

# MESOSCOPIC EIGENVALUE STATISTICS OF WIGNER MATRICES<sup>1</sup>

BY YUKUN HE AND ANTTI KNOWLES

*ETH Zürich*

We prove that the linear statistics of the eigenvalues of a Wigner matrix converge to a universal Gaussian process on all mesoscopic spectral scales, that is, scales larger than the typical eigenvalue spacing and smaller than the global extent of the spectrum.

**1. Introduction.** Let  $H$  be an  $N \times N$  Wigner matrix—a Hermitian random matrix with independent upper-triangular entries with zero expectation and constant variance. We normalize  $H$  so that as  $N \rightarrow \infty$  its spectrum converges to the interval  $[-2, 2]$  and, therefore, its typical eigenvalue spacing is of order  $N^{-1}$ . In this paper, we study linear eigenvalue statistics of  $H$  of the form

$$(1.1) \quad \text{Tr} f\left(\frac{H - E}{\eta}\right),$$

where  $f$  is a test function,  $E \in (-2, 2)$  a fixed reference energy inside the bulk spectrum, and  $\eta$  an  $N$ -dependent spectral scale. We distinguish the *macroscopic regime*  $\eta \asymp 1$ , the *microscopic regime*  $\eta \asymp N^{-1}$ , and the *mesoscopic regime*  $N^{-1} \ll \eta \ll 1$ . The limiting distribution of (1.1) in the macroscopic regime is by now well understood; see [2, 18]. Conversely, in the microscopic regime the limiting distribution of (1.1) is governed by the distribution of individual eigenvalues of  $H$ . This question has recently been the focus of much attention, and the universality of the emerging Wigner–Dyson–Mehta (WDM) microscopic eigenvalue statistics for Wigner matrices has been established in great generality; we refer to the surveys [9, 14] for further details.

In this paper, we focus on the mesoscopic regime. The study of linear eigenvalue statistics of Wigner matrices on mesoscopic scales was initiated in [4, 5]. In [4], the authors consider the case of Gaussian  $H$  (the Gaussian Orthogonal Ensemble) and take  $f(x) = (x - i)^{-1}$ , in which case (1.1) is  $\eta$  times the trace of the resolvent of  $H$  at  $E + i\eta$ . Under these assumptions, it is proved in [4] that, after a centring, the linear statistic (1.1) converges in distribution to a Gaussian random variable

---

Received March 2016.

<sup>1</sup>Supported in part by the Swiss National Science Foundation Grant 144662 and the SwissMAP NCCR grant.

*MSC2010 subject classifications.* Primary 15B52; secondary 60B20.

*Key words and phrases.* Wigner matrices, mesoscopic eigenvalue distribution, linear statistics, universality.

on all mesoscopic scales  $N^{-1} \ll \eta \ll 1$ . In [5], this result was extended to a class of Wigner matrices for the range of mesoscopic scales  $N^{-1/8} \ll \eta \ll 1$ . Recently, the results of [5] were extended in [17] to arbitrary Wigner matrices, mesoscopic scales  $N^{-1/3} \ll \eta \ll 1$ , and general test functions  $f$  subject to mild regularity and decay conditions. Apart from the works [5, 17] on Wigner matrices, mesoscopic eigenvalue statistics have also been analysed for invariant ensembles; see [6, 8] and the references therein.

Let  $Z = (Z(f))_f$  denote the Gaussian process obtained as the mesoscopic limit of a centring of (1.1). From the works cited above, it is known that the variance of  $Z(f)$  is the square of the Sobolev  $H^{1/2}$ -norm of  $f$ :

$$(1.2) \quad \mathbb{E}Z(f)^2 = \frac{1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy = \frac{1}{\pi} \int |\xi| |\hat{f}(\xi)|^2 d\xi,$$

where  $\hat{f}(\xi) := (2\pi)^{-1/2} \int f(x) e^{-i\xi x} dx$ . Hence, a remarkable property of  $Z$  is scale invariance:  $Z \stackrel{d}{=} Z_\lambda$ , where  $Z_\lambda(f) := Z(f_\lambda)$  and  $f_\lambda(x) := f(\lambda x)$ . It may be shown that  $Z$  is obtained by extrapolating the microscopic WDM statistics to mesoscopic scales, and we therefore refer to its behaviour as the WDM mesoscopic statistics. In light of the microscopic universality results for Wigner matrices mentioned above, the emergence of WDM statistics on mesoscopic scales is therefore not surprising.

All of the models described above, including Wigner matrices, correspond to mean-field models without spatial structure. In [10, 11], linear eigenvalue statistics were analysed for *band matrices*, where matrix entries are set to be zero beyond a certain distance  $W \leq N$  from the diagonal. Band matrices are a commonly used model of quantum transport in disordered media. Wigner matrices can be regarded as a special case  $W = N$  of band matrices. Unlike the mean-field Wigner matrices, band matrices possess a nontrivial spatial structure. An important motivation for the study of mesoscopic eigenvalue statistics of band matrices arises from the theory of conductance fluctuations; we refer to [10] for more details. The results of [10, 11] hold in the regime  $W^{-1/3} \ll \eta \ll 1$ , and hence for the special case of Wigner matrices they hold for  $N^{-1/3} \ll \eta \ll 1$ . A key conclusion of [10, 11] is that for band matrices there is a sharp transition in the mesoscopic spectral statistics, predicted in the physics literature [1]: above a certain critical spectral scale  $\eta_c$  the mesoscopic spectral statistics are no longer governed by the Wigner–Dyson–Mehta mesoscopic statistics (1.2), but by new limiting statistics, referred to as Altshuler–Shklovskii (AS) statistics in [10, 11], which are not scale invariant like (1.2). For instance for the  $d$ -dimensional ( $d = 1, 2, 3$ ) AS statistics, the variance of the limiting Gaussian process  $Z(f)$  is  $\frac{1}{\pi} \int |\xi|^{1-d/2} |\hat{f}(\xi)|^2 d\xi$  instead of the right-hand side of (1.2); see [10, 11]. In particular, there is a range of mesoscopic scales such that, although the *microscopic* eigenvalue statistics are expected to satisfy the WDM statistics, the *mesoscopic* statistics do not, and instead satisfy the

AS statistics. Hence, the WDM statistics on microscopic and mesoscopic scales do in general not come hand in hand.

In this paper, we establish the WDM mesoscopic statistics for Wigner matrices in full generality. Our results hold on all mesoscopic scales  $N^{-1} \ll \eta \ll 1$  and all Wigner matrices whose entries have finite moments of order  $4 + o(1)$ . We require our test functions to have  $1 + o(1)$  continuous derivatives and be subject to mild decay assumptions, as in [17]. The precise statements are given in Section 2 below.

Our proof is based on two main ingredients: families of self-consistent equations for moments of linear statistics inspired by [5], and the local semicircle law for Wigner matrices from [13, 16]. Our analysis of the self-consistent equations departs significantly from that of [5], since repeating the steps there, even using the optimal bounds provided by the local semicircle law, requires the lower bound  $\eta \gg N^{-1/2}$  on the spectral scale. In addition, dealing with general test functions  $f$  instead of  $f(x) = (x - i)^{-1}$  requires a new family of self-consistent equations that is combined with the Helffer–Sjöstrand representation for general functions of  $H$ . We perform the proof in two major steps.

In the first step, performed in Section 4, we consider traces of resolvents  $G \equiv G(z) = (H - z)^{-1}$ , corresponding to taking  $f(x) = (x - i)^{-1}$  in (1.1). Denoting by  $\underline{G}$  the normalized trace of  $G$  and  $\langle X \rangle := X - \mathbb{E}X$ , we derive a family of self-consistent equations [see (4.22) below] for the moments  $\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m$  following [5], obtained by expanding one factor inside the expectation using the resolvent identity and then applying a standard cumulant expansion (see Lemma 3.1 below) to the resulting expression of the form  $\mathbb{E}f(h)h$ . The main work of the first step is to estimate the error terms of the self-consistent equation. An important ingredient is a careful estimate of the remainder term in the cumulant expansion [see Lemma 4.6(i) below], which allows us to remove the condition  $\eta \gg N^{-1/2}$  on the spectral scale that would be required if one merely combined the local semicircle law with the approach of [5]. Other important tools behind these estimates are new precise high-probability bounds on the entries of the powers  $G^k$  of the resolvent (see Lemma 4.4 below) and a further family of self-consistent equations for  $\mathbb{E}\underline{G}^k$  (see Lemma 4.8 below).

In the second step, performed in Section 5, we consider general test functions  $f$ . The starting point is the well-known Helffer–Sjöstrand representation of (1.1) as an integral of traces of resolvents. An important ingredient of the proof is a self-consistent equation [see (5.21) below] that is used on the integrand of the Helffer–Sjöstrand representation. Compared to the first step, we face the additional difficulty that the arguments  $z$  of the resolvents are now integrated over, and may in particular have very small imaginary parts. Handling such integrals for arbitrary mesoscopic scales  $\eta$  and comparatively rough test functions in  $\mathcal{C}^{1+o(1)}$  requires some care, and we use two different truncation scales  $N^{-1} \ll \omega \ll \sigma \ll \eta$  in the imaginary part of  $z$ , which allow us to extract the leading term. See Section 5 for a more detailed explanation of the truncation scales. The error terms are estimated by a generalization of the estimates established in the first step.

Finally, in Section 6 we give a simple truncation and comparison argument that allows us to consider without loss of generality Wigner matrices whose entries have finite moments of all order, instead of finite moments of order  $4 + o(1)$ .

*Conventions.* We regard  $N$  as our fundamental large parameter. Any quantities that are not explicitly constant or fixed may depend on  $N$ ; we almost always omit the argument  $N$  from our notation. We use  $C$  to denote a generic large positive constant, which may depend on some fixed parameters and whose value may change from one expression to the next. Similarly, we use  $c$  to denote a generic small positive constant.

**2. Results.** We begin this section by defining the class of random matrices that we consider.

**DEFINITION 2.1 (Wigner matrix).** A *Wigner matrix* is a Hermitian  $N \times N$  matrix  $H = H^* \in \mathbb{C}^{N \times N}$  whose entries  $H_{ij}$  satisfy the following conditions:

- (i) The upper-triangular entries  $(H_{ij} : 1 \leq i \leq j \leq N)$  are independent.
- (ii) We have  $\mathbb{E}H_{ij} = 0$  for all  $i, j$ , and  $\mathbb{E}|\sqrt{N}H_{ij}|^2 = 1$  for  $i \neq j$ .
- (iii) There exist constants  $c, C > 0$  such that  $\mathbb{E}|\sqrt{N}H_{ij}|^{4+c-2\delta_{ij}} \leq C$  for all  $i, j$ .

We distinguish the *real symmetric* case, where  $H_{ij} \in \mathbb{R}$  for all  $i, j$ , and the *complex Hermitian* case, where  $\mathbb{E}H_{ij}^2 = 0$  for  $i \neq j$ .

For conciseness, we state our results for the real symmetric case. Analogous results hold for the complex Hermitian case; see Remark 2.4 below.

Our first result is on the convergence of the trace of the resolvent  $G(z) := (H - z)^{-1}$ , where  $\text{Im } z \neq 0$ . The Stieltjes transform of the empirical spectral measure of  $H$  is

$$(2.1) \quad \underline{G}(z) := \frac{1}{N} \text{Tr } G(z).$$

For  $x \in \mathbb{R}, z \in \mathbb{C}$ , and  $\text{Im } z \neq 0$ , the Wigner semicircle law  $\varrho$  and its Stieltjes transform  $m$  are defined by

$$(2.2) \quad \varrho(x) := \frac{1}{2\pi} \sqrt{(4 - x^2)_+}, \quad m(z) := \int \frac{\varrho(x)}{x - z} dx.$$

Denote by  $\mathbb{H} := \{z \in \mathbb{C} : \text{Im } z > 0\}$  the complex upper half-plane. Let  $(Y(b))_{b \in \mathbb{H}}$  denote the complex-valued Gaussian process with mean zero and covariance

$$(2.3) \quad \mathbb{E}(Y(b_1)\overline{Y(b_2)}) = -\frac{2}{(b_1 - \overline{b_2})^2}, \quad \mathbb{E}(Y(b_1)Y(b_2)) = 0$$

for all  $b_1, b_2 \in \mathbb{H}$ . For instance, we can set

$$(2.4) \quad Y(b) = \frac{1}{\sqrt{2}} \left( \frac{2}{b+i} \right)^2 \sum_{k=0}^{\infty} \sqrt{k+1} \left( \frac{b-i}{b+i} \right)^k \vartheta_k,$$

where  $(\vartheta_k)_{k=0}^{\infty}$  is a family of independent standard complex Gaussians, where by definition, a *standard complex Gaussian* is a mean-zero Gaussian random variable  $X$  satisfying  $\mathbb{E}X^2 = 0$  and  $\mathbb{E}|X|^2 = 1$ . Finally, for  $E \in \mathbb{R}$  and  $\eta > 0$ , we define the process  $(\hat{Y}(b))_{b \in \mathbb{H}}$  through

$$\hat{Y}(b) := N\eta(G(E + b\eta) - m(E + b\eta))$$

for all  $b \in \mathbb{H}$ . We may now state our first result.

**THEOREM 2.2 (Convergence of the resolvent).** *Let  $H$  be a real symmetric Wigner matrix. Fix  $\alpha \in (0, 1)$  and set  $\eta := N^{-\alpha}$ . Fix  $E \in (-2, 2)$ . Then the process  $(\hat{Y}(b))_{b \in \mathbb{H}}$  converges in the sense of finite-dimensional distributions to  $(Y(b))_{b \in \mathbb{H}}$  as  $N \rightarrow \infty$ . That is, for any fixed  $p$  and  $b_1, b_2, \dots, b_p \in \mathbb{H}$ , we have*

$$(2.5) \quad (\hat{Y}(b_1), \dots, \hat{Y}(b_p)) \xrightarrow{d} (Y(b_1), \dots, Y(b_p))$$

as  $N \rightarrow \infty$ .

Our second result is on the convergence of the trace of general functions of  $H$ . For fixed  $r, s > 0$ , denote by  $\mathcal{C}^{1,r,s}(\mathbb{R})$  the space of all real-valued  $\mathcal{C}^1$ -functions  $f$  such that  $f'$  is  $r$ -Hölder continuous uniformly in  $x$ , and  $|f(x)| + |f'(x)| = O((1 + |x|)^{-1-s})$ . Let  $(Z(f))_{f \in \mathcal{C}^{1,r,s}(\mathbb{R})}$  denote the real-valued Gaussian process with mean zero and covariance

$$(2.6) \quad \mathbb{E}(Z(f_1)Z(f_2)) = \frac{1}{2\pi^2} \int \frac{(f_1(x) - f_1(y))(f_2(x) - f_2(y))}{(x - y)^2} dx dy$$

for all  $f_1, f_2 \in \mathcal{C}^{1,r,s}(\mathbb{R})$  [see also (1.2)]. Our next result is the weak convergence of the process

$$(2.7) \quad \hat{Z}(f) := \text{Tr} f\left(\frac{H - E}{\eta}\right) - N \int_{-2}^2 \varrho(x) f\left(\frac{x - E}{\eta}\right) dx,$$

where  $f \in \mathcal{C}^{1,r,s}(\mathbb{R})$ . We may now state our second result.

**THEOREM 2.3 (Convergence of general test functions).** *Let  $H$  be a real symmetric Wigner matrix. Fix  $\alpha \in (0, 1)$  and set  $\eta := N^{-\alpha}$ . Fix  $E \in (-2, 2)$ . Then the process  $(\hat{Z}(f))_{f \in \mathcal{C}^{1,r,s}(\mathbb{R})}$  converges in the sense of finite-dimensional distributions to  $(Z(f))_{f \in \mathcal{C}^{1,r,s}(\mathbb{R})}$  as  $N \rightarrow \infty$ . That is, for any fixed  $p$  and  $f_1, f_2, \dots, f_p \in \mathcal{C}^{1,r,s}(\mathbb{R})$ , we have*

$$(2.8) \quad (\hat{Z}(f_1), \dots, \hat{Z}(f_p)) \xrightarrow{d} (Z(f_1), \dots, Z(f_p))$$

as  $N \rightarrow \infty$ .

REMARK 2.4. In the complex Hermitian case, Theorems 2.2 and 2.3 remain true up to an additional factor 1/2 in the covariances. More precisely, if  $H$  is a complex Wigner matrix then (2.5) is replaced by

$$(2.9) \quad (\hat{Y}(b_1), \dots, \hat{Y}(b_p)) \xrightarrow{d} \frac{1}{\sqrt{2}}(Y(b_1), \dots, Y(b_p))$$

and (2.8) by

$$(2.10) \quad (\hat{Z}(f_1), \dots, \hat{Z}(f_p)) \xrightarrow{d} \frac{1}{\sqrt{2}}(Z(f_1), \dots, Z(f_p)).$$

The minor modifications to the proof in the complex Hermitian case are given in Section 7 below.

**3. Tools.** The rest of this paper is devoted to the proofs of Theorems 2.2 and 2.3. In this section, we collect notation and tools that are used throughout the paper.

Let  $M$  be an  $N \times N$  matrix. We use the notation  $M_{ij}^n \equiv (M_{ij})^n$ ,  $M^{*n} \equiv (M^*)^n$ ,  $M_{ij}^{*n} \equiv (M^*)_{ij} = \overline{M_{ji}}$ . We denote by  $\|M\|$  the operator norm of  $M$ , and abbreviate  $\underline{M} := \frac{1}{N} \text{Tr } M$ . It is easy to see that  $\|G(E + i\eta)\| \leq |\eta|^{-1}$ . For  $\sigma > 0$ , we use  $\mathcal{N}(0, \sigma^2)$  to denote the real Gaussian random variable with mean zero and variance  $\sigma^2$ , and  $\mathcal{N}_{\mathbb{C}}(0, \sigma^2) \stackrel{d}{=} \sigma \mathcal{N}_{\mathbb{C}}(0, 1)$  the complex Gaussian random variable with mean zero and variance  $\sigma^2$ . We abbreviate  $\langle X \rangle := X - \mathbb{E}X$  for any random variable  $X$  with finite expectation. Finally, if  $h$  is a real-valued random variable with finite moments of all order, we denote by  $\mathcal{C}_k(h)$  the  $k$ th cumulant of  $h$ , that is,

$$(3.1) \quad \mathcal{C}_k(h) := (-i)^k \cdot (\partial_\lambda^k \log \mathbb{E}e^{i\lambda h})|_{\lambda=0}.$$

We now state the cumulant expansion formula that is a central ingredient of the proof. The formula is analogous to the corresponding formula in [5], and its proof is obtained as a minor modification whose details we omit.

LEMMA 3.1 (Cumulant expansion). *Let  $h$  be a real-valued random variable with finite moments of all order, and  $f$  a complex-valued smooth function on  $\mathbb{R}$ . Then for any fixed  $l \in \mathbb{N}$  we have*

$$(3.2) \quad \mathbb{E}f(h)h = \sum_{k=0}^l \frac{1}{k!} \mathcal{C}_{k+1}(h) \mathbb{E}f^{(k)}(h) + R_{l+1},$$

provided all expectations in (3.2) exist. For any fixed  $\tau > 0$ , the remainder term  $R_{l+1}$  satisfies

$$(3.3) \quad R_{l+1} = O(1) \cdot \mathbb{E}|h|^{l+2} \cdot \mathbf{1}_{\{|h| > N^{\tau-1/2}\}} \cdot \|f^{(l+1)}\|_\infty + O(1) \cdot \mathbb{E}|h|^{l+2} \cdot \sup_{|x| \leq N^{\tau-1/2}} |f^{(l+1)}(x)|.$$

The bulk of the proof is performed on Wigner matrices satisfying a stronger condition than Definition 2.1(iii) by having entries with finite moments of all order.

DEFINITION 3.2. We consider the subset of Wigner matrices obtained from Definition 2.1 by replacing (iii) with:

(iii)' For each  $p \in \mathbb{N}$  there exists a constant  $C_p$  such that  $\mathbb{E}|\sqrt{N}H_{ij}|^p \leq C_p$  for all  $N, i, j$ .

We focus on Wigner matrices satisfying Definition 3.2 until Section 6, where we explain how to relax the condition (iii)' to (iii) using a Green function comparison argument; see Section 6 for more details.

We shall deduce Theorem 2.3 from Theorem 2.2 using the Helffer–Sjöstrand formula [7], which is summarized in the following result whose standard proof we omit.

LEMMA 3.3 (Helffer–Sjöstrand formula). *Let  $f \in C^{1,r,s}(\mathbb{R})$  with some  $r, s > 0$ . Let  $\tilde{f}$  be the almost analytic extension of  $f$  defined by*

$$(3.4) \quad \tilde{f}(x + iy) := f(x) + i(f(x + y) - f(x)).$$

*If  $f$  is further in  $C^2(\mathbb{R})$ , we can also set*

$$(3.5) \quad \tilde{f}(x + iy) := f(x) + iyf'(x).$$

*Let  $\chi \in C_c^\infty(\mathbb{R})$  be a cutoff function satisfying  $\chi(0) = 1$ , and by a slight abuse of notation write  $\chi(z) \equiv \chi(\text{Im } z)$ . Then for any  $\lambda \in \mathbb{R}$  we have*

$$(3.6) \quad f(\lambda) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial_{\bar{z}}(\tilde{f}(z)\chi(z))}{\lambda - z} d^2z,$$

*where  $\partial_{\bar{z}} := \frac{1}{2}(\partial_x + i\partial_y)$  is the antiholomorphic derivative and  $d^2z$  the Lebesgue measure on  $\mathbb{C}$ .*

The following definition introduces a notion of a high-probability bound that is suited for our purposes. It was introduced (in a more general form) in [13].

DEFINITION 3.4 (Stochastic domination). Let

$$X = (X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}), \quad Y = (Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)})$$

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

$$(3.7) \quad \sup_{u \in U^{(N)}} \mathbb{P}[X^{(N)}(u) > N^\varepsilon Y^{(N)}(u)] \leq N^{-D}$$

for large enough  $N \geq N_0(\varepsilon, D)$ . If  $X$  is stochastically dominated by  $Y$ , we use the notation  $X \prec Y$ . The stochastic domination will always be uniform in all parameters, such as  $z$  and matrix indices, that are not explicitly constant.

We conclude this section with the local semicircle law for Wigner matrices from [13, 16]. For a recent survey of the local semicircle law, see [3], where the following version of the local semicircle law is stated.

**THEOREM 3.5 (Local semicircle law).** *Let  $H$  be a Wigner matrix satisfying Definition 3.2, and define the spectral domain*

$$\mathbf{S} := \{E + i\eta : |E| \leq 10, 0 < \eta \leq 10\}.$$

*Then we have the bounds*

$$(3.8) \quad \max_{i,j} |G_{ij}(z) - \delta_{ij}m(z)| < \sqrt{\frac{\operatorname{Im} m(z)}{N\eta}} + \frac{1}{N\eta}$$

*and*

$$(3.9) \quad |\underline{G}(z) - m(z)| < \frac{1}{N\eta},$$

*uniformly in  $z = E + i\eta \in \mathbf{S}$ . Moreover, outside the spectral domain we have the stronger estimates*

$$(3.10) \quad \max_{i,j} |G_{ij}(z) - \delta_{ij}m(z)| < \frac{1}{\sqrt{N}}$$

*and*

$$(3.11) \quad |\underline{G}(z) - m(z)| < \frac{1}{N},$$

*uniformly in  $z \in \mathbb{H} \setminus \mathbf{S}$ .*

**4. Convergence of the resolvent.** In this section, we prove the following weaker form of Theorem 2.2.

**THEOREM 4.1.** *Theorem 2.2 holds for Wigner matrices  $H$  satisfying Definition 3.2, and the convergence also holds in the sense of moments.*

For the statements of the following results, we abbreviate  $G \equiv G(E + i\eta)$  and  $[\underline{G}] := \underline{G} - m(E + i\eta)$ . The following result is a special case of Theorem 4.1.

**PROPOSITION 4.2.** *Under the assumptions of Theorem 4.1, we have*

$$(4.1) \quad N\eta[\underline{G}] \xrightarrow{d} \mathcal{N}_{\mathbb{C}}\left(0, \frac{1}{2}\right)$$

*as  $N \rightarrow \infty$ . The convergence also holds in the sense of moments.*

The main work in this section is to show the one-dimensional case from Proposition 4.2, whose proof can easily be extended to the general case of Theorem 4.1 (see Section 4.4 below). Recall the notation  $\langle X \rangle := X - \mathbb{E}X$ . Proposition 4.2 is a direct consequence of the following lemma.

LEMMA 4.3. *Under the assumptions of Theorem 4.1, the following hold:*

(i) *For fixed  $m, n \in \mathbb{N}$  satisfying  $m + n \geq 2$  we have*

$$(4.2) \quad \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \begin{cases} \frac{n!}{2^n} N^{2n(\alpha-1)} + O(N^{2n(\alpha-1)-c_0}), & \text{if } m = n, \\ O(N^{(m+n)(\alpha-1)-c_0}), & \text{if } m \neq n, \end{cases}$$

where

$$(4.3) \quad c_0 \equiv c_0(\alpha) := \frac{1}{3} \min\{\alpha, 1 - \alpha\}.$$

(ii) *We have*

$$(4.4) \quad [\underline{G}] - \langle \underline{G} \rangle = \mathbb{E} \underline{G} - m(E + i\eta) = O(N^{\alpha-1-c_0}),$$

with  $c_0$  defined in (4.3).

As advertised, Proposition 4.2 follows immediately from Lemma 4.3. Indeed, suppose that Lemma 4.3 holds. From (4.2), we find that  $N^{1-\alpha} \langle \underline{G} \rangle$  converges to  $\mathcal{N}_{\mathbb{C}}(0, 1/2)$  in the sense of moments, and hence also in distribution. Proposition 4.2 therefore follows from (4.4).

The bulk of this section, Sections 4.1–4.3, is devoted to the proof of Lemma 4.3.

4.1. *Preliminary estimates on  $G$ .* We begin with estimates on the entries of  $G^k$ . For  $k \geq 2$ , the bounds provided by the following lemma are significantly better than those obtained for  $G^k$  by applying the estimate (3.8) to each entry of the matrix product. For instance, a straightforward application of (3.8) yields  $|(G^k)_{ij}| \prec N^{k(1+\alpha)/2-1}$ , which is not enough to conclude the proof of Lemma 4.3. The following result yields bounds that grow slower with  $k$  and in addition provide extra smallness for the off-diagonal entries of  $G^k$ . Both of these features are necessary for the proof of Lemma 4.3.

LEMMA 4.4. *Under the conditions of Theorem 4.1, for any fixed  $k \in \mathbb{N}_+$  we have*

$$(4.5) \quad |\langle \underline{G}^k \rangle| \prec N^{k\alpha-1}$$

as well as

$$(4.6) \quad |(G^k)_{ij}| \prec \begin{cases} N^{(k-1)\alpha}, & \text{if } i = j, \\ N^{(k-1/2)\alpha-1/2}, & \text{if } i \neq j, \end{cases}$$

uniformly in  $i, j$ .

PROOF. We first prove (4.5). The case  $k = 1$  is easy. Indeed, from (3.9) we get

$$(4.7) \quad |\langle \underline{G} \rangle| \leq |\underline{G}| + |\mathbb{E}[\underline{G}]| \prec N^{\alpha-1}$$

as desired, where we used the definition of  $\prec$  combined with the trivial bound  $\|G\| \leq N^\alpha$  to estimate  $\mathbb{E}[\underline{G}]$ .

Next, for  $k \geq 2$  we write

$$(4.8) \quad G^k = \left( \frac{(H - E)/\eta + i}{(H - E)^2/\eta^2 + 1} \right)^k \cdot \eta^{-k} = f\left(\frac{H - E}{\eta}\right) \cdot \eta^{-k},$$

where we defined  $f(x) := \left(\frac{x+i}{x^2+1}\right)^k$ . Note that  $f : \mathbb{R} \rightarrow \mathbb{C}$  is smooth, and for any  $n \in \mathbb{N}$ ,  $|f^{(n)}(x)| = O((1 + |x|)^{-2})$ . We define  $\tilde{f}$  as in (3.5) and let  $\chi$  be as in Lemma 3.3 and satisfy  $\chi(y) = 1$  for  $|y| \leq 1$ . Writing  $f_\eta(x) := f\left(\frac{x-E}{\eta}\right)$ , we obtain from Lemma 3.3 that

$$f_\eta(H) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial_{\bar{z}}(\tilde{f}_\eta(z)\chi(z/\eta))}{H - z} d^2z,$$

so that

$$(4.9) \quad \begin{aligned} \langle \underline{G}^k \rangle &= \frac{1}{2\pi\eta^k} \int_{\mathbb{R}^2} \left( iyf''_\eta(x)\chi(y/\eta) + \frac{i}{\eta} f_\eta(x)\chi'(y/\eta) - \frac{y}{\eta} f'_\eta(x)\chi'(y/\eta) \right) \\ &\quad \times \langle G(x + iy) \rangle dx dy. \end{aligned}$$

In order to estimate the right-hand side, we use (3.9) and (3.11) to obtain

$$(4.10) \quad |\langle G(x + iy) \rangle| \prec \frac{1}{N|y|}$$

uniformly for  $|y| \in (0, 1)$  and  $x$ . Hence,

$$(4.11) \quad \begin{aligned} &\frac{1}{2\pi\eta^k} \int_{\mathbb{R}^2} |iyf''_\eta(x)\chi(y/\eta)\langle G(x + iy) \rangle| dx dy \\ &\prec \frac{1}{2\pi\eta^k} \int_{\mathbb{R}^2} \left| \frac{1}{N} f''_\eta(x)\chi(y/\eta) \right| dx dy = \frac{O(1)}{N\eta^k}. \end{aligned}$$

(Note that the use of stochastic domination inside the integral requires some justification. In fact, we use that a high-probability bound of the form (4.10) holds *simultaneously* for all  $x \in \mathbb{R}$  and  $|y| \in (0, 1)$ . We refer to [3], Remark 2.7 and Lemma 10.2, for further details. Similarly, by our choice of  $\chi$ , we find

$$\begin{aligned} &\frac{1}{2\pi\eta^k} \int_{\mathbb{R}^2} \left| \frac{i}{\eta} f_\eta(x)\chi'(y/\eta)\langle G(x + iy) \rangle \right| dx dy \\ &\prec \frac{1}{2\pi\eta^k} \int_{|y| \geq \eta} \left| \frac{1}{N\eta^2} f_\eta(x)\chi'(y/\eta) \right| dx dy = \frac{O(1)}{N\eta^k}. \end{aligned}$$

An analogous estimate yields

$$\frac{1}{2\pi\eta^k} \int_{\mathbb{R}^2} \left| \frac{y}{\eta} f'_\eta(x) \chi'(y/\eta) \langle \underline{G}(x + iy) \rangle \right| dx dy < \frac{1}{N\eta^k}.$$

Altogether we have  $|\langle \underline{G}^k \rangle| < N^{k\alpha-1}$ , which is (4.5).

Next, we prove (4.6). For  $k = 1$ , we use the well-known bound  $|m(z)| < 1$ , which follows using an elementary estimate from the fact that  $m$  is the unique solution of

$$(4.12) \quad m(z) + \frac{1}{m(z)} + z = 0$$

satisfying  $\text{Im } m(z) \text{Im } z > 0$ . Thus, by (3.8) we have

$$(4.13) \quad |(G)_{ij}| < \begin{cases} 1, & \text{if } i = j, \\ N^{(\alpha-1)/2}, & \text{if } i \neq j, \end{cases}$$

which is (4.6) for  $k = 1$ . The extension to  $k \geq 2$  follows again using Lemma 3.1, and we omit the details.  $\square$

Lemma 4.4 is very useful in estimating the expectations involving entries of  $G$ , in combination with the following elementary result about stochastic domination.

LEMMA 4.5. (i) *If  $X_1 < Y_1$  and  $X_2 < Y_2$  then  $X_1 X_2 < Y_1 Y_2$ .*

(ii) *Suppose that  $X$  is a nonnegative random variable satisfying  $X \leq N^C$  and  $X < \Phi$  for some deterministic  $\Phi \geq N^{-C}$ . Then  $\mathbb{E}X < \Phi$ .*

4.2. *Proof of Lemma 4.3(i).* Abbreviating  $\zeta_i := \mathbb{E}|\sqrt{N}H_{ii}|^2$ , we find from Definition 3.2(iii)' that  $\zeta_i = O(1)$ . We write  $z := E + i\eta$  and often omit the argument  $z$  from our notation. Note that  $G = (H - z)^{-1}$  and  $G^* = (H - \bar{z})^{-1}$ . In particular, Theorem 3.5 also holds for  $G^*$  with obvious modifications accounting for the different sign of  $\eta$ . For  $m, n \geq 1$ , we need to compute

$$(4.14) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \mathbb{E}\langle \underline{G} \rangle^{m-1} \langle \underline{G}^* \rangle^n \underline{G}.$$

By the resolvent identity, we have

$$G = \frac{1}{z}GH - \frac{1}{z}I,$$

so that

$$(4.15) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \frac{1}{z} \mathbb{E}\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle \underline{G}H = \frac{1}{zN} \sum_{i,j} \mathbb{E}\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij} H_{ji}.$$

Since  $H$  is symmetric, for any differentiable  $f = f(H)$  we set

$$(4.16) \quad \frac{\partial}{\partial H_{ij}} f(H) = \frac{\partial}{\partial H_{ji}} f(H) := \frac{d}{dt} \Big|_{t=0} f(H + t\Delta^{(ij)}),$$

where  $\Delta^{(ij)}$  denotes the matrix whose entries are zero everywhere except at the sites  $(i, j)$  and  $(j, i)$  where they are one:  $\Delta_{kl}^{(ij)} = (\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})(1 + \delta_{ij})^{-1}$ . We then compute the last averaging in (4.15) using the formula (3.2) with  $f = f_{ij}(H) := \langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij}$ ,  $h = H_{ji}$ , and obtain

$$\begin{aligned}
 z\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m &= \frac{1}{N^2} \sum_{i,j} \mathbb{E} \frac{\partial(\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij})}{\partial H_{ji}} (1 + \delta_{ji}(\zeta_i - 1)) + L \\
 &= \frac{1}{N^2} \sum_{i,j} \mathbb{E} \langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle \frac{\partial G_{ij}}{\partial H_{ji}} (1 + \delta_{ji}) \\
 &\quad + \frac{1}{N^2} \sum_{i,j} \mathbb{E} \frac{\partial(\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle)}{\partial H_{ji}} G_{ij} (1 + \delta_{ji}) + K + L \\
 &=: (a) + (b) + K + L,
 \end{aligned}
 \tag{4.17}$$

where

$$K = N^{-2} \sum_i \mathbb{E} \frac{\partial(\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ii})}{\partial H_{ii}} (\zeta_i - 2)
 \tag{4.18}$$

and

$$L = N^{-1} \cdot \sum_{i,j} \left[ \sum_{k=2}^l \frac{1}{k!} C_{k+1}(H_{ji}) \mathbb{E} \frac{\partial^k(\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij})}{\partial H_{ji}^k} + R_{l+1}^{(ji)} \right].
 \tag{4.19}$$

Here,  $l$  is a fixed positive integer to be chosen later, and  $R_{l+1}^{(ji)}$  is a remainder term defined analogously to  $R_{l+1}$  in (3.2). More precisely, we have the bound

$$\begin{aligned}
 R_{l+1}^{(ji)} &= O(1) \cdot \mathbb{E} |H_{ji}|^{l+2} \mathbf{1}_{\{|H_{ji}| > N^{\tau-1/2}\}} \cdot \|\partial_{ji}^{l+1} f_{ij}(H)\|_\infty \\
 &\quad + O(1) \cdot \mathbb{E} |H_{ji}|^{l+2} \cdot \mathbb{E} \sup_{|x| \leq N^{\tau-1/2}} |\partial_{ji}^{l+1} f_{ij}(H^{(ij)} + x\Delta^{(ij)})|,
 \end{aligned}
 \tag{4.20}$$

where we define  $H^{(ij)} := H - H_{ij}\Delta^{(ij)}$ , so that the matrix  $H^{(ij)}$  has zero entries at the positions  $(i, j)$  and  $(j, i)$ , and abbreviate  $\partial_{ij} := \frac{\partial}{\partial H_{ij}}$ . Note that for  $G = (H - z)^{-1}$  we have

$$\frac{\partial G_{ij}}{\partial H_{kl}} = -(G_{ik}G_{lj} + G_{il}G_{kj})(1 + \delta_{kl})^{-1},
 \tag{4.21}$$

which gives

$$\begin{aligned}
 (a) &= N^{-2} \sum_{i,j} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} (-G_{ij} G_{ij} - G_{ii} G_{jj}) \\
 &= -N^{-1} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle - \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle \\
 &= -N^{-1} \mathbb{E} \langle \underline{G}^* \rangle^m \langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle - \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m \langle \underline{G} \rangle \\
 &\quad - 2\mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m \mathbb{E} \underline{G} + \mathbb{E} \langle \underline{G}^* \rangle^m \langle \underline{G} \rangle^{m-1} \mathbb{E} \langle \underline{G} \rangle^2.
 \end{aligned}$$

Similarly, a straightforward calculation gives

$$(b) = -\frac{2}{N^2} [n\mathbb{E} \langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} \underline{G} \underline{G}^{*2} + (m-1)\mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-2} \underline{G}^3].$$

Altogether we obtain

$$\begin{aligned}
 \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m &= \frac{1}{T} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m+1} - \frac{1}{T} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \mathbb{E} \langle \underline{G} \rangle^2 \\
 (4.22) \quad &+ \frac{1}{TN} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle + \frac{2m-2}{N^2 T} \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-2} \underline{G}^3 \\
 &- \frac{K}{T} - \frac{L}{T} + \frac{2n}{N^2 T} \mathbb{E} \langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} \underline{G} \underline{G}^{*2},
 \end{aligned}$$

where  $T := -z - 2\mathbb{E} \underline{G}$ . From (3.9), (4.12) and Lemma 4.5 it is easy to see that

$$(4.23) \quad \left| \frac{1}{T} \right| = O(1),$$

and the implicit constant depends only on the distance to the spectral edge

$$(4.24) \quad \kappa := 2 - |E|.$$

In (4.22), the last term is the leading term. The calculation of (4.22) consists of computing the leading term and estimating the subleading terms. We aim to show that the subleading terms are of order  $N^{(m+n)(\alpha-1)-c_0}$ .

We begin with  $L$ . For  $k \geq 2$ , define

$$(4.25) \quad J_k := N^{-(k+3)/2} \sum_{i,j} \left| \mathbb{E} \frac{\partial^k (\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} G_{ij})}{\partial H_{ji}^k} \right|.$$

LEMMA 4.6. *Let  $R_{l+1}^{(ji)}$  be as in (4.19).*

(i) *For any fixed  $D_0 > 0$  there exists some  $l_0 = l_0(D_0) \geq 2$  such that*

$$(4.26) \quad \sum_{i,j} R_{l_0+1}^{(ji)} = O(N^{-D_0}).$$

(ii) For all fixed  $k \geq 2$  we have

$$(4.27) \quad J_k = O(N^{(m+n)(\alpha-1)-c_0}),$$

where  $c_0$  is defined in (4.3).

Before proving Lemma 4.6, we show how to use it to estimate  $L$ . By setting  $D_0 = (m+n)(1-\alpha)$  in (4.26), we obtain

$$(4.28) \quad \sum_{i,j} R_{l_0+1}^{(ji)} = O(N^{(m+n)(\alpha-1)})$$

for some  $l_0 \geq 2$ . From Definition 3.2 (iii)' we get

$$\max_{i,j} |C_k(H_{ji})| = O(N^{-k/2})$$

for all  $k \geq 2$ . Thus, (4.27) and (4.28) together imply

$$(4.29) \quad L = O(N^{(m+n)(\alpha-1)-c_0}),$$

as desired.

PROOF OF LEMMA 4.6(i). Let  $D_0 > 0$  be given. Fix  $i, j$ , and choose  $\tau = \min\{\alpha/2, (1-\alpha)/4\}$  in (4.20). Define  $W := H_{ij}\Delta^{(ij)}$  and  $\hat{H} := H^{(ij)} = H - W$ . Let  $\hat{G} := (\hat{H} - E - i\eta)^{-1}$ . We have the resolvent expansions

$$(4.30) \quad \hat{G} = G + (GW)G + (GW)^2\hat{G}$$

and

$$(4.31) \quad G = \hat{G} - (\hat{G}W)\hat{G} + (\hat{G}W)^2G.$$

Note that only two entries of  $W$  are nonzero, and they are stochastically dominated by  $N^{-1/2}$ . Then the trivial bound  $\max_{a,b} |\hat{G}_{ab}| \leq N^\alpha$  together with (4.13) and (4.30) show that  $\max_{a \neq b} |\hat{G}_{ab}| \prec N^{-(\alpha-1)/2}$ , and  $\max_a |\hat{G}_{aa}| \prec 1$ . Combining with (4.31), the trivial bound  $\max_{a,b} |G_{ab}| \leq N^\alpha$ , and the fact  $\hat{G}$  is independent of  $W$ , we have

$$(4.32) \quad \max_{a \neq b} \sup_{|H_{ji}| \leq N^{\tau-1/2}} |G_{ab}| \prec N^{-(\alpha-1)/2}$$

and

$$(4.33) \quad \max_a \sup_{|H_{ji}| \leq N^{\tau-1/2}} |G_{aa}| \prec 1.$$

Now let us estimate the last term in (4.20). We have the derivatives

$$\partial_{ji} G_{ab} = -(G_{aj}G_{ib} + G_{ai}G_{jb})(1 + \delta_{ji})^{-1}, \quad \partial_{ji} \langle \underline{G} \rangle = \frac{2}{N} (G^2)_{ji} (1 + \delta_{ji})^{-1}$$

and

$$\partial_{ji}(G^2)_{ab} = ((G^2)_{aj}G_{ib} + (G^2)_{ai}G_{jb} + (G^2)_{bj}G_{ia} + (G^2)_{bi}G_{ja})(1 + \delta_{ji})^{-1},$$

where  $\partial_{ij} := \frac{\partial}{\partial H_{ij}}$ . Hence, for any fixed  $l \in \mathbb{N}$ ,  $\partial_{ji}^{l+1} f_{ij}$  is a polynomial in the variables  $\langle \underline{G} \rangle$ ,  $\langle \underline{G}^* \rangle$ ,  $\frac{1}{N}(G^2)_{ab}$ ,  $\frac{1}{N}(G^{*2})_{ab}$ ,  $G_{ab}$  and  $G_{ab}^*$ , with  $a, b \in \{i, j\}$ . Note that in each term of the polynomial, the sum of the degrees of  $\langle \underline{G} \rangle$ ,  $\langle \underline{G}^* \rangle$ ,  $\frac{1}{N}(G^2)_{ab}$  and  $\frac{1}{N}(G^{*2})_{ab}$  is  $m + n - 1$ , so that the product of the factors other than  $G_{ab}$  and  $G_{ab}^*$  is trivially bounded by  $O(N^{(m+n-1)\alpha})$  for all  $H$ . Together with (4.32) and (4.33) we know, for any fixed  $l \in \mathbb{N}$ ,

$$\sup_{|x| \leq N^{\tau-1/2}} |\partial_{ji}^{l+1} f_{ij}(\hat{H} + x\Delta^{(ij)})| \prec N^{(m+n-1)\alpha}.$$

Note that  $\mathbb{E}|H_{ji}|^{l+2} = O(N^{-(l+2)/2})$ , and we can find  $l_0 = l_0(D_0, m, n) \geq 2$  such that

$$(4.34) \quad \mathbb{E}|H_{ji}|^{l_0+2} \cdot \mathbb{E} \sup_{x \leq N^{\tau-1/2}} |\partial_{ji}^{l_0+1} f_{ij}(\hat{H} + x\Delta^{(ij)})| = O(N^{-(D_0+2)}).$$

Finally, we estimate the first term of (4.20). Note that by the trivial bound  $|G_{ij}| \leq N^\alpha$ , we have  $\|\partial_{ji}^{l_0+1} f(H)\|_\infty = O(N^{C(m,n,l_0)})$ . From Definition 3.2(iii)' we find  $\max_{i,j} |H_{ij}| \prec \frac{1}{\sqrt{N}}$ , then by Hölder's inequality we have

$$(4.35) \quad \mathbb{E}|H_{ji}|^{l_0+2} \mathbf{1}_{\{|H_{ji}| > N^{\tau-1/2}\}} \cdot \|\partial_{ji}^{l_0+1} f_{ij}(H)\|_\infty = O(N^{-(D_0+2)}).$$

Combining (4.34) and (4.35), we obtain from (4.20) that  $R_{l_0+1}^{(ji)} = O(N^{-(D_0+2)})$ , from which (4.26) follows.  $\square$

PROOF OF LEMMA 4.6(ii). We begin with the case  $k = 2$ , which gives rise to terms of three types depending on how many derivatives act on  $G_{ij}$ . We deal with each type separately.

Step 1. The first type is

$$J_{2,1} := N^{-5/2} \sum_{i,j} \left| \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \frac{\partial^2 G_{ij}}{\partial H_{ji}^2} \right|.$$

Note that

$$\frac{\partial^2 G_{ij}}{\partial H_{ji}^2} = a_1 \cdot G_{ii} G_{jj} G_{ij} + a_2 \cdot G_{ij}^3,$$

where  $a_1, a_2$  are some constants depending on the value of  $\delta_{ji}$ . Together with (4.6) and Lemma 4.5, we find

$$\begin{aligned} J_{2,1} &\leq N^{-5/2} \cdot N^2 \cdot \mathbb{E} |\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1}| \cdot O(N^{(\alpha-1)/2+\varepsilon}) \\ &= O(N^{\alpha/2+\varepsilon-1}) \cdot \mathbb{E} |\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1}| \end{aligned}$$

for any fixed  $\varepsilon > 0$ . Together with (4.5) and Lemma 4.5, we find

$$(4.36) \quad J_{2,1} = O(N^{\alpha/2+\varepsilon-1+(m+n-1)(\alpha-1)}) = O(N^{(m+n)(\alpha-1)-c_0}),$$

where in the last inequality we chose  $\varepsilon$  small enough depending on  $\alpha$ .

Step 2. The second type is

$$(4.37) \quad J_{2,2} := N^{-\frac{5}{2}} \sum_{i,j} \left| \mathbb{E} \frac{\partial^2 (\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1})}{\partial H_{ji}^2} G_{ij} \right|.$$

Since

$$\begin{aligned} \frac{\partial \langle \underline{G}^* \rangle}{\partial H_{ji}} &= \frac{a_3}{N} \cdot (G^{*2})_{ij} \quad \text{and} \\ \frac{\partial^2 \langle \underline{G}^* \rangle}{\partial H_{ji}^2} &= \frac{a_4}{N} \cdot (G^{*2})_{ii} G_{jj}^* + \frac{a_5}{N} \cdot (G^{*2})_{jj} G_{ii}^* + \frac{a_6}{N} \cdot (G^{*2})_{ij} G_{ij}^* \end{aligned}$$

for some constants  $a_3, a_4$  and  $a_5$ , we see that the most dangerous term of  $J_{2,2}$  is of the form

$$(4.38) \quad P_{2,2} := N^{-7/2} \cdot \sum_{i,j} \left| \mathbb{E} \langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} (G^{*2})_{ii} G_{jj}^* G_{ij} \right|.$$

By (4.5), (4.6) and Lemma 4.5, we have

$$P_{2,2} = O(N^{-7/2+2+(m+n-2)(\alpha-1)+\alpha+(\alpha-1)/2+\varepsilon}) = O(N^{(m+n)(\alpha-1)-\alpha/2+\varepsilon})$$

for any fixed  $\varepsilon > 0$ . The other terms of  $J_{2,2}$  are estimated similarly. By choosing  $\varepsilon$  small enough, we obtain

$$(4.39) \quad J_{2,2} = O(N^{(m+n)(\alpha-1)-c_0}).$$

Step 3. The third type is

$$(4.40) \quad J_{2,3} := N^{-5/2} \sum_{i,j} \left| \mathbb{E} \frac{\partial (\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1})}{\partial H_{ji}} \frac{\partial G_{ij}}{\partial H_{ji}} \right|.$$

The most dangerous term in  $J_{2,3}$  is of the form

$$(4.41) \quad P_{2,3} := N^{-7/2} \cdot \sum_{i,j} \left| \mathbb{E} \langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} (G^{*2})_{ij} G_{ii} G_{jj} \right|.$$

Again by (4.5), (4.6) and Lemma 4.5, we have

$$P_{2,3} = O(N^{-7/2+2+(m+n-2)(\alpha-1)+3\alpha/2-1/2+\varepsilon}) = O(N^{(m+n)(\alpha-1)-c_0})$$

for any fixed  $\varepsilon > 0$ . The other terms of  $J_{2,3}$  are estimated similarly. Thus, we get

$$J_{2,3} = O(N^{(m+n)(\alpha-1)-c_0}).$$

Step 4. Putting the estimates of the three types in Steps 1–3 together, we find

$$J_2 = O(N^{(m+n)(\alpha-1)-c_0}),$$

which concludes the proof of (4.27) for  $k = 2$ .

For  $k \geq 3$ , the estimates are easier than those in  $k = 2$  because of the small prefactor  $N^{-(k+3)/2}$  in the definition of  $J_k$ . Analogously to the case  $k = 2$ , we obtain for any fixed  $k \geq 3$  and  $\varepsilon > 0$ ,

$$J_k = O(N^{(m+n)(\alpha-1)+1-k/2+\varepsilon}),$$

for any  $\varepsilon > 0$ , from which (4.27) follows. We omit further details.  $\square$

Now we look at the term  $K$  defined in (4.18), whose estimate is contained in the next lemma.

LEMMA 4.7. *We have*

$$(4.42) \quad K = O(N^{(m+n)(\alpha-1)-\alpha/2}).$$

PROOF. Let us first consider

$$(4.43) \quad B := N^{-2} \sum_i \mathbb{E} \left| \frac{\partial(\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} G_{ii})}{\partial H_{ii}} \right|.$$

The estimate of  $B$  is similar to that of  $J_2$ , namely we will have terms of two types depending on whether the derivative acts on  $G_{ii}$  or not. We then estimate the terms by Lemmas 4.4 and 4.5, which easily yields

$$B = O(N^{(m+n)(\alpha-1)-\alpha+\varepsilon}) = O(N^{(m+n)(\alpha-1)-\alpha/2}).$$

From Definition 3.2(iii)' we get  $\max_i |\zeta_i - 2| = O(1)$ , hence  $K = O(B)$ . This completes the proof.  $\square$

In order to conclude the proof, we need to use that the expectation of  $\underline{G}^k$  is typically much smaller than  $\underline{G}^k$  itself. Lemma 4.4 implies that  $\mathbb{E}|\underline{G}^k| \prec N^{(k-1)\alpha}$ , which is not enough to conclude the proof. We need some extra decay from the expectation, which is provided by the following result.

LEMMA 4.8. *Let  $c_0$  be defined as in (4.3). We have*

$$(4.44) \quad \mathbb{E}\underline{G}^k = O(N^{(k-1)\alpha-c_0})$$

for  $k = 2, 3$ .

PROOF. The proof is analogous to that of Lemma 4.6. Let us first consider  $\mathbb{E}\underline{G}^2$ . Again by the resolvent identity and the cumulant expansion, we arrive at

$$(4.45) \quad \mathbb{E}\underline{G}^2 = \frac{1}{T}\mathbb{E}\underline{G} + \frac{2}{T}\mathbb{E}\langle \underline{G} \rangle \langle \underline{G}^2 \rangle + \frac{1}{TN}\mathbb{E}\underline{G}^3 - \frac{K^{(2)}}{T} - \frac{L^{(2)}}{T},$$

where

$$(4.46) \quad K^{(2)} = N^{-2} \sum_i \mathbb{E} \frac{\partial(G^2)_{ii}}{\partial H_{ii}} (\zeta_i - 2)$$

and

$$(4.47) \quad L^{(2)} = \frac{1}{N} \sum_{i,j} \left[ \sum_{k=2}^l \frac{1}{k!} C_{k+1}(H_{ji}) \mathbb{E} \frac{\partial^k(G^2)_{ij}}{\partial H_{ji}^k} + R_{l+1}^{(2,ji)} \right],$$

and we recall the definition (4.23) of  $T$ . Here,  $R_{l+1}^{(2,ji)}$  is a remainder term defined analogously to  $R_{l+1}^{(ji)}$  in (4.20). We can argue similarly as in the proof of Lemma 4.6(i) and show that  $R_{l_0+1}^{(2,ji)} = O(N^{-1})$  for some  $l_0 \in \mathbb{N}$ . Thus, we have  $|L^{(2)}| \leq \sum_{k=2}^{l_0} O(J_k^{(2)}) + O(1)$ , where

$$(4.48) \quad J_k^{(2)} := N^{-(k+3)/2} \sum_{i,j} \left| \mathbb{E} \frac{\partial^k(G^2)_{ij}}{\partial H_{ji}^k} \right|.$$

Analogously to the proof of Lemma 4.6(ii), we find

$$J_2^{(2)} = O(N^{-5/2} \cdot N^2 \cdot N^{\alpha+\varepsilon}) = O(N^{\alpha+\varepsilon-1/2}),$$

for any fixed  $\varepsilon > 0$ , and  $J_k^{(2)} = O(N^{\alpha-1/2})$  for  $k \geq 3$ . This shows  $|L^{(2)}| = O(N^{\alpha+\varepsilon-1/2})$  for any fixed  $\varepsilon > 0$ . Similarly, as in Lemma 4.7, one can show  $K^{(2)} = O(N^{\alpha/2})$ . By Lemmas 4.4 and 4.5, we have

$$(4.49) \quad \begin{aligned} |\mathbb{E}\langle \underline{G} \rangle \langle \underline{G}^2 \rangle| &= O(N^{(\alpha-1)+(2\alpha-1)+\varepsilon}) = O(N^{3\alpha-2+\varepsilon}), \\ \frac{1}{N} |\mathbb{E}\underline{G}^3| &= O(N^{-1+2\alpha+\varepsilon}), \end{aligned}$$

for any fixed  $\varepsilon > 0$ . Hence, by using (4.23) and choosing  $\varepsilon$  small enough, we obtain

$$|\mathbb{E}\underline{G}^2| \leq O(N^{2\alpha-1+\varepsilon}) + O(N^{\alpha/2}) + O(N^{\alpha+\varepsilon-1/2}) = O(N^{\alpha-c_0}).$$

The proof of the case  $k = 3$  is similar, and we omit the details.  $\square$

Armed with Lemmas 4.6 and 4.8, we may now conclude the proof of Lemma 4.3(i). We still have to estimate the subleading terms on the right-hand side of (4.22). From (4.5), Lemma 4.5 and Lemma 4.8, we have

$$(4.50) \quad \begin{aligned} \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-2} \underline{G}^3 &= \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-2} (\mathbb{E}\underline{G}^3 + \langle \underline{G}^3 \rangle) \\ &= O(N^{(m+n)(\alpha-1)+2-c_0}). \end{aligned}$$

Moreover, (4.5) and Lemma 4.5 imply

$$(4.51) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle = O(N^{(m+n)(\alpha-1)+1-c_0}),$$

$$(4.52) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m+1} = O(N^{(m+n)(\alpha-1)-c_0}),$$

as well as

$$(4.53) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \mathbb{E}\langle \underline{G} \rangle^2 = O(N^{(m+n)(\alpha-1)-c_0}).$$

Applying (4.29), Lemma 4.8 and (4.50)–(4.53) to (4.22), together with (4.23), we obtain

$$\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \frac{2n}{N^2 T} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} \underline{G} \underline{G}^* 2 + O(N^{(m+n)(\alpha-1)-c_0}).$$

By the resolvent identity,

$$(4.54) \quad \underline{G} \underline{G}^* 2 = \frac{1}{z - \bar{z}} (\underline{G} \underline{G}^* - \underline{G}^* 2) = -\frac{N^{2\alpha}}{4} (\underline{G} - \underline{G}^*) - \frac{N^\alpha}{2i} \underline{G}^* 2.$$

Moreover, (4.5) and Lemmas 4.5 and 4.8 give  $|\mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} \underline{G}^* 2| = O(N^{(m+n-2)(\alpha-1)+\alpha-c_0})$ . We therefore conclude that

$$\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = -\frac{n}{2T} N^{2\alpha-2} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} (\underline{G} - \underline{G}^*) + O(N^{(m+n)(\alpha-1)-c_0}),$$

which yields

$$\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = -\frac{n}{2T} N^{2\alpha-2} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} (\mathbb{E}\underline{G} - \mathbb{E}\underline{G}^*) + O(N^{(m+n)(\alpha-1)-c_0})$$

by (4.5) and Lemma 4.5. Writing  $\xi := \text{Im } \underline{G}$ , we have  $\mathbb{E}\underline{G} - \mathbb{E}\underline{G}^* = 2i\mathbb{E}\xi$ . Moreover, (3.9) and (4.12) imply that  $T = -z - 2\mathbb{E}\underline{G} = -2i\mathbb{E}\xi + O(N^{-c_0})$ . Together with (4.5) and Lemma 4.5, we have

$$(4.55) \quad \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \frac{n}{2} N^{2\alpha-2} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} + O(N^{(m+n)(\alpha-1)-c_0})$$

for  $m, n \geq 1$ .

The preceding argument can also be used to show

$$(4.56) \quad \mathbb{E}\langle \underline{G} \rangle^m = O(N^{m(\alpha-1)-c_0})$$

for all  $m \geq 2$ . In fact, one can start with

$$\mathbb{E}\langle \underline{G} \rangle^m = \frac{1}{zN} \sum_{i,j} \mathbb{E}\langle \underline{G} \rangle^{m-1} G_{ij} H_{ji}$$

and apply Lemma 3.1 to get an analogue of (4.22), which is

$$(4.57) \quad \begin{aligned} \mathbb{E}\langle \underline{G} \rangle^m &= \frac{1}{T} \mathbb{E}\langle \underline{G} \rangle^{m+1} - \frac{1}{T} \mathbb{E}\langle \underline{G} \rangle^{m-1} \mathbb{E}\langle \underline{G} \rangle^2 + \frac{1}{TN} \mathbb{E}\langle \underline{G} \rangle^{m-1} \langle \underline{G}^2 \rangle \\ &+ \frac{2m-2}{N^2 T} \mathbb{E}\langle \underline{G} \rangle^{m-2} \underline{G}^3 - \frac{K^{(3)}}{T} - \frac{L^{(3)}}{T}. \end{aligned}$$

Here,  $T = -z - 2\mathbb{E}\underline{G}$ , and  $K^{(3)}, L^{(3)}$  are defined analogously as  $K$  and  $L$  in (4.22). Due to the absence of  $G^*$ , there is no leading term in (4.57) as the last term in (4.22). One can easily apply our previous techniques and show every term in the right-hand side of (4.57) is bounded by  $O(N^{m(\alpha-1)-c_0})$ .

By taking complex conjugation in (4.56), we also have

$$(4.58) \quad \mathbb{E}\langle \underline{G}^* \rangle^n = O(N^{n(\alpha-1)-c_0})$$

for all  $n \geq 2$ . Now (4.2) follows from (4.55), (4.56) and (4.58) combined with induction. This concludes the proof of Lemma 4.3(i).

4.3. *Proof of Lemma 4.3(ii).* Again by the resolvent identity and the cumulant expansion, we have

$$(4.59) \quad \mathbb{E}\underline{G} = \frac{1}{U} \left( 1 + \mathbb{E}\langle \underline{G} \rangle^2 + \frac{1}{N} \mathbb{E}\underline{G}^2 - K^{(4)} - L^{(4)} \right),$$

where  $U := -z - \mathbb{E}\underline{G}$ ,  $z = E + i\eta$ ,

$$(4.60) \quad K^{(4)} = N^{-2} \sum_i \mathbb{E} \frac{\partial G_{ii}}{\partial H_{ii}} (\zeta_i - 2)$$

and

$$(4.61) \quad L^{(4)} = N^{-1} \sum_{i,j} \left[ \sum_{k=2}^l \frac{1}{k!} C_{k+1}(H_{ji}) \mathbb{E} \frac{\partial^k G_{ij}}{\partial H_{ji}^k} + R_{l+1}^{(4,ji)} \right].$$

Here,  $R_{l+1}^{(4,ji)}$  is a remainder term defined analogously to  $R_{l+1}^{(ji)}$  in (4.20). We can argue similarly as in the proof of Lemma 4.6(i) and show that  $R_{l_0+1}^{(4,ji)} = O(N^{-2})$  for some  $l_0 \in \mathbb{N}$ . Thus, we have  $|L^{(4)}| \leq \sum_{k=2}^{j_0} O(J_k^{(4)}) + O(N^{-1})$ , where

$$(4.62) \quad J_k^{(4)} := N^{-(k+3)/2} \sum_{i,j} \left| \mathbb{E} \frac{\partial^k G_{ij}}{\partial H_{ji}^k} \right|.$$

Analogously to the proof of Lemma 4.6(ii), we find

$$J_2^{(4)} = O(N^{-5/2} \cdot N^2 \cdot N^{(\alpha-1)/2+\varepsilon}) = O(N^{\alpha/2-1+\varepsilon}),$$

for any fixed  $\varepsilon > 0$ , and  $J_k^{(4)} = O(N^{\alpha/2-1})$  for  $k \geq 3$ . This shows that  $|L^{(2)}| = O(N^{\alpha/2-1+\varepsilon})$  for any fixed  $\varepsilon > 0$ . As in Lemma 4.7, one can show that  $K^{(4)} = O(N^{-1})$ . By (4.56) and Lemma 4.8, we have

$$\mathbb{E}\langle \underline{G} \rangle^2 = O(N^{2\alpha-2-c_0}) \quad \text{and} \quad \frac{1}{N} \mathbb{E}\underline{G}^2 = O(N^{\alpha-1-c_0}).$$

Altogether we have

$$(4.63) \quad \mathbb{E}\underline{G}(z + \mathbb{E}\underline{G}) + 1 = O(N^{\alpha-1-c_0}).$$

Recall that  $m(z)$  is the unique solution of  $x^2 + zx + 1 = 0$  satisfying  $\text{sgn}(\text{Im } m(z)) = \text{sgn}(\text{Im } z) = \text{sgn}(\eta) = 1$ . Let  $\tilde{m}(z)$  be the other solution of  $x^2 + zx + 1 = 0$ . An application of Lemma 5.5 in [3] gives

$$(4.64) \quad \min\{|\mathbb{E}\underline{G} - m(z)|, |\mathbb{E}\underline{G} - \tilde{m}(z)|\} = \frac{O(N^{\alpha-1-c_0})}{\sqrt{\kappa}} = O(N^{\alpha-1-c_0}),$$

where we recall the definition (4.24) of  $\kappa$ . Since  $G = (H - z)^{-1}$ , we know that  $\text{sgn}(\text{Im } \underline{G}) = \text{sgn}(\text{Im } z) > 0$ . Also, we have  $\text{Im } \tilde{m}(z) \leq -c$  for some  $c = c(\kappa) > 0$ . This shows  $|\mathbb{E}\underline{G} - \tilde{m}(z)| \geq c$ . Thus, from (4.64) we have

$$|\mathbb{E}\underline{G} - m(z)| = O(N^{\alpha-1-c_0}),$$

which completes the proof.

4.4. *Proof of Theorem 4.1.* Let  $\tilde{Y}(b) := N^{1-\alpha} \langle \underline{G}(E + b\eta) \rangle$ . As in the one-dimensional case, Proposition 4.2, Theorem 4.1 follows from the following lemma, which generalizes Lemma 4.3.

LEMMA 4.9. *Let  $c_0$  be defined as in (4.3). Under the assumptions of Theorem 4.1, the following hold:*

(i) *For fixed  $m, n \geq 1$  and  $i_1, \dots, i_m, j_1, \dots, j_n \in \{1, 2, \dots, p\}$ , we have*

$$(4.65) \quad \begin{aligned} &\mathbb{E}[\tilde{Y}(b_{i_1}) \cdots \tilde{Y}(b_{i_m}) \tilde{Y}(\bar{b}_{j_1}) \cdots \tilde{Y}(\bar{b}_{j_n})] \\ &= \begin{cases} \sum \prod \frac{-2}{(b_{i_l} - \bar{b}_{j_k})^2} + O(N^{-c_0}), & \text{if } m = n, \\ O(N^{-c_0}), & \text{if } m \neq n, \end{cases} \end{aligned}$$

where the notation  $\sum \prod$  means summing over all distinct ways of partitioning  $b_{i_1}, \dots, b_{i_m}, \bar{b}_{j_1}, \dots, \bar{b}_{j_n}$  into pairs  $b_{i_l}, \bar{b}_{j_k}$ , and each summand is the product of the  $n$  pairs.

(ii) *For any fixed  $b \in \mathbb{H}$ , we have*

$$(4.66) \quad \hat{Y}(b) - \tilde{Y}(b) = \mathbb{E}\hat{Y}(b) = O(N^{-c_0}).$$

Suppose Lemma 4.9 holds. Result (4.65) implies

$$(4.67) \quad (\tilde{Y}(b_1), \dots, \tilde{Y}(b_p)) \xrightarrow{d} (Y(b_1), \dots, Y(b_p)).$$

Theorem 4.1 then follows from (4.66).

PROOF OF LEMMA 4.9. The proof is similar to that of Lemma 4.3. Indeed, we see that

$$(4.68) \quad \begin{aligned} &\mathbb{E}[\tilde{Y}(b_{i_1}) \cdots \tilde{Y}(b_{i_m}) \tilde{Y}(\bar{b}_{j_1}) \cdots \tilde{Y}(\bar{b}_{j_n})] \\ &= N^{(m+n)(1-\alpha)} \mathbb{E}[\langle \underline{G}(E + b_{i_1}\eta) \rangle \cdots \langle \underline{G}(E + b_{i_m}\eta) \rangle \langle \underline{G}(E + \bar{b}_{j_1}\eta) \rangle \cdots \\ &\quad \times \langle \underline{G}(E + \bar{b}_{j_n}\eta) \rangle], \end{aligned}$$

which can be computed in the same way as  $\mathbb{E}\langle G \rangle^m \langle G^* \rangle^n = \mathbb{E}\langle G(E + i\eta) \rangle^m \times \langle G(E - i\eta) \rangle^n$  in Section 4.2. Most of our previous techniques and estimates can be applied to the new computation, and the only difference is when using the resolvent identity [e.g., in (4.54)], we now have

$$G(E + b_{i_k}\eta)G(E + \bar{b}_{j_l}\eta) = \frac{1}{b_{i_k} - \bar{b}_{j_l}}(G(E + b_{i_k}\eta) - G(E + \bar{b}_{j_l}\eta))$$

instead of

$$GG^* = \frac{1}{2i}(G - G^*).$$

This will give us different constants in the leading terms, and lead to the induction step

$$\begin{aligned} &\mathbb{E}[\tilde{Y}(b_{i_1}) \cdots \tilde{Y}(b_{i_m}) \tilde{Y}(\bar{b}_{j_1}) \cdots \tilde{Y}(\bar{b}_{j_n})] \\ &= \sum_{k=1}^n \frac{-2}{(b_{i_m} - \bar{b}_{j_k})^2} \mathbb{E}[\tilde{Y}(b_{i_1}) \cdots \tilde{Y}(b_{i_{m-1}}) \tilde{Y}(\bar{b}_{j_1}) \cdots \tilde{Y}(\bar{b}_{j_n}) / \tilde{Y}(\bar{b}_{j_k})] \\ &\quad + O(N^{-c_0}) \end{aligned}$$

for  $m, n \geq 1$ . One can also show that

$$\mathbb{E}[\tilde{Y}(b_{i_1}) \cdots \tilde{Y}(b_{i_m})] = O(N^{-c_0}) \quad \text{and} \quad \mathbb{E}[\tilde{Y}(\bar{b}_{j_1}) \cdots \tilde{Y}(\bar{b}_{j_n})] = O(N^{-c_0})$$

for  $m, n \geq 2$ . These results together imply (4.65).

Moreover, (4.66) says nothing but  $N^{1-\alpha} \mathbb{E}(G(E + b\eta) - m(E + b\eta)) = O(N^{-c_0})$ , and this can be shown using the steps in Section 4.3, in which we proved  $N^{1-\alpha} \mathbb{E}(G(E + i\eta) - m(E + i\eta)) = O(N^{-c_0})$ .  $\square$

**5. Convergence of general functions.** Similarly, as in the resolvent case, in Section 5 we prove the following analogue of Theorem 2.3.

**THEOREM 5.1.** *Theorem 2.3 holds for Wigner matrices  $H$  satisfying Definition 3.2, and the convergence also holds in the sense of moments.*

Let us abbreviate  $f_\eta(x) := f(\frac{x-E}{\eta})$ , and denote  $[\text{Tr } f_\eta(H)] := \text{Tr } f_\eta(H) - N \int_{-2}^2 \varrho(x) f_\eta(x) dx$ . Our next result is a particular case of Theorem 5.1.

**PROPOSITION 5.2.** *Let  $\eta, E, H$  be as in Theorem 5.1. Then*

$$(5.1) \quad [\text{Tr } f_\eta(H)] \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2\pi^2} \int \left(\frac{f(x) - f(y)}{x - y}\right)^2 dx dy\right)$$

as  $N \rightarrow \infty$ . The convergence also holds in the sense of moments.

Our main work in this section will be to show the above one-dimensional case, since the proof can easily be extended to the general case (see Section 5.5). Recall that for a random variable  $X$ ,  $\langle X \rangle := X - \mathbb{E}X$ . Proposition 5.2 is a direct consequence of the following lemma.

LEMMA 5.3. *Under the conditions of Theorem 5.1, we have the following results:*

(i) *For any  $n \geq 2$*

$$(5.2) \quad \mathbb{E}\langle \text{Tr } f_\eta(H) \rangle^n = \frac{n-1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x-y} \right)^2 dx dy \cdot \mathbb{E}\langle \text{Tr } f_\eta(H) \rangle^{n-2} + O(N^{-c/n^2}),$$

where  $c = c(r, s, \alpha) > 0$ .

(ii) *The random variables  $[\text{Tr } f_\eta(H)]$  and  $\langle \text{Tr } f_\eta(H) \rangle$  are close in the sense that*

$$(5.3) \quad [\text{Tr } f_\eta(H)] - \langle \text{Tr } f_\eta(H) \rangle = \mathbb{E}[\text{Tr } f_\eta(H)] = O(N^{-rs^2c_0/16}),$$

with  $c_0$  defined in (4.3).

Assume Lemma 5.3 holds. Then (5.2) and Wick’s theorem imply

$$(5.4) \quad \langle \text{Tr } f_\eta(H) \rangle \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x-y} \right)^2 dx dy\right)$$

as  $N \rightarrow \infty$ . Note that the above result is proved in a stronger sense that we have convergence in moments. Proposition 5.2 then follows from (5.3).

Sections 5.1 to 5.3 are devoted to proving Lemma 5.3(i). Before starting the proof, we give some explanations of the ideas, especially the choice of truncations in the proof. We use Lemma 3.3 to write  $f_\eta(H)$  in the form (5.6) below, where we scale the cutoff function  $\chi$  to be supported in an interval of size  $O(\sigma)$ , with  $N^{-1} \ll \sigma \ll \eta$ . This scaling ensures that when we integrate  $\varphi_f$ , the integral of the last term in (5.7) below dominates over the others.

We then write  $\mathbb{E}\langle \text{Tr } f_\eta(H) \rangle^n$  as an integral over  $\mathbb{C}^n$ , written  $\int F$  in (5.8) below. The leading contribution to  $\int F$  arises from the region  $\{|y_1|, \dots, |y_n| \geq \omega\}$ , where  $\omega \gg N^{-1}$  is a second truncation scale. In order to ensure that  $\int F$  is small in the complementary region, we require that  $\omega \ll \sigma$ . Then, when estimating  $\int_{|y_1| < \omega} F$ , the integral over  $z_1$  yields a factor that is small enough to compensate the integrals from the other variables. We use the notation  $\sigma = N^{-(\alpha+\beta)}$  and  $\omega = N^{-(\alpha+\gamma)}$ , so that  $0 < \beta < \gamma$ . In addition, for all steps of the analysis to work, we have further requirements on the exponents  $\gamma$  and  $\beta$ ; for instance, the last step in (5.18) below requires  $n\beta \leq rs\gamma/4$ . Combining all requirements, we are led to set  $\beta$  as in (5.5) below.

5.1. *Transformation by Helffer–Sjöstrand formula.* Let  $f \in C^{1,r,s}(\mathbb{R})$  with  $r, s > 0$ , and without loss of generality we assume  $s \leq 1$ . Fix  $n \geq 2$ , and define  $\sigma := N^{-(\alpha+\beta)}$ , where

$$(5.5) \quad \beta := \frac{rs^2c_0}{24n^2},$$

and  $c_0$  is defined in (4.3). We define  $\tilde{f}$  as in (3.4). Let  $\chi$  be as in Lemma 3.3 satisfying  $\chi(y) = 1$  for  $|y| \leq 1$ , and  $\chi(y) = 0$  for  $|y| \geq 2$ . An application of Lemma 3.3 gives

$$(5.6) \quad f_\eta(H) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial_{\bar{z}}(\tilde{f}_\eta(z)\chi(z/\sigma))}{H - z} d^2z = \int_{\mathbb{C}} \varphi_f(z)G(z) d^2z,$$

where

$$(5.7) \quad \begin{aligned} \varphi_f(x + iy) = & \frac{1}{2\pi} \left( (i - 1)(f'_\eta(x + y) - f'_\eta(x))\chi(y/\sigma) \right. \\ & \left. - \frac{1}{\sigma}(f_\eta(x + y) - f_\eta(x))\chi'(y/\sigma) + \frac{i}{\sigma}f_\eta(x)\chi'(y/\sigma) \right). \end{aligned}$$

Thus,

$$(5.8) \quad \begin{aligned} \mathbb{E}\langle \text{Tr } f_\eta(H) \rangle^n &= N^n \int \varphi_f(z_1) \cdots \varphi_f(z_n) \mathbb{E}\langle \underline{1}G \rangle \cdots \langle \underline{n}G \rangle d^2z_1 \cdots d^2z_n \\ &=: \int F, \end{aligned}$$

where  ${}^kG := (H - z_k)^{-1}$  for  $i \in \{1, 2, \dots, n\}$ . Note that  $\chi(y/\sigma) \equiv 0$  for  $|y| \geq 2\sigma$ , and we only need to consider the integral for  $|y_1|, \dots, |y_n| \leq 2\sigma$ .

5.2. *The subleading terms.* Let  $\omega := N^{-(\alpha+\gamma)}$  with  $\gamma := 4n\beta/rs$ , and by (5.5) we have  $\alpha + \beta < \alpha + \gamma < 1$ . We define  $X := \{|x_1|, \dots, |x_n| \leq 2 - \frac{\kappa}{2}\}$  and  $Y := \{|y_1|, \dots, |y_n| \in [\omega, 2\sigma]\}$ , where we recall the definition (4.24) of  $\kappa$ . We have a lemma about  $\int F$  outside the region  $X \times Y$ .

LEMMA 5.4. *For  $F$  as in (5.8) we have*

$$(5.9) \quad \int_{(X \times Y)^c} F = O(N^{-\beta/2}).$$

PROOF. We first estimate  $\int_{\mathbb{R}^n \times Y^c} F$ . By the estimates (3.9) and (3.11), we know

$$(5.10) \quad |\langle \underline{1}G \rangle \cdots \langle \underline{n}G \rangle| \prec \frac{1}{|y_1 \cdots y_n| N^n}$$

uniformly in  $\{|y_1|, \dots, |y_n| \leq 2\sigma\}$ . Since  $\chi'(y/\sigma) = 0$  for  $|y| < \sigma$ , we have

$$\begin{aligned}
 & \int_{|y| < \omega} \left| \varphi_f(z) \cdot \frac{1}{y} \right| d^2z \\
 (5.11) \quad &= O(1) \cdot \int_{|y| < \omega} \left| (f'_\eta(x+y) - f'_\eta(x)) \cdot \frac{1}{y} \right| dx dy \\
 &= O(1) \cdot \int_{|b| < N^{\beta-\gamma}} \left| (f'(a + bN^{-\beta}) - f'(a)) \cdot \frac{1}{b} \right| da db,
 \end{aligned}$$

where in the second step we used the change of variables

$$(5.12) \quad a := (x - E)/\eta \quad \text{and} \quad b := y/\sigma.$$

By the Hölder continuity and decay of the function  $f'$ , we know

$$\begin{aligned}
 (5.13) \quad |f'(a + bN^{-\beta}) - f'(a)| &\leq C \min \left\{ (|b|N^{-\beta})^r, \frac{1}{1 + |a|^{1+s}} \right\} \\
 &\leq C (|b|N^{-\beta})^{rq} \left( \frac{1}{1 + |a|^{1+s}} \right)^{1-q}
 \end{aligned}$$

for all  $q \in [0, 1]$ . Choose  $q = q_0(s) := \frac{s}{2(1+s)} \geq \frac{s}{4}$ , so that  $(1 + s)(1 - q_0) = 1 + \frac{s}{2} > 1$ . Thus, we have

$$\begin{aligned}
 & \int_{|y| < \omega} \left| \varphi_f(z) \cdot \frac{1}{y} \right| d^2z \\
 (5.14) \quad &= O(N^{-rq_0\beta}) \cdot \int_{|b| < N^{\beta-\gamma}} |b|^{rq_0-1} \frac{1}{1 + |a|^{1+s/2}} da db \\
 &= O(N^{-rq_0\gamma}).
 \end{aligned}$$

Similarly, one can show that

$$(5.15) \quad \int_{|y| \in [\omega, \sigma]} \left| \varphi_f(z) \cdot \frac{1}{y} \right| d^2z = O(N^{-rq_0\beta}).$$

We also have

$$(5.16) \quad \int_{|y| \geq \sigma} \left| \varphi_f(z) \cdot \frac{1}{y} \right| d^2z = N^\beta \int_{1 \leq |b| \leq 2} \left| \psi_f(a, b) \cdot \frac{1}{b} \right| da db = O(N^\beta),$$

where we used the change of variables (5.12), and abbreviate

$$\begin{aligned}
 (5.17) \quad \psi_f(a, b) &:= \frac{1}{2\pi} (N^{-\beta}(i-1)(f'(a + bN^{-\beta}) - f'(a))\chi(b) \\
 &\quad - (f(a + bN^{-\beta}) - f(a))\chi'(b) + if(a)\chi'(b)).
 \end{aligned}$$

Using Lemma 4.5, we have

$$\begin{aligned}
 \left| \int_{\mathbb{R}^n \times Y^c} F \right| &\leq n \int_{|y_1| < \omega} |F| \\
 &\prec \int_{|y_1| < \omega} \left| \varphi_f(z_1) \cdot \frac{1}{y_1} \cdots \varphi_f(z_n) \cdot \frac{1}{y_n} \right| d^2 z_1 \cdots d^2 z_n \\
 (5.18) \quad &= \int_{|y_1| < \omega} \left| \varphi_f(z_1) \cdot \frac{1}{y_1} \right| dz_1 \\
 &\quad \times \int \left| \varphi_f(z_2) \cdot \frac{1}{y_1} \cdots \varphi_f(z_n) \cdot \frac{1}{y_n} \right| d^2 z_2 \cdots d^2 z_n \\
 &= O(N^{-rq_0\gamma} \cdot N^{(n-1)\beta}) \leq O(N^{-rs\gamma/4} \cdot N^{(n-1)\beta}) \\
 &\leq O(N^{-\beta}).
 \end{aligned}$$

Next, we estimate  $\int_{X^c \times Y} F$ . By the decay of the functions  $f$  and  $f'$ , we have

$$\begin{aligned}
 \int_{|x| > 2 - \frac{\kappa}{2}} \left| \varphi_f(z) \cdot \frac{1}{y} \right| d^2 z \\
 \leq N^\beta \int_{|a| \geq \frac{\kappa}{2\eta}} \left| \psi_f(a, b) \cdot \frac{1}{b} \right| da db \\
 (5.19) \quad = O(N^\beta) \cdot \left( \int_{|a| \geq \frac{\kappa}{2\eta}} \frac{1}{1 + |a|^{1+s/2}} da \right. \\
 \left. + \int_{|a| \geq \frac{\kappa}{2\eta}} \frac{1}{1 + |a|^{1+s}} da + \int_{|a| \geq \frac{\kappa}{2\eta}} |f(a)| da \right) \\
 = O(N^{\beta-s\alpha/2}).
 \end{aligned}$$

where  $a, b$  are defined as in (5.12). Hence,

$$\begin{aligned}
 \left| \int_{X^c \times Y} F \right| &\prec \int_{\{|x_1| > 2 - \frac{\kappa}{2}\} \times Y} \left| \varphi_f(z_1) \cdot \frac{1}{y_1} \cdots \varphi_f(z_n) \cdot \frac{1}{y_n} \right| d^2 z_1 \cdots d^2 z_n \\
 (5.20) \quad &= O(N^{(n-1)\beta}) \cdot \int_{\{|x_1| > 2 - \frac{\kappa}{2}, |y_1| \geq \omega\}} \left| \varphi_f(z_1) \cdot \frac{1}{y_1} \right| d^2 z_1 \\
 &= O(N^{n\beta-s\alpha/2}) \leq O(N^{-\beta}).
 \end{aligned}$$

Combining (5.18) and (5.20), we get

$$\int F = \int_{X \times Y} F + O(N^{-\beta/2}). \quad \square$$

5.3. *The main computation.* Now let us focus on the integral  $\int_{X \times Y} F$ . Note that now we are in the “good” region where it is effective to apply the cumulant expansion to the resolvent. In  $X \cap Y$ , we want to compute the quantity

$\mathbb{E}\langle \underline{1G} \rangle \cdots \langle \underline{nG} \rangle$ . Note that this is very close to the expression we had in (4.14). Let us abbreviate

$$Q_m := \langle \underline{1G} \rangle \cdots \langle \underline{mG} \rangle \quad \text{and} \quad Q_m^{(k)} := Q_m / \langle \underline{kG} \rangle$$

for all  $1 \leq k \leq m \leq n$ , and  $\zeta_i := \mathbb{E}|\sqrt{N}H_{ii}|^2$ . We proceed the computation as in Section 4.2, and get an analogue of (4.22):

$$\begin{aligned} \mathbb{E}Q_n &= \frac{1}{T_n} \mathbb{E}Q_n \langle \underline{nG} \rangle - \frac{1}{T_n} \mathbb{E}Q_{n-1} \mathbb{E}\langle \underline{nG} \rangle^2 + \frac{1}{NT_n} \mathbb{E}Q_{n-1} \langle \underline{nG}^2 \rangle \\ (5.21) \quad &- \frac{\tilde{K}}{T_n} - \frac{\tilde{L}}{T_n} + \frac{2}{N^2 T_n} \sum_{k=1}^{n-1} \mathbb{E}Q_{n-1}^{(k)} \langle \underline{kG}^2 \cdot \underline{nG} \rangle, \end{aligned}$$

where  $T_n := -z_n - 2\mathbb{E}\underline{nG}$ ,

$$\tilde{K} = N^{-2} \sum_i \mathbb{E} \frac{\partial(\langle \langle \underline{1G} \rangle \cdots \langle \underline{n-1G} \rangle \rangle \cdot \underline{nG}_{ii})}{\partial H_{ii}} (\zeta_i - 2)$$

and

$$\tilde{L} = N^{-1} \sum_{i,j} \left[ \sum_{k=2}^l \frac{1}{k!} C_{l+1}(H_{ji}) \mathbb{E} \frac{\partial^k(\langle \langle \underline{1G} \rangle \cdots \langle \underline{n-1G} \rangle \rangle \cdot \underline{nG}_{ij})}{\partial H_{ji}^k} + \tilde{R}_{l+1}^{(ji)} \right].$$

Here,  $\tilde{R}_{l+1}^{(ji)}$  is a remainder term defined analogously to  $R_{l+1}^{(ji)}$  in (4.20). Note in  $Y$ , we have  $|y_1|, \dots, |y_n| \geq \omega$ , and we have estimates analogue to those in Section 4.2. We state these estimates in the next lemma and omit the proof.

LEMMA 5.5. *Let us extend the definition of  $c_0$  in (4.3) to a function  $c_0(\cdot) : (0, 1) \rightarrow \mathbb{R}$  such that*

$$(5.22) \quad c_0(x) := \frac{1}{3} \min\{x, 1 - x\}.$$

The following results hold uniformly in  $X \times Y$ :

(i) *Analogously to Lemma 4.4, for any  $m \in \mathbb{N}_+$  and  $k = 1, 2, \dots, n$ , we have*

$$(5.23) \quad |\langle \underline{kG}^m \rangle| \prec N^{m(\alpha+\gamma)-1}$$

as well as

$$(5.24) \quad |\langle \underline{kG}^m \rangle_{ij}| \prec \begin{cases} N^{(m-1)(\alpha+\gamma)}, & \text{if } i = j, \\ N^{(m-1/2)(\alpha+\gamma)-1/2}, & \text{if } i \neq j. \end{cases}$$

(ii) *Analogously to (4.23), we have*

$$(5.25) \quad \left| \frac{1}{T_n} \right| = O(1).$$

(iii) Analogously to Lemma 4.6, we have

$$(5.26) \quad |\tilde{L}| = O(N^{n(\alpha+\gamma-1)-c_0(\alpha+\gamma)}).$$

(iv) Analogously to Lemma 4.7, we have

$$(5.27) \quad |\tilde{K}| = O(N^{n(\alpha+\gamma-1)-(\alpha+\gamma)/2}).$$

(v) Analogously to Lemma 4.8, we have

$$(5.28) \quad \mathbb{E}^k \underline{G}^2 = O(N^{\alpha+\gamma-c_0(\alpha+\gamma)})$$

for  $k = 1, 2, \dots, n$ .

Applying Lemma 5.5 to (5.21) yields

$$(5.29) \quad \mathbb{E} Q_n = \frac{2}{N^2 T_n} \sum_{k=1}^{n-1} \mathbb{E} Q_{n-1}^{(k)} \underline{k G^2 \cdot n G} + O(N^{n(\alpha+\gamma-1)-c_0(\alpha+\gamma)})$$

uniformly in  $X \times Y$ . Note that by the definition of  $\beta$  and  $\gamma$  we have  $\gamma = sc_0(\alpha)/6n < c_0(\alpha)/2$ , which gives  $c_0(\alpha + \gamma) > 5c_0(\alpha)/6 > 0$ . Since we have the simple estimate

$$(5.30) \quad \int |\varphi_f(z)| d^2z = O(\eta),$$

we know

$$(5.31) \quad \begin{aligned} \int_{X \times Y} F &= N^n \int_{X \times Y} \varphi_f(z_1) \cdots \varphi_f(z_n) \mathbb{E} Q_n d^2z_1 \cdots d^2z_n \\ &= \frac{2N^{n-2}}{T_n} \sum_{k=1}^{n-1} \int_{X \times Y} \varphi_f(z_1) \cdots \varphi_f(z_n) \mathbb{E} Q_{n-1}^{(k)} \underline{k G^2 \cdot n G} d^2z_1 \cdots d^2z_n \\ &\quad + O(N^{-2c_0(\alpha)/3}), \end{aligned}$$

where in the estimate of the error term we implicitly used  $n\gamma = sc_0(\alpha)/6 \leq c_0(\alpha)/6$ . By symmetry, it suffices to fix  $k \in \{1, 2, \dots, n - 1\}$ , and consider the integral over  $X \times Y$  of

$$(5.32) \quad F_{kn} := \frac{2N^{n-2}}{T_n} \varphi_f(z_1) \cdots \varphi_f(z_n) \mathbb{E} Q_{n-1}^{(k)} \underline{k G^2 \cdot n G}.$$

As before, we summarize the necessary estimates into a lemma.

LEMMA 5.6. *Let  $F_{kn}$  be as in (5.32). Then we have the following estimates:*

(i) *Let  $A_1 := \{(x, y) \in X \times Y : y_k y_n > 0, |x_k - x_n| \leq \eta N^{-(n+1)\gamma}\}$ . Then*

$$(5.33) \quad \left| \int_{A_1} F_{kn} \right| < N^{-\gamma}.$$

(ii) Let  $A_2 := \{(x, y) \in X \times Y : y_k y_n > 0, |x_k - x_n| \in (\eta N^{-(n+1)\gamma}, \eta N^{(n+1)\gamma/s})\}$ . Then

$$(5.34) \quad \int_{A_2} F_{kn} = O(N^{-c_0(\alpha)/4}),$$

where the function  $c_0(\cdot)$  is defined in (5.22).

(iii) For  $A_3 := \{(x, y) \in X \times Y : |x_k - x_n| > \eta N^{(n+1)\gamma/s}\}$ , we have

$$(5.35) \quad \left| \int_{A_3} F_{kn} \right| < N^{-\gamma}.$$

PROOF. (i) By Lemma 5.5(i)–(ii), we have

$$\left| \frac{N^{n-2}}{T_n} \mathbb{E} Q_{n-1}^{(k)} \underline{kG^2} \cdot \underline{nG} \right| < \omega^{-n}$$

uniformly in  $A_1$ . Thus, by (5.30) and the decay of  $f$  and  $f'$  we know

$$(5.36) \quad \begin{aligned} \left| \int_{A_1} \tilde{F}_{kn} \right| &< \omega^{-n} \cdot \eta^{n-2} \cdot \int |\varphi_f(z_k) \cdot \varphi_f(z_n)| \mathbf{1}_{\{|x_k - x_n| \leq \eta \cdot N^{-(n+1)\gamma}\}} d^2 z_k d^2 z_n \\ &= O(\omega^{-n} \cdot \eta^n) \int |\psi_f(a_k, b_k) \cdot \psi_f(a_n, b_n)| \\ &\quad \times \mathbf{1}_{\{|a_k - a_n| \leq N^{-(n+1)\gamma}\}} da_k db_k da_n db_n \\ &= O(N^{n\gamma} \cdot N^{-(n+1)\gamma}) = O(N^{-\gamma}), \end{aligned}$$

where we use the change of variables

$$(5.37) \quad a_i = (x_i - E)/\eta \quad \text{and} \quad b_i = y_i/\sigma, \quad i = k, n,$$

and  $\psi_f$  is defined as in (5.17).

(ii) Note that our assumption (5.5) on  $\beta$  shows  $\eta N^{(n+1)\gamma/s} = O(N^{-\alpha/2})$ . By the resolvent identity, the semicircle law (3.9), and Lemma 5.5 we know

$$(5.38) \quad \begin{aligned} &\left| \frac{N^{n-2}}{T_n} \mathbb{E} Q_{n-1}^{(k)} \underline{kG^2} \cdot \underline{nG} \right| \\ &= \left| N^{n-2} \mathbb{E} Q_{n-1}^{(k)} \left( \frac{\langle \underline{kG^2} \rangle + \mathbb{E} \underline{kG^2}}{T_n(z_k - z_n)} + \frac{\langle \underline{nG} \rangle - \langle \underline{kG} \rangle + \mathbb{E}(\underline{nG} - \underline{kG})}{T_n(z_k - z_n)^2} \right) \right| \\ &< \omega^{-(n-2)} \cdot \left( \frac{\omega^{-1} N^{-c_0(\alpha+\gamma)}}{\eta N^{-(n+1)\gamma}} + \frac{(N\omega)^{-1} + (N\omega)^{-1} + N^{-c_0(\alpha+\gamma)}}{\eta^2 N^{-2(n+1)\gamma}} \right) \\ &= O(\eta^{-n} \cdot N^{3n\gamma - c_0(\alpha+\gamma)}) \leq O(\eta^{-n} \cdot N^{-c_0(\alpha)/3}) \end{aligned}$$

uniformly in  $A_2$ . Hence, (5.30) yields

$$(5.39) \quad \int_{A_2} F_{kn} = O(N^{-c_0(\alpha)/4}).$$

(iii) Similarly, as in (5.36), we know

$$\begin{aligned}
 & \left| \int_{A_3} F_{kn} \right| < \omega^{-n} \cdot \eta^{n-2} \cdot \int |\varphi_f(z_k) \cdot \varphi_f(z_n)| \mathbf{1}_{\{|x_k - x_n| > \eta N^{(n+1)\gamma/s}\}} d^2 z_k d^2 z_n \\
 (5.40) \quad & = O(\omega^{-n} \cdot \eta^n) \int |\psi_f(a_k, b_k) \cdot \psi_f(a_n, b_n)| \\
 & \quad \times \mathbf{1}_{\{|a_k - a_n| > N^{(n+1)\gamma/s}\}} da_k db_k da_n db_n.
 \end{aligned}$$

Note that in  $\{|a_k - a_n| > N^{(n+1)\gamma/s}\}$ , either  $|a_k| > \frac{1}{2}N^{(n+1)\gamma/s}$  or  $|a_n| > \frac{1}{2}N^{(n+1)\gamma/s}$ . Hence, by the decay conditions of  $f$  and  $f'$ , we have

$$\left| \int_{A_3} F_{kn} \right| < \omega^{-n} \cdot \eta^n \cdot N^{-(n+1)\gamma} = N^{-\gamma}. \quad \square$$

Let  $A_4 := \{(x, y) \in (X \times Y) : y_k y_n < 0, |x_k - x_n| \leq \eta N^{(n+1)\gamma/s}\}$ . Note that  $C^n$  is the disjoint union of  $A_1, \dots, A_4$ . The next result is about the integral of  $F_{kn}$  in  $A_4$ , which gives the leading contribution.

LEMMA 5.7. *We have*

$$\begin{aligned}
 (5.41) \quad \int_{A_4} F_{kn} &= \frac{1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy \cdot \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} \\
 & \quad + O(N^{-rs\beta/9}),
 \end{aligned}$$

where  $\beta$  is defined in (5.5).

PROOF. Step 1. By symmetry, let us consider  $A_5 := \{(x, y) \in A_4 : y_k \geq \omega, y_n \leq -\omega, |x_k - x_n| \leq \eta N^{(n+1)\gamma/s}\}$ . Similarly, as in (5.38), we have

$$\begin{aligned}
 (5.42) \quad & \frac{N^{n-2}}{T_n} \mathbb{E} Q_{n-1}^{(k)} \mathbb{E}^k G^2 \cdot {}^n G \\
 & = N^{n-2} \mathbb{E} Q_{n-1}^{(k)} \frac{\mathbb{E}^n G - \mathbb{E}^k G}{T_n(z_k - z_n)^2} + O(\eta^{-n} \cdot N^{-c_0(\alpha)/3})
 \end{aligned}$$

uniformly in  $A_5$ . Note the semicircle law (3.9) now gives

$$(5.43) \quad \frac{\mathbb{E}^n G - \mathbb{E}^k G}{T_n} = \frac{\mathbb{E}^n G - \mathbb{E}^k G}{-z_n - 2\mathbb{E}^n G} = -1 + O(N^{-c_0(\alpha+\gamma)})$$

uniformly in  $A_5$ . By (5.30), we know

$$\begin{aligned}
 \int_{A_5} F_{kn} &= N^{n-2} \int_{A_5} \varphi_f(z_1) \cdots \varphi_f(z_n) \mathbb{E} Q_{n-1}^{(k)} \frac{-2}{(z_k - z_n)^2} d^2 z_1 \cdots d^2 z_n \\
 & \quad + O(N^{-c_0(\alpha)/3}) \\
 & = -2 \int_{A'_5} \frac{1}{(z_k - z_n)^2} \varphi_f(z_k) \varphi_f(z_n) d^2 z_k d^2 z_n \cdot \int_{A''_5} \hat{F}_{kn} + O(N^{-c_0(\alpha)/3}),
 \end{aligned}$$

where we decompose  $A_5 = A'_5 \times A''_5$ , with  $A'_5$  depends on  $(x_k, y_k, x_n, y_n)$ . Here,  $\hat{F}_{kn}$  is defined as

$$\hat{F}_{kn} := N^{n-2} \varphi_f(z_1) \cdots \varphi_f(z_{n-1}) / \varphi_f(z_k) \mathbb{E} Q_{n-1}^{(k)}.$$

Let  $X^{(k,n)} := \{|x_1|, \dots, |x_{k-1}|, |x_{k+1}|, \dots, |x_{n-1}| \leq 2 - \frac{\kappa}{2}\}$ , and  $Y^{(k,n)} := \{|y_1|, \dots, |y_{k-1}|, |y_{k+1}|, \dots, |y_{n-1}| \geq \omega\}$ , and note that  $A''_5 = X^{(k,n)} \times Y^{(k,n)}$ . Applying Lemma 5.4 with  $n$  replaced by  $n - 2$ , we get

$$(5.44) \quad \int_{A''_5} \hat{F}_{kn} = \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} + O(N^{-\beta/2}).$$

By the decay conditions of  $f$  and  $f'$ ,

$$\begin{aligned} & \int_{A'_5} \frac{1}{(z_k - z_n)^2} \varphi_f(z_k) \varphi_f(z_n) \, d^2 z_k \, d^2 z_n \\ &= \int \frac{1}{(z_k - z_n)^2} \varphi_f(z_k) \varphi_f(z_n) \mathbf{1}_{\{|y_k| \geq \omega, |y_n| \leq -\omega\}} \, d^2 z_k \, d^2 z_n + O(N^{-\gamma}) \\ &= \int \frac{\psi_f(a_k, b_k) \psi_f(a_n, b_n)}{(a_k - a_n + i(b_k - b_n)N^{-\beta})^2} \mathbf{1}_{\{|b_k| \geq N^{\beta-\gamma}, |b_n| \leq -N^{\beta-\gamma}\}} \, d^2 z_k \, d^2 z_n \\ &\quad + O(N^{-\gamma}) \\ &=: \int \Psi + O(N^{-\gamma}), \end{aligned}$$

where in the second-to-last step we use the change of variables in (5.37), and  $\psi_f$  is as in (5.17). Note that one can repeat the steps in the proof of Lemma 4.4 for any  $f \in C^{1,r,s}(\mathbb{R})$  instead of  $f(x) = (\frac{x+i}{x^2+1})^k$ , and get

$$(5.45) \quad | \langle f_\eta(H) \rangle | < 1.$$

Together with Lemma 4.5, we know

$$(5.46) \quad \int_{A_5} F_{kn} = -2 \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} \int \Psi + \left| \int \Psi \right| \cdot O(N^{-\beta/2}) + O(N^{-\gamma/2}).$$

Step 2. We now compute  $\int \Psi$ . Let us set

$$(5.47) \quad \begin{aligned} \psi_{f,1}(a, b) &:= \frac{i-1}{2\pi} N^{-\beta} (f'(a + bN^{-\beta}) - f'(a)) \chi(b), \\ \psi_{f,2}(a, b) &:= -\frac{1}{2\pi} (f(a + bN^{-\beta}) - f(a)) \chi'(b) \quad \text{and} \\ \psi_{f,3}(a, b) &:= \frac{i}{2\pi} f(a) \chi'(b), \end{aligned}$$

which gives  $\psi_f(a, b) = \psi_{f,1}(a, b) + \psi_{f,2}(a, b) + \psi_{f,3}(a, b)$ . Let

$$\int \Psi_{i,j} := \int \frac{\psi_{f,i}(a_k, b_k) \psi_{f,j}(a_n, b_n)}{(a_k - a_n + i(b_k - b_n)N^{-\beta})^2} \mathbf{1}_{\{|b_k| \geq N^{\beta-\gamma}, |b_n| \leq -N^{\beta-\gamma}\}} \, da_k \, db_k \, da_n \, db_n,$$

with  $1 \leq i, j \leq 3$ . We will calculate  $\int \Psi$  by calculating 6 different integrals  $\int \Psi_{i,j}$ , subject to symmetry.

Let us first consider  $\int \Psi_{1,1}$ . Note that, by (5.13),

$$(5.48) \quad |f'(a + bN^{-\beta}) - f'(a)| \leq C(|b|N^{-\beta})^{r_{q_0}} \left( \frac{1}{1 + |a|^{1+s/2}} \right),$$

where  $q_0 = q_0(s) = \frac{s}{2(s+1)} \geq \frac{s}{4} > 0$ . Thus,

$$\begin{aligned} \int \Psi_{1,1} &\leq C \int \frac{1}{|b_k b_n| N^{-2\beta}} N^{-2\beta - 2r_{q_0}\beta} |b_k b_n|^{r_{q_0}} \chi(b_k) \chi(b_n) \\ &\quad \times \mathbf{1}_{\{b_k \geq N^{\beta-\gamma}, b_n \leq -N^{\beta-\gamma}\}} db_k db_n \\ &= O(N^{-2r_{q_0}\beta}). \end{aligned}$$

Now we consider  $\int \Psi_{2,1}$ . Note that  $\chi'(b) = 0$  for  $|b| < 1$ . Using integration by parts on the variable  $a_k$ , we know

$$\begin{aligned} \int \Psi_{2,1} &= \int \frac{\psi_{f,2}(a_k, b_k) \psi_{f,1}(a_n, b_n)}{(a_k - a_n + i(b_k - b_n)N^{-\beta})^2} \mathbf{1}_{\{b_k \geq 1, b_n \leq -N^{\beta-\gamma}\}} da_k db_k da_n db_n \\ &= \int \frac{\tilde{\psi}_{f,2}(a_k, b_k) \psi_{f,1}(a_n, b_n)}{a_k - a_n + i(b_k - b_n)N^{-\beta}} \mathbf{1}_{\{b_k \geq 1, b_n \leq -N^{\beta-\gamma}\}} da_k db_k da_n db_n, \end{aligned}$$

where  $\tilde{\psi}_{f,2}(a, b) := -\frac{1}{2\pi} (f'(a + bN^{-\beta}) - f'(a)) \chi'(b)$ . Then by (5.48) we know

$$(5.49) \quad \int \Psi_{2,1} = O(N^\beta \cdot N^{-r_{q_0}\beta} \cdot N^{-\beta - r_{q_0}\beta}) = O(N^{-2r_{q_0}\beta}).$$

Similarly,  $\int \Psi_{3,1} = O(N^{-r_{q_0}\beta})$ .

Now we move to  $\int \Psi_{2,2}$ . Using integration by parts on the variables  $a_k$  and  $a_n$ , we know

$$\begin{aligned} \int \Psi_{2,2} &= \int \log(a_k - a_n + i(b_k - b_n)N^{-\beta}) \tilde{\psi}_{f,2}(a_k, b_k) \tilde{\psi}_{f,2}(a_n, b_n) \\ &\quad \times \mathbf{1}_{\{b_k \geq 1, b_n \leq -1\}} da_k db_k da_n db_n \\ &= O(\log N \cdot N^{-2r_{q_0}\beta}) = O(N^{-r_{q_0}\beta}). \end{aligned}$$

Similarly,  $\int \Psi_{3,2} = O(N^{-r_{q_0}\beta/2}) \leq O(N^{-rs\beta/8})$ .

The leading contribution comes from  $\Psi_{3,3}$ . Note that

$$\begin{aligned} \int \Psi_{3,3} &= -\frac{1}{4\pi^2} \int \frac{f(a_k)f(a_n)}{(a_k - a_n + i(b_k - b_n)N^{-\beta})^2} \chi'(b_k)\chi'(b_n) \\ &\quad \times \mathbf{1}_{\{b_k \geq 1, b_n \leq -1\}} da_k db_k da_n db_n \\ &= \frac{1}{8\pi^2} \int \left( \frac{f(a_k) - f(a_n)}{(a_k - a_n + i(b_k - b_n)N^{-\beta})} \right)^2 \chi'(b_k)\chi'(b_n) \\ &\quad \times \mathbf{1}_{\{b_k \geq 1, b_n \leq -1\}} da_k db_k da_n db_n \\ &= \frac{1}{8\pi^2} \int \left( \frac{f(a_k) - f(a_n)}{a_k - a_n} \right)^2 \chi'(b_k)\chi'(b_n) \mathbf{1}_{\{b_k \geq 1, b_n \leq -1\}} da_k db_k da_n db_n \\ &\quad + O(N^{-\beta/3}) \\ &= -\frac{1}{8\pi^2} \int \left( \frac{f(a_k) - f(a_n)}{a_k - a_n} \right)^2 da_k da_n + O(N^{-\beta/3}), \end{aligned}$$

where the second-to-last last step is an elementary estimate whose details we omit. Hence, by (5.45) and Lemma 4.5 we have

$$(5.50) \quad \int_{A_5} F_{kn} = \frac{1}{4\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy \cdot \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} + O(N^{-rs\beta/9}).$$

Similarly, let  $A_6 := \{(x, y) \in A_4 : y_k \leq -\omega, y_n \geq \omega, |x_k - x_n| \leq \eta N^{(n+1)\gamma/s}\}$ , and we have

$$(5.51) \quad \int_{A_6} F_{kn} = \frac{1}{4\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy \cdot \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} + O(N^{-rs\beta/9}).$$

Thus, by (5.50) and (5.51) we conclude the proof.  $\square$

Note that Lemmas 5.6 and 5.7 imply

$$\int F_{kn} = \frac{n-1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy \cdot \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} + O(N^{-rs\beta/9}).$$

Together with Lemma 5.4 and (5.31), we have

$$(5.52) \quad \mathbb{E}(\text{Tr } f_\eta(H))^n = \frac{n-1}{2\pi^2} \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy \cdot \mathbb{E}(\text{Tr } f_\eta(H))^{n-2} + O(N^{-rs\beta/9}),$$

which completes the proof of Lemma 5.3(i).

5.4. *Proof of Lemma 5.3(ii).* Let  $f \in C^{1,r,s}(\mathbb{R})$  with  $r, s > 0$ , and without loss of generality we assume  $s \leq 1$ . We define  $\tilde{f}$  as in (3.4). Let  $\sigma = N^{-(\alpha+\beta)}$ , where we define  $\beta = sc_0/4$  instead in (5.5). Let  $\chi$  be as in Lemma 3.3 satisfying  $\chi(y) = 1$  for  $|y| \leq 1$ , and  $\chi(y) = 0$  for  $|y| \geq 2$ . An application of Lemma 3.3 gives

$$(5.53) \quad \mathbb{E}[\text{Tr } f_\eta(H)] = N \int \varphi_f(x + iy)(\underline{\mathbb{E}G(x + iy)} - m(x + iy)) \, dx \, dy =: \int \tilde{F},$$

where  $\varphi_f$  is defined as in (5.7). Note that (3.9) and (3.11) imply

$$|\underline{G(x + iy)} - m(x + iy)| < \frac{1}{N|y|}$$

uniformly in  $x \in \mathbb{R}, |y| \leq 1$ . Then we have

$$(5.54) \quad \left| \int_{|y| \leq \sigma} \tilde{F} \right| < \int_{|y| \leq \sigma} \left| \varphi_f(z) \cdot \frac{1}{y} \right| \, d^2z = O(N^{-rq_0\beta}),$$

where  $q_0 = q_0(s) = \frac{s}{2(1+s)} \geq \frac{s}{4}$ . Also, we have

$$(5.55) \quad \begin{aligned} \left| \int_{|y| \geq \sigma, |x| > 2 - \frac{\kappa}{2}} \tilde{F} \right| &< \int_{|y| \geq \sigma, |x| > 2 - \frac{\kappa}{2}} \left| \varphi_f(z) \cdot \frac{1}{y} \right| \, d^2z \\ &= O(N^{\beta-s\alpha/2}) = O(N^{-sc_0/2}). \end{aligned}$$

An analogue of (4.4) yields

$$(5.56) \quad \underline{\mathbb{E}G(x + iy)} - m(x + iy) = O(N^{(\alpha+\beta)-1-c_0(\alpha+\beta)})$$

uniformly in  $|x| \leq 2 - \frac{\kappa}{2}, |y| \geq \sigma$ , where the function  $c_0(\cdot)$  is defined as in (5.22). Thus,

$$(5.57) \quad \begin{aligned} \int_{|y| \geq \sigma, |x| \leq 2 - \frac{\kappa}{2}} \tilde{F} &= O(N^{(\alpha+\beta)-c_0(\alpha+\beta)} \cdot \eta) = O(N^{\beta-c_0(\alpha+\beta)}) \\ &= O(N^{-c_0/2}). \end{aligned}$$

Altogether we have (5.3).

5.5. *Remark on the general case.* Let us turn to Theorem 5.1. As in the one-dimensional case, we first show that

$$(5.58) \quad (\tilde{Z}(f_1), \dots, \tilde{Z}(f_p)) \xrightarrow{d} (Z(f_1), \dots, Z(f_p))$$

as  $N \rightarrow \infty$ , where  $\tilde{Z}(f_i) := \langle \text{Tr } f_i \left( \frac{H-E}{\eta} \right) \rangle$  for  $1 \leq i \leq p$ . In order to show (5.58), it suffices to compute  $\mathbb{E}[\tilde{Z}(f_{i_1}) \cdots \tilde{Z}(f_{i_n})]$ ,  $i_1, \dots, i_n \in \{1, 2, \dots, p\}$ , and this follows exactly the same way as we compute  $\mathbb{E}\langle \text{Tr } f_\eta(H) \rangle^n$ . Theorem 5.1 then follows from the estimate  $\mathbb{E}\hat{Z}(f_i) = O(N^{-rs^2c_0/16})$  for all  $i \in \{1, 2, \dots, p\}$ , which is Lemma 5.3(ii).

**6. Relaxing the moment condition.** In this section, we use a Green function comparison argument to pass from Theorems 4.1 and 5.1 to Theorems 2.2 and 2.3.

Recall  $G := G(E + i\eta) = (H - E - i\eta)^{-1}$ , and  $[\underline{G}] := \underline{G} - m(E + i\eta)$  with  $E, \eta$  defined in Theorem 2.2. Similarly, as in Section 4, we have a particular case of Theorem 2.2.

PROPOSITION 6.1. *Let  $\eta, E, H$  be as in Theorem 2.2. Then*

$$(6.1) \quad N^{1-\alpha}[\underline{G}] \xrightarrow{d} \mathcal{N}_{\mathbb{C}}\left(0, \frac{1}{2}\right)$$

as  $N \rightarrow \infty$ .

In this section, we only sketch a proof of Proposition 6.1, and the other results can be proved analogously. We begin with the following lemma.

LEMMA 6.2. *Fix  $m > 2$  and let  $X$  be a real random variable, with absolutely continuous law, satisfying*

$$(6.2) \quad \mathbb{E}X = 0, \quad \mathbb{E}X^2 = \sigma^2, \quad \mathbb{E}|X|^m \leq C_m$$

for some constant  $C_m > 0$ . Let  $\lambda > 2\sigma$ . Then there exists a real random variable  $Y$  that satisfies

$$(6.3) \quad \mathbb{E}Y = 0, \quad \mathbb{E}Y^2 = \sigma^2, \quad |Y| \leq \lambda, \quad \mathbb{P}(X \neq Y) \leq 2C_m\lambda^{-m}.$$

In particular,  $\mathbb{E}|Y|^m \leq 3C_m$ . Moreover, if  $m > 4$  and  $\sigma = 1$ , then there exists a real random variable  $Z$  matching the first four moments of  $Y$ , and satisfies  $|Z| \leq 6C_m$ .

The existence of  $Y$  is a slight modification of Lemma 7.6 in [12], and the construction of  $Z$  is contained in the proof of Theorem 2.5 in [12]; we omit further details.

The next lemma is an easy application of Lemma 6.2.

LEMMA 6.3. *Let  $H$  be a real symmetric Wigner matrix, whose entries have absolutely continuous law. Let  $c$  be as in Definition 2.1. Then there exists a real symmetric Wigner matrix  $H^{(1)}$  satisfying Definition 2.1 and*

$$(6.4) \quad \max_{i,j} \mathbb{P}(H_{ij}^{(1)} \neq H_{ij}) = O(N^{-2-c/4+\delta_{ij}}), \quad \max_{i,j} |H_{ij}^{(1)}| \leq N^{-\varepsilon},$$

where  $\varepsilon = \varepsilon(c) := \frac{c}{4(4+c)} > 0$ . Moreover, there exists a real symmetric Wigner matrix  $H^{(2)}$  satisfying Definition 3.2, such that for all  $i, j$ ,

$$(6.5) \quad \mathbb{E}(H_{ij}^{(2)})^k = \mathbb{E}(H_{ij}^{(1)})^k,$$

where  $1 \leq k \leq 4 - 2\delta_{ij}$ .

PROOF. Fix  $c, C > 0$  such that  $\mathbb{E}|\sqrt{N}H_{ij}|^{4+c-2\delta_{ij}} \leq C$  for all  $i, j$ . By using Lemma 6.2 with  $m := 4 + c - 2\delta_{ij}$ ,  $X := \sqrt{N}H_{ij}$ ,  $\lambda := N^{1/2-\varepsilon}$ ,  $C_m := C$ , we construct, for each  $H_{ij}$ , a random variable  $H_{ij}^{(1)} := N^{-1/2}Y$  such that the family  $\{H_{ij}^{(1)}\}_{i \leq j}$  is independent and

$$\begin{aligned} \mathbb{E}H_{ij}^{(1)} &= 0, & \mathbb{E}(H_{ij}^{(1)})^2 &= \mathbb{E}H_{ij}^2, & |H_{ij}^{(1)}| &\leq N^{-\varepsilon}, \\ \mathbb{P}(H_{ij} \neq H_{ij}^{(1)}) &\leq 2CN^{-2-c/4+\delta_{ij}}, \end{aligned}$$

and we also have

$$\mathbb{E}|\sqrt{N}H_{ij}^{(1)}|^{4+c-2\delta_{ij}} \leq 3C.$$

Hence, we have proved the existence of  $H^{(1)}$ .

For  $i < j$ , by using the second part of Lemma 6.2 on  $Y = \sqrt{N}H_{ij}^{(1)}$ , we construct, for each  $H_{ij}^{(1)}$ , a random variable  $H_{ij}^{(2)} := N^{-1/2}Z$  matching the first four moments of  $H_{ij}^{(1)}$ , and the family  $\{H_{ij}^{(2)}\}_{i < j}$  is independent. Moreover, we have the bound  $|\sqrt{N}H_{ij}^{(2)}| \leq 6C$ , which ensures  $\sqrt{N}H_{ij}^{(2)}$  has uniformly bounded moments of all order. Let us denote  $\zeta_i := \mathbb{E}|\sqrt{N}H_{ii}^{(1)}|^2$ . Then we can construct random variables  $H_{ii}^{(2)}$  such that  $\sqrt{N}H_{ii}^{(2)} \stackrel{d}{\sim} \mathcal{N}(0, \zeta_i)$  and the family  $\{H_{ij}^{(2)}\}_{i \leq j}$  is independent. This completes the proof.  $\square$

Now we look at Proposition 6.1.

PROOF OF PROPOSITION 6.1. Let  $H$  be as in Theorem 2.2. Note that it suffices to consider the case that the entries of  $H$  have absolutely continuous law. Otherwise, consider the matrix

$$H' := (1 - e^{-2N})^{1/2} \cdot H + e^{-N}V,$$

where  $V$  is a GOE matrix independent of  $H$ . Then  $H'$  also satisfies Definition 2.1. Let  $G' := (H' - E - i\eta)^{-1}$ , and  $[G'] := \underline{G}' - m(E + i\eta)$  with  $E, \eta$  defined in Theorem 2.2. The resolvent identity  $G' - G = G(H - H')G'$  implies

$$|[G'] - [G]| \prec e^{-N/2}.$$

We can then construct  $H^{(1)}$  and  $H^{(2)}$  from  $H$ , as in Lemma 6.3.

Let  $z := E + i\eta$ . We have already obtained from Proposition 4.2 that

$$(6.6) \quad N^{1-\alpha} \left( \frac{1}{N} \text{Tr} \frac{1}{H^{(2)} - z} - m(z) \right) \xrightarrow{d} \mathcal{N}_{\mathbb{C}} \left( 0, \frac{1}{2} \right),$$

and we need to show (6.6) holds with  $H^{(2)}$  replaced by  $H$ . We first compare the local spectral statistics of  $H^{(2)}$  and  $H^{(1)}$  using the Green function comparison

method from [15]; see also Section 4 of [9] for an overview. Fix a bijective ordering map on the index set of the independent matrix entries,

$$\phi : \{(i, j) : 1 \leq i \leq j \leq N\} \longrightarrow \{1, \dots, \gamma(N)\}, \quad \gamma(N) := \frac{N(N+1)}{2},$$

and we assume  $\phi(i, i) = i$  for  $i = 1, 2, \dots, N$ . Denote by  $H_\gamma$  the Wigner matrix whose matrix entries  $h_{ij} = H_{ij}^{(2)}$  if  $\phi(i, j) \leq \gamma$  and  $h_{ij} = H_{ij}^{(1)}$ ; otherwise; in particular  $H^{(2)} = H_0$  and  $H^{(1)} = H_{\gamma(N)}$ . Let  $F = F(x + iy)$  be a complex-valued, smooth, bounded function, with bounded derivatives. Then

$$\begin{aligned} & \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H^{(2)} - z} - m(z)\right)\right) \\ & - \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H^{(1)} - z} - m(z)\right)\right) \\ & = \sum_{\gamma=1}^{\gamma(N)} \left[ \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_{\gamma-1} - z} - m(z)\right)\right) \right. \\ & \quad \left. - \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_\gamma - z} - m(z)\right)\right) \right]. \end{aligned}$$

Now we focus on the term

$$\begin{aligned} & \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_{\gamma-1} - z} - m(z)\right)\right) - \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_\gamma - z} - m(z)\right)\right) \\ & =: \varepsilon_\gamma. \end{aligned}$$

For  $\gamma > N$ , let  $(i, j) = \phi^{-1}(\gamma)$ . Note that  $i < j$ , and we define

$$\tilde{V} = H_{ij}^{(2)} \Delta^{(ij)}, \quad \hat{V} = H_{ij}^{(1)} \Delta^{(ij)},$$

and recall from Section 4.2 that the matrix  $\Delta^{(ij)}$  satisfies  $\Delta_{kl}^{(ij)} = (\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})(1 + \delta_{ij})^{-1}$ . Denote  $Q := H_{\gamma-1} - \tilde{V}$ . Then  $H_{\gamma-1} = Q + \tilde{V}$ , and  $\tilde{V}$  is independent of  $Q$ . Also,  $H_\gamma = Q + \hat{V}$ . Define the Green functions

$$R := \frac{1}{Q - z}, \quad S := \frac{1}{H_{\gamma-1} - z}, \quad T := \frac{1}{H_\gamma - z},$$

and the resolvent expansion gives

$$(6.7) \quad S = R - R\tilde{V}R + (R\tilde{V})^2 - (R\tilde{V})^3R + (R\tilde{V})^4R - (R\tilde{V})^5S.$$

Since  $\tilde{V}$  has only at most two nonzero entries, when computing the  $(k, l)$  matrix entry of this matrix identity, each term is a finite sum involving matrix entries of  $S$  or  $R$  and  $H_{ij}^{(2)}$ , for example,  $(S\tilde{V}S)_{kl} = S_{ki}H_{ij}^{(2)}S_{jl} + S_{kj}H_{ji}^{(2)}S_{il}$ . Let  $\mathring{S} := N^{1-\alpha}(\underline{S} - m(z))$ , and  $\mathring{R}, \mathring{T}$  are defined analogously. Set  $\xi := \mathring{S} - \mathring{R}$ , and note that

one can easily obtain  $\xi$  from (6.7). Similarly,  $\mu := \mathring{T} - \mathring{R}$ , and we have an explicit expansion for

$$(6.8) \quad \varepsilon_\gamma = \mathbb{E}F(\mathring{S}) - \mathbb{E}F(\mathring{T}) = \mathbb{E}F(\mathring{R} + \xi) - \mathbb{E}F(\mathring{R} + \mu).$$

Now we expand  $F(\mathring{R} + \xi)$  and  $F(\mathring{R} + \mu)$  around  $\mathring{R}$  using the Taylor expansion. The detailed formulas of the expansion can be found in Section 4.1 of [9], and we omit them here. Since the first four moments of the entries of  $H^{(1)}$  and  $H^{(2)}$  coincide, the error is bounded by the terms with factors  $(H_{ij}^{(1)})^{m_1}(H_{ji}^{(1)})^{n_1}$  or  $(H_{ij}^{(2)})^{m_2}(H_{ji}^{(2)})^{n_2}$  in the expansion, where  $m_1 + n_1, m_2 + n_2 \geq 5$ . Since

$$\mathbb{E}|H_{ij}^{(1)}|^m = N^{-(m-4-c)\varepsilon} \cdot \mathbb{E}|H_{ij}^{(1)}|^{4+c} = O(N^{-2-c/2})$$

and

$$\mathbb{E}|H_{ij}^{(2)}|^m = O(N^{-\frac{m}{2}})$$

for all  $m \geq 5$ , a routine estimate shows the rest terms are bounded by  $O(N^{-2-c/2})$ . Thus, we have  $\varepsilon_\gamma = O(N^{-2-c/2})$  uniformly in  $\gamma > N$ . Similarly, one can show  $\varepsilon_\gamma = O(N^{-1-c/2})$  uniformly for  $\gamma \leq N$ . Thus,

$$\begin{aligned} & \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N}\text{Tr}\frac{1}{H^{(2)}-z} - m(z)\right)\right) - \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N}\text{Tr}\frac{1}{H^{(1)}-z} - m(z)\right)\right) \\ &= O(N^{-c/2}). \end{aligned}$$

The transition from  $H^{(1)}$  to  $H$  is immediate, since we have

$$(6.9) \quad \begin{aligned} & \left| \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N}\text{Tr}\frac{1}{H^{(1)}-z} - m(z)\right)\right) \right. \\ & \quad \left. - \mathbb{E}F\left(N^{1-\alpha}\left(\frac{1}{N}\text{Tr}\frac{1}{H-z} - m(z)\right)\right) \right| \\ & \leq O(\mathbb{P}(H^{(1)} \neq H)) \leq \sum_{i,j} \mathbb{P}(H_{ij}^{(1)} \neq H_{ij}) = O(N^{-c/4}). \end{aligned}$$

Note that by an approximation argument, for  $F$  in the above class,  $\lim_{N \rightarrow \infty} \mathbb{E}F(X_N) = \mathbb{E}F(X)$  is sufficient in showing  $X_N \xrightarrow{d} X$ . Thus, we have finished the proof.  $\square$

**7. The complex Hermitian case.** We conclude the paper with a remark on the complex Hermitian case. As mentioned in Remark 2.4, in the complex case we now have (2.9) and (2.10) instead of Theorems 2.2 and 2.3. We omit the complete statements of the results here. The proof in complex Hermitian case relies on the complex cumulant expansion, which we state in the lemma below, whose proof is omitted.

LEMMA 7.1 (Complex cumulant expansion). *Let  $h$  be a complex random variable with all its moments exist. The  $(p, q)$ -cumulant of  $h$  is defined as*

$$\mathcal{C}^{(p,q)}(h) := (-i)^{p+q} \cdot \left( \frac{\partial^{p+q}}{\partial s^p \partial t^q} \log \mathbb{E} e^{ish + i\bar{h}t} \right) \Big|_{s=t=0}.$$

Let  $f : \mathbb{C}^2 \rightarrow \mathbb{C}$  be a smooth function, and we denote its holomorphic 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

$$(7.1) \quad \mathbb{E} f(h, \bar{h}) \bar{h} = \sum_{p+q=0}^l \frac{1}{p!q!} \mathcal{C}^{(p,q+1)}(h) \mathbb{E} f^{(p,q)}(h, \bar{h}) + R_{l+1},$$

given all integrals in (7.1) exist. Here,  $R_{l+1}$  is the remainder term depending on  $f$  and  $h$ , and for any  $\tau > 0$ , we have the estimate

$$\begin{aligned} R_{l+1} &= O(1) \cdot \mathbb{E} |h|^{l+2} \cdot \mathbf{1}_{\{|h| > N^{\tau-1/2}\}} \cdot \max_{p+q=l+1} \|f^{(p,q)}(z, \bar{z})\|_\infty \\ &\quad + O(1) \cdot \mathbb{E} |h|^{l+2} \cdot \max_{p+q=l+1} \|f^{(p,q)}(z, \bar{z}) \cdot \mathbf{1}_{\{|z| \leq N^{\tau-1/2}\}}\|_\infty. \end{aligned}$$

Using Lemma 7.1, it is not hard to extend the argument of Sections 4–5 to the complex case. We sketch the required modifications.

Let  $H$  be a complex Wigner matrix. An argument analogous to Section 6 shows that it suffices to consider  $H$  satisfying Definition 3.2. Let  $G := G(E + i\eta) = (H - E - i\eta)^{-1}$  with  $E, \eta$  defined in Theorem 2.2. Let  $m, n \geq 1$ . Since  $H$  is complex hermitian, for any differentiable  $f = f(H)$  we set

$$(7.2) \quad \frac{\partial}{\partial H_{ij}} f(H) := \frac{d}{dt} \Big|_{t=0} f(H + t\tilde{\Delta}^{(ij)}),$$

where  $\tilde{\Delta}^{(ij)}$  denotes the matrix whose entries are zero everywhere except at the site  $(i, j)$  where it is one:  $\tilde{\Delta}_{kl}^{(ij)} = \delta_{ik}\delta_{jl}$ . Then by using Lemma 7.1 with  $h = H_{ij}$ , we have

$$\begin{aligned} z \mathbb{E} \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m &= \frac{1}{N} \sum_{i,j} \mathbb{E} \langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij} H_{ji} \\ (7.3) \quad &= \frac{1}{N^2} \sum_{i,j} \mathbb{E} \frac{\partial (\langle \langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \rangle G_{ij})}{\partial H_{ij}} + \hat{K} + \hat{L}, \end{aligned}$$

where  $\hat{K}$  and  $\hat{L}$  are defined analogously to  $K$  and  $L$  in (4.17). Note that

$$(7.4) \quad \frac{\partial G_{ij}}{\partial H_{kl}} = -G_{ik} G_{lj},$$

and by (7.3) we have

$$(7.5) \quad \begin{aligned} \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m &= \frac{1}{T} \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m+1} - \frac{1}{T} \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-1} \mathbb{E}\langle \underline{G} \rangle^2 \\ &+ \frac{m-1}{N^2 T} \mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^{m-2} \underline{G}^3 \\ &- \frac{\hat{K}}{T} - \frac{\hat{L}}{T} + \frac{n}{N^2 T} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} \underline{G} \underline{G}^* 2, \end{aligned}$$

where  $T = -z - 2\mathbb{E}\underline{G}$ . By a comparison of (7.5) and its real analogue (4.22), we see that the leading term is now halved. By estimating the subleading terms in a similar fashion, one can show that instead of (4.55), we have

$$\mathbb{E}\langle \underline{G}^* \rangle^n \langle \underline{G} \rangle^m = \frac{n}{4} N^{2\alpha-2} \mathbb{E}\langle \underline{G}^* \rangle^{n-1} \langle \underline{G} \rangle^{m-1} + O(N^{(m+n)(\alpha-1)-c_0}),$$

which agrees with our statement that we have an additional factor of 1/2 in the covariances.

## REFERENCES

- [1] ALTSHULER, B. L. and SHKLOVSKII, B. I. (1986). Repulsion of energy levels and the conductance of small metallic samples. *Zh. Eksp. Teor. Fiz. (Sov. Phys. JETP)* **91** 220.
- [2] BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. [MR2189081](#)
- [3] BENAYCH-GEORGES, F. and KNOWLES, A. Lectures on the local semicircle law for Wigner matrices. Preprint, [arXiv:1601.04055](#).
- [4] BOUTET DE MONVEL, A. and KHORUNZHY, A. (1999). Asymptotic distribution of smoothed eigenvalue density. I. Gaussian random matrices. *Random Oper. Stoch. Equ.* **7** 1–22.
- [5] BOUTET DE MONVEL, A. and KHORUNZHY, A. (1999). Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices. *Random Oper. Stoch. Equ.* **7** 149–168.
- [6] BREUER, J. and DUIJS, M. (2016). Universality of mesoscopic fluctuations for orthogonal polynomial ensembles. *Comm. Math. Phys.* **342** 491–531. [MR3459158](#)
- [7] DAVIES, E. B. (1995). The functional calculus. *J. Lond. Math. Soc.* (2) **52** 166–176. [MR1345723](#)
- [8] DUIJS, M. and JOHANSSON, K. On mesoscopic equilibrium for linear statistics in Dyson’s Brownian motion. Preprint, [arXiv:1312.4295](#).
- [9] ERDŐS, L. (2011). Universality of Wigner random matrices: A survey of recent results. *Russian Math. Surveys* **66** 67–198.
- [10] ERDŐS, L. and KNOWLES, A. (2015). The Altshuler–Shklovskii formulas for random band matrices I: The unimodular case. *Comm. Math. Phys.* **333** 1365–1416.
- [11] ERDŐS, L. and KNOWLES, A. (2015). The Altshuler–Shklovskii formulas for random band matrices II: The general case. *Ann. Henri Poincaré* **16** 709–799. [MR3311888](#)
- [12] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640.
- [13] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** no. 59, 58. [MR3068390](#)
- [14] ERDŐS, L. and YAU, H.-T. (2012). Universality of local spectral statistics of random matrices. *Bull. Amer. Math. Soc.* **49** 377–414.

- [15] ERDŐS, L., YAU, H.-T. and YIN, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.* **2** 15–81. [MR2847916](#)
- [16] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515.
- [17] LODHIA, A. and SIMM, N. J. Mesoscopic linear statistics of Wigner matrices. Preprint, [arXiv:1503.03533](#).
- [18] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840.

DEPARTEMENT MATHEMATIK  
ETH ZÜRICH  
8092 ZÜRICH  
SWITZERLAND  
E-MAIL: [yukun.he@math.ethz.ch](mailto:yukun.he@math.ethz.ch)  
[knowles@math.ethz.ch](mailto:knowles@math.ethz.ch)