_1}^{\kappa_2}\frac{\partial(\operatorname{Im} G_{ii})}{\partial h_{ba}}\mathrm{d}x \\ &= -F''(\mathcal{X})\sum_{i=1}^{N}\int_{\kappa_1}^{\kappa_2}\widetilde{\operatorname{Im}}\,(G_{ib}G_{ai})\mathrm{d}x = -F''(\mathcal{X})\Delta\widetilde{\operatorname{Im}}\,G_{ab}. \end{aligned} \tag{3.26}$$

We first apply cumulant expansions to the complex-valued off-diagonal entries h_{ab}, *i.e.,* let $a \neq b$ in the summations in (3.25). Then by direct computations and using Assumption 1.1 (b.), the second order terms in the cumulant expansions corresponding to $s_{ab}^{(1,1)}(t) \equiv 1$ are canceled with the second term on the right of (3.25) with $a \neq b$. The third and fourth order terms in the cumulant expansions, corresponding to $p+q+1 \in \{3,4\}$, are given in (3.19). We stop the cumulant expansion at $l = 4$ and the corresponding truncation error $R_5 = \sum_{a\neq b} R_5^{(ab)}$ is estimated as follows.

We have from (2.26) that

$$\begin{aligned}|R_5^{(ab)}| \le{}& C\mathbb{E}[|h_{ab}|^5]\mathbb{E}\Big[\max_{p+q=4}\Big\{\sup_{|w|\le N^{-1/2+\gamma}}\Big|\frac{\partial^{p+q}}{\partial h_{ba}^p\partial h_{ab}^q}f_{ab}\Big(H^{(ab)}+wE^{(ba)}+\bar{w}E^{(ab)}\Big)\Big|\Big\}\Big] \\ &+ C\mathbb{E}\Big[|h_{ab}|^5 1_{|h_{ab}|>N^{-1/2+\gamma}}\Big] \\ &\mathbb{E}\Big[\max_{p+q=4}\Big\{\sup_{w\in\mathbb{C}}\Big|\frac{\partial^{p+q}}{\partial h_{ba}^p\partial h_{ab}^q}f_{ab}\Big(H^{(ab)}+wE^{(ba)}+\bar{w}E^{(ab)}\Big)\Big|\Big\}\Big], \end{aligned} \tag{3.27}$$

with a fixed small $\gamma > 0$, and where we use the notation $E^{(ab)} := (\delta_{ab})_{i,j=1}^N$, $H^{(ab)} := H - h_{ab}E^{(ab)} - h_{ba}E^{(ba)}$, as well as

$$f_{ab}(H) := F'(\mathcal{X})\Delta\widetilde{\operatorname{Im}}\,G_{ba}. \tag{3.28}$$

Using the second resolvent identity, we can write

$$G_{ij}^{H^{(ab)}} = G_{ij}^H + \Big(G^{H^{(ab)}}(h_{ab}E^{(ab)} + h_{ba}E^{(ba)})G^H\Big)_{ij}. \tag{3.29}$$

From the local law in (3.10), we have $\max_{i\neq j}|G_{ij}^H| \prec \Psi$ and $\max_i|G_{ii}^H| \prec 1$. In addition, we have $|h_{ij}| \prec \frac{1}{\sqrt{N}}$ from the moment condition (1.3). Therefore, we have from (3.29) that $\max_{i\neq j}|G_{ij}^{H^{(ab)}}| \prec \Psi$ and $\max_i|G_{ii}^{H^{(ab)}}| \prec 1$. Similarly, we have

$$G_{ij}^{H^{(ab)}+wE^{(ab)}+\bar{w}E^{(ba)}} = G_{ij}^{H^{(ab)}} - \Big(G^{H^{(ab)}+wE^{(ab)}+\bar{w}E^{(ba)}}(wE^{(ab)}+\bar{w}E^{(ba)})G^{H^{(ab)}}\Big)_{ij}, \tag{3.30}$$

and thus

$$\sup_{|w|<N^{-1/2+\gamma}}\Big\{\max_{i,j}\Big|G_{ij}^{H^{(ab)}+wE^{(ab)}+\bar{w}E^{(ba)}}\Big|\Big\} \prec 1. \tag{3.31}$$

Combining with (3.21), (3.26), and the fact that F in (3.28) has bounded derivatives, we obtain that

$$\sup_{|w|<N^{-1/2+\gamma}} \left| \frac{\partial^{p+q}}{\partial h_{ba}^p \partial h_{ab}^q} f_{ab}\Big(H^{(ab)} + wE^{(ba)} + \bar{w}E^{(ab)}\Big)\right| \prec 1.$$

Together with $\mathbb{E}|h_{ij}|^5 \le CN^{-5/2}$ under Assumption 1.1 and Lemma 1.1, the first term on the right side of (3.27) is bounded by $O_\prec(N^{-5/2})$. Note that for $z = E + \mathrm{i}\eta$ with $\eta \ge N^{-1+\epsilon}$, we have the deterministic upper bound for $\max_{i,j}|G_{ij}| \le \|G\|_2 \le \frac{1}{\eta} = O(N^{1-\epsilon})$. So the conditions of statement (3) of Lemma 1.1 are satisfied, and we can directly bound the expectation of the first term on the right side of (3.27).

We next estimate the second term on the right side of (3.27). Using the deterministic bound $\max_{i,j}|G_{ij}| = O(N^{1-\epsilon})$, we have from (3.21), (3.26) and the fact that F in (3.28) has bounded derivatives that

$$\max_{p+q=4}\left\{\sup_{w\in\mathbb{C}}\left|\frac{\partial^{p+q}}{\partial h_{ba}^p \partial h_{ab}^q} f_{ab}\Big(H^{(ab)} + wE^{(ba)} + \bar{w}E^{(ab)}\Big)\right|\right\} = O(N^{5-5\epsilon}).$$

Combining with the moment condition (1.3) and Hölder's inequality, the second term on the right side of (3.27) can also be bounded by $O_\prec(N^{-5/2})$. Thus the truncation error R_5 in the cumulant expansions satisfies $|R_5| = O_\prec(N^{-1/2})$. Throughout the paper, we will use similar arguments as above to estimate the error terms stemming from cutting cumulant expansions at some fixed order without specifically mentioning it.

Concerning the terms involving the diagonal entries h_{aa} in (3.25), we apply the cumulant expansion for real-valued random variables in Lemma 2.4 and stop at the second order $l = 2$. The resulting second order term in combination with the second sum in (3.25) with $a \equiv b$ is given by E_2 in (3.20) and the truncation error is similarly bounded by $O_\prec(N^{-1/2})$. We have hence finished the proof of Lemma 3.1.

Having established Lemma 3.1, we next estimate the expectation of the drift term $\mathbb{E}[\Theta]$ in (3.18) in the next proposition, whose proof is postponed to Sect. 7.

Proposition 3.1. *The drift term* $\mathbb{E}[\Theta]$ *in* (3.18) *has the following bound:*

$$|\mathbb{E}[\Theta(t, z_1, z_2)]| \le N^{-1/3+c\epsilon}, \tag{3.32}$$

uniformly in $t \ge 0$ *and* z_1, z_2 *given in* (3.15)*, for a numerical constant* $c > 0$ *that does not depend on* ϵ *and sufficiently large* $N \ge N_0(\epsilon, C_0)$.

In order to finish the proof of Theorem 1.4, we now choose $T := 8\log N$ and integrate (3.16) over $[0, T]$. Then taking the expectation, the diffusion term vanishes (see Remark 3.1) and the drift term is bounded using (3.32). We hence find by writing out $\mathcal{X}$ in (3.12) that

$$\begin{aligned}&\left|\mathbb{E}\Big[F\Big(N\int_{\kappa_1}^{\kappa_2} \operatorname{Im} m_N(0, 2+x+\mathrm{i}\eta)\mathrm{d}x\Big)\Big] - \mathbb{E}\Big[F\Big(N\int_{\kappa_1}^{\kappa_2} \operatorname{Im} m_N(T, 2+x+\mathrm{i}\eta)\mathrm{d}x\Big)\Big]\right| \\ &\quad = O(N^{-\frac{1}{3}+c\epsilon}\log N).\end{aligned} \tag{3.33}$$

Using the inequality $\|A\|_{\max} \le \|A\|_2 \le N\|A\|_{\max}$, the second resolvent identity, that $\|G(E+\mathrm{i}\eta)\|_2 \le \frac{1}{\eta}$, and (3.8), one shows that $G(T, z)$ is sufficiently close to the Green function of the GUE, *i.e.,*

$$\|G(T,z)-G^{\mathrm{GUE}}(z)\|_{\max} \le \|G(T,z)(\mathrm{GUE}-H(T))G^{\mathrm{GUE}}(z)\|_2 \le \frac{N}{\eta^2}\|(\mathrm{GUE}-H(T))\|_{\max} \prec \frac{1}{N^3\eta^2}. \tag{3.34}$$

Since F is a smooth function with uniformly bounded derivatives, we have

$$\left|F\Big(N\int_{\kappa_1}^{\kappa_2}\mathrm{Im}\, m_N(T,2+x+\mathrm{i}\eta)\mathrm{d}x\Big) - F\Big(N\int_{\kappa_1}^{\kappa_2}\mathrm{Im}\, m_N^{\mathrm{GUE}}(2+x+\mathrm{i}\eta)\mathrm{d}x\Big)\right| \prec \frac{N^{\epsilon}}{N^{8/3}\eta^2}. \tag{3.35}$$

Combining (3.33) and (3.35), we conclude the proof of Theorem 1.4.

Remark 3.3. In the traditional approach to the Green function comparison theorem [19] a Lindeberg type replacement strategy is used. In (3.8) we use a continuous flow to interpolate between Wigner matrices and the invariant ensembles. This is notationally easier than the Lindeberg replacement, especially when we do recursive comparisons to estimate the contributions from the fourth order cumulants in Sect. 5.

4. A Special Case: Estimates on $\mathbb{E}[\mathrm{Im}\, m_N]$

In this section, we prove the simplest version of the Green function comparison theorem, Theorem 1.4, when $F(x)=x$. It then suffices to compare the expected normalized trace of the Green function of a Wigner matrix $\mathbb{E}[m_N(z)]$ with $\mathbb{E}^{\mathrm{GUE}}[m_N(z)]$. The ideas in this section will also be used to prove Proposition 3.1, which is a key ingredient to establish the Green function comparison theorem for a general function F. The proof for general functions F will rely on the estimate (4.3) in Proposition 4.1 below as an input.

Proposition 4.1. *Let H_N be a complex Wigner matrix satisfying Assumption 1.1 and recall the time dependent matrix $H(t)$ in (3.7). For any $\epsilon>0$ and $C_0>0$, define the domain of the spectral parameter z near the upper edge,*

$$S_{\mathrm{edge}} \equiv S_{\mathrm{edge}}(\epsilon,C_0) := \{z=E+\mathrm{i}\eta\in S : |E-2|\le C_0N^{-2/3+\epsilon},\ N^{-1+\epsilon}\le\eta\le N^{-2/3+\epsilon}\}, \tag{4.1}$$

with S given in (2.7). Then for any $\tau>0$, we have

$$\left|\mathbb{E}[m_N(t,z)]-\mathbb{E}^{\mathrm{GUE}}[m_N(z)]\right| \le N^{-1/3+\tau}, \tag{4.2}$$

uniformly in $z\in S_{\mathrm{edge}}$ and $t\ge 0$, for sufficiently large $N\ge N_0(C_0,\epsilon,\tau)$. Furthermore, there exists some $C>0$ independent of ϵ, such that

$$\mathbb{E}[\mathrm{Im}\, m_N(t,z)] \le CN^{-1/3+\epsilon}, \tag{4.3}$$

uniformly in $z\in S_{\mathrm{edge}}$ and $t\ge 0$, for sufficiently large $N\ge N_0'(C_0,\epsilon)$.

In the rest of this section we prove Proposition 4.1; its proof is split into several parts organized in subsections.

4.1. Interpolation between a Wigner matrix and the GUE. Following the proof of Lemma 3.1 in Sect. 3, we start by applying Ito's lemma to the time dependent normalized trace of the Green function, $m_N(t,z)$, from (3.9). We find using (3.22) that

$$\begin{aligned} \mathrm{d}(m_N(t,z)) &= -\frac{1}{N^{3/2}}\sum_{v,a,b=1}^{N} G_{va}G_{bv}\mathrm{d}\beta_{ab} \\ &\quad + \frac{1}{2N}\sum_{v,a,b=1}^{N}\Big(h_{ab}G_{va}G_{bv} + \frac{1}{N}G_{vb}G_{bv}G_{aa} + \frac{1}{N}G_{va}G_{av}G_{bb}\Big)\mathrm{d}t \\ &:= \mathrm{d}M_0 + \Theta_0\mathrm{d}t, \end{aligned} \tag{4.4}$$

with diffusion term $\mathrm{d}M_0$ and drift term $\Theta_0\mathrm{d}t \equiv \Theta_0(t,z)\mathrm{d}t$; here we use the subscript 0 to indicate that we are considering the simple case $F(x) = x$. The diffusion term $\mathrm{d}M_0$ yields a martingale after integration; see Remark 3.1. Taking the expectation of the drift term and applying the cumulant expansions in Lemma 2.4, we have the analogue of (3.18),

$$\begin{aligned} \mathbb{E}[\Theta_0] &= \frac{1}{2N^2}\sum_{v,a=1}^{N}(s_{aa}^{(2)}-1)\mathbb{E}\Big[\frac{\partial(G_{av}G_{va})}{\partial h_{aa}}\Big] \\ &\quad + \frac{1}{2N}\sum_{\substack{v,a,b=1\\ a\neq b}}^{N}\sum_{p+q+1=3}^{4}\frac{1}{p!q!}\frac{s_{ab}^{(p,q+1)}}{N^{\frac{p+q+1}{2}}}\mathbb{E}\Big[\frac{\partial^{p+q}(G_{bv}G_{va})}{\partial h_{ba}^{p}\partial h_{ab}^{q}}\Big] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big) \\ &= -\frac{1}{2N^2}\sum_{v,a=1}^{N}(s_{aa}^{(2)}-1)\mathbb{E}\Big[\frac{\partial^2 G_{vv}}{\partial h_{aa}^2}\Big] \\ &\quad - \sum_{p+q+1=3}^{4}\frac{1}{2p!q!N^{\frac{p+q+3}{2}}}\sum_{\substack{v,a,b=1\\ a\neq b}}^{N}s_{ab}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^{p+q+1}G_{vv}}{\partial h_{ba}^{p}\partial h_{ab}^{q+1}}\Big] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big), \end{aligned} \tag{4.5}$$

where the error stems from the truncation of the cumulant expansions at fourth order. Recalling the arguments in Sect. 3, in order to establish Proposition 4.1 it suffices to show that for any $\tau > 0$,

$$|\mathbb{E}[\Theta_0(t,z)]| \le N^{-1/3+\tau}, \tag{4.6}$$

uniformly in $z \in S_{\mathrm{edge}}(\epsilon, C_0)$ and $t \ge 0$, for sufficiently large $N \ge N_0(C_0,\epsilon,\tau)$.

Admitting (4.6), for $T = 8\log N$ and any $0 \le t' \le T$, we integrate (4.4) over $[t', T]$ and take the expectation to get

$$\Big|\mathbb{E}\Big[m_N(t',z)\Big] - \mathbb{E}\Big[m_N(T,z)\Big]\Big| = O\Big(N^{-1/3+\tau}\log N\Big). \tag{4.7}$$

Combining with (3.34), we obtain the comparison estimate in (4.2) between the GUE and the time dependent $H(t)$ in (3.7) staring from the Wigner matrix H. The bound (4.3) will follow directly from the comparison result (4.2) and the corresponding estimate for the GUE in Lemma 5.2 below.

In the remaining part of this section, we will hence prove (4.6). For that it suffices to estimate the terms on the right side of (4.5).

4.2. Third and fourth order terms with unmatched indices. Using the differential rule for the Green function entries in (3.21), each term in the cumulant expansion (4.5) can be written out in terms of an averaged product of the Green function entries. The first group of terms on the right side of (4.5) is given by

$$-\frac{2}{N^2}\sum_{v,a=1}^{N}(s_{aa}^{(2)}-1)\mathbb{E}\Big[G_{va}G_{av}G_{aa}\Big].$$

In the second group of terms on the right side of (4.5), one example of a third order term with $p=1, q=1$ and one example of a fourth order term with $p=2, q=1$ are given by,

$$\sqrt{N}\frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(1,2)}}{2}\mathbb{E}\Big[G_{va}G_{bv}G_{aa}G_{bb}\Big],\quad -\frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(2,2)}}{4}\mathbb{E}\Big[G_{va}G_{av}G_{aa}G_{bb}G_{bb}\Big].$$

We remark that the third order terms with $p+q+1=3$ are averaged products of Green function entries with an additional leading factor $\sqrt{N}$.

To study these averaged products of the Green function entries in (4.5), we introduce the general form in (4.8) below. We will use the letters v_j to denote the free summation indices running from 1 to N, and the letters x_i, y_i as the row and column indices of the Green function entries. In order to avoid confusion, we clarify that $x_i = y_i = v_j$ means that both x_i and y_i represent the same summation index v_j. Further we write $x_i \neq y_i$ if x_i and y_i represent two distinct summation indices, say v_j and $v_{j'}$. They could have the same value as the summation indices v_j and $v_{j'}$ run from 1 to N.

We are now ready to introduce the general form of averaged products of the Green function entries:

$$\frac{1}{N^m}\sum_{v_1=1}^{N}\cdots\sum_{v_m=1}^{N}c_{v_1,\ldots,v_m}\Big(\prod_{i=1}^{n}G_{x_iy_i}(t,z)\Big)=:\frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}}c_{\mathcal{I}}\Big(\prod_{i=1}^{n}G_{x_iy_i}(t,z)\Big),$$
$$t\in\mathbb{R}^+, z\in\mathbb{C}^+, \tag{4.8}$$

for $m,n\in\mathbb{N}$, where $\mathcal{I}:=\{v_j\}_{j=1}^m$ is a free summation index set which may include a,b,v from (4.5), $m:=\#\{\mathcal{I}\}$ is the number of the free summation indices, and the coefficients $\{c_{\mathcal{I}}:=c_{v_1,\ldots,v_m}\}$ are uniformly bounded complex numbers. Moreover, n is the number of Green function entries in the product, and each row index x_i and column index y_i $(1\le i\le n)$ of the Green function entries represent some element in the free summation index set $\mathcal{I}$.

We further define the degree of such a term in (4.8) to be the number of off-diagonal terms in the product of the Green function entries, *i.e.,*

$$d:=\#\{1\le i\le n: x_i\neq y_i\}. \tag{4.9}$$

In particular, we have $0\le d\le n$. We use $\mathcal{Q}_d\equiv\mathcal{Q}_d(t,z)$ to denote the collection of the averaged products of the Green function entries of the form in (4.8) of degree d. For any $Q_d\equiv Q_d(t,z)\in\mathcal{Q}_d$, it is clear from the local law in (3.10) that

$$|Q_d(t,z)|\prec\Psi^d+\frac{1}{N}, \tag{4.10}$$

uniformly in $z \in S$ given in (2.7) and $t \geq 0$. We will often omit the parameters z and t for notational simplicity. The last error N^{-1} is from the coincidence of distinct summation indices.

Now we first look at the third order terms in the cumulant expansion (4.5) with $p + q + 1 = 3$. Using the differential rule for the Green function entries in (3.21), all the third order terms with $p + q + 1 = 3$ can be written out in the form in (4.8), with an extra factor $\sqrt{N}$ in front. We observe that these terms are unmatched, see Definition 4.1 below, since the indices a, b both appear an odd number of times in the product of the Green function entries.

In a similarly way, the fourth order terms in the cumulant expansion (4.5) with $p+q+1 = 4$, except the ones corresponding to $p = 2, q = 1$, are also unmatched terms of the form in (4.8) from Definition 4.1, since the number of times the index a (or b) appears in the row index set does not agree with the number of times it appears in the column index set of the product of Green function entries.

Definition 4.1 (*Terms with unmatched indices*). Given any $Q_d \in \mathcal{Q}_d$ of the form in (4.8) of degree d, let $\nu_j^{(r)}$, $\nu_j^{(c)}$, be the number of times the free summation index $v_j \in \mathcal{I}$ appears as the row, respectively column, index in the product of the Green function entries, *i.e.,*

$$\nu_j^{(r)} := \#\{1 \leq i \leq n : x_i = v_j\}, \quad \nu_j^{(c)} := \#\{1 \leq i \leq n : y_i = v_j\}, \quad 1 \leq j \leq m. \tag{4.11}$$

We define the set of the unmatched summation indices as

$$\mathcal{I}^o := \{1 \leq j \leq m : \nu_j^{(r)} \neq \nu_j^{(c)}\} \subset \mathcal{I}.$$

If $\mathcal{I}^o$ is empty, *i.e.,* all the free summation indices appear the same number of times in the row index set $\{x_i\}$ and the row column index set $\{y_i\}$, then we say that Q_d is matched. Otherwise, we say Q_d is an unmatched term, denoted by Q_d^o. The collection of the unmatched terms of the form in (4.8) of degree d is denoted by $\mathcal{Q}_d^o \subset \mathcal{Q}_d$.

Given any unmatched term $Q_d^o \in \mathcal{Q}_d^o$, we define the unmatched index set for both row and column as

$$\mathcal{R}^o := \{1 \leq j \leq m : \nu_j^{(r)} > \nu_j^{(c)}\} \subset \mathcal{I}^o; \quad \mathcal{C}^o := \{1 \leq j \leq m : \nu_j^{(r)} < \nu_j^{(c)}\} \subset \mathcal{I}^o. \tag{4.12}$$

Neither of $\mathcal{R}^o$ and $\mathcal{C}^o$ is empty. Moreover, $\mathcal{R}^o \cap \mathcal{C}^o$ is empty, and $\mathcal{R}^o \cup \mathcal{C}^o = \mathcal{I}^o$.

Next, we give two examples of unmatched terms, which appear as fourth order terms in (4.5),

$$\begin{aligned} &-\frac{1}{N^3} \sum_{v,a,b} \frac{s_{ab}^{(1,3)}}{4} \mathbb{E}\Big[G_{va} G_{bv} G_{ba} G_{aa} G_{bb}\Big] \in \mathcal{Q}_3^o; \\ &-\frac{1}{N^3} \sum_{v,a,b} \frac{s_{ab}^{(0,4)}}{12} \mathbb{E}\Big[G_{va} G_{bv} G_{ba} G_{ba} G_{ba}\Big] \in \mathcal{Q}_5^o; \end{aligned} \tag{4.13}$$

and two examples of the unmatched terms from the third order terms on the right side of (4.5),

$$\frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(1,2)}}{2}\mathbb{E}\Big[G_{va}G_{bv}G_{aa}G_{bb}\Big]\in\mathcal{Q}_2^o;\quad \frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(0,3)}}{4}\mathbb{E}\Big[G_{va}G_{bv}G_{ba}G_{ba}\Big]\in\mathcal{Q}_4^o, \tag{4.14}$$

up to a factor of $\sqrt{N}$.

The following proposition states that the expectations of the unmatched terms are much smaller than their naive size obtained by the power counting from the local law as in (4.10). The proof is postponed to Sect. 6.

Proposition 4.2. *Consider any unmatched term $Q_d^o\in\mathcal{Q}_d^o$ of degree d with fixed n (the number of Green function entries in the product) given in (4.8). For any fixed $D\in\mathbb{N}$, we have*

$$\mathbb{E}[Q_d^o(t,z)]=O_{\prec}\Big(\frac{1}{N}+\Psi^D\Big), \tag{4.15}$$

uniformly in $z\in S$ given in (2.7) and $t\geq 0$.

Remark 4.1. In the observable $Q_d^o(t,z)$ in (4.15) the Green function entries from (4.8) are all chosen at the same spectral parameter $z\in S$. Our proofs can be extended to the setting where the Green function entries are evaluated at different spectral parameters in the domain S with the estimate in (4.15) holding true. As we do not require this generalization to prove Proposition 4.1 we do not pursue this direction here.

Therefore, using Proposition 4.2, the third order terms in the cumulant expansion (4.5) are all bounded as $O_{\prec}(N^{-1/2}+\sqrt{N}\Psi^D)$. Moreover all the fourth order terms in the cumulant expansion (4.5), except the one corresponding to $p=2,q=1$, are bounded by $O_{\prec}(N^{-1}+\Psi^D)$. By choosing $D\geq\frac{1}{\epsilon}$ with $\epsilon>0$ as in (2.9), we hence obtain from (4.5) that

$$\begin{aligned}\mathbb{E}[\Theta_0]=&-\frac{1}{2N^2}\sum_{v,a=1}^{N}(s_{aa}^{(2)}-1)\mathbb{E}\Big[\frac{\partial^2 G_{vv}}{\partial h_{aa}^2}\Big]-\frac{1}{4N^3}\sum_{\substack{v,a,b=1\\ a\neq b}}^{N}s_{ab}^{(2,2)}\mathbb{E}\Big[\frac{\partial^4 G_{vv}}{\partial h_{ba}^2\partial h_{ab}^2}\Big]\\ &+O_{\prec}(\frac{1}{\sqrt{N}}).\end{aligned} \tag{4.16}$$

The remaining terms on the right side of (4.16) are matched under Definition 4.1. It is thus sufficient to estimate these matched terms, as presented in the next subsection.

4.3. Terms with matched indices. Applying the differentiation rule (3.21) to the right side of (4.16), the index v appears once as a row index and once as a column index of the Green function entries of the resulting terms on the right side of (4.16). In addition, the indices a,b from (4.16) will take a special role and appear twice as a row index and twice as a column index of the Green function entries. After differentiation by (3.21), we write out these products of Green function entries and observe that they are of the following form which we call type-AB terms.

Definition 4.2 (*Type-AB terms, type-A terms, type-0 terms*). For arbitrary $m, n \in \mathbb{N}$, we consider averaged products of Green functions of the form

$$\frac{1}{N^{m+2}} \sum_{v_1=1}^{N} \cdots \sum_{v_m=1}^{N} \sum_{a=1}^{N} \sum_{b=1}^{N} c_{a,b,v_1,\ldots,v_m} \Big(\prod_{i=1}^{n} G_{x_i y_i}(t,z) \Big) =: \frac{1}{N^{\#\mathcal{I}+2}} \sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}} \Big(\prod_{i=1}^{n} G_{x_i y_i} \Big), \tag{4.17}$$

for $t \in \mathbb{R}^+$, $z \in \mathbb{C}^+$, where each x_i and y_i represent the free summation indices a, b or v_j $(1 \le j \le m)$. Here the coefficients $\{c_{a,b,\mathcal{I}} := c_{a,b,v_1,\ldots,v_m}\}$ are uniformly bounded complex numbers. Note that the form in (4.17) is a special case of the form given in (4.8) with the two indices a and b singled out. The degree, denoted by d, of such a term is defined as in (4.9) by counting the number of the off-diagonal Green function entries. Recall $\nu_j^{(r)}$, $\nu_j^{(c)}$ defined in (4.11). We further define similarly

$$\nu_a^{(r)} := \#\{i : x_i = a\},\ \nu_a^{(c)} := \#\{i : y_i = a\},\ \nu_b^{(r)} := \#\{i : x_i = b\},\ \nu_b^{(c)} := \#\{i : y_i = b\},$$

for the special indices a, b.

A *type-AB term*, denoted by P_d^{AB}, is of the form in (4.17) with each v_j appearing once in the row index set $\{x_i\}$ and once in the column index set $\{y_i\}$ in the product of the Green function entries, *i.e.*, $\nu_j^{(r)} = \nu_j^{(c)} = 1$. The indices a and b both appear the same number of times (more than once) in the row index set $\{x_i\}$ and column index set $\{y_i\}$ in the product of the Green function entries, *i.e.*, $\nu_a^{(r)} = \nu_a^{(c)} \ge 2$ and $\nu_b^{(r)} = \nu_b^{(c)} \ge 2$. We denote by $\mathcal{P}_d^{AB} \equiv \mathcal{P}_d^{AB}(t,z)$ the collection of the type-AB terms of degree d. We remark that type-AB terms are matched in the sense of Definition 4.1.

A *type-A term*, denoted by P_d^{A}, is of the form in (4.17) with $\nu_a^{(r)} = \nu_a^{(c)} \ge 2$, and $\nu_b^{(r)} = \nu_b^{(c)} = \nu_j^{(r)} = \nu_j^{(c)} = 1$ for $1 \le j \le m$. We denote the collection of the type-A terms of degree d by $\mathcal{P}_d^{A} \equiv \mathcal{P}_d^{A}(t,z)$.

Finally, a *type-0 term*, denoted by P_d, is of the form in (4.17) with all the free summation indices appearing once in the row index set $\{x_i\}$ and once in the column index set $\{y_i\}$ in the product of the Green function entries, *i.e.*, $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = \nu_j^{(r)} = \nu_j^{(c)} = 1$ for $1 \le j \le m$. We denote the collection of the type-0 terms of degree d by $\mathcal{P}_d \equiv \mathcal{P}_d(t,z)$.

We remark that the index b does no longer play a special role in type-A terms, as well as the indices a and b are not special in type-0 terms. We keep them in the notation in order to emphasize the inheritance from the form in (4.17).

Next, we give two examples for type-AB terms, which are generated from the fourth order expansion terms in (4.16) corresponding to the (2, 2)-cumulants,

$$-\frac{1}{4N^3} \sum_{v,a,b} s_{ab}^{(2,2)} \Big(G_{va} G_{aa} G_{av} G_{bb} G_{bb} \Big) \in \mathcal{P}_2^{AB};$$

$$-\frac{1}{4N^3} \sum_{v,a,b} s_{ab}^{(2,2)} \Big(G_{va} G_{ab} G_{bv} G_{aa} G_{bb} \Big) \in \mathcal{P}_3^{AB};$$

and an example of a type-A term, which is from the second order terms of diagonal entries in the cumulant expansion (4.16),

$$-\frac{1}{2N^2} \sum_{v,a} (s_{aa}^{(2)} - 1) \Big(G_{va} G_{aa} G_{av} \Big) \in \mathcal{P}_2^{A},$$

where the index b no longer takes the special role.

In the following, we only consider special type-AB terms with both indices a and b appearing in the product of the Green function entries four times (*i.e.*, $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = 2$) and the corresponding type-A terms. For the general case, see Remark 4.2.

The next proposition claims that, in expectation, any type-AB term as well as any type-A term of degree d can be expanded into linear combinations of type-0 terms of degrees at least d up to negligible error. The proof of Proposition 4.3 is presented in Sect. 5.2.

Proposition 4.3. *Consider any type-AB term $P_d^{AB} \in \mathcal{P}_d^{AB}$ of the form in (4.8) of degree d with fixed $n \in \mathbb{N}$, and $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = 2$. Then for any fixed $D \in \mathbb{N}$, we have*

$$\mathbb{E}[P_d^{AB}(t,z)] = \sum_{\substack{P_{d'} \in \mathcal{P}_{d'} \\ d \le d' < D}} \mathbb{E}[P_{d'}(t,z)] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{4.18}$$

uniformly in $z \in S$ (see (2.7)), $t \in \mathbb{R}^+$, where we use $\sum_{P_{d'} \in \mathcal{P}_{d'}, d \le d' < D} \mathbb{E}[P_{d'}(t,z)]$ to denote a sum of finitely many type-0 terms of the form in (4.17) of degrees d' satisfying $d \le d' < D$. Moreover, the number of type-0 terms in the sum above is bounded by $(6(n+8D))^{2D}$ and the number of the Green function entries in each type-0 term is bounded by $n + 8D$.

Similarly, for any type-A term $P_d^A \in \mathcal{P}_d^A$ of the form in (4.17) with $\nu_a^{(r)} = \nu_a^{(c)} = 2$, we have

$$\mathbb{E}[P_d^{A}(t,z)] = \sum_{\substack{P_{d'} \in \mathcal{P}_{d'} \\ d \le d' < D}} \mathbb{E}[P_{d'}(t,z)] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{4.19}$$

uniformly in $z \in S$ and $t \in \mathbb{R}^+$. The number of the type-0 terms in the sum above is at most $(6(n+4D))^D$, and the number of the Green function entries in each type-0 term is at most $n + 4D$.

Remark 4.2. The above expansions also hold true if we consider a slightly generalized setup when both indices a and b appear arbitrary even number of times, not limited to $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = 2$. Then the number of expansions generated on the right side also depends on the values $\nu_a^{(r)} (= \nu_a^{(c)})$ and $\nu_b^{(r)} (= \nu_b^{(c)})$; see also Remark 5.1. Furthermore, in the above all the Green function entries are taken at the same spectral parameter $z \in S$. The expansion results may be generalized to the setting when the Green functions are taken at different spectral parameters in the domain S, c.f. Remark 4.1.

Armed with Proposition 4.3, we return to (4.16). Recalling Definition 4.2 and using (3.21), the second group of terms on the right side of (4.16) can be written out as type-AB terms of the form in (4.17) of degrees satisfying $d \ge 2$, where the number of Green function entries in each type-AB term is $n = 5$, the summation index set $\mathcal{I} = \{v\}$ and the coefficients $c_{a,b,v} = s_{ab}^{(2,2)}$. Similarly, the first group of terms on the right side of (4.16) can be written as a type-A term with degree $d = 2$ and the number of Green

function entries $n = 3$. Therefore, from Proposition 4.3, we can expand (4.16) as a sum of finitely many type-0 terms of degrees at least two, *i.e.,*

$$\mathbb{E}[\Theta_0(t,z)] = \sum_{\substack{P_d \in \mathcal{P}_d \\ 2 \le d \le D-1}} \mathbb{E}[P_d(t,z)] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{4.20}$$

uniformly in $z \in S$ and $t \in \mathbb{R}^+$, where the number of type-0 terms in the sum above can be bounded by $(CD)^{cD}$, for some numerical constants C, c.

Having expanded $\mathbb{E}[\Theta_0(t,z)]$ into type-0 terms, we next estimate the size of type-0 terms of the form in (4.17) of degree $d \ge 2$ in the edge scaling, *i.e.,* when the spectral parameter z is chosen in the domain S_{edge} defined in (4.1). The proof of Lemma 4.1 is presented in Sect. 5.3.

Lemma 4.1. *For any type-0 term $P_d \in \mathcal{P}_d$ of the form in (4.17) of degree $d \ge 2$ with fixed $n \in \mathbb{N}$, we have*

$$|\mathbb{E}[P_d(t,z)]| = O_{\prec}(N^{-1/3}), \tag{4.21}$$

uniformly in $z \in S_{\text{edge}}$ given by (4.1) and $t \ge 0$.

We hence obtain the estimate of $\mathbb{E}[\Theta_0(t,z)]$ in (4.6) by combining (4.20) and (4.21), and by choosing $D \ge \frac{1}{\epsilon}$ and using the upper bound in (2.9). This yields the proof of Proposition 4.1.

5. Product of Green Function Entries with Matched Indices

In this section, we prove Proposition 4.3 and Lemma 4.1. Before diving into their proofs, we outline in the next subsection the intuition stemming from the GUE.

5.1. Intuition from the GUE. In this subsection, we focus on the special case of the GUE. The idea of eliminating the indices appearing more than twice and reducing type-AB to type-0 terms as in Proposition 4.3 stems from explicit computations for the GUE based on the Weingarten calculus for Haar unitary matrices. To simplify the arguments, we only consider the following example of a type-AB term of the form in (4.17),

$$\frac{1}{N^2}\sum_{a,b}(G_{aa}(z))^2(G_{bb}(z))^2 \in \mathcal{P}_0^{AB}. \tag{5.1}$$

Thanks to the unitary conjugation invariance, we know that the eigenvalues (λ_i) and the corresponding orthonormal eigenvectors $(\mathbf{u}_i)$ of a GUE matrix are independent, and that the collection of eigenvectors $U := (\mathbf{u}_1, \ldots, \mathbf{u}_N)$ is distributed according to Haar measure on the unitary group $U(N)$.

Further, using the spectral decomposition

$$G(z) = \frac{1}{H - z} = \sum_{j=1}^{N} \frac{\mathbf{u}_j \mathbf{u}_j^*}{\lambda_j - z}, \quad z \in S, \tag{5.2}$$

we write the expectation of (5.1) as

$$\begin{aligned}
&\frac{1}{N^2}\sum_{a,b}\mathbb{E}[(G_{aa}(z))^2(G_{bb}(z))^2] \\
&= \frac{1}{N^2}\sum_{a,b}\sum_{j,k,p,q}\mathbb{E}\Big[\frac{\mathbf{u}_j(a)\overline{\mathbf{u}_j(a)}\mathbf{u}_k(a)\overline{\mathbf{u}_k(a)}\mathbf{u}_p(b)\overline{\mathbf{u}_p(b)}\mathbf{u}_q(b)\overline{\mathbf{u}_q(b)}}{(\lambda_j-z)(\lambda_k-z)(\lambda_p-z)(\lambda_q-z)}\Big] \\
&= \frac{1}{N^2}\sum_{a,b}\sum_{j,k,p,q}\mathbb{E}\Big[\frac{1}{(\lambda_j-z)(\lambda_k-z)(\lambda_p-z)(\lambda_q-z)}\Big] \\
&\quad\times\mathbb{E}[U_{aj}U_{ak}U_{bp}U_{bq}\overline{U_{aj}U_{ak}U_{bp}U_{bq}}]. \qquad (5.3)
\end{aligned}$$

In order to estimate the expectations of the eigenvectors above, we use the following result for the Weingarten calculus on the unitary groups [10,11].

Lemma 5.1 (Corollary 2.4, Proposition 2.6 in [11]). *Let $U = (U_{ij})_{i,j=1}^N$ be a Haar unitary random matrix of size N. Let $n \in \mathbb{N}$ and denote by S_n the symmetric group of order n. Then, for arbitrary column and row indices $i_k, i'_k, j_k, j'_k \in [\![1, N]\!]$, $1 \le k \le n$, we have*

$$\begin{aligned}
&\mathbb{E}[U_{i_1j_1}\cdots U_{i_nj_n}\overline{U_{i'_1j'_1}}\cdots\overline{U_{i'_nj'_n}}] \\
&\quad= \sum_{\alpha,\beta\in S_n}\delta_{i_1,i'_{\alpha(1)}}\cdots\delta_{i_n,i'_{\alpha(n)}}\delta_{j_1,j'_{\beta(1)}}\cdots\delta_{j_n,j'_{\beta(n)}}\mathrm{Wg}(N,\alpha^{-1}\beta), \qquad (5.4)
\end{aligned}$$

where $\mathrm{Wg}(N,\gamma)$ is the Weingarten function given by

$$\mathrm{Wg}(N,\gamma) := \mathbb{E}[U_{11}\cdots U_{nn}\overline{U_{1\gamma(1)}}\cdots\overline{U_{n,\gamma(n)}}], \quad \gamma\in S_n. \qquad (5.5)$$

In the limit of large N, the Weingarten function $\mathrm{Wg}(N,\gamma)$ has the following asymptotic behavior: Let $\{c_i\}_{i=1}^{\#(\gamma)}$ denotes the cycles of $\gamma \in S_n$, with $\#(\gamma)$ the total number of cycles. Then

$$\mathrm{Wg}(N,\gamma) = N^{\#(\gamma)-2n}\prod_{i=1}^{\#(\gamma)}(-1)^{|c_i|-1}\mathrm{Cat}(|c_i|-1) + O(N^{\#(\gamma)-2n-2}), \qquad (5.6)$$

where $|c_i|$ denotes the length of the cycle c_i and $\mathrm{Cat}(k) = \frac{(2k)!}{k!(k+1)!}$ is the k-th Catalan number.

Now we are ready to evaluate, for large N, $\mathbb{E}[U_{aj}U_{ak}U_{bp}U_{bq}\overline{U_{aj}U_{ak}U_{bp}U_{bq}}]$ from (5.3) using Lemma 5.1 with $n = 4$. We may assume that $a \neq b$, as the case $a = b$ only contributes $O(N^{-1})$ to the expectation of (5.1) uniformly for $z \in S$, using the local law (3.10) and Lemma 2.1. We set $n = 4$, $i_1 = i_2 = i'_1 = i'_2 = a$, $i_3 = i_4 = i'_3 = i'_4 = b$, $j_1 = j'_1 = j$, $j_2 = j'_2 = k$, $j_3 = j'_3 = p$, and $j_4 = j'_4 = q$. Since $\max_{\gamma\in S_n}\#(\gamma) = 4$, the leading term in (5.4), corresponding to $\mathrm{Wg}(N,\gamma)$ with $\gamma = \mathbb{1}$ ($\alpha^{-1}\beta = \mathbb{1}$), is of size $O(\frac{1}{N^4})$ from (5.6) and the rest terms are bounded by $O(\frac{1}{N^5})$. Moreover, the coefficient in front of $\mathrm{Wg}(N,\mathbb{1})$ is given by the number of permutations $\sigma \in S_4$ such that

$$i_l = i'_{\sigma(l)}, \quad j_l = j'_{\sigma(l)}, \quad l = 1,2,3,4. \qquad (5.7)$$

We then separate into the following five cases: (1.) all indices j, k, p, q are distinct, (2.) only two of them coincide while the other two are distinct, (3.) two pairs of them coincide, (4.) three of them coincide and the rest one is different, and (5.) all the indices are the same. As $a \neq b$, the number of permutations satisfying (5.7) is given by 1, 8, 6, 8 and 4, respectively. Therefore, for $a \neq b$, we obtain

$$\begin{aligned}\mathbb{E}[(G_{aa}(z))^2(G_{bb}(z))^2] = \frac{1}{N^4} \sum_{\substack{j,k,p,q \\ \text{all distinct}}} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)(\lambda_k - z)(\lambda_p - z)(\lambda_q - z)}\Big] \\ \times \Big(1 + O\big(\frac{1}{N}\big)\Big) \\ + \frac{8}{N^4} \sum_{\substack{j,p,q \\ \text{all distinct}}} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)^2(\lambda_p - z)(\lambda_q - z)}\Big]\Big(1 + O\big(\frac{1}{N}\big)\Big) \\ + \frac{6}{N^4} \sum_{j \neq q} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)^2(\lambda_q - z)^2}\Big]\Big(1 + O\big(\frac{1}{N}\big)\Big) \\ + \frac{8}{N^4} \sum_{j \neq q} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)^3(\lambda_q - z)}\Big]\Big(1 + O\big(\frac{1}{N}\big)\Big) \\ + \frac{4}{N^4} \sum_{j} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)^4}\Big]\Big(1 + O\big(\frac{1}{N}\big)\Big). \end{aligned} \tag{5.8}$$

For example, by direct computation, the first term on the right side of (5.8) can be written using the spectral decomposition (5.2) as

$$\begin{aligned}&\frac{1}{N^4} \sum_{\substack{j,k,p,q \\ \text{all distinct}}} \mathbb{E}\Big[\frac{1}{(\lambda_j - z)(\lambda_k - z)(\lambda_p - z)(\lambda_q - z)}\Big] \\ &= \frac{1}{N^4}\mathbb{E}[(\mathrm{Tr}G)^4] - \frac{6}{N^4}\mathbb{E}[(\mathrm{Tr}G^2)(\mathrm{Tr}G)^2] \\ &\quad + \frac{8}{N^4}\mathbb{E}[(\mathrm{Tr}G^3)(\mathrm{Tr}G)] - \frac{6}{N^4}\mathbb{E}[\mathrm{Tr}G^4] + \frac{3}{N^4}\mathbb{E}[(\mathrm{Tr}G^2)(\mathrm{Tr}G^2)]. \end{aligned} \tag{5.9}$$

Observe that the resulting terms on the right side are type-0 terms under Definition 4.2. We further write the other terms on the right side of (5.8) similarly by type-0 terms using the spectral decomposition. To sum up, averaging over a, b and adding the subleading diagonal terms, (5.3) eventually becomes,

$$\begin{aligned}\frac{1}{N^2}\sum_{a,b}\mathbb{E}[(G_{aa})^2(G_{bb})^2] = \frac{1}{N^4}\mathbb{E}[(\mathrm{Tr}G)^4] + \frac{2}{N^4}\mathbb{E}[(\mathrm{Tr}G^2)(\mathrm{Tr}G)^2] \\ + \frac{1}{N^4}\mathbb{E}[(\mathrm{Tr}G^2)(\mathrm{Tr}G^2)] + O(N^{-1}),\end{aligned}$$

uniformly in $z \in S$, after exact cancellations between the terms.

In this way, we have eliminated one pair of a-indices and b-indices from the type-AB term (5.1) and shown that they can be written as linear combinations of type-0 terms, which involves only products of traces.

For Wigner matrices, the above does not apply anymore as the eigenvectors are no longer exactly Haar distributed on $U(N)$, further the expectation in (5.3) does not factorize. Yet successively applying cumulant expansions, we can reduce type-AB terms to sums of type-A terms up to negligible error, and then finally reduce type-A terms to sums of type-0 terms. This procedure is explained in the next subsection.

5.2. Proof of Proposition 4.3. In this subsection, we give the proof of Proposition 4.3 for arbitrary Wigner matrices using cumulant expansions.

Proof of Proposition 4.3. We consider a type-AB term of the form in (4.17) with both indices a and b appearing twice as a row index and twice as a column index in the product of the Green function entries. There are two steps as follows. We first expand the type-AB term as a linear combination of type-A terms by eliminating one pair of the index b. Then in a second step we expand the resulting type-A terms as linear combinations of type-0 terms by further eliminating a pair of the index a.

Step 1: Reduction to type-A terms. Given a type-AB term, we will eliminate one pair of the index b using the relation

$$G_{ij} = \delta_{ij}\underline{G} + G_{ij}\underline{HG} - \underline{G}(HG)_{ij}, \tag{5.10}$$

and then applying cumulant expansions. The identity may be checked directly from the definition of the Green function. In (5.10) we use the notation $\underline{A} := \frac{1}{N}\mathrm{Tr}A$, for any $A \in \mathbb{C}^{N\times N}$, to denote the normalized trace. Similar ideas were used in [22,30].

Consider now a type-AB term $P_d^{AB} \in \mathcal{P}_d^{AB}$ of the form in (4.17). We split into the following two cases.

Case 1: If there exists some i such that $x_i = y_i = b$, *i.e.,* there is a factor G_{bb} in the product of Green function entries, we may then assume $i = 1$. Applying (5.10) to G_{bb} and performing cumulant expansions for the resulting terms $\underline{HG}$ and $(HG)_{bb}$, we obtain

$$\begin{aligned}
\mathbb{E}[P_d^{AB}] &= \frac{1}{N^{\#\mathcal{I}+2}}\sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}}\mathbb{E}\Big[(\underline{G} + G_{bb}\underline{HG} - \underline{G}(HG)_{bb})\prod_{2\le i\le n} G_{x_i y_i}\Big] \\
&= \frac{1}{N^{\#\mathcal{I}+2}}\sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}}\mathbb{E}\Big[\underline{G}\prod_{2\le i\le n} G_{x_i y_i}\Big] \\
&\quad + \frac{1}{N^{\#\mathcal{I}+4}}\sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}}\mathbb{E}\Big[\frac{\partial G_{bb}G_{jk}\prod_{2\le i\le n} G_{x_i y_i}}{\partial h_{jk}}\Big] \\
&\quad - \frac{1}{N^{\#\mathcal{I}+4}}\sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}}\mathbb{E}\Big[\frac{\partial G_{jj}G_{kb}\prod_{2\le i\le n} G_{x_i y_i}}{\partial h_{kb}}\Big] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big),
\end{aligned} \tag{5.11}$$

where the error $O_{\prec}(\frac{1}{\sqrt{N}})$ is from the truncation of the cumulant expansions. Using (3.21), the first order of the second group of terms above corresponding to $\frac{\partial}{\partial h_{jk}}G_{jk}$ is precisely canceled by that of the third group of terms corresponding to $\frac{\partial}{\partial h_{kb}}G_{kb}$. Then we write

$$\mathbb{E}[P_d^{AB}] = \frac{1}{N^{\#\mathcal{I}+2}}\sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}}\mathbb{E}\Big[\underline{G}\prod_{2\le i\le n} G_{x_i y_i}\Big]$$

$$
\begin{aligned}
&- \frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{bb} \prod_{2\le i\le n} G_{x_i y_i}}{\partial h_{jk}} G_{jk} \Big] \\
&+ \frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{jj} \prod_{2\le i\le n} G_{x_i y_i}}{\partial h_{kb}} G_{kb} \Big] \\
&+ O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big). \qquad (5.12)
\end{aligned}
$$

The first term on the right side above is obtained by replacing G_{bb} by the normalized trace $\underline{G}$ in the expression of P_d^{AB}. In this way we have eliminated one pair of the index b. Since the index b originally appeared twice as a row index and twice as a column index in the product of the Green function entries, the first term has become a type-A term of degree d. Moreover, from (3.21) and the fact that j, k are fresh indices, the other terms on the right side of (5.12) can be written out as a sum of $2n$ type-AB terms of the form in (4.17), where the corresponding free summation index set is $\mathcal{I}' = \{\mathcal{I}, j, k\}$, $m' = \#\mathcal{I}' = m + 2$, and the number of Green function entries is $n' = n + 2$.

We next study the degrees of these terms in detail. In the second group of summation in (5.12), if $\partial/\partial h_{jk}$ acts on G_{bb}, then the degree of the resulting term is increased by three, since j and k are fresh indices. If $\partial/\partial h_{jk}$ acts on $G_{x_i y_i}$ $(2 \le i \le n)$, then the degree is increased by at least two for the same reason. In the last group of summation in (5.12), if $\partial/\partial h_{kb}$ acts on G_{jj}, then the degree is increase by three. When $\partial/\partial h_{jk}$ acts on $G_{x_i y_i}$ $(2 \le i \le n)$, we split the discussion into three cases: 1) if $G_{x_i y_i}$ is diagonal and $x_i = y_i \ne b$, then the degree is increased by three; 2) if $G_{x_i y_i}$ is off-diagonal with $y_i \ne b$, then the degree is increased by two; 3) if $G_{x_i y_i}$ is off-diagonal with $y_i = b$, then the degree is increased by one.

Hence the degrees, denoted by d', of all the terms on the right side of (5.12) except the first one, satisfy $d' \ge d + 1$. We use $\sum_{P_{d'}^{AB} \in \mathcal{P}_{d'}^{AB}, d' \ge d+1} \mathbb{E}[P_{d'}^{AB}]$ to denote the finite sum of these terms, *i.e.,* we write

$$
\mathbb{E}[P_d^{AB}] = \frac{1}{N^{\#\mathcal{I}+2}} \sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\underline{G} \prod_{i=2}^{n} G_{x_i y_i} \Big] + \sum_{\substack{P_{d'}^{AB} \in \mathcal{P}_{d'}^{AB} \\ d' \ge d+1}} \mathbb{E}[P_{d'}^{AB}] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big). \qquad (5.13)
$$

Therefore, the combination of the identity (5.10) and the cumulant expansion gives a cancellation to first order, and the only leading term left is obtained by replacing a factor G_{bb} with the normalized trace $\underline{G}$ of the product of Green function entries in the expression of the original P_d^{AB}.

Case 2: If there is no i such that $x_i = y_i = b$, *i.e.,* there is no factor as G_{bb} in the product of Green function entries in (4.17), we may then assume that $x_1 = b$ and $y_1 \ne b$. Since the index b appears exactly twice in $\{y_i\}_{i=2}^n$, we may assume that $y_2 = y_3 = b$ and $x_2 \ne b$ and $x_3 \ne b$. Then there is no b in the remaining column index set $\{y_i\}_{i=4}^n$. Using the identity (5.10) on G_{by_1} and applying cumulant expansions, we find

$$
\begin{aligned}
\mathbb{E}[P_d^{AB}] = &\frac{1}{N^{\#\mathcal{I}+2}} \sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}} \mathbb{E}\Big[(G_{by_1} \underline{HG} - \underline{G}(HG)_{by_1}) G_{x_2 b} G_{x_3 b} \prod_{i=4}^{n} G_{x_i y_i} \Big] \\
&+ \frac{1}{N^{\#\mathcal{I}+2}} \sum_{\mathcal{I},a,b} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\delta_{by_1} \underline{G} \prod_{i=2}^{n} G_{x_i y_i} \Big]
\end{aligned}
$$

$$= -\frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{by_1} G_{x_2 b} G_{x_3 b} \prod_{i=4}^{n} G_{x_i y_i}}{\partial h_{jk}} G_{jk} \Big]$$
$$+ \frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{jj} G_{x_2 b} G_{x_3 b} \prod_{i=4}^{n} G_{x_i y_i}}{\partial h_{kb}} G_{ky_1} \Big] + O_{\prec}(\frac{1}{\sqrt{N}}), \tag{5.14}$$

where in the second step, we observe a cancellation to first order similarly as in (5.12), and the last error $O_{\prec}(N^{-1/2})$ is from the truncation of the cumulant expansions at the third order, while the contribution from the diagonal case $b \equiv y_1$, *i.e.,* the second line of (5.14), can be bounded by $O_{\prec}(N^{-1})$ using the local law in (3.10). From (3.21), the right side of (5.14) can again be written as a sum of $2n$ type-AB terms of the form in (4.17) with $\mathcal{I}' = \{\mathcal{I}, j, k\}$, $m' = m + 2$, and $n' = n + 2$. Since j, k are fresh indices, the resulting type-AB terms have degrees $d' \geq d + 1$ (the finite sum of such terms is denoted by $\sum_{P_{d'}^{AB} \in \mathcal{P}_{d'}^{AB}, d' \geq d+1} \mathbb{E}[P_{d'}^{AB}]$), except the following two terms corresponding to taking $\frac{\partial}{\partial h_{kb}}$ of a Green function entry whose column index coincides with b, *i.e.,*

$$\frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[G_{jj} G_{x_2 k} G_{bb} G_{x_3 b} \prod_{i=4}^{n} G_{x_i y_i} G_{ky_1} \Big] \tag{5.15}$$

and

$$\frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[G_{jj} G_{x_2 b} G_{x_3 k} G_{bb} \prod_{i=4}^{n} G_{x_i y_i} G_{ky_1} \Big]. \tag{5.16}$$

Compared with the original term P_d^{AB}, one observes that the terms in (5.15) and (5.16) are obtained by replacing one pair of the index b by a fresh index k and adding a factor G_{bb} for the replaced index b. These terms are again type-AB terms in $\mathcal{P}_d^{AB}$ with a factor G_{bb} in the product of Green function entries considered in Case 1. Using (5.13) on these terms and combining with (5.14), we hence obtain

$$\mathbb{E}[P_d^{AB}] = \frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[G_{jj} \underline{G} G_{ky_1} G_{x_2 k} G_{x_3 b} \prod_{i=4}^{n} G_{x_i y_i} \Big]$$
$$+ \frac{1}{N^{\#\mathcal{I}+4}} \sum_{\mathcal{I},a,b,j,k} c_{a,b,\mathcal{I}} \mathbb{E}\Big[G_{jj} \underline{G} G_{ky_1} G_{x_2 b} G_{x_3 k} \prod_{i=4}^{n} G_{x_i y_i} \Big]$$
$$+ \sum_{\substack{P_{d'}^{AB} \in \mathcal{P}_{d'}^{AB} \\ d' \geq d+1}} \mathbb{E}[P_{d'}^{AB}] + \sum_{\substack{P_{d''}^{AB} \in \mathcal{P}_{d''}^{AB} \\ d'' \geq d+1}} \mathbb{E}[P_{d''}^{AB}] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big), \tag{5.17}$$

where the first two lines above are type-A terms in $\mathcal{P}_d^A$, obtained from the original term P_d^{AB} by replacing a pair of index b, *i.e.,* (x_1, y_2) or (x_1, y_3) by a fresh index k and multiplied by $(\underline{G})^2$. The first group of sum on the last line of (5.17) comes from (5.14) excluding two terms (5.15) and (5.16), and the number of the type-AB terms in the sum is at most $2n - 2$. The second group of sum on the last line of (5.17) is obtained from expanding (5.15) and (5.16) by (5.13). The corresponding type-AB terms are of the form in (4.17) with $m'' = m' + 2$ and $n'' = n' + 2$, and the number of the terms in the sum is at most $4n'$.

Combining with Case 1, for any type-AB term $P_d^{AB} \in \mathcal{P}_d^{AB}$, we rewrite (5.13) and (5.17) in the short form

$$\mathbb{E}[P_d^{AB}] = \sum_{P_d^A \in \mathcal{P}_d^A} \mathbb{E}[P_d^A] + \sum_{\substack{P_{d'}^{AB} \in \mathcal{P}_{d'}^{AB} \\ d' \geq d+1}} \mathbb{E}[P_{d'}^{AB}] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big), \tag{5.18}$$

where the summations above denote a sum of at most two type-A terms of degree d and a sum of at most $(6n + 8)$ type-AB terms of degree not less than $d + 1$. The number of the Green function entries in the product (see (4.17)) of each term is at most $n + 4$.

Remark 5.1. In general, if the number of the index b appearing in the Green function entries of P_d^{AB} is not limited to four, *i.e.*, $\nu_b^{(r)} = \nu_b^{(c)} = s \geq 3$, then the terms in the first group of sum on the right side of (5.18) are of the form in (4.17) with $\nu_b^{(r)} = \nu_b^{(c)} = s - 1 \geq 2$. Moreover, the number of such terms in the first group of the sum is at most s. We can repeat the expansion procedure in (5.18) for s times until $\nu_b^{(r)} = \nu_b^{(c)} = 1$. We then end up with at most $s!$ type-A terms in $\mathcal{P}_d^A$, and at most $6s^s(n+4s)$ type-AB terms of degrees not less than $d + 1$ generated in the above expansion procedures.

Iterating the expansion procedure (5.18) $D - d$ times, the resulting type-AB terms have degrees at least D. Using the local law in (3.10), we expand an arbitrary type-AB term $P_d^{AB} \in \mathcal{P}_d^{AB}$ as a finite sum of type-A terms of degrees at least d, up to negligible error. We hence arrive at

$$\mathbb{E}[P_d^{AB}] = \sum_{d \leq d' < D} \sum_{P_{d'}^A \in \mathcal{P}_{d'}^A} \mathbb{E}[P_{d'}^A] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big). \tag{5.19}$$

The number of the Green function entries in the product of each type-A term above is bounded by $n+4D$, and the number of these type-A terms is bounded by $(6(n+4D))^D$.

Step 2: Reduction to type-0 terms. For the expanded type-A terms on the right side of (5.19), we follow the idea in Step 1 to expand the resulting type-A terms as linear combinations of type-0 terms by further eliminating one pair of the index a.

Given a type-A term $P_d^A \in \mathcal{P}_d^A$ of the form in (4.17), we split into two cases: 1) there exists a factor G_{aa} in the product of Green function entries; 2) there is no factor G_{aa} in the product of the Green function entries. We utilize similar arguments as in Case 1 and Case 2 of Step 1 above and obtain the analogue of (5.18), namely that

$$\mathbb{E}[P_d^A] = \sum_{P_d \in \mathcal{P}_d} P_d + \sum_{\substack{P_{d''}^A \in \mathcal{P}_{d''}^A \\ d'' \geq d+1}} \mathbb{E}[P_{d''}^A] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big), \tag{5.20}$$

where the summations above denote a sum of at most two type-0 terms of degree d and a sum of at most $(6n + 8)$ type-A terms of degrees at least $d + 1$. The number of the Green function entries in the product of each term is bound by $n + 4$.

Iterating the above expansion for $D - d$ times, we then expand an arbitrary type-A term $P_d^A \in \mathcal{P}_d^A$ as a sum of at most $(6(n + 4D))^D$ type-0 terms of degree d' satisfying $d \leq d' < D$, up to negligible error. As the analogue of (5.19), we write

$$\mathbb{E}[P_d^A] = \sum_{d \leq d' < D} \sum_{P_{d'} \in \mathcal{P}_{d'}} \mathbb{E}[P_{d'}] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{5.21}$$

where the number of the Green function entries in the product of each type-0 term above is bounded by $n + 4D$.

Combining with the first step (5.19), we finish the proof of Proposition 4.3.

Remark 5.2. The expansion procedures in the proof of Proposition 4.3 are not unique in the sense that at each step the resulting expansion will depend on the choice of the Green function entry we pick to be replaced by the identity (5.10) and perform cumulant expansions. However in view of Lemma 4.1 this is not pertinent to the proof of Proposition 4.1. This arbitrariness can be used to derive relations among Green function correlation functions. We finally remark that the errors $O_\prec(N^{-1/2})$ in (5.19) and (5.21) stemming from truncating the cumulant expansions at second order, can be improved to $O_\prec(N^{-1})$ because of the negligibility of third order terms; see Proposition 4.2.

5.3. Proof of Lemma 4.1. In the subsection, we estimate the expectations of the type-0 terms and prove Lemma 4.1. We start with the following lemma for the GUE.

Lemma 5.2. *Let H belong to the GUE. For any $\epsilon > 0$ and $C_0 > 0$, recall the domain $S_{\text{edge}} \equiv S_{\text{edge}}(\epsilon, C_0)$ defined in (4.1). Then there exists a constant C independent of ϵ such that*

$$\frac{1}{N}\mathbb{E}^{\text{GUE}}\Big[\operatorname{Im}\operatorname{Tr} G(z)\Big] \le CN^{-1/3+\epsilon}, \tag{5.22}$$

holds uniformly for all $z \in S_{\text{edge}}$, for sufficiently large $N \ge N_0(C_0, \epsilon)$. Furthermore, for any $\tau > 0$, all the type-0 terms $P_d \in \mathcal{P}_d$ ($d \ge 2$) of the form in (4.17) have the upper bound

$$|\mathbb{E}^{\text{GUE}}[P_d(z)]| \le N^{-1/3+\tau}, \tag{5.23}$$

uniformly for all $z \in S_{\text{edge}}$, for sufficiently large $N \ge N_0'(C_0, \epsilon, \tau)$.

The proof of Lemma 5.2 is postponed to Sect. 5.4. Using the above lemma for the GUE and the comparison method, we are now ready to prove Lemma 4.1 for arbitrary Wigner matrices.

Proof of Lemma 4.1. Consider any type-0 term $P_d \in \mathcal{P}_d$ of the form in (4.17) of degree $d \ge 2$. If $d \ge D$ for some large D, then by the local law in (3.10), $|\mathbb{E}[P_d]| = O_\prec(\Psi^D + N^{-1})$. Else, if d is smaller, we estimate $\mathbb{E}[P_d]$ using the comparison method iteratively and the corresponding estimates for the GUE in (5.23).

We start the iteration by denoting the type-0 term P_d of the form in (4.17) as $P_d \equiv P_{d_1}^{(1)}$, where the superscript (1) and degree $d \equiv d_1$ will be used to indicate the iteration step. We hence consider a term of the form

$$P_{d_1}^{(1)} \equiv P_{d_1}^{(1)}(t,z) : \quad \frac{1}{N^{\#\mathcal{I}_1+2}} \sum_{\mathcal{I}_1, a_1, b_1} c_{a_1,b_1,\mathcal{I}_1}\Big(\prod_{i=1}^{n_1} G_{x_i y_i}(t,z)\Big), \quad t \in \mathbb{R}^+,\ z \in S_{\text{edge}}, \tag{5.24}$$

with $n_1 = \#\mathcal{I}_1 + 2$, where each summation index in $\{a_1, b_1, \mathcal{I}_1\}$ appears exactly once in the row index set $\{x_i\}$ and exactly once in the column index set $\{y_i\}$. In the following, we often omit the parameters t, z and the errors below are always bounded uniformly in $z \in S_{\text{edge}}$ and $t \ge 0$.

We next derive the stochastic differential equation for the type-0 term $P_{d_1}^{(1)}$ under the Ornstein–Uhlenbeck flow in (3.7), similarly to (4.4). In general, for any $\{x_i, y_i\}_{i=1}^n$ with some $n \in \mathbb{N}$, using Ito's formula and the stochastic differential equation for the Green function entries in (3.22), we have

$$\begin{aligned}
\mathrm{d}\Big(\prod_{i=1}^n G_{x_i y_i}\Big) &= \sum_{j=1}^n \prod_{i\neq j} G_{x_i y_i} \mathrm{d}G_{x_j y_j} + \frac{1}{2}\sum_{j,k=1}^n \prod_{i\neq j,k} G_{x_i y_i} \mathrm{d}G_{x_j y_j} \mathrm{d}G_{x_k y_k} \\
&= -\frac{1}{\sqrt{N}} \sum_{a,b=1}^N \sum_{j=1}^n G_{x_j a} G_{b y_j} \prod_{i\neq j} G_{x_i y_i} \mathrm{d}\beta_{ab} \\
&+ \frac{1}{2}\sum_{a,b=1}^N \sum_{j=1}^n \Big(h_{ab} G_{x_j a} G_{b y_j} + \frac{1}{N} G_{x_j b} G_{b y_j} G_{aa} + \frac{1}{N} G_{x_j a} G_{a y_j} G_{bb} \Big) \\
&\prod_{i\neq j} G_{x_i y_i} \mathrm{d}t \\
&+ \frac{1}{2N} \sum_{a,b=1}^N \sum_{j,k=1}^n G_{x_j a} G_{b y_j} G_{x_k b} G_{a y_k} \prod_{i\neq j,k} G_{x_i y_i} \mathrm{d}t := \mathrm{d}\widehat{M} + \widehat{\Theta}\,\mathrm{d}t,
\end{aligned} \tag{5.25}$$

with diffusion term $\mathrm{d}\widehat{M}$ and drift term $\widehat{\Theta}\,\mathrm{d}t$. Applying cumulant expansions to the drift term, we observe cancellations of the second order expansions as in (4.5) and obtain that

$$\begin{aligned}
\mathbb{E}[\widehat{\Theta}] &= \frac{1}{2N}\sum_{a=1}^N (s_{aa}^{(2)} - 1) \sum_{j=1}^n \mathbb{E}\Big[\frac{\partial (G_{x_j a} G_{a y_j} \prod_{i\neq j} G_{x_i y_i})}{\partial h_{aa}}\Big] \\
&+ \frac{1}{2}\sum_{\substack{a,b=1 \\ a\neq b}}^N \sum_{p+q+1=3}^4 \frac{s_{ab}^{(p,q+1)}}{p!q!N^{\frac{p+q+1}{2}}} \\
&\times \sum_{j=1}^n \mathbb{E}\Big[\frac{\partial^{p+q}(G_{x_j a} G_{b y_j} \prod_{i\neq j} G_{x_i y_i})}{\partial h_{ba}^p \partial h_{ab}^q}\Big] + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big) \\
&= -\frac{1}{2N}\sum_{a=1}^N (s_{aa}^{(2)} - 1)\mathbb{E}\Big[\frac{\partial^2 (\prod_{i=1}^n G_{x_i y_i})}{\partial h_{aa}^2}\Big] \\
&- \sum_{p+q+1=3}^4 \frac{1}{2p!q!N^{\frac{p+q+1}{2}}} \sum_{\substack{a,b=1 \\ a\neq b}}^N s_{ab}^{(p,q+1)} \mathbb{E}\Big[\frac{\partial^{p+q+1}(\prod_{i=1}^n G_{x_i y_i})}{\partial h_{ba}^p \partial h_{ab}^{q+1}}\Big] \\
&+ O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big).
\end{aligned} \tag{5.26}$$

From (5.25) and (5.26), we find that $P_{d_1}^{(1)}$ in (5.24) satisfies the stochastic differential equation

$$\mathrm{d}(P_{d_1}^{(1)}) = \mathrm{d}M_{d_1}^{(1)} + \Theta_{d_1}^{(1)}\mathrm{d}t, \tag{5.27}$$

where the diffusion term $\mathrm{d}M_{d_1}^{(1)}$ yields a martingale after integration (see Remark 3.1) and the drift term $\Theta_{d_1}^{(1)}\mathrm{d}t$ satisfies the following analogue of (4.5),

$$\begin{aligned}\mathbb{E}[\Theta_{d_1}^{(1)}] = &-\frac{1}{2N}\sum_{a_2=1}^{N}(s_{a_2a_2}^{(2)}-1)\mathbb{E}\Big[\frac{\partial^2(P_{d_1}^{(1)})}{\partial h_{a_2a_2}^2}\Big]\\ &-\sum_{p+q+1=3}^{4}\frac{1}{2p!q!N^{\frac{p+q+1}{2}}}\sum_{\substack{a_2,b_2=1\\ a_2\neq b_2}}^{N}s_{a_2b_2}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^{p+q+1}(P_{d_1}^{(1)})}{\partial h_{b_2a_2}^p\partial h_{a_2b_2}^{q+1}}\Big]\\ &+O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big),\end{aligned}\tag{5.28}$$

where a_2, b_2 are fresh summation indices, as a, b in (5.26). The subscript 2 is used to indicate the iteration step and distinguish from a_1, b_1 in (5.24).

From (3.21), all the third order terms for $p+q+1=3$ in the cumulant expansion above can be written out in the form in (4.8), with an extra factor $\sqrt{N}$ in front. Since the fresh indices a_2, b_2 both appear an odd number of times in the product of the Green function entries, they are unmatched from Definition 4.1. Using Proposition 4.2, these term are bounded by $O_{\prec}(N^{-1/2}+\sqrt{N}\Psi^D)$.

The fourth order terms in the cumulant expansion with $p+q+1=4$ in (5.28), with the exception of those corresponding to $p=2, q=1$, are also unmatched terms of the form in (4.8), since the number of times the index a_2 (or b_2) appears in the row index set $\{x_i\}$ does not agree with the number of times it appears in the column index set $\{y_i\}$. Using Proposition 4.2, these term are bounded by $O_{\prec}(N^{-1}+\Psi^D)$.

By choosing $D \geq \frac{1}{\epsilon}$ with ϵ as in (2.9), we hence obtain the following analogue of (4.16)

$$\begin{aligned}\mathbb{E}[\Theta_{d_1}^{(1)}] = &-\frac{1}{2N}\sum_{a_2=1}^{N}(s_{a_2a_2}^{(2)}-1)\mathbb{E}\Big[\frac{\partial^2(P_{d_1}^{(1)})}{\partial h_{a_2a_2}^2}\Big]\\ &-\frac{1}{4N^2}\sum_{\substack{a_2,b_2=1\\ a_2\neq b_2}}^{N}s_{a_2b_2}^{(2,2)}\mathbb{E}\Big[\frac{\partial^4(P_{d_1}^{(1)})}{\partial h_{b_2a_2}^2\partial h_{a_2b_2}^2}\Big]+O_{\prec}(N^{-1/2})\,.\end{aligned}\tag{5.29}$$

It then suffices to estimate the remaining matched terms above. Using (3.21) and (5.24), the second group of terms on the right side of (5.29) can be written out in the form:

$$\frac{1}{N^{\#\mathcal{I}_1+4}}\sum_{\mathcal{I}_1,a_1,b_1,a_2,b_2}c_{a_1,b_1,a_2,b_2,\mathcal{I}_1}\Big(\prod_{i=1}^{n_1+4}G_{x_iy_i}\Big),\tag{5.30}$$

where the coefficients $\{c_{a_1,b_1,a_2,b_2,\mathcal{I}_1}\}$ are determined by $\{c_{a_1,b_1,\mathcal{I}_1}\}$ and $\{s_{a_2,b_2}^{(2,2)}\}$, and each summation index in $\{a_1, b_1, \mathcal{I}_1\}$ appears once in the row index set $\{x_i\}$ and once in the column index set $\{y_i\}$. Moreover, both indices a_2, b_2 appear exactly twice in the row index set $\{x_i\}$ and exactly twice in the column index set $\{y_i\}$. We define the degree of the form in (5.30) as in (4.9) by counting the number of off-diagonal Green function entries. Recall the definition of the type-AB, type-A and type-0 terms from Definition 4.2. The definitions can be adapted naturally with respect to the fresh indices a_2 and b_2, for the form given in (5.30).

Thus the second group of terms on the right side of (5.29) are $n_1(n_1+1)(n_1+2)(n_1+3)$ type-AB terms considered in Proposition 4.3 of degrees not less than d_1+1, from (3.21) and the fact that a_2, b_2 are fresh indices. Similarly, the first group of terms on the right side of (5.29) are $n_1(n_1+1)$ type-A terms of degrees not less than d_1+1. Using Proposition 4.3, we expand each of these terms as a sum of finitely many type-0 terms of degrees at least d_1+1, which are in the form:

$$\mathcal{P}_{d_2}^{(2)}: \quad \frac{1}{N^{\#\mathcal{I}_2+4}} \sum_{\mathcal{I}_2,a_1,b_1,a_2,b_2} c_{a_1,b_1,a_2,b_2,\mathcal{I}_2}\Big(\prod_{i=1}^{n_2} G_{x_i y_i}\Big), \tag{5.31}$$

where $\mathcal{I}_2$ is a set of free summation indices, the coefficients $\{c_{a_1,b_1,a_2,b_2,\mathcal{I}_2}\}$ are uniformly bounded complex numbers, and each index in $\{a_2, b_2, a_1, b_1, \mathcal{I}_2\}$ appears once in $\{x_i\}$ and once in $\{y_i\}$. In particular, $n_2 = \#\mathcal{I}_2+4$. The degree of such a term, denoted by d_2, is given as in (4.9). The collection of the type-0 terms of the form in (5.31) of degree d_2 is denoted by $\mathcal{P}_{d_2}^{(2)}$. Here we use the subscript 2 to indicate the iteration step. Note that the form in (5.31) is a special case of the form given in (4.8) and the indices a_1, b_1, a_2, b_2 do not take special roles. We keep them in the notation to emphasize the inheritance from (5.30). Then from Proposition 4.3, we expand (5.29) and write for short

$$\mathbb{E}[\Theta_{d_1}^{(1)}] = \sum_{\substack{P_{d_2}^{(2)}\in\mathcal{P}_{d_2}^{(2)} \\ d_1+1\le d_2<D}} \mathbb{E}[P_{d_2}^{(2)}] + O_{\prec}\big(N^{-1/2}+\Psi^D\big), \tag{5.32}$$

uniformly in $t \ge 0$ and $z \in S_{\text{edge}}$, where the summation above is over finitely many type-0 terms of the form in (5.31), and the number of these terms is determined by D and n_1.

We now return to the stochastic differential equation for $P_{d_1}^{(1)}$ in (5.27). Integrating (5.27) over $[t', T]$ for any $0 \le t' \le T = 8\log N$ and taking the expectation similarly to (4.7), we find from (5.32) that

$$\begin{aligned}\mathbb{E}[P_{d_1}^{(1)}(T,z)] - \mathbb{E}[P_{d_1}^{(1)}(t',z)] = &\sum_{\substack{P_{d_2}^{(2)}\in\mathcal{P}_{d_2}^{(2)} \\ d_1+1\le d_2<D}} \int_{t'}^{T} \mathbb{E}[P_{d_2}^{(2)}(t,z)]\mathrm{d}t \\ &+ O_{\prec}\big(\log N(N^{-1/2}+\Psi^D)\big).\end{aligned} \tag{5.33}$$

Using the local law in (3.10), (3.34) and (5.23), $\mathbb{E}[P_{d_1}^{(1)}(T,z)]$ is sufficiently close (up to an error $O(N^{-1})$) to $\mathbb{E}^{\text{GUE}}[P_{d_1}^{(1)}(z)]$, which can be bounded by $O_{\prec}(N^{-1/3})$. Hence it suffices to estimate $\mathbb{E}[P_{d_2}^{(2)}(t,z)]$ on the right side of (5.33), for $t \in [0,T]$, $z \in S_{\text{edge}}$ in (4.1).

Given any $P_{d_2}^{(2)} \in \mathcal{P}_{d_2}^{(2)}$ ($d_2 \ge d_1+1$) of the form in (5.31), if $d_1 = D-1$, we find $|\mathbb{E}[P_{d_2}^{(2)}(t,z)]| = O_{\prec}(\Psi^D + N^{-1})$ using the local law in (3.10). We then obtain from (5.33) that

$$|\mathbb{E}[P_{d_1}^{(1)}(t',z)]| = O_{\prec}\Big(\log N(N^{-1/2}+\Psi^D) + N^{-1/3}\Big), \tag{5.34}$$

uniformly in $t' \in [0,T]$ and $z \in S_{\text{edge}}$.

Else, if $d_1 \le D-2$, we repeat the above arguments for the resulting type-0 terms $P^{(2)}_{d_2} \in \mathcal{P}^{(2)}_{d_2}$ $(d_2 \ge d_1+1)$ on the right side of (5.33) as in (5.27). Using (5.25) and (5.26), we then create two fresh summation indices, denoted by a_3, b_3, to derive the evolution under the Ornstein–Uhlenbeck flow of any $P^{(2)}_{d_2} \in \mathcal{P}^{(2)}_{d_2}$. Similarly as in (5.29), the expectation of the corresponding drift terms is given by

$$\mathbb{E}[\Theta^{(2)}_{d_2}] = -\frac{1}{2N}\sum_{a_3=1}^{N}(s^{(2)}_{a_3a_3}-1)\mathbb{E}\Big[\frac{\partial^2(P^{(2)}_{d_2})}{\partial h^2_{a_3a_3}}\Big] - \frac{1}{4N^2}\sum_{\substack{a_3,b_3=1\\ a_3\neq b_3}}^{N} s^{(2,2)}_{a_3b_3}\mathbb{E}\Big[\frac{\partial^4(P^{(2)}_{d_2})}{\partial h^2_{b_3a_3}\partial h^2_{a_3b_3}}\Big] + O_{\prec}(N^{-1/2}). \tag{5.35}$$

From Definition 4.2, the right side above can be written out as linear combinations of type-A terms and type-AB terms, with respect to fresh summation indices a_3 and b_3, of degrees not less than d_2+1. Using Proposition 4.3, these terms can further be expanded by the type-0 terms of degrees at least d_2+1. In this way, we obtain an estimate similar to (5.34) for $d_1 = D-2$.

Next, we discuss the iterative mechanism to extend to any small $d_1 \ge 2$. In general, for any $s \ge 1$, we define a type-0 term in the s-th iteration step to be in the form of

$$\mathcal{P}^{(s)}_{d_s}: \quad \frac{1}{N^{\#\mathcal{I}_s+2s}}\sum_{\mathcal{I}_s,a_1,b_1,\ldots,a_s,b_s} c_{a_1,b_1,\ldots,a_s,b_s,\mathcal{I}_s}\mathbb{E}\Big[\prod_{i=1}^{n_s} G_{x_iy_i}(t,z)\Big], \tag{5.36}$$

where $\mathcal{I}_s$ is a set of free summation indices, the coefficients $\{c_{a_1,b_1,\ldots,a_s,b_s,\mathcal{I}_s}\}$ are uniformly bounded complex numbers, and each free summation index in $\{a_1,b_1,\ldots,a_s,b_s,\mathcal{I}_s\}$ appears once in $\{x_i\}$ and once in $\{y_i\}$. In particular, we have $n_s = \#\mathcal{I}_s + 2s$. The degree, denoted by d_s, of such a term in (5.36) is given as in (4.9) by counting the number of off-diagonal Green function entries. We denote by $\mathcal{P}^{(s)}_{d_s}$ the collection of the type-0 terms in the s-th step of the form in (5.36) of degree d_s. Note that the form in (5.36) is a special case of the form given in (4.8), in order to emphasize the s-th iteration step and the dependence on $\{a_s,b_s\}$.

We then derive the stochastic evolution for any $P^{(s)}_{d_s} \in \mathcal{P}^{(s)}_{d_s}$ $(s \ge 1)$, using (5.25) and (5.26) similarly as in (5.27) and (5.32). That is,

$$\mathrm{d}(P^{(s)}_{d_s}) = \mathrm{d}M^{(s)}_{d_s} + \Theta^{(s)}_{d_s}\mathrm{d}t, \tag{5.37}$$

where $\mathrm{d}M^{(s)}_{d}$ yields a martingale after integration, and $\mathbb{E}[\Theta^{(s)}_{d}]$ satisfies

$$\mathbb{E}[\Theta^{(s)}_{d_s}(t,z)] = \sum_{\substack{P^{(s+1)}_{d_{s+1}}\in\mathcal{P}^{(s+1)}_{d_{s+1}}\\ d_s+1\le d_{s+1}<D}} \mathbb{E}[P^{(s+1)}_{d_{s+1}}(t,z)] + O_{\prec}(N^{-1/2}+\Psi^D), \tag{5.38}$$

uniformly in $t \ge 0$ and $z \in S_{\mathrm{edge}}$, where the sums in (5.38) are over finitely many type-0 terms in the $(s+1)$-th step given in (5.36) and the number of such terms is determined by D and n_s. Moreover, the number of Green function entries in the product of each type-0 term is finite and determined by D, n_s.

We run the dynamics of $P_{d_s}^{(s)}$ in (5.37) up to $T = 8 \log N$ as chosen previously. We next estimate the size of $\mathbb{E}[P_{d_s}^{(s)}(t,z)]$ at the terminal time T for any $P_{d_s}^{(s)} \in \mathcal{P}_{d_s}^{(s)}$, with $s \geq 1$ and $d_s \geq 2$. Indeed, from (3.34) and the local law in (3.10), we have

$$\left|\mathbb{E}[P_{d_s}^{(s)}(T,z)] - \mathbb{E}^{\mathrm{GUE}}[P_{d_s}^{(s)}(z)]\right| = O(N^{-1}). \tag{5.39}$$

Together with the estimate (5.23) for the GUE, we obtain that, for any $s \geq 1$ and $d_s \geq 2$,

$$\left|\mathbb{E}[P_{d_s}^{(s)}(T,z)]\right| = O_{\prec}(N^{-1/3}). \tag{5.40}$$

Next, we return to the stochastic differential equation of $P_{d_s}^{(s)}$ in (5.37). Integrating (5.37) over $[t', T]$ for any $0 \leq t' \leq T$ and taking the expectation as in (5.33), we have from (5.38) and (5.40) that

$$\begin{aligned}\mathbb{E}[P_{d_s}^{(s)}(t',z)] = &\sum_{\substack{P_{d_{s+1}}^{(s+1)} \in \mathcal{P}_{d_{s+1}}^{(s+1)} \\ d_s+1 \leq d_{s+1} < D}} \int_{t'}^{T} \mathbb{E}[P_{d_{s+1}}^{(s+1)}(t,z)]\mathrm{d}t \\ &+ O_{\prec}\Big(\log N(N^{-1/2} + \Psi^D) + N^{-1/3}\Big). \end{aligned} \tag{5.41}$$

Now, we are ready to iterate using (5.41). In the first step, we start by $P_{d_1}^{(1)}(t,z)$ in (5.24) and have (5.41) for $s = 1$. The number of the terms $P_{d_2}^{(2)} \in \mathcal{P}_{d_2}^{(2)}$ with $d_2 \geq d_1 + 1$ on the right side of (5.41) is finite and depends on n_1 and D. Then we further estimate these type-0 terms $P_{d_2}^{(2)}$ using (5.41) for $s = 2$ as the second step. The resulting type-0 terms $P_{d_3}^{(3)} \in \mathcal{P}_{d_3}^{(2)}$ with $d_3 \geq d_2 + 1 \geq d_1 + 2$ will be estimated again using (5.41) for $s = 3$ as the third step. Since in each step of using (5.41), the degrees of the corresponding type-0 terms $P_{d_{s+1}}^{(s+1)} \in \mathcal{P}_{d_{s+1}}^{(s+1)}$ on the right side of (5.41) are increased by at least one, we have $d_{s+1} \geq d_1 + s$. We hence stop at step $s = s_0 := D - d_1$. For any $P_{d_{s_0}}^{(s_0)} \in \mathcal{P}_{d_{s_0}}^{(s_0)}$ with $d_{s_0} \geq D - 1$, the resulting terms $P_{d_{s_0+1}}^{(s_0+1)} \in \mathcal{P}_{d_{s_0+1}}^{(s_0+1)}$ on the right side of (5.41) have degrees $d_{s_0+1} \geq D$. The number of these terms is finite and depends on D, n_1. Using the local law in (3.10), all these terms can be bounded by $O_{\prec}(\Psi^D + N^{-1})$. This implies that the finite sum of these terms after integration over $[t', T]$ can be absorbed into the error term on the right side of (5.41). That is, for any $P_{d_{s_0}}^{(s_0)} \in \mathcal{P}_{d_{s_0}}^{(s_0)}$ with $d_{s_0} \geq D - 1$,

$$\left|\mathbb{E}[P_{d_{s_0}}^{(s_0)}(t',z)]\right| = O_{\prec}\Big(\log N(N^{-1/2} + \Psi^D) + N^{-1/3}\Big).$$

We hence plug the above estimate back to the previous step, *i.e.,* (5.41) for $s = s_0 - 1$. We then obtain a similar estimate for any $P_{d_{s_0-1}}^{(s_0-1)} \in \mathcal{P}_{d_{s_0-1}}^{(s_0-1)}$ with $d_{s_0-1} \geq D - 2$,

$$\left|\mathbb{E}[P_{d_{s_0-1}}^{(s_0-1)}(t',z)]\right| = O_{\prec}\Big(\log^2 N(N^{-1/2} + \Psi^D) + N^{-1/3}\log N\Big).$$

Repeating the above process until $s = 1$, we hence obtain that, for $d_1 \geq 2$,

$$\left|\mathbb{E}[P_{d_1}^{(1)}(t,z)]\right| = O_{\prec}\Big((N^{-1/3} + \Psi^D)\log^D N\Big),$$

uniformly in $t \in [0, T]$ and $z \in S_{\mathrm{edge}}$. By choosing $D \geq \frac{1}{\epsilon}$ with $\epsilon > 0$ as in (2.9), we prove (4.21) for $t \in [0, T]$. If $t \geq T$, a similar estimate can be obtained by using (5.39) and (5.40). We have hence finished the proof of Lemma 4.1.

5.4. Proof of Lemma 5.2. We end this section with the proof of Lemma 5.2 considering the GUE.

Proof of Lemma 5.2. Using the spectral decomposition (5.2), we write

$$\frac{1}{N}\mathbb{E}^{\mathrm{GUE}}\Big[\operatorname{Im}\operatorname{Tr}G(z)\Big]=\frac{N\eta}{N^2}\mathbb{E}^{\mathrm{GUE}}\Big[\sum_{j=1}^{N}\frac{1}{|\lambda_j-z|^2}\Big],\quad z\in S_{\mathrm{edge}}. \tag{5.42}$$

Then it suffices to estimate the following linear eigenvalue statistics, which can be written from (2.29), (2.36) and then (2.37) as

$$\begin{aligned}\frac{1}{N^2}\mathbb{E}^{\mathrm{GUE}}\Big[\sum_{i=1}^{N}\frac{1}{|\lambda_i-z|^2}\Big]&=\frac{1}{N^2}\int_{\mathbb{R}}\frac{\widetilde{K}_N(x,x)}{|x-2-\kappa-\mathrm{i}\eta|^2}\,\mathrm{d}x\\&=\frac{1}{N^{\frac{2}{3}}}\int_{\mathbb{R}}\frac{K_N^{\mathrm{edge}}(x,x)}{|x-N^{2/3}\kappa-\mathrm{i}N^{2/3}\eta|^2}\,\mathrm{d}x,\end{aligned} \tag{5.43}$$

where $z=2+\kappa+\mathrm{i}\eta\in S_{\mathrm{edge}}$, with $|\kappa|\le C_0N^{-2/3+\epsilon}$ and $N^{-1+\epsilon}\le\eta\le N^{-2/3+\epsilon}$.

To control the integral on the right side of (5.43), we choose a fixed $L_0<0$ (see Lemma 2.5 and Theorem 2.3) and split the real line in the parts, $(-\infty,-N^{2/3}]$, $(-N^{2/3},L_0]$ and $(L_0\,\infty)$.

For the integration domain $(-\infty,-N^{2/3}]$, we find that

$$\frac{1}{N^{\frac{2}{3}}}\int_{x<-N^{2/3}}\frac{K_N^{\mathrm{edge}}(x,x)}{|x-N^{2/3}\kappa-\mathrm{i}N^{2/3}\eta|^2}\,\mathrm{d}x=O(N^{-1}), \tag{5.44}$$

using the trace identity (2.34) for the kernel K_N and that $|\kappa|\le C_0N^{-2/3+\epsilon}$.

Moreover, from Theorem 2.3 and Lemma 2.5, we have on (L_0,∞), that

$$\begin{aligned}\frac{1}{N^{\frac{2}{3}}}\int_{x>L_0}\frac{K_N^{\mathrm{edge}}(x,x)}{|x-N^{2/3}\kappa-\mathrm{i}N^{2/3}\eta|^2}\,\mathrm{d}x&=\frac{1}{N^{\frac{2}{3}}}\int_{x>L_0}\frac{K_{\mathrm{airy}}(x,x)+O(N^{-2/3})}{|x-N^{2/3}\kappa-\mathrm{i}N^{2/3}\eta|^2}\,\mathrm{d}x\\&=O\Big(\frac{1}{N^{\frac{4}{3}}\eta}\Big).\end{aligned} \tag{5.45}$$

It hence suffices to focus on the regime $(-N^{2/3},L_0]$. Recall from (2.32) and (2.37) that

$$K_N(x,x)=\sum_{k=0}^{N-1}\phi_k^2(x);\quad K_N^{\mathrm{edge}}(x,x)=\frac{1}{N^{1/6}}K_N\Big(2\sqrt{N}+\frac{x}{N^{1/6}},2\sqrt{N}+\frac{x}{N^{1/6}}\Big). \tag{5.46}$$

From (2.31) and (2.33), the derivative of $K_N(x,x)$ is given by

$$K_N'(x,x)=-\sqrt{N}\phi_{N-1}(x)\phi_N(x)\,.$$

The Hermite functions satisfy, for all k,

$$\sup_{x\in\mathbb{R}}|\phi_k(x)|\le Ck^{-1/12}\,. \tag{5.47}$$

for some constant C independent of k, as was proved in [4]. Therefore, the derivative of the edge kernel $K_N^{\mathrm{edge}}(x,x)$ is given by

$$\Big(K_N^{\mathrm{edge}}(x,x)\Big)' = \frac{1}{N^{1/3}} K_N'\Big(2\sqrt{N} + \frac{x}{N^{1/6}}, 2\sqrt{N} + \frac{x}{N^{1/6}}\Big) = O(1)\,. \tag{5.48}$$

For any $x \in (-N^{2/3}, L_0]$, we have from (5.48) and Lemma 2.5 that

$$K_N^{\mathrm{edge}}(x,x) = K_N^{\mathrm{edge}}(L_0,L_0) - \int_x^{L_0} \big(K_N^{\mathrm{edge}}(x,x)\big)' \mathrm{d}x \le C'(1+|x|). \tag{5.49}$$

Therefore, we obtain from (5.49) that

$$\begin{aligned} &\frac{1}{N^{\frac{2}{3}}} \int_{-N^{2/3}<x<L_0} \frac{K_N^{\mathrm{edge}}(x,x)}{|x - N^{2/3}\kappa - \mathrm{i}N^{2/3}\eta|^2}\,\mathrm{d}x \\ &\quad \le \frac{C'}{N^{\frac{2}{3}}} \int_{-N^{2/3}<x<L_0} \frac{1+|x|}{(x-N^{2/3}\kappa)^2 + (N^{2/3}\eta)^2}\,\mathrm{d}x \\ &\quad = O\Big(N^{-2/3}\log N + \frac{N^{\epsilon}}{N^{\frac{4}{3}}\eta}\Big) = O\Big(\frac{1}{N^{\frac{4}{3}-\epsilon}\eta}\Big), \end{aligned} \tag{5.50}$$

where we used that $|\kappa| \le C_0 N^{-2/3+\epsilon}$.

Plugging (5.44), (5.45) and (5.50) into (5.43), there exists some constant C independent of ϵ such that

$$\frac{1}{N^2}\mathbb{E}^{\mathrm{GUE}}\Big[\sum_{j=1}^{N} \frac{1}{|\lambda_j - z|^2}\Big] \le \frac{CN^{\epsilon}}{N^{\frac{4}{3}}\eta}, \tag{5.51}$$

uniformly in $z \in S_{\mathrm{edge}}$, for sufficiently large $N \ge N_0(\epsilon, C_0)$. In combination with (5.42), we hence have proved (5.22).

Finally, we consider any type-0 term $P_d(z) \in \mathcal{P}_d(z)$ of the form in (4.17) of degree $d \ge 2$ for the GUE. For notational simplicity, we no longer emphasize the indices a,b and write

$$P_d(z) = \frac{1}{N^n} \sum_{v_1=1}^{N} \cdots \sum_{v_n=1}^{N} c_{v_1,\ldots,v_n} \Big(\prod_{i=1}^{n} G_{x_i y_i}(z)\Big), \tag{5.52}$$

with $n \ge 2$, where each summation index v_j $(1 \le j \le n)$ appears once in the row index set $\{x_i\}_{i=1}^n$ and once in the column index set $\{y_i\}_{i=1}^n$ and the coefficients $\{c_{v_1,\ldots,v_n}\}$ are uniformly bounded complex numbers. For any $1 \le j \le n$, if there exists $1 \le i \le n$ such that $x_i = y_i = v_j$, then we say that v_j is isolated. For any $1 \le j \ne j' \le n$, if there exists $1 \le i \le n$ such that either $x_i = v_j$, $y_i = v_{j'}$ or $y_i = v_j$, $x_i = v_{j'}$, then we say that v_j and $v_{j'}$ are connected indices. Because the degree of (5.52) is at least two, there exists at least one cluster of connected indices containing at least two elements. We may assume that $v_1, \ldots, v_{n_0}$ $(2 \le n_0 \le n)$ form a cluster of connected indices. Using the local law in (3.10), we have

$$|P_d(z)| \prec \frac{1}{N^{n_0}} \sum_{v_1=1}^{N} \cdots \sum_{v_{n_0}=1}^{N} \big|G_{v_1 v_2} G_{v_2 v_3} \cdots G_{v_{n_0} v_1}(z)\big|\,.$$

If $n_0 = 2$, from Young's inequality and the Ward identity

$$\frac{1}{N^2}\sum_{i,j}|G_{ij}(z)|^2 = \frac{\operatorname{Im} m_N(z)}{N\eta}, \quad z = E + \mathrm{i}\eta \in \mathbb{C}^+, \tag{5.53}$$

which follows from the spectral decomposition (5.2), we then obtain

$$\begin{aligned}|P_d(z)| &\prec \frac{1}{N^2}\sum_{v_1,v_2}\left|G_{v_1v_2}(z)G_{v_2v_1}(z)\right| \le \frac{1}{2N^2}\sum_{v_1,v_2}\left(|G_{v_1v_2}(z)|^2 + |G_{v_2v_1}(z)|^2\right)\\ &= \frac{\operatorname{Im} m_N(z)}{N\eta}.\end{aligned} \tag{5.54}$$

For $n_0 \ge 3$, we have similarly from the local law (3.10) that

$$\begin{aligned}|P_d(z)| &\prec \Psi^{n_0-2}\frac{1}{N^3}\sum_{v_1,v_2,v_3}\left|G_{v_1v_2}(z)G_{v_2v_3}(z)\right|\\ &\le \Psi^{n_0-2}\frac{1}{2N^3}\sum_{v_1,v_2,v_3}\left(|G_{v_1v_2}(z)|^2 + |G_{v_2v_3}(z)|^2\right) = O\Big(\frac{\operatorname{Im} m_N(z)}{(N\eta)^{n_0-1}}\Big),\end{aligned} \tag{5.55}$$

where in the last two steps we use Young's inequality, the Ward identity (5.53), and that $\Psi(z) = O(\frac{1}{N\eta})$ for any $z \in S_{\text{edge}}$. Therefore, combining with the estimate (5.22) for the expectation of $\operatorname{Im} m_N(z)$, the properties of stochastic domination in Lemma 1.1, and that $\eta \ge N^{-1+\epsilon}$, we have, for any $\tau > 0$,

$$\mathbb{E}^{\text{GUE}}\big[|P_d^{(s)}(z)|\big] \le N^{-1/3+\tau}, \quad d \ge 2,$$

uniformly in $z \in S_{\text{edge}}$, for sufficiently large $N \ge N_0'(C_0, \epsilon, \tau)$. This completes the proof of (5.23), and hence the proof of Lemma 5.2.

6. Product of Green Function Entries with Unmatched Indices

In this section, we prove Proposition 4.2. Before stating the proof for Wigner matrices, we first consider the GUE for the intuition why expectations of unmatched terms are much smaller than the naive size obtained using power counting and the local law as in (4.10).

6.1. Intuition from the GUE. In this subsection, we focus on the special case of the GUE, as in Sect. 5.1. Consider any $Q_d^o \in \mathcal{Q}_d^o$ of the form (4.8). Using the spectral decomposition (5.2) and the unitary invariance of the GUE similarly as in (5.3), we write the expectation of the unmatched Q_d^o as

$$\mathbb{E}[Q_d^o] = \frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}} \sum_{j_1,\dots,j_n=1}^{N} \mathbb{E}\Big[\prod_{i=1}^{n}\frac{1}{(\lambda_{j_i} - z)}\Big] \times \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{u}_{j_i}(x_i)\overline{\mathbf{u}_{j_i}(y_i)}\Big], \tag{6.1}$$

with (λ_j) the eigenvalues and the corresponding normalized eigenvectors $(\mathbf{u}_j)$, and each x_i, y_i represent some free summation index in $\mathcal{I}$. In order to estimate the expectations of the eigenvectors, we recall the Weingarten calculus formula in Lemma 5.1. Under

Definition 4.1 for unmatched indices, if the values of the free summation indices in $\mathcal{I}$ are distinct, then $\delta_{x_1,y_{\sigma(1)}}\cdots\delta_{x_n,y_{\sigma(n)}}=0$, for any permutation $\sigma\in S_n$. Thus from (5.4), for any $1\le j_1,\dots,j_n\le N$, we have

$$\mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{u}_{j_i}(x_i)\overline{\mathbf{u}_{j_i}(y_i)}\Big]=0.$$

The non-vanishing contributions come from the diagonal cases when the values of some free summation indices in $\mathcal{I}$ coincide. Because of the averaged form of Q_d^o in (4.8) and the local law in (3.10) one works out that, for any $z\in S$ and $t\ge 0$,

$$\mathbb{E}[Q_d^o]=O(N^{-1})\,. \tag{6.2}$$

For Wigner matrices, the above argument does not apply anymore. We hence use similar expansions as in Sect. 5.2 to extend to arbitrary Wigner matrices. Before we give the proof of Proposition 4.2, we start by considering an example of the unmatched term in $\mathcal{Q}_d^o$ to illustrate the mechanism.

6.2. Example of an unmatched term. We look at the following example of an unmatched term

$$\frac{1}{N^2}\sum_{a,b}G_{ab}G_{ba}G_{ab}\in\mathcal{Q}_3^o, \tag{6.3}$$

with $a\in\mathcal{R}^o$ and $b\in\mathcal{C}^o$; see (4.12) in Definition 4.1. Using the local law in (3.10), the expectation of this term can be naively bounded by $O_\prec(\Psi^3+N^{-1})$. The idea to improve this bound is similar to the proof of Proposition 4.3. Note that the combination of the identity (5.10) and the cumulant expansion gives a cancellation to the leading order. Thus we can improve the upper bound to $O_\prec(\Psi^4+\frac{\Psi^3}{\sqrt{N}}+N^{-1})$. We next discuss the details.

Using the identity (5.10) on the off-diagonal entry G_{ab} with unmatched a as the row index and applying cumulant expansions, we have

$$\begin{aligned}
&\frac{1}{N^2}\sum_{a,b}\mathbb{E}[G_{ab}G_{ba}G_{ab}]\\
&=\frac{1}{N^2}\sum_{a\ne b}\mathbb{E}\Big[\Big(G_{ab}\underline{HG}-\underline{G}(HG)_{ab}\Big)G_{ba}G_{ab}\Big]+\frac{1}{N^2}\sum_{a=1}^{N}\mathbb{E}[(G_{aa})^3]\\
&=\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{ab}G_{jk}G_{ba}G_{ab}}{\partial h_{jk}}\Big]-\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{jj}G_{kb}G_{ba}G_{ab}}{\partial h_{ka}}\Big]\\
&\quad+\frac{1}{\sqrt{N}}\frac{1}{N^4}\sum_{p+q+1=3}\frac{1}{p!q!}\sum_{a,b,j,k}s_{jk}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^2 G_{ab}G_{jk}G_{ba}G_{ab}}{\partial h_{jk}^p\partial h_{kj}^q}\Big]\\
&\quad-\frac{1}{\sqrt{N}}\frac{1}{N^4}\sum_{p+q+1=3}\frac{1}{p!q!}\sum_{a,b,j,k}s_{ak}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^2 G_{jj}G_{kb}G_{ba}G_{ab}}{\partial h_{ka}^p\partial h_{ak}^q}\Big]+O_\prec\Big(\frac{1}{N}\Big),
\end{aligned} \tag{6.4}$$

where the last error term comes from the truncation of the cumulant expansions at the third order and the diagonal case $a = b$.

Using (3.21) and that j, k are fresh summation indices, all the third order expansions for $\{p + q + 1 = 3\}$ can be written out using the terms of the form in (4.8) of degree at least three, with an additional factor $\frac{1}{\sqrt{N}}$ in front. Since both the fresh indices j, k appear in the product of the Green function entries for an odd number of times, the resulting terms are unmatched from Definition 4.1. From the local law in (3.10), they are bounded by $O_{\prec}\Big(\frac{\Psi^3}{\sqrt{N}} + \frac{1}{N^{3/2}}\Big)$.

Now we return to the second order terms in the cumulant expansions in (6.4), *i.e.,*

$$\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{ab}G_{jk}G_{ba}G_{ab}}{\partial h_{jk}}\Big] - \frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{jj}G_{kb}G_{ba}G_{ab}}{\partial h_{ka}}\Big]. \tag{6.5}$$

Using (3.21), the fresh indices j, k are then matched and the index a remains to be an unmatched row index. The key observation here is that the leading sub-term from the first term above, corresponding to taking $\frac{\partial}{\partial h_{jk}}$ of G_{jk}, will be canceled precisely by the leading sub-term from the second term above, resulting from taking $\frac{\partial}{\partial h_{ka}}$ of G_{kb}. We hence rewrite (6.5) as

$$\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{ab}G_{ba}G_{ab}}{\partial h_{jk}}G_{jk}\Big] - \frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[\frac{\partial G_{jj}G_{ba}G_{ab}}{\partial h_{ka}}G_{kb}\Big]. \tag{6.6}$$

The degrees of the resulting terms from the first part above are five as j, k are fresh indices. Similarly, the ones from the second part have degrees at least four, except one sub-term from taking $\frac{\partial}{\partial h_{ka}}$ of G_{ba}, whose column index coincides with the unmatched row index a:

$$\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[G_{jj}G_{bk}G_{aa}G_{ab}G_{kb}\Big].$$

Compared with the original term in (6.3), one replaces one pair of the index a by a fresh index k and adds a factor G_{aa} for the replaced index a. The good news is that this leading term of degree three remains unmatched with an unmatched row index a. We then expand it further as in (6.4) and obtain that

$$\begin{aligned}
&\frac{1}{N^4}\sum_{a,b,j,k}\mathbb{E}\Big[G_{jj}G_{aa}G_{ab}G_{bk}G_{kb}\Big] \\
&= \frac{1}{N^6}\sum_{a,b,j,k,j',k'}\mathbb{E}\Big[\frac{\partial G_{jj}G_{aa}G_{ab}G_{bk}G_{kb}}{\partial h_{j'k'}}G_{j'k'}\Big] \\
&- \frac{1}{N^6}\sum_{a,b,j,k,j',k'}\mathbb{E}\Big[\frac{\partial G_{jj}G_{aa}G_{j'j'}G_{bk}G_{kb}}{\partial h_{k'a}}G_{k'b}\Big] \\
&+ \{\text{third order terms}\} + O_{\prec}\Big(\frac{1}{N}\Big),
\end{aligned} \tag{6.7}$$

with j', k' another two fresh summation indices. Here, the third order terms are also unmatched terms of the form in (4.8) of degree at least three with an extra $\frac{1}{\sqrt{N}}$ in front, similarly as in (6.4). From (3.21), the resulting terms from the first part on the right side of (6.7) have degrees at least five. As for the second part above, even though the column

index of the diagonal entry G_{aa} coincides with the unmatched row index a, the resulting terms have degrees at least four.

In this way, we improve the upper bound of the unmatched term given in (6.3) to

$$\Big|\frac{1}{N^2}\sum_{a,b}\mathbb{E}\Big[G_{ab}G_{ba}G_{ab}\Big]\Big| \prec \Psi^4 + \frac{\Psi^3}{\sqrt{N}} + N^{-1}.$$

Indeed, we expand this unmatched term as

$$\frac{1}{N^2}\sum_{a,b}\mathbb{E}\Big[G_{ab}G_{ba}G_{ab}] = \sum_{\substack{Q^o_{d_1'}\in\mathcal{Q}^o_{d_1'}\\ d_1'\geq 4}}\mathbb{E}[Q^o_{d_1'}] + \frac{1}{\sqrt{N}}\sum_{\substack{Q^o_{d_2'}\in\mathcal{Q}^o_{d_2'}\\ d_2'\geq 3}}\mathbb{E}[Q^o_{d_2'}] + O_{\prec}(N^{-1}), \tag{6.8}$$

where we write $\sum_{Q^o_{d_1'}\in\mathcal{Q}^o_{d_1'},d_1'\geq 4} Q^o_{d_1'}$ as a sum of finitely many unmatched terms of the form in (4.8) of degrees increased by at least one, which comes from the second order expansions. Moreover, we write $\frac{1}{\sqrt{N}}\sum_{Q^o_{d_2'}\in\mathcal{Q}^o_{d_2'},d_2'\geq 3} Q^o_{d_2'}$ as a finite sum of unmatched terms of the form in (4.8) with an extra factor $\frac{1}{\sqrt{N}}$ in front, which corresponds to the third order expansions. The last error term $O_{\prec}(N^{-1})$ is from the truncation of the cumulant expansion and the diagonal cases. By repeating the above expansion procedure in (6.8) for arbitrary D times, we improve the upper bound to $O_{\prec}\big(\Psi^D + \frac{\Psi^{D-1}}{\sqrt{N}} + N^{-1}\big)$. The full proof is presented in the following section.

6.3. Proof of Proposition 4.2. In this section, we give the proof of Proposition 4.2 for Wigner matrices using the cumulant expansions as explained above.

Proof of Proposition 4.2. Consider an arbitrary unmatched term $Q^o_d \in \mathcal{Q}^o_d$ of the form (4.8). Because it is equivalent to expand a Green function entry G_{xy} in the row index x or column index y, we focus on the unmatched row indices in the following.

We may assume that the index v_1 belongs to the unmatched row index set $\mathcal{R}^o$ (which cannot be empty) from Definition 4.1. Then there exists an off-diagonal factor in the product of Green function entries with v_1 as the row index. Without loss of generality, we set $x_1 = v_1$, and $y_1 \neq v_1$. Using (5.10) on the off-diagonal entry $G_{v_1y_1}$ and applying cumulant expansions similarly as in (6.4), we have

$$\begin{aligned}
\mathbb{E}[Q^o_d] &= \frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}}\mathbb{E}\Big[G_{v_1y_1}\prod_{i=2}^{n}G_{x_iy_i}\Big]\\
&= \frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}}\mathbb{E}\Big[\delta_{v_1y_1}G_{v_1y_1}\prod_{i=2}^{n}G_{x_iy_i}\Big]\\
&\quad + \frac{1}{N^{2+\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}}\sum_{j,k}\mathbb{E}\Big[\frac{\partial G_{v_1y_1}G_{jk}\prod_{i=2}^{n}G_{x_iy_i}}{\partial h_{jk}}\Big]\\
&\quad - \frac{1}{N^{2+\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}}\sum_{j,k}\mathbb{E}\Big[\frac{\partial G_{jj}G_{ky_1}\prod_{i=2}^{n}G_{x_iy_i}}{\partial h_{kv_1}}\Big]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{N^{2+\#\mathcal{I}}}\frac{1}{\sqrt{N}}\sum_{\mathcal{I}}c_{\mathcal{I}}\sum_{p+q+1=3}\frac{1}{p!q!}\sum_{j,k}s_{jk}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^2 G_{x_1y_1}G_{jk}\prod_{i=2}^n G_{x_iy_i}}{\partial h_{jk}^p\partial h_{kj}^q}\Big]\\
&-\frac{1}{N^{2+\#\mathcal{I}}}\frac{1}{\sqrt{N}}\sum_{\mathcal{I}}c_{\mathcal{I}}\sum_{p+q+1=3}\frac{1}{p!q!}\sum_{j,k}s_{v_1k}^{(p,q+1)}\mathbb{E}\Big[\frac{\partial^2 G_{jj}G_{ky_1}\prod_{i=2}^n G_{x_iy_i}}{\partial h_{kv_1}^p\partial h_{v_1k}^q}\Big]\\
&+O_{\prec}(\frac{1}{N})\\
=&\frac{1}{N^{2+\#\mathcal{I}}}\sum_{\mathcal{I}}c_{\mathcal{I}}\sum_{j,k}\mathbb{E}\Big[\frac{\partial G_{v_1y_1}\prod_{i=2}^n G_{x_iy_i}}{\partial h_{jk}}G_{jk}\Big]\\
&-\frac{1}{N^{2+\#\mathcal{I}}}\sum_{\mathcal{I}}c_{\mathcal{I}}\sum_{j,k}\mathbb{E}\Big[\frac{\partial G_{jj}\prod_{i=2}^n G_{x_iy_i}}{\partial h_{kv_1}}G_{ky_1}\Big]\\
&+\{\text{ third order terms for } p+q+1=3\}+O_{\prec}\Big(\frac{1}{N}\Big),
\end{aligned}
\tag{6.9}
$$

where j,k are fresh summation indices, the last error $O_{\prec}(\frac{1}{N})$ is from the truncation of the cumulant expansions at the third order and the diagonal case $v_1\equiv y_1$.

We first look at the third order expansions for $p+q+1=3$, which are much smaller because we gain an extra $\frac{1}{\sqrt{N}}$ from the third order cumulants. Since both j,k are fresh indices, it is straightforward to check from (3.21) that the resulting terms are also of the form in (4.8) with an extra $\frac{1}{\sqrt{N}}$ in front. Their degrees, denoted by d', satisfy $d'\geq d$, the corresponding free summation index set is $\mathcal{I}'=\{\mathcal{I},j,k\}$ and the number of Green function entries is $n'=n+3$. In addition, the number of such terms is at most $6(n+3)^2$. Comparing these terms with the original Q_d^o, we add in total an odd number of j's (or k's) into the original row index set and column index set of the product of the Green function entries. Then all these terms are unmatched terms from Definition 4.1. We use $\frac{1}{\sqrt{N}}\sum_{Q_{d'}^o\in\mathcal{Q}_{d'}^o;d'\geq d}\mathbb{E}[Q_{d'}^o]$ to denote the finite sum of these unmatched terms from the third order expansions.

Next, we estimate the second order expansion terms, *i.e.,* the second but last line on the right side of (6.9). Using (3.21) we write them as a sum of at most $2n$ terms of the form in (4.8) with $\mathcal{I}'=\{\mathcal{I},j,k\}$ and $n'=n+2$. The degrees of these terms are estimated as follows.

For the first group of terms in the second but last line of (6.9), comparing with the original Q_d^o, we have added one fresh index j and one fresh index k into both the original row index set and column index set. Then j and k are both matched indices. Moreover, v_1 from $G_{v_1y_1}$ remains an unmatched row index. After taking $\frac{\partial}{\partial h_{jk}}$ by (3.21), the degrees are then increased by at least two.

Similarly, we compare the second group in the second but last line of (6.9) with the original Q_d^o. We find again that both j and k are matched, and the index v_1 is still an unmatched row index. However, the degrees of the resulting terms from taking $\frac{\partial}{\partial h_{kv_1}}$ may not be increased. This is because the column index of some Green function entry $G_{x_iy_i}(2\leq i\leq n)$ may coincide with the unmatched row index v_1. The number of such Green function entries with v_1 as column index is given by $\nu_1^c(\leq n)$ from Definition 4.1. So we split the discussion into three cases.

Case 1: If $y_i\neq v_1$, then after taking $\frac{\partial}{\partial h_{kv_1}}$ of $G_{x_iy_i}$, the degree of the resulting term is increased by at least one.

Case 2: If $y_i = x_i = v_1$, then after taking $\frac{\partial}{\partial h_{kv_1}}$ of $G_{x_i y_i}$, the degree is then increased by exactly one.

Case 3: If $y_i = v_1$, but $x_i \neq v_1$, then, for simplicity, we may assume that $y_2 = v_1$ and $x_2 \neq v_1$. From Definition 4.1 for unmatched indices, there exists some $3 \leq i' \leq n$ such that $x_{i'} = v_1$ and $y_{i'} \neq v_1$, because else v_1 cannot be an unmatched row index of the original Q_d^o. We may assume $x_3 = v_1$ and $y_3 \neq v_1$. Then the corresponding term after taking $\frac{\partial}{\partial h_{kv_1}}$ of G_{x_2,v_1} becomes

$$(*) := \frac{1}{N^{2+\#\mathcal{I}}} \sum_{\mathcal{I},j,k} c_{\mathcal{I}} \mathbb{E}\Big[G_{jj} G_{v_1 v_1} G_{k y_1} G_{x_2 k} G_{v_1, y_3} \prod_{i=4}^{n} G_{x_i y_i} \Big], \tag{6.10}$$

with $y_1 \neq v_1$, $x_2 \neq v_1$, and $y_3 \neq v_1$, and the degree of this term is still d. Compared with the original Q_d^o, we have replaced one pair of the index v_1, *i.e.,* the row index of $G_{x_1 y_1}$ and the column index of $G_{x_2 y_2}$, by the fresh index k. Further we get an additional diagonal Green function entry $G_{v_1 v_1}$ for the replaced pair of index v_1. Since the index v_1 from $G_{v_1 y_3}$ remains an unmatched row index, we can further expand the term in (6.10) using the unmatched row index v_1, as in (6.9). We write

$$\begin{aligned}(*) = &- \frac{1}{N^{4+\#\mathcal{I}}} \sum_{\mathcal{I},j,k,j',k'} c_{\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{jj} G_{v_1 v_1} G_{v_1 y_3} G_{x_2 k} G_{k y_1} \Big(\prod_{i=4}^{n} G_{x_i y_i} \Big)}{\partial h_{j'k'}} G_{j'k'} \Big] \\ &+ \frac{1}{N^{4+\#\mathcal{I}}} \sum_{\mathcal{I},j,k,j',k'} c_{\mathcal{I}} \mathbb{E}\Big[\frac{\partial G_{jj} G_{v_1 v_1} G_{j'j'} G_{x_2 k} G_{k y_1} \Big(\prod_{i=4}^{n} G_{x_i y_i} \Big)}{\partial h_{k'v_1}} G_{k' y_3} \Big] \\ &+ \{ \text{ third order expansions for } p+q+1 = 3 \} + O_{\prec}\Big(\frac{1}{N}\Big). \end{aligned} \tag{6.11}$$

Similar as (6.9), the third order expansions contains at most $6(n+5)^2$ unmatched terms of the form in (4.8) with an additional factor $\frac{1}{\sqrt{N}}$ in front, of degrees $d'' \geq d$, with $\mathcal{I}'' = \{\mathcal{I}, j, k, j', k'\}$ and $n'' = n+5$. We next estimate the second order expansions on the right side of (6.11). From (3.21), they become a sum of at most $2n$ terms of the form in (4.8), with $\mathcal{I}'' = \{\mathcal{I}, j, k, j', k'\}$ and $n'' = n+4$.

If for any $4 \leq i \leq n$, either $y_i \neq v_1$ or $x_i = y_i = v_1$ holds, as considered in Cases 1 and 2 above, then the degrees of these resulting terms are increased by at least one, *i.e.,* $d'' \geq d+1$.

Else we may assume that $y_4 = v_1$ and $x_4 \neq v_1$. The resulting leading term of degree d, as the analogue of (6.10), is obtained from replacing one pair of the index v_1, *i.e.,* the row index of $G_{x_3 y_3}$ and the column index of $G_{x_4 y_4}$, by the fresh index k' and adding an additional diagonal Green function entry $G_{v_1 v_1}$. Moreover, there exists some $5 \leq i'' \leq n$ such that $x_{i''} = v_1$ and $y_{i''} \neq v_1$ to make sure v_1 is an unmatched row index of the original Q_d^o in (6.9), as explained at the beginning of Case 3. We may assume $i'' = 5$ for simplicity. Then the index v_1 from $G_{v_1 y_5}$ is again unmatched. We can expand this leading term of degree d for the third time by applying (5.10) on $G_{v_1 y_5}$ and applying cumulant expansions, similarly as in (6.11).

We continue this procedure of expanding in the unmatched row index v_1 repeatedly for s times, until there is no off-diagonal Green function entry with column index $y_i = v_1$ in the remaining product of the Green function entries $\prod_{i=2s}^{n} G_{x_i y_i}$. Then from Case 1

and Case 2 above, the resulting terms have degrees increased by at least one. The number of iteration s is at most $\nu_1^{(c)} (\leq n)$, where $\nu_1^{(c)}$ defined in (7.5) is the number of times the unmatched row index v_1 appears in the column index set of the original Q_d^o.

In this way, we expand the original unmatched Q_d^o in terms of finitely many unmatched terms in the form (4.8) of degrees at least $d+1$, as well as the third order cumulant expansion terms generated in the iterations, plus an error $O_\prec(N^{-1})$ from the truncation of the cumulant expansion and the diagonal cases. In summary, for any unmatched $Q_d^o \in \mathcal{Q}_d^o$, we write the following expansions for short:

$$\mathbb{E}[Q_d^o] = \sum_{\substack{Q_{d_1'}^o \in \mathcal{Q}_{d_1'}^o \\ d_1' \geq d+1}} \mathbb{E}[Q_{d_1'}^o] + \frac{1}{\sqrt{N}} \sum_{\substack{Q_{d_2'}^o \in \mathcal{Q}_{d_2'}^o \\ d_2' \geq d}} \mathbb{E}[Q_{d_2'}^o] + O_\prec(\frac{1}{N}), \tag{6.12}$$

where the number of unmatched terms in the summations above is bounded by $(Cn)^{cn}$, and the number of the Green function entries in the product of each the unmatched term is bounded by Cn for some numerical constants $C, c > 0$.

We finally iterate the expansion in (6.12) for $D-d$ times. Then the unmatched terms in the first summation have degrees at least D, and the unmatched terms with $\frac{1}{\sqrt{N}}$ in the second summation have degrees at least $D-1$. Note that the total number of the terms generated in the iteration of the expansions is bounded by $\big((C^D n)^{c^D n}\big)^D$, and the number of the Green function entries in the product of each term is bounded by $C^D n$. We hence obtain from the local law in (3.10) that

$$\mathbb{E}[Q_d^o] = O_\prec\big(\Psi^D + \frac{\Psi^{D-1}}{\sqrt{N}} + \frac{1}{N}\big) = O_\prec\big(\Psi^D + \frac{1}{N}\big)\,. \tag{6.13}$$

We hence have finished the proof of Proposition 4.2.

7. Proof of Proposition 3.1

In this section, we prove Proposition 3.1, which is a key ingredient in the proof the Green function comparison theorem, Theorem 1.4. The special case of Proposition 3.1 considering $F(x) = x$ was stated in (4.6), which leads to the corresponding Green function comparison theorem for $F(x) = x$ in Proposition 4.1. The proof of Proposition 3.1 relies on the analogues of Proposition 4.3 (expansion in type-0 terms) and Proposition 4.2 (the negligibility of unmatched terms), as well as the estimate (4.3) obtained in Proposition 4.1 to bound the resulting type-0 terms.

Proof of Proposition 3.1. We extend the ideas from the proofs of (4.6) to the setup of Proposition 3.1. Recall $\mathbb{E}[\Theta(t, z_1, z_2)]$ from (3.18), *i.e.,*

$$\mathbb{E}[\Theta(t, z_1, z_2)] \equiv \mathbb{E}[\Theta] = \sum_{\substack{p+q+1=3 \\ p,q \in \mathbb{N}}}^{4} K_{p,q+1} + E_2 + O_\prec(N^{-1/2}), \tag{7.1}$$

with $K_{p,q+1}$ given in (3.19) and E_2 given in (3.20).

Using the differentiation rules (3.21) and (3.26), each term on the right side of (7.1) can be written out in terms of an average product of Green function entries with $\Delta\widetilde{\mathrm{Im}}$

acting on it and multiplied by derivatives of F. We give one example of a third order term with $p = 1, q = 1$,

$$\sqrt{N}\frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(1,2)}}{2}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left(G_{va}G_{bv}G_{aa}G_{bb}\right)\Big],$$

and one example of a fourth order terms with $p = 2, q = 1$,

$$-\frac{1}{N^3}\sum_{v,a,b}\frac{s_{ab}^{(2,2)}}{4}\mathbb{E}\Big[F''(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left(G_{aa}G_{bb}\right)\Delta\widetilde{\mathrm{Im}}\left(G_{aa}G_{bb}\right)\Big].$$

We point out that the third order terms with $p + q + 1 = 3$ have an additional leading factor $\sqrt{N}$.

To estimate these averaged products of Green function entries multiplied by derivatives of F, we introduce the following form of terms generalizing the definition in (4.8):

$$\widetilde{\mathcal{Q}}(t, z_1, z_2): \quad \frac{1}{N^m}\sum_{v_1=1}^{N}\cdots\sum_{v_m=1}^{N}c_{v_1,\ldots,v_m}\mathbb{E}\Big[F^{(\alpha)}(\mathcal{X})\prod_{i=1}^{i_0}\Delta\widetilde{\mathrm{Im}}\Big(\prod_{l=1}^{n_i}G_{x_l^{(i)}y_l^{(i)}}\Big)\Big], \tag{7.2}$$

with $\alpha, m, i_0, n_i \in \mathbb{N}$, $F^{(\alpha)}$ be the α-th derivative of a smooth function F which has uniformly bounded derivatives, $\Delta\widetilde{\mathrm{Im}} : \mathbb{R}^+ \times (\mathbb{C}\setminus\mathbb{R})^2 \to \mathbb{C}$ defined in (3.14), where $\mathcal{I} := \{v_j\}_{j=1}^m$ is a free summation index set, and the v_j's may also represent a, b from (3.19) and (3.20). The coefficients $\{c_{\mathcal{I}} := c_{v_1,\ldots,v_m}\}$ are uniformly bounded complex numbers, and each $x_l^{(i)}$ and $y_l^{(i)}$ represent some element in the free summation index set $\mathcal{I}$. The total number of the Green function entries in (7.2) is then given by

$$n := \sum_{i=1}^{i_0} n_i. \tag{7.3}$$

We further define the degree of a term in the form (7.2) by counting the number of off-diagonal Green function entries, *i.e.,*

$$d := \sum_{i=1}^{i_0}\#\left\{1 \le l \le n_i : x_l^{(i)} \ne y_l^{(i)}\right\}. \tag{7.4}$$

In particular, we have $0 \le d \le n$. The collection of the terms in the form (7.2) of degree d is denoted by $\widetilde{\mathcal{Q}}_d \equiv \widetilde{\mathcal{Q}}_d(t, z_1, z_2)$. From the definition of $\Delta\widetilde{\mathrm{Im}}$ in (3.14), the local law in (3.10) and the fact that F has bounded derivatives, we have, for any term $\widetilde{Q}_d \equiv \widetilde{Q}_d(t, z_1, z_2) \in \widetilde{\mathcal{Q}}_d$,

$$|\widetilde{Q}_d(t, z_1, z_2)| = O_{\prec}\Big(\Psi^d + \frac{1}{N}\Big),$$

uniformly in $t \in \mathbb{R}^+$, and $z_1, z_2 \in S$ given in (2.7). In the following, we often omit the parameters t, z_1, z_2 for notational simplicity.

7.1. Unmatched terms $K_{p,q+1}$ in (3.19). In this subsection, we follow the idea in Sect. 6 to show the negligibility of the terms $K_{p,q+1}$ given in (3.19) with unmatched indices as defined next, *c.f.,* Proposition 4.2. Recall Definition 4.1 for unmatched terms of the form in (4.8).

Definition 7.1. Given any $\widetilde{Q}_d \in \widetilde{\mathcal{Q}}_d$ of the form in (7.2), let $\nu_j^{(r)}$, $\nu_j^{(c)}$, be the number of times the free summation index $v_j \in \mathcal{I}$ appears in the the row index set $\{x_l^{(i)}\}$ and the column index set $\{y_l^{(i)}\}$ of the Green function entries, *i.e.,*

$$\nu_j^{(r)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : x_l^{(i)} = v_j\}, \quad \nu_j^{(c)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : y_l^{(i)} = v_j\}. \tag{7.5}$$

Definition 4.1 for unmatched terms can be adapted naturally to the general form given in (7.2). Define the set of unmatched summation indices as

$$\mathcal{I}^o := \{1 \le j \le m : \nu_j^{(r)} \ne \nu_j^{(c)}\} \subset \mathcal{I}.$$

If $\mathcal{I}^o$ is not empty, then we say $\widetilde{Q}_d$ is an unmatched term, denoted by $\widetilde{Q}_d^o$. We denote by $\widetilde{\mathcal{Q}}_d^o \subset \widetilde{\mathcal{Q}}_d$ the collection of unmatched terms in the form (7.2) of degree d.

The combination of the identity (5.10) and the cumulant expansion formula Lemma 2.4 used previously in the proof of Proposition 4.2 still applies similarly to the form in (7.2), using that $\{h_{ij}\}$ commute with $\Delta\mathrm{Im}$ given in (3.14), the differentiation rules (3.21) and (3.26), and the assumption that the function F has bounded derivatives. Therefore, for fixed $D \ge 1$ and any unmatched term $\widetilde{Q}_d^o \in \widetilde{\mathcal{Q}}_d^o$ of the form in (7.2) with fixed n given in (7.3),

$$\mathbb{E}[\widetilde{Q}_d^o(t, z_1, z_2)] = O_{\prec}\big(\frac{1}{N} + \Psi^D\big), \tag{7.6}$$

holds uniformly in $t \in \mathbb{R}^+$ and $z_1, z_2 \in S$, as in Proposition 4.2.

Now we return to the right side of (7.1). Using (3.21) and (3.26), all the third order expansion terms $K_{p,q+1}$ in (3.19) for $p+q+1 = 3$ can be written out as a sum of finitely many unmatched terms of the form in (7.2) with an extra factor $\sqrt{N}$ in front, since both the indices a and b appear an odd number of times in the product of the Green function entries. We hence have from (7.6) that

$$|K_{2,1} + K_{1,2} + K_{0,3}| = O_{\prec}(N^{-1/2} + \sqrt{N}\Psi^D)\,. \tag{7.7}$$

Similarly, the fourth order expansion terms $K_{p,q+1}$, $p+q+1 = 4$, in (3.19), with the exception of $K_{2,2}$, can also be written as a finite sum of unmatched terms of the form in (7.2), since the number of times the index a (or b) appears in the row index set $\{x_l^{(i)}\}$ does not agree with the number of times it appears in the column index set $\{y_l^{(i)}\}$. We then find from (7.6) that

$$|K_{3,1} + K_{1,3} + K_{0,4}| = O_{\prec}\big(N^{-1} + \Psi^D\big)\,. \tag{7.8}$$

It hence suffices to estimate the remaining matched terms $K_{2,2}$ and E_2 on the right side of (7.1) as follows. We first consider $K_{2,2}$ given in (3.19), E_2 in (3.20) can then be estimated similarly. The proof contains two parts: 1) expanding matched terms into type-0 terms defined as below (*c.f.,* Proposition 4.3); 2) estimating the resulting type-0 terms whose degrees are at least two (*c.f.,* Lemma 4.1) and the rest type-0 terms of degree zero using the estimate (4.3) in the edge scaling.

7.2. Expanding $K_{2,2}$. We start by $K_{2,2}$ given in (3.19), corresponding to the (2,2)-cumulants. Using the differentiation rules (3.21) and (3.26), we first write $K_{2,2}$ as the following sum

$$K_{2,2} = \sum_{k=1}^{8} I_k, \tag{7.9}$$

with

$$\begin{aligned}
I_1 &:= -\frac{1}{2N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2(G_{bb})^2\right)\Big];\\
I_2 &:= -\frac{1}{N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left(G_{ab}G_{ba}G_{aa}G_{bb}\right)\Big];\\
I_3 &:= -\frac{2}{N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F''(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\,(G_{ab})\Delta\widetilde{\mathrm{Im}}\left(G_{aa}G_{bb}G_{ba}\right)\Big];\\
I_4 &:= -\frac{1}{2N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\,(G_{aa}G_{bb})\Big)^2\Big];\\
I_5 &:= -\frac{1}{N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'''(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\,(G_{ab})\Delta\widetilde{\mathrm{Im}}\,(G_{ba})\Delta\widetilde{\mathrm{Im}}\left(G_{aa}G_{bb}\right)\Big];\\
I_6 &:= -\frac{1}{4N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F''(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{ab})^2\right)\Delta\widetilde{\mathrm{Im}}\left((G_{ba})^2\right)\Big];\\
I_7 &:= -\frac{1}{2N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'''(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\,(G_{ab})\Big)^2\Delta\widetilde{\mathrm{Im}}\left((G_{ba})^2\right)\Big];\\
I_8 &:= -\frac{1}{4N^2}\sum_{a\neq b} s_{ab}^{(2,2)}\mathbb{E}\Big[F''''(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\,(G_{ab})\Big)^2\Big(\Delta\widetilde{\mathrm{Im}}\,(G_{ba})\Big)^2\Big],
\end{aligned} \tag{7.10}$$

where $s_{ab}^{(2,2)}$ $(a \neq b)$ are the (2,2)-cumulants of the rescaled entries $\sqrt{N}h_{ab}$ given in (2.24).

Observe that for the terms given in (7.10), both indices a and b appear exactly twice as the row index and exactly twice as the column index of a Green function entry. We hence consider the special case of the form in (7.2) with the two indices a, b singled out, namely,

$$\frac{1}{N^{\#\mathcal{I}+2}}\sum_{a,b,\mathcal{I}} c_{a,b,\mathcal{I}}\mathbb{E}\Big[F^{(\alpha)}(\mathcal{X})\prod_{i=1}^{i_0}\Delta\widetilde{\mathrm{Im}}\left(\prod_{l=1}^{n_i} G_{x_l^{(i)}y_l^{(i)}}\right)\Big], \tag{7.11}$$

where each $x_l^{(i)}$ and $y_l^{(i)}$ represent a, b or some element in the free summation index set $\mathcal{I} = \{v_j\}_{j=1}^{m}$, and $\{c_{a,b,\mathcal{I}}\}$ are uniformly bounded complex numbers. The number of Green function entries in the product, denoted by n, is given as in (7.3). The degree, denoted by d, is given as in (7.4) by counting the number of off-diagonal Green function entries in the product.

Definition 7.2. Given any term of the form in (7.11), Definition 4.2 for the type-AB, type-A and Type-0 terms of the form in (4.17) can be adapted naturally. Recall $\nu_j^{(r)}$, $\nu_j^{(c)}$ given in (7.5) for any free summation index $v_j \in \mathcal{I}$. We further define similarly for the special summation indices a and b, *i.e.,*

$$\nu_a^{(r)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : x_l^{(i)} = a\}, \quad \nu_a^{(c)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : y_l^{(i)} = a\};$$
$$\nu_b^{(r)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : x_l^{(i)} = b\}, \quad \nu_b^{(c)} := \sum_{i=1}^{i_0} \#\{1 \le l \le n_i : y_l^{(i)} = b\}.$$

If the following two conditions are satisfied,

1. all the free summation indices in $\{\mathcal{I}\}$ appear once in the row index set $\{x_l^{(i)}\}$ and once in the column index set $\{y_l^{(i)}\}$ of the Green function entries, *i.e.,* $\nu_j^{(r)} = \nu_j^{(c)} = 1$ $(1 \le j \le m)$;
2. both the special indices a and b appear twice in the row index set $\{x_l^{(i)}\}$ and twice in the column index set $\{y_l^{(i)}\}$ of the Green function entries, *i.e.,* $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = 2$,

then such a term is a type-AB term. We denote a type-AB term in the form (7.11) of degree d by $T_d^{AB} \equiv T_d^{AB}(t, z_1, z_2)$. The collection of all the type-AB terms of degree d is denoted by $\mathcal{T}_d^{AB} \equiv \mathcal{T}_d^{AB}(t, z_1, z_2)$.

A type-A term in the form (7.11) of degree d, denoted by T_d^A, has $\nu_a^{(r)} = \nu_a^{(c)} = 2$, and $\nu_b^{(r)} = \nu_b^{(c)} = \nu_j^{(r)} = \nu_j^{(c)} = 1$ $(1 \le j \le m)$. Moreover, a type-0 term, denoted by T_d, is of the form (7.11) of degree d with $\nu_a^{(r)} = \nu_a^{(c)} = \nu_b^{(r)} = \nu_b^{(c)} = \nu_j^{(r)} = \nu_j^{(c)} = 1$ $(1 \le j \le m)$. In addition, the collections of the type-A terms and the type-0 terms of the form in (7.2) of degree d are denoted by $\mathcal{T}_d^A \equiv \mathcal{T}_d^A(t, z_1, z_2)$ and $\mathcal{T}_d \equiv \mathcal{T}_d(t, z_1, z_2)$, respectively. We finally remark that the index b in a type-A term, as well as both indices a, b in a type-0 term, do not take special roles. We keep them in the notation in order to emphasize the inheritance from the form (7.11).

Under Definition 7.2, we observe that all the terms given in (7.10) are type-AB terms in the form (7.11) with $\mathcal{I} = \emptyset$ and the coefficients given by $c_{a,b} = s_{ab}^{(2,2)} \delta_{a \ne b}$. In particular, we have that $I_1, I_4 \in \mathcal{T}_0^{AB}$, $I_2, I_3, I_5 \in \mathcal{T}_2^{AB}$, and $I_6, I_7, I_8 \in \mathcal{T}_4^{AB}$. In the following, we use, as in the proof of Proposition 4.3, the combination of the identity (5.10) and cumulant expansion formula Lemma 2.4 to eliminate one pair of the index b and also one pair of the index a, and thus expand the type-AB terms as linear combinations of type-0 terms up to negligible error.

Lemma 7.1. *For any fixed $D \in \mathbb{N}$, we have*

$$K_{2,2} = -\frac{s_4}{2}\Big\{\mathbb{E}\big[F'(\mathcal{X})\big(\Delta\widetilde{\mathrm{Im}}\,(\underline{G})^4\big)\big] + \mathbb{E}\big[F''(\mathcal{X})\big(\Delta\widetilde{\mathrm{Im}}\,(\underline{G})^2\big)^2\big]\Big\}$$
$$+ \sum_{\substack{T_d \in \mathcal{T}_d \\ 2 \le d < D}} T_d + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{7.12}$$

uniformly in $t \in \mathbb{R}^+$, *and* $z_1, z_2 \in S$ *given in (2.7), with*

$$s_4 \equiv s_4(t) := \frac{1}{N^2} \sum_{a \neq b} s_{ab}^{(2,2)}(t), \tag{7.13}$$

where $s_{ab}^{(2,2)}(t)$ *are the (2,2)-cumulants defined in (2.24) of the time-dependent scaled off-diagonal entries* $\sqrt{N} h_{ab}$ *given in (3.7). In addition, the number of type-0 terms appearing in the sum in (7.12) can be bounded by* $(CD)^{cD}$, *for some numerical constants* $C, c > 0$.

Proof. We first consider $I_1 \in \mathcal{T}_0^{AB}$ given in (7.10) and expand it into a sum of finitely many type-0 terms. The expansion procedure consists of two steps: (1) eliminating one pair of the index b and expanding I_1 in terms of type-A terms; (2) further eliminating one pair of the index a in the resulting type-A terms from (1) and then expanding them in terms of type-0 terms.

Recall the definition of $\Delta\widetilde{\mathrm{Im}}$ in (3.14). Replacing G_{bb} by the relation (5.10) and using the cumulant expansion formula in Lemma 2.4, since $\{h_{ij}\}$ commute with $\Delta\widetilde{\mathrm{Im}}$, we have

$$\begin{aligned}
I_1 &= -\frac{1}{2N^2} \sum_{a,b} s_{ab}^{(2,2)} \mathbb{E}\Big[F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}} \Big((G_{aa})^2 G_{bb} \Big(\underline{G} + G_{bb} \underline{HG} - \underline{G}(HG)_{bb} \Big) \Big) \Big] \\
&= -\frac{1}{2N^2} \sum_{a,b} s_{ab}^{(2,2)} \mathbb{E}\Big[F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}} \Big((G_{aa})^2 G_{bb} \underline{G} \Big) \Big] \\
&\quad - \frac{1}{2N^4} \sum_{a,b,j,k} s_{ab}^{(2,2)} \mathbb{E}\Big[\frac{\partial F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}} \Big((G_{aa})^2 (G_{bb})^2 G_{jk} \Big)}{\partial h_{jk}} \Big] \\
&\quad + \frac{1}{2N^4} \sum_{a,b,j,k} s_{ab}^{(2,2)} \mathbb{E}\Big[\frac{\partial F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}} \Big((G_{aa})^2 G_{bb} G_{jj} G_{kb} \Big)}{\partial h_{kb}} \Big] \\
&\quad + O_{\prec}\Big(\frac{1}{\sqrt{N}} \Big),
\end{aligned} \tag{7.14}$$

where the error is from the truncation of the cumulant expansion, as in the proof of Lemma 3.1. The first term on the right side of (7.14) is a type-A term in $\mathcal{T}_0^A$ of the form (7.11) obtained by replacing G_{bb} with $\underline{G}$ in the product of the Green function entries. We observe as in (5.12), the leading sub-term from the second term above, corresponding to taking $\frac{\partial}{\partial h_{jk}}$ of G_{jk}, is exactly canceled by the leading sub-term from the third term resulting from taking $\frac{\partial}{\partial h_{kb}}$ of G_{kb}. Thus using the differentiation rules (3.21) and (3.26), the second and third term on the right side of (7.14) can be written as a sum of at most ten type-AB terms of the form in (7.11) with degrees $d' \geq 2$, the number of Green function entries $n' = 6$, and $\mathcal{I}' = \{j, k\}$. We denote the finite sum as $\sum_{T_{d'}^{AB} \in \mathcal{T}_{d'}^{AB}; d' \geq 2} T_{d'}^{AB}$, and write

$$\begin{aligned}
I_1 &= -\frac{1}{2N^2} \sum_{a,b} s_{ab}^{(2,2)} \mathbb{E}\Big[F'(\mathcal{X}) \Delta\widetilde{\mathrm{Im}} \Big((G_{aa})^2 G_{bb} \underline{G} \Big) \Big] \\
&\quad + \sum_{T_{d'}^{AB} \in \mathcal{T}_{d'}^{AB}; d' \geq 2} T_{d'}^{AB} + O_{\prec}\Big(\frac{1}{\sqrt{N}} \Big).
\end{aligned} \tag{7.15}$$

Next, we further replace G_{bb} in the first terms on the right side of (7.15) by $\underline{G}$ using (5.10) and the cumulant expansion formula as in (7.14) to obtain

$$
\begin{aligned}
&-\frac{1}{2N^2}\sum_{a,b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2 G_{bb}\underline{G}\right)\Big] \\
&\quad = -\frac{1}{2N^2}\sum_{a,b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2(\underline{G})^2\right)\Big] \\
&\qquad -\frac{1}{2N^4}\sum_{a,b,j,k} s_{ab}^{(2,2)}\mathbb{E}\Big[\frac{\partial F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2 G_{bb}\underline{G}G_{jk}\right)}{\partial h_{jk}}\Big] \\
&\qquad +\frac{1}{2N^4}\sum_{a,b,j,k} s_{ab}^{(2,2)}\mathbb{E}\Big[\frac{\partial F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2\underline{G}G_{jj}G_{kb}\right)}{\partial h_{kb}}\Big] \\
&\qquad + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big)\,. \qquad (7.16)
\end{aligned}
$$

Observe similarly to above that the leading sub-term from the second term will be canceled exactly by the leading sub-term from the third term. The remaining sub-terms form a sum of at most ten type-A terms of degrees at least two, denoted as $\sum_{T_{d'}^A\in\mathcal{T}_{d'}^A;d'\geq 2} T_{d'}^A$. Combining with (7.15), we have

$$
\begin{aligned}
I_1 = &-\frac{1}{2N^2}\sum_{a,b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2(\underline{G})^2\right)\Big] \\
&+\sum_{\substack{T_{d'}^A\in\mathcal{T}_{d'}^A\\ d'\geq 2}} T_{d'}^A + \sum_{\substack{T_{d'}^{AB}\in\mathcal{T}_{d'}^{AB}\\ d'\geq 2}} T_{d'}^{AB} + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big)\,. \qquad (7.17)
\end{aligned}
$$

In general, for an arbitrary type-AB term $T_d^{AB}\in\mathcal{T}_d^{AB}$ of the form (7.11) with fixed n given in (7.3), we extend the arguments as in Step 1 in Sect. 5.2, using the differentiation rules (3.21) and (3.26) and that $\{h_{ij}\}$ commute with $\Delta\widetilde{\mathrm{Im}}$ in (3.14). We hence obtain the analogue of (5.18),

$$
T_d^{AB} = \sum_{T_d^A\in\mathcal{T}_d^A} T_d^A + \sum_{\substack{T_{d'}^{AB}\in\mathcal{T}_{d'}^{AB}\\ d'\geq d+1}} T_{d'}^{AB} + O_{\prec}\Big(\frac{1}{\sqrt{N}}\Big), \qquad (7.18)
$$

where the summations above denote a sum of at most two type-A terms of degree d and a sum of at most $6(n+4)$ type-AB terms of degrees not less than $d+1$. The number of the Green function entries in each term above is at most $n+4$. Iterating the expansion procedure (7.18) $D-d$ times and using the local law in (3.10), we expand $T_d^{AB}\in\mathcal{T}_d^{AB}$ as a sum of at most $(6(n+4D))^D$ type-A terms of degrees at least d, up to negligible error. We write for short

$$
T_d^{AB} = \sum_{d\leq d'<D}\sum_{T_{d'}^A\in\mathcal{T}_{d'}^A} T_{d'}^A + O_{\prec}\Big(\frac{1}{\sqrt{N}}+\Psi^D\Big), \qquad (7.19)
$$

where the number of the Green function entries in each type-A term above is bounded by $(n+4D)$.

Therefore, from (7.17) and (7.19), the first term $I_1 \in \mathcal{T}_0^{AB}$ given in (7.10) can be reduced into the following sum of type-A terms,

$$I_1 = -\frac{1}{2N^2}\sum_{a,b} s_{ab}^{(2,2)}\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\mathrm{Im}}\left((G_{aa})^2(\underline{G})^2\right)\Big] + \sum_{2\le d<D}\sum_{T_d^A\in\mathcal{T}_d^A} T_d^A + O_{\prec}\big(\frac{1}{\sqrt{N}}+\Psi^D\big), \tag{7.20}$$

where the number of type-A terms above is bounded by $(C_1D)^{c_1D}$ and the number of the Green function entries in each type-A term is bounded by C_1D for some constants $c_1, C_1 > 0$.

Next, we expand the resulting type-A terms on the right side of (7.20) into linear combinations of type-0 terms by further eliminating one pair of the index a. In general, for any type-A term $T_d^A \in \mathcal{T}_d^A$ of the form (7.11), using similar arguments as in Step 2 in Sect. 5.2, we obtain the analogue of (5.21),

$$T_d^A = \sum_{d\le d'<D}\sum_{T_{d'}\in\mathcal{T}_{d'}} T_{d'} + O_{\prec}\big(\frac{1}{\sqrt{N}}+\Psi^D\big), \tag{7.21}$$

where the number of these type-0 terms is bounded by $(6(n+4D))^D$, and the number of the Green function entries in each type-0 term is bounded by $(n+4D)$.

Similar to (7.15) and (7.17), we further eliminate the index a and expand $I_1 \in \mathcal{T}_0^{AB}$ in (7.20) into type-0 terms using (7.21), *i.e.,*

$$I_1 = -\frac{s_4}{2}\mathbb{E}\Big[F'(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\ (\underline{G})^4\Big)\Big] + \sum_{2\le d<D}\sum_{T_d\in\mathcal{T}_d} T_d + O_{\prec}\big(\frac{1}{\sqrt{N}}+\Psi^D\big), \tag{7.22}$$

with s_4 given in (7.13), where the number of the type-0 terms in the sum above is bounded by $(C_2D)^{c_2D}$.

We now turn to the remaining terms in (7.10). We only sketch the arguments for sake of brevity. We start with $I_4 \in \mathcal{T}_0^{AB}$ in (7.10). Similarly to $I_1 \in \mathcal{T}_0^{AB}$, I_4 can be expanded as

$$I_4 = -\frac{s_4}{2}\mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\ (\underline{G})^2\Big)^2\Big] + \sum_{2\le d<D}\sum_{T_d\in\mathcal{T}_d} T_d + O_{\prec}\big(\frac{1}{\sqrt{N}}+\Psi^D\big). \tag{7.23}$$

Further, using (7.19) and (7.21), $I_2, I_3, I_5 \in \mathcal{T}_2^{AB}$ from (7.10) can also be expanded as sums of finitely many type-0 terms of degrees at least two up to negligible error. Moreover, the last three terms $I_6, I_7, I_8 \in \mathcal{T}_4^{AB}$ can be expanded similarly into type-0 terms of degrees at least four.

In sum, we have expanded $K_{2,2}$ given in (7.9) as a finite sum of type-0 terms,

$$\begin{aligned} K_{2,2} = &-\frac{s_4}{2}\Big\{\mathbb{E}\Big[F'(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\ (\underline{G})^4\Big)\Big] + \mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\mathrm{Im}}\ (\underline{G})^2\Big)^2\Big]\Big\} \\ &+ \sum_{2\le d<D}\sum_{T_d\in\mathcal{T}_d} T_d + O_{\prec}\big(\frac{1}{\sqrt{N}}+\Psi^D\big), \end{aligned}$$

where the number of the type-0 terms in the sum above is bounded by $(C_3D)^{c_3D}$ for some $c_3, C_3 > 0$. This completes the proof of Lemma 7.1.

It then suffices to estimate the resulting type-0 terms on the right side of (7.12) in Lemma 7.1.

7.3. Estimate of type-0 terms. In this subsection, we first show that all the resulting type-0 terms of degrees $d \geq 2$ on the right side of (7.12) are bounded by $O_{\prec}(N^{-1/3})$. Using the estimate in (4.3) and similar arguments as in the proof of (5.23) in Lemma 5.2, we establish the following analogue of Lemma 4.1.

Lemma 7.2. *For any type-0 term $T_d \in \mathcal{T}_d$ of the form (7.11) of degree $d \geq 2$, we have*

$$|T_d(t, z_1, z_2)| = O_{\prec}\big(N^{-1/3}\big), \tag{7.24}$$

uniformly in $t \in \mathbb{R}^+$, $z_1, z_2 \in S_{\mathrm{edge}}$ given in (4.1).

Proof. Given any type-0 term $T_d \in \mathcal{T}_d$ of the form (7.11), we no longer emphasize the indices a,b for notational simplicity. We then write T_d from the definition of $\Delta\widetilde{\mathrm{Im}}$ in (3.14) as

$$\begin{aligned}
&\mathbb{E}\Big[F^{(\alpha)}(\mathcal{X})\frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}} \\
&\prod_{i=1}^{i_0}\Big(\prod_{l=1}^{n_i} G_{x_l^{(i)}y_l^{(i)}}(t, z_1) - \prod_{l=1}^{n_i} G_{x_l^{(i)}y_l^{(i)}}(t, \overline{z_1}) - \prod_{l=1}^{n_i} G_{x_l^{(i)}y_l^{(i)}}(t, z_2) \\
&+\prod_{l=1}^{n_i} G_{x_l^{(i)}y_l^{(i)}}(t, \overline{z_2})\Big],
\end{aligned}$$

with $t \geq 0$, $z_1, z_2 \in S_{\mathrm{edge}}$, and $\alpha, m, i_0, n_i \in \mathbb{N}$, where each summation index $v_j \in \mathcal{I} := \{v_j\}_{j=1}^m$ appears exactly once in the row index set $\{x_l^{(i)}\}$ and once in the column index set $\{x_l^{(i)}\}$ of the Green function entries. In particular, we have $\#\mathcal{I} = n = \sum_{i=1}^{i_0} n_i$. For $1 \leq j \leq m$, if there exist $x_l^{(i)} = y_l^{(i)} = v_j$, then we say v_j is isolated. For any $1 \leq j \neq j' \leq m$, if there exist $1 \leq i \leq i_0$, $1 \leq l \leq n_i$ such that either $x_l^{(i)} = v_j$, $y_l^{(i)} = v_{j'}$ or $y_l^{(i)} = v_j$, $x_l^{(i)} = v_{j'}$, then we say that v_j and $v_{j'}$ are connected indices. We then write out T_d as a linear combination of the terms in the following form, which are rearranged using clusters of connected indices, denoted by $\{v_1^{(q)}, \ldots, v_{l_q}^{(q)}\}_q$,

$$\begin{aligned}
(**) := \mathbb{E}\Big[&F^{(\alpha)}(\mathcal{X})\frac{1}{N^{\#\mathcal{I}}}\sum_{\mathcal{I}} c_{\mathcal{I}} \\
&\prod_q \Big(G_{v_1^{(q)}v_2^{(q)}}(t, z_1^{(q)})G_{v_2^{(q)}v_3^{(q)}}(t, z_2^{(q)})\cdots G_{v_{l_q}^{(q)}v_1^{(q)}}(t, z_{l_q}^{(q)})\Big)\Big],
\end{aligned} \tag{7.25}$$

where $\sum_q l_q = n$, $z_l^{(q)}$ for any q and $1 \leq l \leq l_q$ takes the values $z_1, \overline{z_1}, z_2$, or $\overline{z_2}$. Because the degree $d \geq 2$, there exists at least one cluster of connected indices such that $l_q \geq 2$. We may assume that $q = 1$. Recall that the coefficients $\{c_{\mathcal{I}}\}$ are uniformly bounded and that the function F has bounded derivatives. Then using the local law in (3.10) and the properties of stochastic domination in Lemma 1.1, we have that

$$|(**)| \prec \mathbb{E}\left[\frac{1}{N^{l_1}}\sum_{v_1^{(1)},\ldots,v_{l_1}^{(1)}=1}^{N}\left|G_{v_1^{(1)}v_2^{(1)}}(t, z_1^{(1)})G_{v_2^{(1)}v_3^{(1)}}(t, z_2^{(1)})\cdots G_{v_{l_1}^{(1)}v_1^{(1)}}(t, z_{l_1}^{(1)})\right|\right].$$

In combination with Young's inequality and the Ward identity (5.53), we find, similarly to (5.55), that

$$|(**)| \prec \frac{\mathbb{E}[\operatorname{Im} m_N(t,z_1)]}{(N\eta)^{l_1-1}} + \frac{\mathbb{E}[\operatorname{Im} m_N(t,z_2)]}{(N\eta)^{l_1-1}}, \quad l_1 \geq 2,\ z_1, z_2 \in S_{\text{edge}}. \tag{7.26}$$

Together with the estimate (4.3) on $\mathbb{E}[\operatorname{Im} m_N(t,z)]$ in the edge scaling and the fact that $\eta \geq N^{-1+\epsilon}$, we obtain the estimate in (7.24).

Applying Lemma 7.2 to (7.12), we find that

$$\begin{aligned} K_{2,2} = &-\frac{s_4}{2}\Big\{\mathbb{E}\Big[F'(\mathcal{X})\Big(\Delta\widetilde{\operatorname{Im}}\,(\underline{G})^4\Big)\Big] + \mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\operatorname{Im}}\,(\underline{G})^2\Big)^2\Big]\Big\} \\ &+ O_{\prec}(N^{-1/3} + \Psi^D), \end{aligned} \tag{7.27}$$

uniformly in $t \geq 0$ and $z_1, z_2 \in S_{\text{edge}}$. It then suffices to estimate the remaining type-0 terms of degree zero on the right side of (7.27). Using the definition of $\Delta\widetilde{\operatorname{Im}}$ in (3.14), the estimate in (4.3) of $\mathbb{E}[\operatorname{Im} \underline{G}(t,z)]$ for $z \in S_{\text{edge}}$ and $t \geq 0$, the properties of stochastic domination Lemma 1.1 and that the function F has bounded derivatives, we conclude, for any fixed $D \geq 1$, that

$$|K_{2,2}| = O_{\prec}(N^{-1/3+\epsilon} + \Psi^D), \tag{7.28}$$

uniformly in $t \in \mathbb{R}^+$ and $z_1, z_2 \in S_{\text{edge}}$.

7.4. Estimate of E_2. In this subsection, we estimate the second order term E_2 given in (3.20) similarly as $K_{2,2}$. Using (3.26) and (3.21), we write E_2 as

$$\begin{aligned} E_2 = &-\frac{1}{2N}\sum_{a=1}^{N}(s_{aa}^{(2)} - 1)\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\operatorname{Im}}\,(G_{aa})^2\Big] \\ &-\frac{1}{2N}\sum_{a=1}^{N}(s_{aa}^{(2)} - 1)\mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\operatorname{Im}}\,(G_{aa})\Big)^2\Big]. \end{aligned} \tag{7.29}$$

Observe that the above two terms are both type-A terms in $\mathcal{T}_0^A$ of the form (7.11), where the index b no longer plays a special role. Using the combination of the identity (5.10) and the cumulant expansion formula, we expand E_2 into a sum of finitely many type-0 terms, similarly to (7.12). That is,

$$\begin{aligned} E_2 = &-\frac{s_2 - 1}{2}\Big\{\mathbb{E}\Big[F'(\mathcal{X})\Delta\widetilde{\operatorname{Im}}\,(\underline{G})^2\Big] + \mathbb{E}\Big[F''(\mathcal{X})\Big(\Delta\widetilde{\operatorname{Im}}\,(\underline{G})\Big)^2\Big]\Big\} \\ &+ \sum_{\substack{T_d \in \mathcal{T}_d \\ 2 \leq d < D}} T_d + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \end{aligned} \tag{7.30}$$

uniformly in $t \geq 0$ and $z_1, z_2 \in S$, with

$$s_2 \equiv s_2(t) := \frac{1}{N}\sum_{a=1}^{N} s_{aa}^{(2)}(t), \tag{7.31}$$

where $s_{aa}^{(2,2)}(t)$ are the second order cumulants given in (2.27) of the time-dependent scaled entries $\sqrt{N}h_{aa}$. Moreover, the number of the type-0 terms in the summation on the right side of (7.30) is bounded by $(CD)^{cD}$ for some numerical constants $c, C > 0$.

Similarly to (7.28), we conclude from Lemma 7.2 and the estimate in (4.3) that, for any $D \geq 1$,

$$|E_2| = O_{\prec}\big(N^{-1/3+\epsilon} + \Psi^D\big), \tag{7.32}$$

uniformly in $t \in \mathbb{R}^+$ and $z_1, z_2 \in S_{\text{edge}}$.

Plugging (7.28), (7.32), (7.7), and (7.8) into (3.18) and by choosing $D \geq \frac{1}{\epsilon}$ with $\epsilon > 0$ as in (2.9), we hence finish the proof of Proposition 3.1.

8. Real Symmetric Wigner Matrices

In this section, we prove the Green function comparison theorem, Theorem 1.4, for real Wigner matrices, using similar ideas as for the complex Hermitian case. To simplify the discussion, we will only address the differences.

Consider the real-valued matrix Ornstein–Uhlenbeck process $\big(h_{ab}(t)\big)_{a,b=1}^N$:

$$\mathrm{d}h_{ab}(t) = \sqrt{\frac{1+\delta_{ab}}{N}}\mathrm{d}\beta_{ab}(t) - \frac{1}{2}h_{ab}(t)\mathrm{d}t, \quad h_{ab}(0) = (H_N)_{ab}, \tag{8.1}$$

where $\big(\beta_{ab}(t)\big)_{a\leq b}$ are independent real standard Brownian motions with $\beta_{ba}(t) = \beta_{ab}(t)$. The initial condition H_N is a real symmetric Wigner matrix satisfying Assumption 1.1. In distribution this is equivalent to writing

$$H(t) = \mathrm{e}^{-\frac{t}{2}}H_N + \sqrt{1-\mathrm{e}^{-t}}\mathrm{GOE}_N, \quad t \in \mathbb{R}^+. \tag{8.2}$$

As the analogue of (3.21), we have a new differentiation rule for the Green function entry of a real symmetric matrix,

$$\frac{\partial G_{ij}}{\partial h_{ab}} = -\frac{G_{ia}G_{bj} + G_{ib}G_{aj}}{1+\delta_{ab}}. \tag{8.3}$$

Then using Ito's formula similarly to (3.22), we obtain

$$\mathrm{d}G_{ij}(t,z) = \mathrm{d}M_{ij} + \Theta_{ij}\mathrm{d}t, \tag{8.4}$$

where the diffusion term $\mathrm{d}M_{ij} := -\frac{1}{\sqrt{N}}\sum_{a\leq b}\frac{1}{\sqrt{1+\delta_{ab}}}\Big(G_{ia}G_{bj} + G_{ib}G_{aj}\Big)\mathrm{d}\beta_{ab}$, and the drift term

$$\Theta_{ij} := \frac{1}{2}\sum_{a,b}h_{ab}G_{ia}G_{bj} + \frac{1}{2N}\sum_{a,b}\Big(2G_{ia}G_{ab}G_{bj} + G_{ib}G_{bj}G_{aa} + G_{ia}G_{aj}G_{bb}\Big).$$

Recall F in (2.21) and $\mathcal{X}$ in (3.12). Applying Ito's formula on $F(\mathcal{X})$ and using (8.4), we derive the dynamics of $F(\mathcal{X})$ in the real symmetric case,

$$\mathrm{d}F(\mathcal{X}) = \mathrm{d}M + \Theta\mathrm{d}t,$$

where the diffusion term $\mathrm{d}M$ yields a martingale after integration, see Remark 3.1, and the drift term is given by (we omit the parameters t and $2+x+\mathrm{i}\eta$ of the following Green function entries)

$$\begin{aligned}\Theta = F'(\mathcal{X})\mathrm{Im}\int_{\kappa_1}^{\kappa_2}\frac{1}{2}\sum_{i,a,b}\Big(h_{ab}G_{ia}G_{bi}+\frac{2}{N}G_{ia}G_{ab}G_{bi}\\ +\frac{1}{N}G_{ib}G_{bi}G_{aa}+\frac{1}{N}G_{ia}G_{ai}G_{bb}\Big)\mathrm{d}x\\ +F''(\mathcal{X})\frac{1}{N}\sum_{i,j}\sum_{a,b}\Big(\mathrm{Im}\int_{\kappa_1}^{\kappa_2}G_{ia}G_{bi}\mathrm{d}x\Big)\Big(\mathrm{Im}\int_{\kappa_1}^{\kappa_2}G_{jb}G_{aj}\mathrm{d}x\Big)\\ =\frac{1}{2}\sum_{a,b}h_{ab}\Big(F'(\mathcal{X})\Delta\mathrm{Im}\,G_{ba}\Big)+\frac{1}{N}\sum_{a,b}\Big(F'(\mathcal{X})\Delta\mathrm{Im}\,(G_{aa}G_{bb})\Big)\\ +\frac{1}{N}\sum_{a,b}\Big(F'(\mathcal{X})\Delta\mathrm{Im}\,(G_{ab})^2\Big)\\ +\frac{1}{N}\sum_{a,b}\Big(F''(\mathcal{X})(\Delta\mathrm{Im}\,G_{ab})(\Delta\mathrm{Im}\,G_{ba})\Big).\end{aligned}\tag{8.5}$$

where we abbreviate, for any function $P:\mathbb{R}^+\times\mathbb{C}\setminus\mathbb{R}\longrightarrow\mathbb{C}$,

$$\Delta\mathrm{Im}\,P\equiv(\Delta\mathrm{Im}\,P)(t,z_1,z_2):=\mathrm{Im}\,P(t,z_2)-\mathrm{Im}\,P(t,z_1),\tag{8.6}$$

with $t\in\mathbb{R}^+$, $z_1=2+\kappa_1+\mathrm{i}\eta$, $z_2=2+\kappa_2+\mathrm{i}\eta\in S_{\mathrm{edge}}$, as in (3.15). In fact, comparing with the drift term in (3.24) for complex Hermitian matrices, the notation $\widetilde{\mathrm{Im}}$ in (3.13) is replaced with the imaginary part Im . This is because $\{h_{ab}\}$ commute with taking the imaginary part, and the Green function of a real symmetric matrix satisfies

$$G_{ij}(z)=G_{ji}(z),\quad z\in\mathbb{C}\setminus\mathbb{R}.\tag{8.7}$$

Moreover, using (8.3), it is easy to find the analogous differentiation rule to (3.26),

$$\begin{aligned}\frac{\partial F'(\mathcal{X})}{\partial h_{ab}}&=-\frac{2}{1+\delta_{ab}}F''(\mathcal{X})\sum_{i=1}^{N}\mathrm{Im}\Big(\int_{\kappa_1}^{\kappa_2}G_{ia}G_{bi}(2+x+\mathrm{i}\eta)\mathrm{d}x\Big)\\ &=-\frac{2}{1+\delta_{ab}}F''(\mathcal{X})\Delta\mathrm{Im}\,G_{ab},\end{aligned}\tag{8.8}$$

with $\Delta\mathrm{Im}$ given in (8.6).

Next, we return to the right side of (8.5). Applying the real cumulant expansion formula in Lemma 2.4 for the independent entries $\{h_{ab}\}_{a\le b}$ in the first term up to the fourth order and using the differentiation rules (8.3) and (8.8), the second order terms in the cumulant expansions are canceled exactly by the last three terms on the right side of (8.5). We hence obtain the real analogue of (3.18),

$$\mathbb{E}[\Theta]=\frac{1}{2N}\sum_{a=1}^{N}(s_{aa}^{(2)}-2)\mathbb{E}\Big[\frac{\partial F'(\mathcal{X})\Delta\mathrm{Im}\,G_{aa}}{\partial h_{aa}}\Big]+\frac{1}{4N^{3/2}}\sum_{a,b}s_{ab}^{(3)}\mathbb{E}\Big[\frac{\partial^2F'(\mathcal{X})\Delta\mathrm{Im}\,G_{ba}}{\partial h_{ab}^2}\Big]$$

$$+\frac{1}{12N^2}\sum_{a,b}s_{ab}^{(4)}\mathbb{E}\Big[\frac{\partial^3 F'(\mathcal{X})\Delta\mathrm{Im}\,G_{ba}}{\partial h_{ab}^3}\Big]+O_{\prec}\big(\frac{1}{\sqrt{N}}\big), \tag{8.9}$$

where the error $O_{\prec}(\frac{1}{\sqrt{N}})$ is from the truncation of the cumulant expansion, and $s_{ab}^{(k)}$ is the k-th cumulant defined in (2.27) of the rescaled entries $\sqrt{N}h_{ab}$.

We now claim that Proposition 3.1 holds true in the real case, which leads to Theorem 1.4 for $\beta = 1$. The arguments in the complex case discussed before can be applied similarly, using the modified differentiation rules (8.3) and (8.8), and the real cumulant expansion formula in Lemma 2.4.

To simplify the statement, we only consider the simplest version of the Green function comparison theorem for $F(x) = x$, as proved in Proposition 4.1 for complex Hermitian Wigner matrices. The Green function comparison theorem for general functions F can be proved using the same idea, following the arguments in Sect. 7 for the complex Hermitian case.

Applying (8.4) to the time dependent normalized trace of the Green function, $m_N(t,z)$, we find the real analogue of (4.4), *i.e.,*

$$\mathrm{d}(m_N(t,z)) = \mathrm{d}M_0 + \Theta_0\mathrm{d}t, \tag{8.10}$$

with the diffusion term $\mathrm{d}M_0 := \frac{1}{N}\sum_{v=1}^N \mathrm{d}M_{vv}$ which yields a martingale term after integration; see Remark 3.1, and the drift term $\Theta_0\mathrm{d}t := \frac{1}{N}\sum_{v=1}^N \Theta_{vv}\mathrm{d}t$. Applying the real cumulant expansion formula as in (8.9), the drift term satisfies the real analogue of (4.5), *i.e.,*

$$\begin{aligned}\mathbb{E}[\Theta_0] = {}& \frac{1}{2N^2}\sum_{v,a}(s_{aa}^{(2)}-2)\mathbb{E}\Big[\frac{\partial(G_{va}G_{bv})}{\partial h_{aa}}\Big]+\frac{1}{4N^{5/2}}\sum_{v,a,b}s_{ab}^{(3)}\mathbb{E}\Big[\frac{\partial^2(G_{va}G_{bv})}{\partial h_{ab}^2}\Big]\\ &+\frac{1}{12N^3}\sum_{v,a,b}s_{ab}^{(4)}\mathbb{E}\Big[\frac{\partial^3(G_{va}G_{bv})}{\partial h_{ab}^3}\Big]+O_{\prec}\big(\frac{1}{\sqrt{N}}\big) =: J_2+J_3+J_4+O_{\prec}\big(\frac{1}{\sqrt{N}}\big).\end{aligned} \tag{8.11}$$

It then suffices to prove the estimate (4.6) in the real symmetric case. Using (8.3), the terms J_2, J_3, J_4 above can be written out again in the form (4.8). The degree of a term in the form (4.8) is defined as in (4.9). We recall from (8.7) that the row and column index of a Green function entry can be switched.

Following the idea from complex Hermitian case, the proof of (4.6) consists of three steps: 1) the third order terms from J_3 are unmatched and thus negligible (*c.f.,* Proposition 4.2); 2) expanding the fourth order terms from J_4 (as well as the second order terms in J_2) as linear combinations of type-0 terms of degrees at least two up to arbitrary order (*c.f.,* Proposition 4.3); 3) estimating the resulting type-0 terms in 2) of degrees at least two (*c.f.,* Lemma 4.1).

We start with the first step. Recall Definition 4.1 for unmatched terms in the complex Hermitian case. Because of (8.7), we can ignore the difference from the row and column index of a Green function entry of a real symmetric matrix.

Definition 8.1 (*Terms with unmatched indices in the real case.*) Given any term, denoted by Q_d, of the form (4.8) of degree d, let v_j be the number of times the free summation

index $v_j \in \mathcal{I}$ appears as the row or column index in the product of the Green function entries, *i.e.*,

$$v_j := \#\{1 \le i \le n : x_i = v_j\} + \#\{1 \le i \le n : y_i = v_j\}, \quad 1 \le j \le m. \tag{8.12}$$

We define the set of the unmatched summation indices as

$$\mathcal{I}^o := \{1 \le j \le m : v_j \text{ is odd }\} \subset \mathcal{I}.$$

Note that $\#\mathcal{I}^o$ is even. If $\mathcal{I}_0 = \emptyset$, then we say Q_d is matched. Otherwise, Q_d is an unmatched term, denoted by Q_d^o. The collection of the unmatched terms in the form (4.8) of degree d is denoted by $\mathcal{Q}_d^o$.

Then the third order terms from J_3 on the right side of (8.11) are of the form (4.8) with an extra $\sqrt{N}$ in front and are unmatched with $v_a = v_b = 3$ defined in (8.14) below. Following the arguments in Sect. 6, using the relation (5.10), the real cumulant expansion formula, and the new differentiation rule of the Green function entry (8.3), we observe a similar cancellation to the first order and then expand a unmatched term of the form (4.8) iteratively and prove that Proposition 4.2 holds true in the real symmetric case. Therefore, we have

$$|J_3| = O_{\prec}\big(N^{-1/2} + \sqrt{N}\Psi^D\big). \tag{8.13}$$

Next, in the second step, we expand the remaining terms of the form (4.8) from J_2 and J_4 that are matched. Recall a special case of matched terms as in (4.17) with two summation indices a, b singled out and Definition 4.2 for type-AB, type-A, type-0 terms in the complex case.

Definition 8.2 (*Type-AB terms, type-A terms, type-0 terms*.) Given any term of the form in (4.17) of degree d with two special indices a and b, recall v_j in (8.12) for any $v_j \in \mathcal{I}$ and define similarly

$$\begin{aligned} v_a &:= \#\{1 \le i \le n : x_i = a\} + \#\{1 \le i \le n : y_i = a\}, \\ v_b &:= \#\{1 \le i \le n : x_i = b\} + \#\{1 \le i \le n : y_i = b\}. \end{aligned} \tag{8.14}$$

If for any $1 \le j \le m$, $v_j = 2$ and $v_a = v_b = 4$, then such a term is a type-AB term. A type-A term has $v_a = 4$, and $v_b = v_j = 2$ $(1 \le j \le m)$. Finally, a type-0 term is defined to be in the form (4.17) with $v_a = v_b = v_j = 2$ $(1 \le j \le m)$. The collection of the type-AB, type-A, type-0 terms of degree d is denoted by $\mathcal{P}_d^{AB}$, $\mathcal{P}_d^A$, and $\mathcal{P}_d$, respectively.

Following the arguments in Sect. 5, using the relations (5.10) and (8.3), and the real cumulant expansion formula, we expand any type-AB (or type-A) term iteratively and prove that Proposition 4.3 holds true in the real symmetric case. Therefore, expanding the type-AB terms from J_2 and the type-A terms from J_4 and then combining with (8.13), we write (8.11) as

$$\mathbb{E}[\Theta_0(t,z)] = \sum_{\substack{P_d \in \mathcal{P}_d \\ 2 \le d \le D-1}} \mathbb{E}[P_d(t,z)] + O_{\prec}\Big(\frac{1}{\sqrt{N}} + \Psi^D\Big), \tag{8.15}$$

where the summation on the right side above denotes a linear combination of at most $(CD)^{cD}$ type-0 terms of degrees at least two, for some numerical constants C, c.

In the last step, we aim to show that any type-0 term of degree $d \ge 2$ can be bounded by $O_{\prec}(N^{-1/3})$ for real symmetric Wigner matrices, as in Lemma 4.1. This reduces to prove Lemma 5.2 for the GOE.

Lemma 8.1. *For any $z \in S_{\mathrm{edge}}(\epsilon, C_0)$ given in (4.1) and $t \geq 0$, we have the following uniform estimate:*

$$\frac{1}{N}\mathbb{E}^{\mathrm{GOE}}\Big[\operatorname{Im}\operatorname{Tr} G(z)\Big] = O\big(N^{-1/3+\epsilon}\big)\,. \tag{8.16}$$

The corresponding estimate (5.23) of the type-0 terms of degree $d \geq 2$ considering the GOE follows directly from Lemma 8.1. Following the iterative comparison idea in the proof of Lemma 4.1, one proves Lemma 4.1 similarly in the real case, using (8.3), (8.4) and the real cumulant expansion formula. Therefore, we obtain from (8.15) that (4.6) holds true in the real case and we hence finish the proof of Proposition 4.1 for real Wigner matrices.

Proof of Lemma 8.1. The proof is similar to that of Lemma 5.2. For the one-point correlation function of the GOE and the corresponding diagonal kernel $K_{N,1}$, we refer to [3,34]. From Chapter 3.9 in [3], we write

$$\begin{aligned} K_{N,1}(x,x) = K_{N,2}(x,x) + \frac{\sqrt{N}}{4}\phi_{N-1}(x)\Big(\int_{-\infty}^{\infty}\operatorname{sgn}(x-t)\phi_N(t)\mathrm{d}t\Big) \\ + \frac{1}{2I_{N-1}}\phi_{N-1}(x)\mathbb{1}_{N=2m+1}\,, \end{aligned} \tag{8.17}$$

where $K_{N,2}(x,x)$ is the one-point correlation function for the GUE given by (2.33), $\{\phi_k\}$ are the Hermite functions in (2.30), and we use $\beta = 1, 2$ to denote the symmetry class. Moreover, we set

$$I_{2m} := \int_0^{\infty}\phi_{2m}(t)\mathrm{d}t = \frac{1}{2}\int_{\mathbb{R}}\phi_{2m}(t)\mathrm{d}t = 2^{-1/4}\pi^{1/4}\sqrt{\frac{(2m)!}{2^{2m}(m!)^2}} \sim m^{-1/4}, \tag{8.18}$$

by the Stirling approximation; see Proposition 3.9.28 in [3]. In addition, from Lemma 1 in [26], we have

$$I_{2m+1} := \int_0^{\infty}\phi_{2m+1}(t)\mathrm{d}t = O(m^{-1/4})\,. \tag{8.19}$$

Note that the trace identity for the kernel $K_{N,1}$ still holds as in (2.34). Next, we change the variable as in (2.37) and define

$$K_{N,1}^{\mathrm{edge}}(x,x) := \frac{1}{N^{1/6}}K_{N,1}\Big(2\sqrt{N} + \frac{x}{N^{1/6}}, 2\sqrt{N} + \frac{x}{N^{1/6}}\Big)\,. \tag{8.20}$$

From Theorem 1.1 in [13], as the real analogue of Theorem 2.3, for any $L_0 \in \mathbb{R}$, we have, in the limit of large N, that

$$K_{N,1}^{\mathrm{edge}}(x,x) = K_{\mathrm{airy}}(x,x) + \frac{1}{2}\mathrm{Ai}(x)\int_{-\infty}^{x}\mathrm{Ai}(t)\mathrm{d}t + o(1), \tag{8.21}$$

uniformly in $x \in [L_0, \infty)$. In addition, the right side of (8.21) is uniformly bounded for $x > L_0$; see Chapter 3 in [3] for a reference. Now we are ready to estimate

$$\frac{1}{N}\mathbb{E}^{\mathrm{GOE}}\Big[\operatorname{Im}\operatorname{Tr} G(z)\Big] = \frac{N\eta}{N^2}\mathbb{E}^{\mathrm{GOE}}\Big[\sum_{j=1}^{N}\frac{1}{|\lambda_j - z|^2}\Big]$$

$$= \frac{N\eta}{N^{\frac{2}{3}}} \int_{\mathbb{R}} \frac{K_{N,1}^{\text{edge}}(x,x)}{|x - N^{2/3}\kappa - \mathrm{i}N^{2/3}\eta|^2} \mathrm{d}x, \tag{8.22}$$

for $z = 2+\kappa+\mathrm{i}\eta \in S_{\text{edge}}$, in a similar way as in the proof of Lemma 5.2. Note that (5.44) and (5.45) still hold true for the GOE. We will focus on the regime $-N^{2/3} < x \leq L_0$, for some fixed $L_0 < 0$. Recalling the estimate (5.50) for the GUE, it suffices to prove, for any $x \in (-N^{2/3}, L_0]$, that

$$\left| K_{N,1}^{\text{edge}}(x,x) - K_{N,2}^{\text{edge}}(x,x) \right| = O(1), \tag{8.23}$$

which then leads to

$$\frac{1}{N^{2/3}} \int_{-N^{2/3}}^{L_0} \frac{K_{N,1}^{\text{edge}}(x,x)}{|x - N^{2/3}\kappa + \mathrm{i}N^{2/3}\eta|^2} \mathrm{d}x = O\Big(\frac{1}{N^{\frac{4}{3}-\epsilon}\eta}\Big). \tag{8.24}$$

We hence obtain (8.16) for the GOE. In order to prove (8.23), we split into two cases below and follow ideas from [26].

Case 1: N **is even.** Let $N = 2m$ and the last term in (8.17) is vanishing. Since ϕ_N is even, we write

$$K_{N,1}^{\text{edge}}(x,x) = K_{N,2}^{\text{edge}}(x,x) + \frac{1}{2} N^{1/3} \phi_{N-1}(y) \int_0^y \phi_N(t) \mathrm{d}t, \tag{8.25}$$

where we set for simplicity,

$$y = 2\sqrt{N} + \frac{x}{N^{1/6}}, \quad \text{with} \quad -N^{2/3} < x \leq L_0, \tag{8.26}$$

which implies that $\sqrt{N} < y < 2\sqrt{N} + L_0 N^{-1/6}$. From [26] and references therein, we have the following asymptotic formula of $\phi_N(t)$. In the domain

$$|t| \leq \sqrt{2}\Big((2N+1)^{1/2} - (2N+1)^{-1/6}\Big), \tag{8.27}$$

we have as $N \to \infty$,

$$\phi_N(t) = A_N(t) + O\Big(N^{1/2}(4N+2-t^2)^{-7/4}\Big), \tag{8.28}$$

with

$$A_N(t) := \sqrt{\frac{2}{\pi}} (4N+2-t^2)^{-1/4} \cos\Big(\frac{(2N+1)(2\alpha_N - \sin 2\alpha_N) - \pi}{4}\Big), \tag{8.29}$$

and $\alpha_N := \arccos(t(4N+2)^{-1/2})$. We choose $L_0 < 0$ in (8.26) sufficiently small so that the upper bound y of the integral in (8.25) satisfies (8.27). Thus we have from (8.28) that

$$\begin{aligned} \int_0^y \phi_N(t) \mathrm{d}t &= \int_0^y A_N(t) \mathrm{d}t + O\Big(\sqrt{N} \int_0^y (4N+2-t^2)^{-7/4} \mathrm{d}t\Big) \\ &= \int_0^y A_N(t) \mathrm{d}t + O(N^{-1/4}). \end{aligned} \tag{8.30}$$

Integrating A_N given in (8.29) and using integration by parts, it was shown in (14) in [26] that

$$\left| \int_0^y A_N(t)\mathrm{d}t \right| \leq C(4N + 2 - y^2)^{-3/4} = O(N^{-1/4}), \tag{8.31}$$

with $\sqrt{N} < y < 2\sqrt{N} + L_0 N^{-1/6}$. Thus we have from (8.30) that $\left| \int_0^y \phi_N(t)\mathrm{d}t \right| = O(N^{-1/4})$, for y given in (8.26). Combining with (5.47), the estimate (8.23) then follows from (8.25).

Case 2: N is odd. Let $N = 2m + 1$. Since ϕ_N is an odd function, we write

$$\begin{aligned} K^{\mathrm{edge}}_{N,1}(x,x) =& K^{\mathrm{edge}}_{N,2}(x,x) + \frac{1}{2} N^{1/3} \phi_{N-1}(y) \int_0^y \phi_N(t)\mathrm{d}t \\ & - \frac{1}{2} N^{1/3} \phi_{2m}(y) I_{2m+1} + \frac{1}{2N^{1/6} I_{2m}} \phi_{2m}(y), \end{aligned}$$

with y given in (8.26). Using (5.47), (8.18), and (8.19), the last two terms above are bounded by $O(1)$. The second term can be estimated similarly as in the case $N = 2m$. Thus (8.23) also hold true for $N = 2m + 1$.

We hence have finished the proof of Lemma 8.1.

Acknowledgements. We thank Paul Bourgade, Maurice Duits, Peter J. Forrester and Rong Ma for useful comments and suggestions.

Funding Open access funding provided by Royal Institute of Technology.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Funding Kevin Schnelli is supported in parts by the Swedish Research Council Grant VR-2017-05195, and the Knut and Alice Wallenberg Foundation. Yuanyuan Xu is supported by the Swedish Research Council Grant VR-2017-05195 and the ERC Advanced Grant "RMTBeyond" No. 101020331.

Availability of data and materials Not applicable.

Code availability Not applicable.

Appendix

In this appendix we prove Lemma 2.2 and Lemma 2.3. To prove Lemma 2.2, we follow the arguments in [19].

Proof of Lemma 2.2. Recall the mollifier θ_η given in (2.18) and the indicator function χ_E given in (2.17), where $N^{-1} \ll \eta \ll E_L - E \le CN^{-2/3+\epsilon}$, with $\epsilon > 0$ as in (2.7). It suffices to estimate the linear eigenvalue statistics

$$\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_E \star \theta_\eta(H) = \mathrm{Tr}g(H) = \sum_{j=1}^{N} g(\lambda_j),$$

where

$$g(x) := \chi_E(x) - \chi_E \star \theta_\eta(x) = \Big(\int_{\mathbb{R}} \mathbb{1}_{[E,E_L]}(x) - \int_{E-x}^{E_L-x} \Big)\theta_\eta(y)\mathrm{d}y\,. \tag{A.1}$$

We first consider the function g. Note that for any $E > 0$, we have

$$\frac{c\eta}{E+\eta} \le \int_E^\infty \theta_\eta(y)\mathrm{d}y = \frac{1}{\pi}\int_E^\infty \frac{\eta}{y^2+\eta^2}\mathrm{d}y \le \frac{C\eta}{E+\eta}.$$

Because of the symmetry of the integrand, we have a similar estimate for the integral over $(-\infty, E]$ with $E < 0$. Thus, if $x \in [E, E_L]$, we have from (A.1) that

$$|g(x)| = \Big(\int_{-\infty}^{E-x} + \int_{E_L-x}^{\infty} \Big)\theta_\eta(y)\mathrm{d}y \le C\eta\Big(\frac{1}{|x-E|+\eta} + \frac{1}{|x-E_L|+\eta}\Big).$$

Else, if $x \in [E, E_L]^c$, we have from the positiveness of $\theta_\eta(y)$ that

$$|g(x)| = \int_{E-x}^{E_L-x} \theta_\eta(y)\mathrm{d}y \le \begin{cases} \frac{C\eta}{|x-E|+\eta}, & \text{if} \quad x < E, \\ \frac{C\eta}{|x-E_L|+\eta}, & \text{if} \quad x > E_L, \end{cases} \tag{A.2}$$

It is easy to check that

$$|g(x)| \le 2C, \quad \text{for} \quad x \in \mathbb{R}. \tag{A.3}$$

Now we choose a parameter l_1 such that $\eta \ll l_1 \ll E_L - E \le CN^{-2/3+\epsilon}$. If we further assume $\min\{|x-E|, |x-E_L|\} \ge l_1$, then we have

$$|g(x)| \le \frac{2C\eta}{l_1}, \quad \text{for} \quad |x-E| > l_1, \quad |x-E_L| < l_1. \tag{A.4}$$

Plugging (A.3) and (A.4) into (A.1), we hence obtain

$$\begin{aligned}\Big|\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_E \star \theta_\eta(H)\Big| \le C\Big(&\mathcal{N}(E-l_1, E+l_1) + \mathcal{N}(E_L-l_1, \infty) \\ &+ \frac{\eta}{l_1}\mathcal{N}(E, E_L) + \mathrm{Tr}f(H)\Big),\end{aligned}$$

where

$$f(x) := \big(\chi_E \star \theta_\eta\big)(x)\,\mathbb{1}_{x \le E-l_1}\,.$$

Using the rigidity of eigenvalues in (2.15), we obtain that

$$\Big|\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_E \star \theta_\eta(H)\Big| \le C\Big(\mathcal{N}(E-l_1, E+l_1) + \frac{\eta}{l_1}N^{2\epsilon} + \mathrm{Tr}f(H)\Big), \tag{A.5}$$

with high probability, *i.e.,* with probability bigger than $1 - N^{-\Gamma}$ for any large $\Gamma > 0$, for N sufficiently large. It is then sufficient to estimate $\mathrm{Tr}\, f(H)$. We write

$$\mathrm{Tr}\, f(H) = \sum_{\lambda_i \le E - l_1} f(\lambda_i) = \sum_{k=0}^{\infty} \sum_{\lambda_i \in \mathcal{I}_k} f(\lambda_i), \quad \mathcal{I}_k := (E - 3^{k+1} l_1, E - 3^k l_1]. \tag{A.6}$$

If $x \le E - l_1$, then $E_L - x \ge E - x \ge l_1 \gg \eta$, and we have

$$\begin{aligned} f(x) = \int_{E-x}^{E_L - x} \theta_\eta(y) \mathrm{d}y &= \arctan\left(\frac{E_L - x}{\eta}\right) - \arctan\left(\frac{E - x}{\eta}\right) \\ &= \arctan\left(\frac{\eta}{E - x}\right) - \arctan\left(\frac{\eta}{E_L - x}\right) \le \frac{C\eta(E_L - E)}{(E_L - x)(E - x)} \\ &\le C \min\left\{\frac{(E_L - E)\eta}{(E - x)^2}, \frac{\eta}{E - x}\right\}. \end{aligned}$$

In combination with (A.6), we have

$$\mathrm{Tr}\, f(H) \le C \sum_{k=0}^{\infty} \min\left\{\frac{(E_L - E)\eta}{3^{2k} l_1^2}, \frac{\eta}{3^k l_1}\right\} \mathcal{N}_k, \quad \mathcal{N}_k := \#\{i : \lambda_i \in \mathcal{I}_k\}. \tag{A.7}$$

We next estimate $\mathcal{N}_k$ using the local law in (3.10). Consider

$$\mathrm{Im}\, m_N(E - 2 \cdot 3^k l_1 + \mathrm{i} 3^k l_1) = \frac{1}{N} \sum_{i=1}^{N} \frac{3^k l_1}{|\lambda_i - (E - 2 \cdot 3^k l_1)|^2 + (3^k l_1)^2} \ge \frac{1}{N} \frac{\mathcal{N}_k}{2 \cdot 3^k l_1}. \tag{A.8}$$

Using the local law in (3.10) and (2.5), for any small $\tau > 0$ and large $\Gamma > 0$, we find an upper bound for the left hand side above as

$$\begin{aligned} \mathrm{Im}\, m_N(E - 2 \cdot 3^k l_1 + \mathrm{i} 3^k l_1) &\le \mathrm{Im}\, m_{sc}(E - 2 \cdot 3^k l_1 + \mathrm{i} 3^k l_1) + \frac{N^{\epsilon+\tau}}{N 3^k l_1} \\ &\le C \sqrt{3^k l_1 + |E - 2 \cdot 3^k l_1 - 2|} + \frac{N^{\epsilon+\tau}}{N 3^k l_1} \\ &\le C\left(\sqrt{3^k l_1} + \frac{N^{\epsilon+\tau}}{N 3^k l_1} + N^{-1/3+\epsilon}\right), \end{aligned}$$

with probability bigger than $1 - N^{-\Gamma}$. By choosing $\tau < \epsilon$, we hence obtain from (A.8) that

$$\mathcal{N}_k \le C\left((3^k l_1)^{3/2} N + N^{2\epsilon} + 3^k l_1 N^{2/3+\epsilon}\right),$$

with high probability. Combining with (A.7), we have

$$\begin{aligned} \mathrm{Tr}\, f(H) &\le C \sum_{k=0}^{\infty} \min\left\{\frac{(E_L - E)\eta}{3^{2k} l_1^2}, \frac{\eta}{3^k l_1}\right\}\left((3^k l_1)^{3/2} N + N^{2\epsilon} + 3^k l_1 N^{2/3+\epsilon}\right) \\ &\le \frac{C N^{1/3+\epsilon} \eta}{\sqrt{l_1}} + \frac{C N^{2\epsilon} \eta}{l_1} \le \frac{C' N^{2\epsilon} \eta}{l_1}, \end{aligned}$$

with high probability. Together with (A.5), we hence obtain

$$\left|\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_E \star \theta_\eta(H)\right| \le C'\Big(\mathcal{N}(E-l_1, E+l_1) + \frac{\eta}{l_1}N^{2\epsilon}\Big),$$

with high probability. This completes the proof of Lemma 2.2. □

Next, we use Lemma 2.2 to prove Lemma 2.3.

Proof of Lemma 2.3. Under the same assumption in Lemma 2.2, we choose a parameter l satisfying $N^{-1} \ll \eta \ll l_1 \ll l \ll E_L - E \le CN^{-2/3+\epsilon}$. We have from Lemma 2.3 that

$$\begin{aligned}\mathrm{Tr}\chi_E(H) &\le l^{-1}\int_{E-l}^{E}\mathrm{Tr}\chi_y(H)\mathrm{d}y\\ &\le l^{-1}\int_{E-l}^{E}\mathrm{Tr}\chi_y\star\theta_\eta(H)\mathrm{d}y + Cl^{-1}\int_{E-l}^{E}\Big(\mathcal{N}(y-l_1,y+l_1)+\frac{\eta}{l_1}N^{2\epsilon}\Big)\mathrm{d}y\\ &\le \mathrm{Tr}\chi_{E-l}\star\theta_\eta(H) + C\Big(N^{2\epsilon}\frac{\eta}{l_1}+\frac{l_1}{l}\mathcal{N}(E-2l,E+l)\Big), \qquad (A.9)\end{aligned}$$

with high probability. Using the rigidity result (2.13) and $l \ll N^{-2/3+\epsilon}$, we have

$$\mathcal{N}(E-2l,E+l) \le \int_{E-2l}^{E+l} N\rho_{sc}(x)\mathrm{d}x + N^{\epsilon} \le CN^{\epsilon},$$

with high probability. Thus we obtain from (A.9) that with high probability

$$\mathrm{Tr}\chi_E(H) - \mathrm{Tr}\chi_{E-l}\star\theta_\eta(H) \le CN^{2\epsilon}\Big(\frac{\eta}{l_1}+\frac{l_1}{l}\Big).$$

One obtains a lower bound similarly. Therefore, for any large $\Gamma > 0$, we have

$$\mathrm{Tr}\chi_{E+l}\star\theta_\eta(H) - CN^{2\epsilon}\Big(\frac{\eta}{l_1}+\frac{l_1}{l}\Big) \le \mathrm{Tr}\chi_E(H) \le \mathrm{Tr}\chi_{E-l}\star\theta_\eta(H) + CN^{2\epsilon}\Big(\frac{\eta}{l_1}+\frac{l_1}{l}\Big),$$

with probability bigger than $1 - N^{-\Gamma}$. We pick $l_1 = N^{3\epsilon}\eta$ and $l = N^{3\epsilon}l_1$ such that $N^{2\epsilon}\left(\frac{\eta}{l_1}+\frac{l_1}{l}\right) = N^{-\epsilon}$. Since the counting function $\mathcal{N}(E, E_L) = \mathrm{Tr}\chi_E(H)$ is integer valued, we have

$$\begin{aligned}\mathbb{P}\Big(\mathcal{N}(E,E_L)=0\Big) &\le \mathbb{P}\Big(\mathrm{Tr}\chi_{E+l}\star\theta_\eta(H)\le 1/9\Big) + N^{-\Gamma}\\ &\le \mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E+l}\star\theta_\eta(H)\Big)\Big] + N^{-\Gamma},\end{aligned}$$

where F is the cut-off function given in (2.21). In the other direction, we have

$$\mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E-l}\star\theta_\eta(H)\Big)\Big] \le \mathbb{P}\Big(\mathrm{Tr}\chi_{E-l}\star\theta_\eta(H)\le 2/9\Big) \le \mathbb{P}\Big(\mathcal{N}(E,E_L)=0\Big) + N^{-\Gamma}.$$

Therefore, together with (2.15), we obtain

$$\mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E-l}\star\theta_\eta(H)\Big)\Big] - N^{-\Gamma} \le \mathbb{P}\Big(\mathcal{N}(E,\infty)=0\Big) \le \mathbb{E}\Big[F\Big(\mathrm{Tr}\chi_{E+l}\star\theta_\eta(H)\Big)\Big] + N^{-\Gamma}.$$

This completes the proof of Lemma 2.3. □

References

1. Adhikari, A., Huang, J.: Dyson Brownian motion for general β and potential at the edge. Probab. Theory Rel. Fields **178**(3), 893–950 (2020)
2. Alt, J., Erdős, L., Krüger, T., Schröder, D.: Correlated random matrices: band rigidity and edge universality. Ann. Probab. **48**(2), 963–1001 (2020)
3. Anderson, G., Guionnet, A., Zeitouni, O: An introduction to random matrices. In: Cambridge Studies in Advanced Mathematics, vol. 118. Cambridge University Press, Cambridge (2010)
4. Bonan, S.S., Clark, D.S.: Estimates of the Hermite and the Freud polynomials. J. Approx. Theory **63**, 210–224 (1990)
5. Bourgade, P.: Extreme gaps between eigenvalues of Wigner matrices. J. Eur. Math. Soc. (2021)
6. Bourgade, P., Erdős, L., Yau, H.-T.: Edge universality of beta ensembles. Commun. Math. Phys. **332**(1), 261–353 (2014)
7. Boutet de Monvel, A., Khorunzhy, A.: Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices. Random Oper. Stoch. Equ. **7**(2), 149–168 (1999)
8. Chatterjee, S.: A generalization of the Lindeberg principle. Ann. Probab. **34**(6), 2061–2076 (2006)
9. Choup, L.: Edgeworth expansion of the largest eigenvalue distribution function of Gaussian orthogonal ensemble. J. Math. Phys. **50**(1), 013512 (2009)
10. Collins, B.: Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. Int. Math. Res. Not. IMRN **17**, 953–982 (2003)
11. Collins, B., Śniady, P.: Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. Commun. Math. Phys. **264**, 773–795 (2006)
12. Deift, P., Gioev, D.: Random matrix theory: invariant ensembles and universality. In: Courant Lecture Notes in Mathematics. Vol. 18. American Mathematical Society (2009)
13. Deift, P., Gioev, D.: Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. Commun. Pure Appl. Math. **60**(6), 867–910 (2007)
14. El Karoui, N.: A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. Ann. Probab. **34**(6), 2077–2117 (2006)
15. Erdős, L., Knowles, A., Yau, H.-T.: Averaging fluctuations in resolvents of random band matrices. Ann. Henri Poincaré **14**, 1837–1926 (2013)
16. Erdős, L., Krüger, T., Schröder, D.: Random matrices with slow correlation decay. Forum Math. Sigma (2019), **7**(8) (2019)
17. Erdős, L., Yau, H.-T.: A dynamical approach to random matrix theory. In: Courant Lecture Notes in Mathematics, vol. **28**. American Mathematical Society, Providence (2017)
18. Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Probab. Theory Rel. Fields **154**(1–2), 341–407 (2012)
19. Erdős, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized Wigner matrices. Adv. Math. **229**(3), 1435–1515 (2012)
20. Forrester, Peter J., Trinh, Allan K.: Functional form for the leading correction to the distribution of the largest eigenvalue in the GUE and LUE. J. Math. Phys. **59**(5), 053302 (2018)
21. He, Y., Knowles, A.: Mesoscopic eigenvalue statistics of Wigner matrices. Ann. Appl. Probab. **27**(3), 1510–1550 (2017)
22. He, Y., Knowles, A.: Fluctuations of extreme eigenvalues of sparse Erdős-Rényi graphs. arXiv:2005.02254 (2020)
23. Huang, J., Landon, B., Yau, H.-T.: Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdös–Rényi graphs. Ann. Probab. **48**(2), 916–962 (2020)
24. Johansson, K.: Random matrices and determinantal processes. arXiv:math-ph/0510038 (2005)
25. Johnstone, I.M., Ma, Z.: Fast approach to the Tracy–Widom law at the edge of GOE and GUE. Ann. Appl. Prob. **22**(5), 1962–1988 (2012)
26. Kholopov, A.A., Tikhomirov, A.N., Timushev, D.A.: Rate of convergence to the semicircle law for the Gaussian orthogonal ensemble. Theory Probab. Appl. **52**(1), 171–177 (2008)
27. Khorunzhy, A., Khoruzhenko, B., Pastur, L.: Asymptotic properties of large random matrices with independent entries. J. Math. Phys. **37**(10), 5033–5060 (1996)
28. Landon, B, Yau, H-T.: Edge statistics of Dyson Brownian motion. arXiv:1712.03881 (2017)
29. Lee, J.O., Schnelli, K.: Edge universality for deformed Wigner matrices. Rev. Math. Phys. **27**(8) (2015)
30. Lee, J.O., Schnelli, K.: Local law and Tracy–Widom limit for sparse random matrices. Probab. Theory Relat. Fields **171**(1), 543–616 (2018)
31. Lee, J.O., Yin, J.: A necessary and sufficient condition for edge universality of Wigner matrices. Duke Math. J. **163**(1), 117–173 (2014)
32. Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Ann. Probab. **37**, 1778–1840 (2009)

33. Ma, Z.: Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. Bernoulli **18**(1), 322–359 (2012)
34. Mehta, M.: Random matrices. In: Pure and Applied Mathematics, vol. **142**, 3rd version. Academic Press (2004)
35. Péché, S., Soshnikov, A.: On the lower bound of the spectral norm of symmetric random matrices with independent entries. Electron. Commun. Probab. **13**, 280–290 (2008)
36. Péché, S., Soshnikov, A.: Wigner random matrices with non-symmetrically distributed entries. J. Stat. Phys. **129**, 857–884 (2007)
37. Pillai, N., Yin, J.: Universality of covariance matrices. Ann. Appl. Probab. **24**(3), 935–1001 (2014)
38. Schnelli, K., Xu, Y.: Convergence rate to the Tracy–Widom laws for the largest eigenvalue sample covariance matrices. arXiv:2108.02728 (2021)
39. Schnelli, K., Xu, Y.: Quantitative Tracy–Widom law for generalized Wigner matrices. In preparation
40. Sinai, Y., Soshnikov, A.: A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge. Funct. Anal. Appl. **32**, 114–131 (1998)
41. Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. Commun. Math. Phys. **207**, 697–733 (1999)
42. Soshnikov, A.: Determinantal random point fields. Russ. Math. Surv. **55**(5), 923–975 (2000)
43. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics up to the edge. Commun. Math. Phys. **298**, 549–572 (2010)
44. Tracy, C., Widom, H.: Level-spacing distributions and the airy kernel. Commun. Math. Phys. **159**, 151–174 (1994)
45. Tracy, C., Widom, H.: On orthogonal and symplectic matrix ensembles. Commun. Math. Phys. **177**, 727–754 (1996)
46. Wang, H.: Quantitative universality for the largest eigenvalue of sample covariance matrices. arXiv:1912.05473 (2019)

Communicated by L. Erdos