



RESEARCH ARTICLE

Equipartition principle for Wigner matrices

Zhigang Bao¹, László Erdős² and Kevin Schnelli³

¹Hong Kong University of Science and Technology, Hong Kong, China; E-mail: mazgbao@ust.hk.

²IST Austria, Klosterneuburg, Austria; E-mail: lerdos@ist.ac.at.

³KTH Royal Institute of Technology, Stockholm, Sweden; E-mail: schnelli@kth.se.

Received: 16 August 2020; **Accepted:** 11 April 2021

2020 Mathematics Subject Classification: *Primary* – 60B20; *Secondary* – 82B10

Abstract

We prove that the energy of any eigenvector of a sum of several independent large Wigner matrices is equally distributed among these matrices with very high precision. This shows a particularly strong microcanonical form of the equipartition principle for quantum systems whose components are modelled by Wigner matrices.

1. Introduction

Equipartition of energy is a general principle in classical statistical physics stating that in an ergodic system at equilibrium, the total energy is shared equally among the elementary degrees of freedom. In quantum systems, equipartition breaks down at very low temperatures. Even at higher temperatures there is no general quantum counterpart of this principle, apart from the standard quantum virial theorem, which only relates the total kinetic energy to a certain derivative of the potential. Nevertheless, in some special cases this principle could be verified; see [4] and references therein for an extensive physics literature on the popular model of a single quantum particle in contact with a quantum heat bath consisting of infinitely many harmonic oscillators. In this paper we show that for Wigner random matrices – that is, for a mean-field quantum system with random quantum transition rates – a particularly strong microcanonical form of the quantum equipartition holds: it is valid separately for every eigenvector.

More precisely, suppose that the total Hamiltonian of a quantum system is represented by a sum of independent $N \times N$ Wigner matrices $H = H_1 + H_2 + \dots + H_k$, where each H_ℓ represents the Hamiltonian of a subsystem. Let $w = (w(1), \dots, w(N))^T \in \mathbb{C}^N$ be an ℓ^2 -normalised eigenvector of H with eigenvalue λ – that is, $Hw = \lambda w$. The eigenvalue λ is the total energy of w :

$$\lambda = E(w) := (w, Hw) = \sum_{\ell=1}^k (w, H_\ell w).$$

The energy of the ℓ th subsystem H_ℓ in the state given by w is $E_\ell(w) := (w, H_\ell w)$. Our main result, formulated precisely in Theorem 3.4, asserts that

$$E_\ell(w) \approx \frac{E(w)}{k}, \quad \forall \ell = 1, 2, \dots, k, \tag{1.1}$$

with very high precision and very high probability. In other words, the total energy is equally distributed among the k subsystems.

Fine properties of eigenvectors of large Wigner matrices have been extensively studied in recent years. They are *delocalised* – that is, $\max_i |w(i)| \leq N^{-1/2+\epsilon}$ for any fixed $\epsilon > 0$, with very high probability as N tends to infinity. Delocalisation follows directly from the optimal local law (see, e.g., [8], and [2] for an optimal rate). Moreover, the eigenvectors are asymptotically normal, in the sense that for any fixed deterministic vector $q \in \mathbb{C}^N$, the moments of $\sqrt{N}|(q, w)|$ coincide with those of the modulus of a standard Gaussian [5, 14, 17]. A multivariate extension involving the joint moments of several eigenvectors also holds [5]. Furthermore, *quantum unique ergodicity* is also valid, stating that

$$\sum_{i \in J} |w(i)|^2 \approx \frac{|J|}{N} \quad (1.2)$$

for any deterministic subset $J \subset \{1, 2, \dots, N\}$ [1, 5, 16]. More recently, quantum unique ergodicity with an optimal $N^{-1/2+\epsilon}$ speed of convergence was proven in [6] for any quadratic forms (w, Aw) with a general *deterministic* matrix A :

$$\left| (w, Aw) - \frac{1}{N} \text{Tr } A \right| \leq N^{-1/2+\epsilon} \quad (1.3)$$

with very high probability. Gaussian fluctuations around this quantum unique ergodicity limit,

$$\sqrt{N} \left[(w, Aw) - \frac{1}{N} \text{Tr } A \right] \implies \text{Normal}, \quad (1.4)$$

have also been shown in [7] for general A with full rank; see also [3] for Gaussianity around the limit in formula (1.2) for mesoscopic subsets, $N^\epsilon \leq |J| \leq N^{1-\epsilon}$. The key difficulty in all these latter results was to prove them *microcanonically* – that is, for each eigenvector; this required the sophisticated equilibration mechanism of Dyson Brownian motion. In contrast, the local law (see Theorem 4.1) directly implies the analogous results for a spectral projection on mesoscopic scale – for example,

$$\frac{1}{2N^\epsilon} \sum_{|\alpha-\alpha_0| \leq N^\epsilon} \sum_{i \in J} |w_\alpha(i)|^2 \approx \frac{|J|}{N}, \quad (1.5)$$

instead of formula (1.2) – involving an average over many eigenvectors w_α with eigenvalues λ_α near λ_{α_0} with a fixed α_0 . Here the eigenvalues λ_α are indexed in an increasing order, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$.

In all these previous results the eigenvector was tested against a specific *deterministic* observable; but in the equipartition relation (1.1) we consider the quadratic form of w with a *random* H_t that is far from being independent of w . Given the complicated dependence between w and H_t , it is somewhat surprising that the proof of formula (1.1) is simpler than those of formulas (1.2) and (1.3). In fact, despite this dependence, we can still directly handle $(w, H_t w)$ for an individual eigenvector – that is, we *do not need* to first establish a spectrally local-averaged version of formula (1.1) in the form

$$\frac{1}{2N^\epsilon} \sum_{|\alpha-\alpha_0| \leq N^\epsilon} (w_\alpha, H_t w_\alpha) \approx \frac{\lambda_{\alpha_0}}{k}$$

and then prove that $(w_\alpha, H_t w_\alpha)$ does not change much if the eigenvalue λ_α remains close to a fixed energy.

The main reason for the simple proof is algebraic. Consider $k = 2$ for simplicity. It turns out that the quadratic forms of $\mathcal{H} := H_1 - H_2$ are especially small due to a strong algebraic cancellation in the cumulant expansion. Once the smallness of $(w, \mathcal{H}w) = (w, H_1 w) - (w, H_2 w)$ is established, formula (1.1) follows from $\lambda = (w, H_1 w) + (w, H_2 w)$.

We remark that this algebraic cancellation holds only if both H_1 and H_2 are Wigner matrices. Equipartition analogous to formula (1.1) is also expected to hold for deformed Wigner matrices (i.e.,

when $\mathbb{E}H_l \neq 0$), for Wigner-type matrices (when the matrix elements of H_l are still independent but not identically distributed) and even for Hermitian random matrices with some nontrivial correlation among their matrix elements; however, the proof of these generalisations will be more complicated. Furthermore, Gaussian fluctuation around formula (1.1) in the spirit of formula (1.4) is also expected, but the proof would require a detailed analysis of Dyson Brownian motion as in [7].

To demonstrate the central simplifying role of \mathcal{H} , in the next section we first give the proof of formula (1.1) for $k = 2$ in the Gaussian case, where the mechanism is especially elementary. In this case we can even prove the Gaussian fluctuation of $(w, H_1 w)$ in the spirit of formula (1.4) in an elementary way. Then we introduce the general model and properly state our main result, Theorem 3.4, in section 3. After collecting some preliminaries from earlier papers in section 4, we will prove our main theorem starting in section 5 for the complex Hermitian case under the additional condition $\mathbb{E}h_{l,ij}^2 = 0$ on the entries of each matrix H_l . This condition is removed in section 9. The necessary modifications for the real symmetric case are presented in section 8.

2. A simple proof of formula (1.1) for the Gaussian case and $k = 2$

Assume we are given two independent Gaussian random matrices H_1 and H_2 of size $N \times N$ – that is, their entries are two sets of independent complex centered Gaussian random variables of variance $\frac{1}{2N}$ subject to the symmetry constraint $H_1 = H_1^*$ and $H_2 = H_2^*$. Then clearly the sum

$$H := H_1 + H_2 \tag{2.1}$$

also belongs to the standard Gaussian unitary ensemble. Denote by $(\lambda_\alpha)_\alpha$ the eigenvalues in ascending order of H and let $(w_\alpha)_\alpha$ be an associated normalised eigenbasis – that is, we have $(w_\alpha, H w_\beta) = \delta_{\alpha\beta} \lambda_\alpha$, for any choice of indices α and β .

Consider now the random variables

$$(w_\alpha, H_1 w_\beta) - \frac{\delta_{\alpha\beta}}{2} \lambda_\alpha.$$

We claim that for any N , these random variables are Gaussian.

Lemma 2.1. *For any choice of index α , the random variable*

$$(w_\alpha, H_1 w_\alpha) - \frac{1}{2} \lambda_\alpha \tag{2.2}$$

is a centered real Gaussian random variable with variance $\frac{1}{4N}$, for any N . Moreover, for any choice of indices α and β , with $\alpha \neq \beta$, the random variable

$$(w_\alpha, H_1 w_\beta) \tag{2.3}$$

is a centered complex Gaussian random variable of variance $\frac{1}{4N}$, for any N .

Proof. Introduce the auxiliary matrix

$$\mathcal{H} := H_1 - H_2, \tag{2.4}$$

whose entries are also independent centered Gaussian random variables, up to the symmetry constraint, with variance $\mathbb{E}|\mathcal{H}_{ij}|^2 = \frac{1}{N}$. A simple calculation then shows that $\mathbb{E}[H_{ij}\mathcal{H}_{ab}] = 0$, for all $i, j, a, b \in \llbracket 1, N \rrbracket$, and hence the matrices H and \mathcal{H} are independent. In particular, \mathcal{H} is independent from w_α and w_β , for any choice of α and β .

Observe now that we can write the random variables in expressions (2.2) and (2.3) as

$$(w_\alpha, H_1 w_\beta) - \frac{1}{2} \lambda_\alpha \delta_{\alpha\beta} = (w_\alpha, H_1 w_\beta) - \frac{1}{2} (w_\alpha, H_1 w_\beta) - \frac{1}{2} (w_\alpha, H_2 w_\beta) = \frac{1}{2} (w_\alpha, \mathcal{H} w_\beta). \tag{2.5}$$

Hence by the independence of H and \mathcal{H} , we conclude that $\frac{1}{2} w_\alpha^* \mathcal{H} w_\beta$ is a Gaussian random variable. Since $\mathbb{E} \mathcal{H}_{ij} = 0$, it follows that $\mathbb{E} w_\alpha^* \mathcal{H} w_\beta = 0$. Further, we have

$$\mathbb{E} |w_\alpha^* \mathcal{H} w_\beta|^2 = \sum_{ijab} \overline{\mathbb{E} w_\alpha(i) \mathcal{H}_{ij} w_\beta(j) w_\alpha(a) \mathcal{H}_{ab} w_\beta(b)} = \sum_{ijab} \frac{1}{N} \delta_{ia} \delta_{jb} |w_\alpha(a)|^2 |w_\beta(j)|^2 = \frac{1}{N},$$

where we used independence and the fact that the eigenvectors are ℓ^2 -normalised. The notation \sum_{ijab} means that we sum over all indices from 1 to N . This shows expressions (2.2) and (2.3). \square

Notation

The symbol $O(\cdot)$ stands for the standard big-O notation. We use c and C to denote positive finite constants that do not depend on the matrix size N . Their values may change from line to line. We use double brackets to denote index sets – that is, for $n_1, n_2 \in \mathbb{R}$, $\llbracket n_1, n_2 \rrbracket := [n_1, n_2] \cap \mathbb{Z}$. For vectors $v, w \in \mathbb{C}^N$, we write $v^* w = (v, w)$ for their scalar product. For an $N \times N$ matrix A , we denote by $\|A\|$ its operator norm and by $\|A\|_\infty := \max_{ij} |A_{ij}|$ its max-norm. We use $\langle A \rangle := \frac{1}{N} \sum_i A_{ii}$ to denote the normalised trace of an $N \times N$ matrix $A = (A_{ij})_{N,N}$.

3. Definitions and results

In this section we introduce the model and state our main result on equipartition.

Assumption 3.1. Fix an integer $k \geq 2$. Let $H_\iota := (h_{\iota,ij})$, $\iota = 1, 2, \dots, k$, be k independent complex Hermitian Wigner matrices of size $N \times N$ – that is, we assume that their entries are independent centered random variables, up to the symmetry constraints $h_{\iota,ij} = \overline{h_{\iota,ji}}$ – satisfying

$$\mathbb{E} |h_{\iota,ij}|^2 = \frac{1}{N}, \quad 1 \leq i, j \leq N, \quad \iota = 1, \dots, k, \tag{3.1}$$

and the families of random variables $\{h_{\iota,ij}\}$ have finite moments to all order – that is, for each $m \geq 3$ there is a positive constant C_m such that

$$\mathbb{E} \left| \sqrt{N} h_{\iota,ij} \right|^m \leq C_m, \quad m \geq 3, \quad \iota = 1, 2, \dots, k. \tag{3.2}$$

For the main part of the paper we assume that H_ι are complex Hermitian matrices. This assumption is only for simplicity of the presentation; our result holds and the proof also applies with minor changes to the real symmetric setup as well (see Remark 3.5).

Choose now k possibly N -dependent numbers $\sigma_\iota \geq 0$ such that

$$\sum_{\iota=1}^k \sigma_\iota^2 = 1, \tag{3.3}$$

and consider the random matrix

$$H := \sum_{\iota=1}^k \sigma_\iota H_\iota. \tag{3.4}$$

To present our results, we use the following definition of high-probability estimates:

Definition 3.2. Let $\mathcal{X} \equiv \mathcal{X}^{(N)}$ and $\mathcal{Y} \equiv \mathcal{Y}^{(N)}$ be two sequences of nonnegative random variables. We say that \mathcal{Y} stochastically dominates \mathcal{X} if, for all (small) $\epsilon > 0$ and (large) $D > 0$,

$$\mathbb{P}\left(\mathcal{X}^{(N)} > N^\epsilon \mathcal{Y}^{(N)}\right) \leq N^{-D} \tag{3.5}$$

for sufficiently large $N \geq N_0(\epsilon, D)$, and we write $\mathcal{X} < \mathcal{Y}$ or $\mathcal{X} = O_{<}(\mathcal{Y})$. When $\mathcal{X}^{(N)}$ and $\mathcal{Y}^{(N)}$ depend on a parameter $\nu \in \mathcal{V}$ (typically an index label or a spectral parameter), then $\mathcal{X}(\nu) < \mathcal{Y}(\nu)$, uniformly in $\nu \in \mathcal{V}$, means that the threshold $N_0(\epsilon, D)$ can be chosen independently of ν .

We often use the notation $<$ also for deterministic quantities, and then $\mathcal{X}^{(N)} \leq N^\epsilon \mathcal{Y}^{(N)}$ holds with probability 1. Stochastic domination has the following properties:

Lemma 3.3 (Proposition 6.5 [9]).

1. $X < Y$ and $Y < Z$ imply $X < Z$.
2. If $X_1 < Y_1$ and $X_2 < Y_2$, then $X_1 + X_2 < Y_1 + Y_2$ and $X_1 X_2 < Y_1 Y_2$.
3. If $X < Y$, $\mathbb{E}Y \geq N^{-c_1}$ and $|X| \leq N^{c_2}$ almost surely with fixed constants c_1 and c_2 , then we have $\mathbb{E}X < \mathbb{E}Y$.

Let $(\lambda_\alpha)_\alpha$ be the eigenvalues of the matrix H in ascending order and let $(w_\alpha)_\alpha$ be a basis of associated normalised eigenvectors. In this paper we are interested in estimating

$$w_\alpha^* H_\iota w_\beta - \sigma_\iota \lambda_\alpha \delta_{\alpha\beta}, \quad \iota = 1, \dots, k, \tag{3.6}$$

for any choice of $\alpha, \beta \in \llbracket 1, N \rrbracket$.

Theorem 3.4. Let H be given by formula (3.4), and assume that H_ι , $\iota = 1, \dots, k$, satisfy Assumption 3.1 and that σ_ι , $\iota = 1, \dots, k$, satisfy equation (3.3). Then

$$\left|w_\alpha^* H_\iota w_\beta - \sigma_\iota \lambda_\alpha \delta_{\alpha\beta}\right| < \frac{1}{\sqrt{N}} \tag{3.7}$$

for all $\alpha, \beta \in \llbracket 1, N \rrbracket$ and $\iota \in \llbracket 1, k \rrbracket$.

Remark 3.5. We formulated Theorem 3.4 for complex Hermitian Wigner matrices, but with some modifications our method and results carry over to the real symmetric case (see Theorem 8.2). The details are given in section 8.

We further remark that one may also consider a mixed-symmetry setup, where some H_ι s are complex Hermitian Wigner matrices and the remaining H_ι s are real symmetric Wigner matrices. The arguments in section 8 can be extended to such a setting, and formula (3.7) indeed holds under this setup as well.

4. Preliminaries

In this section we collect some essential tools used in the proof of Theorem 3.4. We start with the Green function of the random matrix H and the corresponding local laws.

4.1. Local law for the Green function and rigidity of eigenvalues

For any probability measure μ on \mathbb{R} , its Stieltjes transform is defined as

$$m_\mu(z) := \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x), \quad z \in \mathbb{C} \setminus \mathbb{R}. \tag{4.1}$$

We denote the Stieltjes transform of the standard semicircle law by $m_{sc}(z)$.

Let G denote the Green function or resolvent of H :

$$G(z) := \frac{1}{H - z}, \quad z \in \mathbb{C} \setminus \mathbb{R}. \tag{4.2}$$

We refer to $z = E + i\eta$ in definitions (4.1) and (4.2) as a spectral parameter. We denote by $m(z)$ the normalised trace of the Green function $G(z)$:

$$m(z) = \frac{1}{N} \text{Tr} G(z) = \langle G(z) \rangle, \quad z \in \mathbb{C} \setminus \mathbb{R}, \tag{4.3}$$

and note that by the spectral calculus $m(z)$ is the Stieltjes transform of the empirical eigenvalue distribution of H . Finally, we recall the deterministic estimate $\|G(z)\|_\infty \leq \|G(z)\| \leq |\eta|^{-1}$, with $\eta = \text{Im}z$.

We are interested for energies E in a neighbourhood of the support of the semicircle law – that is, $|E| < 2 + \varrho$, for some fixed $\varrho > 0$. Further, fix a small $\epsilon > 0$ and introduce the spectral domain

$$\mathcal{E} := \{z = E + i\eta \in \mathbb{C} : E \in [-2 - \varrho, 2 + \varrho], N^{-1+\epsilon} \leq |\eta| \leq 1\}. \tag{4.4}$$

For $z, z' \in \mathcal{E}$, let $\Psi(z, z')$ denote the deterministic control parameter

$$\Psi(z, z') := \frac{1}{\sqrt{N\eta_0}}, \quad \eta_0 = \min\{|\text{Im}z|, |\text{Im}z'|\}. \tag{4.5}$$

We use the convention $\Psi(z, z) \equiv \Psi(z)$.

Let γ_α be the α th N -quantile of the semicircle law – that is, γ_α is determined by

$$\int_{-2}^{\gamma_\alpha} \frac{1}{2\pi} \sqrt{4 - x^2} dx = \frac{\alpha - 1/2}{N}. \tag{4.6}$$

The quantile γ_α is often also referred to as the *classical location* of the eigenvalue λ_α .

One ingredient for our work is the following strong local law for the Green function and the eigenvalue rigidity estimate:

Theorem 4.1 (Theorem 2.1 [10], Theorem 2.3 [8]). *Let H be as in definition (3.4) satisfying Assumption 3.1. Then we have the uniform estimates*

$$|G_{ij}(z) - \delta_{ij}m_{sc}(z)| < \left(\frac{\text{Im}m_{sc}(z)}{N\eta}\right)^{1/2} + \frac{1}{N|\eta|} < \Psi(z), \quad |m(z) - m_{sc}(z)| < \Psi(z)^2, \tag{4.7}$$

for all $z = E + i\eta \in \mathcal{E}$.

Moreover, we have the eigenvalue rigidity estimate

$$|\lambda_\alpha - \gamma_\alpha| < \frac{1}{N^{2/3} \min\{\alpha, N - \alpha + 1\}^{1/3}}, \tag{4.8}$$

for all $\alpha \in \llbracket 1, N \rrbracket$.

4.2. Cumulant expansion

A second main tool in the proof of Theorem 3.4 are cumulant expansions, which were used, for example, in [13, 15] to study linear eigenvalue statistics of random matrices. For our purposes, the following version from [11, 12] is very suitable.

Lemma 4.2 (Lemma 7.1 [11], Lemma 2.4 [12]). Let h be a complex-valued random variable with finite moments. Let $\kappa^{(p,q)}$ be the (p, q) cumulant of h , which is defined as

$$\kappa^{(p,q)} := (-i)^{p+q} \left(\frac{\partial^{p+q}}{\partial s^p \partial t^q} \log \mathbb{E} e^{ish+it\bar{h}} \right) \Big|_{s,t=0}. \tag{4.9}$$

Set $f \in C^\infty(\mathbb{C}^2; \mathbb{C})$; then for any fixed $l \in \mathbb{N}$, we have

$$\mathbb{E} h f(h, \bar{h}) = \sum_{p+q=0}^l \frac{1}{p!q!} \kappa^{(p+1,q)} f^{(p,q)}(h, \bar{h}) + \Omega_{l+1}, \tag{4.10}$$

where

$$f^{(p,q)}(w_1, w_2) := \partial_{w_1}^p \partial_{w_2}^q f(w_1, w_2), \quad w_1, w_2 \in \mathbb{C},$$

and the error term Ω_{l+1} satisfies

$$\begin{aligned} |\Omega_{l+1}| &\leq C_l \mathbb{E} [|h|^{l+2}] \max_{p+q=l+1} \sup_{|w| \leq M} |f^{(p,q)}(w, \bar{w})| \\ &+ C_l \left[\mathbb{E} [|h|^{2l+4} \mathbb{1}_{|h|>M}] \mathbb{E} \left[\max_{p+q=l+1} \sup_{|w| \leq |h|} |f^{(p,q)}(w, \bar{w})|^2 \right] \right]^{1/2}, \end{aligned} \tag{4.11}$$

where $M > 0$ is an arbitrary cutoff.

We remark that Lemma 4.2 is a combination of [12, Lemma 2.4] and [11, Lemma 7.1]; the combinatoric part comes from [11] and the error estimate is taken from [12].

From definition (4.9), the first few cumulants of a complex random variable h are given by

$$\kappa^{(1,0)} = \mathbb{E} h, \quad \kappa^{(1,1)} = \mathbb{E} |h|^2 - |\mathbb{E} h|^2, \quad \kappa^{(2,0)} = \mathbb{E} h^2 - (\mathbb{E} h)^2,$$

and so on, with $\kappa^{(q,p)} = \overline{\kappa^{(p,q)}}$.

5. Proof of Theorem 3.4

The proof of Theorem 3.4 is based on an essentially optimal estimate on a distinguished observable we introduce in this section (see equation (5.2)). We are going to prove Theorem 3.4 for the case $k = 2$; the case of general k then follows easily by grouping all but one summand in definition (3.4) together and viewing it as a single Wigner matrix.

Generalising definition (2.4), we introduce the auxiliary matrix

$$\mathcal{H} := \sigma_2 H_1 - \sigma_1 H_2, \tag{5.1}$$

whose entries are independent centered random variables, up to the symmetry constraint, with variance $\mathbb{E} |\mathcal{H}_{ij}|^2 = \frac{1}{N} (\sigma_1^2 + \sigma_2^2) = \frac{1}{N}$ (see equation (3.3)). In order to prove Theorem 3.4, we derive a high moment estimate for observables of the form

$$\frac{1}{N} \text{Tr} \mathcal{H} \text{Im} G(z_1) \mathcal{H} \text{Im} G(z_2) = \langle \mathcal{H} \text{Im} G(z_1) \mathcal{H} \text{Im} G(z_2) \rangle, \quad z_1, z_2 \in \mathcal{E}, \tag{5.2}$$

where G denotes the Green function of H (see definition (4.2)) and the domain \mathcal{E} was defined in definition (4.4). The main technical result of this paper is the following proposition:

Proposition 5.1. *Under the assumptions of Theorem 3.4, we have the estimate*

$$\langle \mathcal{H}\text{Im}G(z_1)\mathcal{H}\text{Im}G(z_2) \rangle < 1, \tag{5.3}$$

uniformly in $z_1, z_2 \in \mathcal{E}$.

Remark 5.2. Using the deterministic bound $\|G(z)\| \leq \frac{1}{|\eta|}$ and the bounds $\|H_1\|, \|H_2\| < 1$, which follow from formula (4.8), we get the a priori bound

$$\langle \mathcal{H}\text{Im}G(z_1)\mathcal{H}\text{Im}G(z_2) \rangle < \frac{1}{|\text{Im}z_1||\text{Im}z_2|} < N^2 \tag{5.4}$$

on the spectral domain \mathcal{E} . Thus formula (5.3) is an improvement of two orders in N and gives the correct size, up to factors of N^ϵ .

The proof of Proposition 5.1 is postponed to section 7, and we next show instead how it implies Theorem 3.4.

Proof of Theorem 3.4. In order to link formula (5.3) to formula (3.7), we observe that by spectral decomposition we have

$$\langle \mathcal{H}\text{Im}G(z_1)\mathcal{H}\text{Im}G(z_2) \rangle = \frac{1}{N} \sum_{\alpha, \beta=1}^N |w_\alpha^* \mathcal{H}w_\beta|^2 \frac{\eta_1}{(\lambda_\alpha - E_1)^2 + \eta_1^2} \frac{\eta_2}{(\lambda_\beta - E_2)^2 + \eta_2^2}, \tag{5.5}$$

where $z_1 = E_1 + i\eta_1, z_2 = E_2 + i\eta_2, \eta_1 \neq 0, \eta_2 \neq 0$.

Fix now indices α, β and choose $E_1 = \lambda_\alpha$ and $E_2 = \lambda_\beta$, as well as $\eta_1 = \eta_2 = N^{-1+\epsilon}$ such that $z_1, z_2 \in \mathcal{E}$ with very high probability by formula (4.8). Then we obtain from the uniform bound in formula (5.3), combined with the representation (5.5), the estimate

$$|w_\alpha^* \mathcal{H}w_\beta|^2 < N\eta_1\eta_2 < N^{-1}, \tag{5.6}$$

for all $\alpha, \beta \in \llbracket 1, N \rrbracket$.

Next, similar to equation (2.5), we conclude by noticing that

$$\begin{aligned} w_\alpha^* H_1 w_\beta - \sigma_1 \lambda_\alpha \delta_{\alpha\beta} &= w_\alpha^* H_1 w_\beta - \sigma_1^2 w_\alpha^* H_1 w_\beta - \sigma_1 \sigma_2 w_\alpha^* H_2 w_\beta \\ &= \sigma_2^2 w_\alpha^* H_1 w_\beta - \sigma_1 \sigma_2 w_\alpha^* H_2 w_\beta \\ &= \sigma_2 w_\alpha^* \mathcal{H} w_\beta \\ &= O_{<} \left(\frac{1}{\sqrt{N}} \right), \end{aligned}$$

where we used formula (5.6). □

6. Computation of the expectation

In this section we compute the expectation of the observable $\langle \mathcal{H}\text{Im}G(z_1)\mathcal{H}\text{Im}G(z_2) \rangle$. Since this random variable is positive for $z_1, z_2 \in \mathbb{C}^+$, the expectation already indicates its correct size. Also, the estimation of the expectation unveils the cancellation mechanism from which Theorem 3.4 eventually results.

Lemma 6.1. *Set $z_1, z_2 \in \mathcal{E}$. Then*

$$\mathbb{E} \langle \mathcal{H}\text{Im}G(z_1)\mathcal{H}\text{Im}G(z_2) \rangle = \text{Im}m_{sc}(z_1)\text{Im}m_{sc}(z_2) + O_{<} \left(\frac{1}{\sqrt{N}} \right) + O_{<} \left(\Psi^2(z_1, z_2) \right), \tag{6.1}$$

where $\Psi(z_1, z_2)$ is defined in definition (4.5).

Proof. We start by noticing that it suffices to estimate

$$\mathcal{X}(z, z') := \langle \mathcal{H}G(z)\mathcal{H}G(z') \rangle \tag{6.2}$$

for $z = z_1, \bar{z}_1$ and $z' = z_2, \bar{z}_2$. Further introduce the shorthand notations

$$G \equiv G(z), \quad G' \equiv G(z'). \tag{6.3}$$

Moreover, note that we can write

$$\mathcal{X}(z, z') = \frac{1}{N} \sum_{ijab} \mathcal{H}_{ij} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab}, \tag{6.4}$$

where the matrix (Δ^{ij}) is defined to have entries $(\Delta^{ij})_{nm} = \delta_{in}\delta_{jm}$, or using rank 1 operators,

$$\Delta^{ij} := |e_i\rangle\langle e_j|, \tag{6.5}$$

where $(e_i)_i$ is the canonical basis in \mathbb{C}^N . Recall that \sum_{ijab} indicates a sum over all indices from 1 to N .

Our task is to compute

$$\mathbb{E}\mathcal{X}(z, z') = \frac{1}{N} \sum_{ijab} \mathbb{E} \left[\mathcal{H}_{ij} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right]. \tag{6.6}$$

For this we use the cumulant expansions of Lemma 4.2. To get started, we need more notation. Let $\kappa_{\iota,ij}^{(p,q)} = \overline{\kappa_{\iota,ji}^{(p,q)}}$ denote the cumulants of the matrix entries $h_{\iota,ij}$, $\iota = 1, 2$. We will for simplicity assume for the moment that $\mathbb{E}h_{\iota,ij}^2 = 0$; this condition can easily be relaxed (see section 9). Together with Assumption 3.1, this implies

$$\kappa_{\iota,ij}^{(1,0)} = \kappa_{\iota,ij}^{(0,1)} = 0, \quad \kappa_{\iota,ij}^{(1,1)} = \frac{1}{N}, \quad \kappa_{\iota,ij}^{(2,0)} = \kappa_{\iota,ij}^{(0,2)} = 0. \tag{6.7}$$

Further, from formula (3.2) we have the estimates

$$\left| \kappa_{\iota,ij}^{(p,q)} \right| \leq \frac{C_{p+q}}{N^{\frac{p+q}{2}}}, \quad p + q \geq 3. \tag{6.8}$$

Next, introduce the derivation operator

$$\mathcal{D}_{ji} := (\sigma_2 \partial_{1,ji} - \sigma_1 \partial_{2,ji}), \tag{6.9}$$

where $\partial_{\iota,ji} \equiv \frac{\partial}{\partial h_{\iota,ji}}$, $\iota = 1, 2$.

With these notations we have the computational rules

$$\mathcal{D}_{ji} \mathcal{H}_{ab} = \delta_{ja} \delta_{ib}, \tag{6.10}$$

where we used equation (3.3), and

$$\mathcal{D}_{ji} G(z) = -\sigma_2 \sigma_1 G(z) \Delta^{ji} G(z) + \sigma_1 \sigma_2 G(z) \Delta^{ji} G(z) = 0, \tag{6.11}$$

where we used the basic differential rule

$$\partial_{\iota,ji} G(z) = -G(z) \sigma_{\iota} \Delta^{ji} G(z), \quad \iota = 1, 2. \tag{6.12}$$

We will also require a higher-order analogue of \mathcal{D} : for $p, q \in \mathbb{N}$, define

$$\mathcal{D}_{ji}^{(p,q)} := \frac{1}{p!q!} N^{\frac{p+q+1}{2}} \left(\sigma_2 \kappa_{1,ji}^{(p,q+1)} \partial_{1,ji}^p \partial_{1,ij}^q - \sigma_1 \kappa_{2,ji}^{(p,q+1)} \partial_{2,ji}^p \partial_{2,ij}^q \right). \tag{6.13}$$

With this notation we have $\mathcal{D}_{ji} = \mathcal{D}_{ji}^{(1,0)}$ and record that

$$\mathcal{D}_{ji}^{(1,0)} G = 0, \quad \mathcal{D}_{ji}^{(0,1)} = 0, \tag{6.14}$$

where the first relation follows from equation (6.11) and the second follows from $\kappa^{(0,2)} = 0$ (see equation (6.7)). With the notation in definition (6.13), we next recall Lemma 4.2 to obtain the following cumulant-expansion lemma:

Lemma 6.2. *Fix indices i, j and integers d, d' . Let F be a monomial in the Green-function entries $(G_{nm}(z))_{nm}$, $(G_{nm}(z'))_{nm}$, and matrix entries $(\mathcal{H}_{nm})_{nm}$ of total degree d in the Green-function entries and total degree d' in \mathcal{H}_{nm} , where $d' \leq d$. Then for any fixed $l \in \mathbb{N}$,*

$$\mathbb{E}_{ij} \mathcal{H}_{ij} F = \sum_{p+q=1}^l N^{-\frac{p+q+1}{2}} \mathbb{E}_{ij} \mathcal{D}_{ji}^{(p,q)} F + \Omega_{l+1}(F), \tag{6.15}$$

where \mathbb{E}_{ij} denotes the expectation with respect to the random variables $h_{1,ij}$ and $h_{2,ij}$. The error term satisfies the bound

$$|\Omega_{l+1}(F)| < N^{-(l+2)/2}, \tag{6.16}$$

where the explicit constants depend on d and d' but are uniform in the matrix indices.

The proof of Lemma 6.2 is postponed to Appendix A. Lemma 6.2 has the following direct corollary, whose proof is also postponed to Appendix A:

Corollary 6.3. *Fix indices i, j . Let F be a monomial in the Green-function entries $(G_{nm}(z))_{nm}$, $(G_{nm}(z'))_{nm}$, and matrix entries $(\mathcal{H}_{nm})_{nm}$ of total degree d in Green-function entries and total degree d' in \mathcal{H}_{nm} , where $d' \leq d$. Then for any fixed $l \in \mathbb{N}$,*

$$\mathbb{E} \mathcal{H}_{ij} F = \sum_{p+q=1}^l N^{-\frac{p+q+1}{2}} \mathbb{E} \mathcal{D}_{ji}^{(p,q)} F + \mathbb{E} \Omega_{l+1}(F), \tag{6.17}$$

where the error term satisfies the bound

$$|\mathbb{E} \Omega_{l+1}(F)| < N^{-(l+2)/2}, \tag{6.18}$$

where the explicit constants depend on d and d' but are uniform in the matrix indices.

With Corollary 6.3 and the computational rules (6.10) and (6.14) at hand, we begin to compute the expectation of $\mathcal{X}(z, z')$:

$$\begin{aligned} \mathbb{E} \mathcal{X}(z, z') &= \frac{1}{N} \sum_{ijab} \mathbb{E} \mathcal{H}_{ij} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \\ &= \frac{1}{N} \sum_{p+q=1}^l -N^{\frac{p+q+1}{2}} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(p,q)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] + O_{<} \left(N^{-(l+4)/2} \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N^2} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(1,0)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] \\
 &\quad + \frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(p,q)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] + O_{<} \left(N^{(-l+4)/2} \right), \tag{6.19}
 \end{aligned}$$

where we used Corollary 6.3 together with formula (6.8) and power counting to estimate the error term from cutting the cumulant expansion at order l to be $N^{-1}N^4O_{<} \left(N^{-(l+2)/2} \right) = O_{<} \left(N^{(-l+4)/2} \right)$.

We first focus on the first term on the right side of equation (6.19). Using equations (6.10) and (6.11), we get

$$\begin{aligned}
 \frac{1}{N^2} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(1,0)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] &= \frac{1}{N^2} \sum_{ijab} \mathbb{E} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \delta_{ja} \delta_{ib} \\
 &\quad + \frac{1}{N^2} \sum_{ijab} \mathbb{E} \text{Tr} \left(\Delta^{ij} \left(\mathcal{D}_{ji} G \right) \Delta^{ab} G' \right) \mathcal{H}_{ab} \\
 &\quad + \frac{1}{N^2} \sum_{ijab} \mathbb{E} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} \left(\mathcal{D}_{ji} G' \right) \right) \mathcal{H}_{ab} \\
 &= \frac{1}{N^2} \sum_{ij} \mathbb{E} G_{jj} G'_{ii} = \mathbb{E} m(z) m(z'). \tag{6.20}
 \end{aligned}$$

Note that the only nonzero term is when \mathcal{D}_{ji} acts on \mathcal{H}_{ab} . By the local law in formula (4.7) and the deterministic estimate $|m(z)| \leq \frac{1}{|\eta|} \leq N$, together with Lemma 3.3(3), the first term on the right side of equation (6.19) is thus given by

$$\frac{1}{N^2} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(1,0)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] = m_{sc}(z) m_{sc}(z') + O_{<} \left(\Psi(z, z')^2 \right). \tag{6.21}$$

Consider next the second term on the right of equation (6.19). We are going to use yet another cumulant expansion with respect to \mathcal{H}_{ab} to exploit further cancellation based on equation (6.14). For this purpose we first note that if $\{a, b\} \neq \{i, j\}$ as sets, then

$$\mathcal{D}_{ji}^{(p,q)} \left(\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right) = \mathcal{H}_{ab} \mathcal{D}_{ji}^{(p,q)} \left(\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \right), \tag{6.22}$$

because then $\partial_{1,ij} \mathcal{H}_{ab} = \partial_{2,ij} \mathcal{H}_{ab} = 0$. If $\{a, b\} = \{i, j\}$, then by power counting using $|\mathcal{H}_{ab}| < 1$ and the boundedness of the Green-function entries, we can estimate

$$\left| \frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{ijab} \mathbb{1}_{\{\{a,b\}=\{i,j\}\}} \mathbb{E} \mathcal{D}_{ji}^{(p,q)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] \right| < \frac{1}{N} \frac{1}{N^{3/2}} N^2 < \frac{1}{\sqrt{N}},$$

where we tacitly used Lemma 3.3(3), together with Hölder’s inequality, the deterministic estimate $\|G(z)\| \leq |\eta|^{-1} \leq N$ and the moment bounds in formula (3.2). Hence, we have for the second term on the right side of equation (6.19) that

$$\begin{aligned}
 &\frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(p,q)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \\
 &= \frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{H}_{ab} \mathcal{D}_{ji}^{(p,q)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) + O_{<} \left(\frac{1}{\sqrt{N}} \right). \tag{6.23}
 \end{aligned}$$

Next, using a cumulant expansion to order l with respect to \mathcal{H}_{ab} , we get

$$\begin{aligned} & \frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ji}^{(p_1, q_1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \\ &= \frac{1}{N} \sum_{\substack{p_1+q_1=2 \\ p_2+q_2=1}}^l N^{-\frac{p_1+q_1+p_2+q_2+2}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(p_2, q_2)} \mathcal{D}_{ji}^{(p_1, q_1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) + O_{<} \left(N^{(-l+1)/2} \right). \end{aligned}$$

By equation (6.14), we see that the terms with $p_2 + q_2 = 1$ yield a zero contribution, so we have

$$\begin{aligned} & \frac{1}{N} \sum_{p+q=2}^l N^{-\frac{p+q+1}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ji}^{(p, q)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \\ &= \frac{1}{N} \sum_{\substack{p_1+q_1=2 \\ p_2+q_2=2}}^l N^{-\frac{p_1+q_1+p_2+q_2+2}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(p_2, q_2)} \mathcal{D}_{ji}^{(p_1, q_1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) + O_{<} \left(N^{(-l+1)/2} \right) \\ &= \frac{1}{N^4} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(1,1)} \mathcal{D}_{ji}^{(1,1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \\ &+ \frac{1}{N} \sum_{p_1+q_1+p_2+q_2 \geq 5}^l N^{-\frac{p_1+q_1+p_2+q_2+2}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(p_2, q_2)} \mathcal{D}_{ji}^{(p_1, q_1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \\ &+ O_{<} \left(N^{(-l+1)/2} \right). \tag{6.24} \end{aligned}$$

Using the local law for the Green-function entries in formula (4.7) and Lemma 3.3, we can easily bound the second term on the right side by

$$\left| \frac{1}{N} \sum_{p_1+q_1+p_2+q_2 \geq 5}^l N^{-\frac{p_1+q_1+p_2+q_2+2}{2}} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(p_2, q_2)} \mathcal{D}_{ji}^{(p_1, q_1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \right| < \frac{1}{\sqrt{N}}. \tag{6.25}$$

For the first term on the right side of equation (6.24), we observe that $\mathcal{D}_{ba}^{(1,1)} \mathcal{D}_{ji}^{(1,1)}$ contains four partial derivatives. When those act on the Green-function entries $\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) = G_{ja} G'_{bi}$, they create by equation (6.12) monomials of degree 6 in the Green-function entries. Assuming that a, b, i, j are all distinct, the four partial derivatives will create diagonal as well as off-diagonal Green-function entries when acting on $G_{ja} G'_{bi}$, since, for example, $\partial_{1,ba} G_{ja} = -\sigma_1 G_{jb} G_{aa}$. Note that the total number of off-diagonal entries does not decrease, and hence each resulting monomial contains at least two off-diagonal entries. In power counting we count diagonal entries as $O_{<}(1)$ and off-diagonal entries as $O_{<}(\Psi)$. If there are coincidences among the indices, we gain a factor $1/N$ in the summation for each coincidence, and hence those are negligible when compared with Ψ^2 . We thus have the estimate

$$\left| \frac{1}{N^4} \sum_{\{i,j\} \neq \{a,b\}} \mathbb{E} \mathcal{D}_{ba}^{(1,1)} \mathcal{D}_{ji}^{(1,1)} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \right| < (\Psi(z, z'))^2. \tag{6.26}$$

In sum, choosing $l \geq 5$, we get from formulas (6.21), (6.23), (6.25) and (6.26) that

$$\mathbb{E} \mathcal{X}(z, z') = \frac{1}{N} \mathbb{E} \text{Tr} \mathcal{H}(G(z)) \mathcal{H}(G(z')) = m_{sc}(z) m_{sc}(z') + O_{<} \left((\Psi(z, z'))^2 \right) + O_{<} \left(\frac{1}{\sqrt{N}} \right). \tag{6.27}$$

Using linear combinations, Lemma 6.1 follows directly from equation (6.27). □

7. Proof of Proposition 5.1

In the previous section we identified the expectation of $\langle \mathcal{H}G\mathcal{H}G' \rangle$ in Lemma 6.1. In this section we will control the higher moments of $\langle \mathcal{H}G\mathcal{H}G' \rangle$ to obtain a high-probability bound required to prove Proposition 5.1.

Proposition 7.1. *Under the assumptions of Theorem 3.4, we have*

$$\mathcal{X}(z, z') = m(z)m(z') + O_{\prec}(1) \tag{7.1}$$

uniformly in $z, z' \in \mathcal{E}$.

Proof. We rewrite \mathcal{X} as

$$\mathcal{X}(z, z') = \frac{1}{N} \sum_{ijab} \mathcal{H}_{ij} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} =: \frac{1}{N} \sum_{ijab} \mathcal{H}_{ij} X^{ijab} \mathcal{H}_{ab}, \tag{7.2}$$

where we introduced

$$X^{ijab} \equiv \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) = G_{ja} G'_{bi}. \tag{7.3}$$

As in section 6, we assume for the moment that $\mathbb{E}h_{\iota,ij}^2 = 0$, $\iota = 1, 2$, $i, j \in \llbracket 1, N \rrbracket$. This implies that $\kappa_{\iota,ij}^{(0,2)} = \kappa_{\iota,ij}^{(2,0)} = 0$ as well as $\mathcal{D}_{ji}^{(0,1)} \equiv 0$. We will explain in section 9 how this additional assumption can easily be dropped.

Next, we observe from equations (6.10) and (6.14) that

$$\frac{1}{N^2} \sum_{ijab} \mathcal{D}_{ji}^{(1,0)} X^{ijab} \mathcal{H}_{ab} = \frac{1}{N^2} \sum_{ij} G_{ii} G'_{jj} = m(z)m(z') := \varkappa(z, z'), \tag{7.4}$$

where we introduce the shorthand \varkappa . For $n, m \in \mathbb{N}$, define

$$P(n, m) := (\mathcal{X} - \varkappa)^n \overline{(\mathcal{X} - \varkappa)^m}. \tag{7.5}$$

Fix a (large) $D \in \mathbb{N}$, $z, z' \in \mathcal{E}$, and consider

$$\begin{aligned} \mathbb{E}P(D, D) &= \sum_{\mathbf{i}\mathbf{j}\mathbf{a}\mathbf{b}} \mathbb{E} \left[\prod_{n=1}^D \left(\frac{1}{N} \mathcal{H}_{i_n j_n} X^{i_n j_n a_n b_n} \mathcal{H}_{a_n b_n} - \frac{1}{N^2} \delta_{j_n a_n} \delta_{i_n b_n} G_{i_n i_n} G_{j_n j_n} \right) \right. \\ &\quad \times \left. \prod_{n=D+1}^{2D} \left(\frac{1}{N} \mathcal{H}_{j_n i_n} \overline{X^{i_n j_n a_n b_n}} \mathcal{H}_{b_n a_n} - \frac{1}{N^2} \delta_{j_n a_n} \delta_{i_n b_n} \overline{G_{i_n i_n} G_{j_n j_n}} \right) \right], \end{aligned} \tag{7.6}$$

where $\mathbf{i} = (i_1, i_2, \dots, i_{2D}) \in \llbracket 1, N \rrbracket^{2D}$, and similarly $\mathbf{j}, \mathbf{a}, \mathbf{b} \in \llbracket 1, N \rrbracket^{2D}$ are $8D$ free summation indices corresponding to $4D$ factors of \mathcal{H} s. In this expression, for each n , we call $\mathcal{H}_{i_n j_n}$ and $\mathcal{H}_{a_n b_n}$ ‘twins’.

We now successively use the cumulant expansions from Corollary 6.3 to expand the summands in equation (7.6) in all the factors of \mathcal{H} s. We start by expanding in the variable $\mathcal{H}_{i_1 j_1}$ to obtain

$$\begin{aligned} &\mathbb{E}[(\mathcal{X}(z, z') - \varkappa(z, z'))P(D - 1, D)] \\ &= \frac{1}{N} \sum_{p_1+q_1=1}^l \sum_{i_1 j_1 a_1 b_1} N^{-\frac{p_1+q_1+1}{2}} \mathbb{E} \left[\mathcal{D}_{j_1 i_1}^{(p_1, q_1)} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} P(D - 1, D) \right] \right] \\ &\quad - \mathbb{E}[\varkappa P(D - 1, D)] + \mathbb{E} \left[O_{\prec} \left(N^{-\frac{l+4}{2}} \right) P(D - 1, D) \right]. \end{aligned} \tag{7.7}$$

First, using the facts that $|\mathcal{X}| \leq \eta_0^{-2} \leq N^2$ and $|\kappa| \leq \eta_0^{-2} \leq N^2$, with $\eta_0 = \min\{|\text{Im}z|, |\text{Im}z'|\}$, the third term on the right of equation (7.7) is bounded as $O_{<}\left(N^{-\frac{l+4}{2}+(4D-2)}\right)$, and hence for $l \geq 10D$, that error term is bounded as $O_{<}\left(\left(\frac{1}{\sqrt{N}}\right)^{2D}\right)$. Here we also tacitly used, as we will do repeatedly later, Lemma 3.3(3) to justify the estimate. Second, in the first term on the right, for $p + q = 1$ we consider the derivation $\mathcal{D}_{j_1 i_1}^{(1,0)}$ (recall from equation (6.7) that $\mathcal{D}_{j_1 i_1}^{(0,1)} = 0$). When $\mathcal{D}_{j_1 i_1}^{(1,0)}$ acts on a Green function in $X^{i_1 j_1 a_1 b_1}$, we get a zero contribution thanks to equation (6.14). If $\mathcal{D}_{j_1 i_1}^{(1,0)}$ acts on its twin $\mathcal{H}_{a_1 b_1}$, we generate by equation (7.4) the term $\mathbb{E}[\kappa P(D - 1, D)]$, which will precisely cancel with the second term on the right side of equation (7.7).

Thus, choosing $l \geq 10D$, we have

$$\begin{aligned} \mathbb{E}P(D, D) &= \frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \mathbb{E} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right] \\ &\quad + \frac{1}{N} \sum_{p_1+q_1=2}^l \sum_{i_1 j_1 a_1 b_1} N^{-\frac{p_1+q_1+1}{2}} \mathbb{E} \left[\mathcal{D}_{j_1 i_1}^{(p_1, q_1)} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} P(D - 1, D) \right] \right] \\ &\quad + O_{<}\left(\left(\frac{1}{\sqrt{N}}\right)^{2D}\right). \end{aligned} \tag{7.8}$$

In the following we consider the terms on the right side of equation (7.8) separately.

First term on the right of equation (7.8).

We start with the term

$$\frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \mathbb{E} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right]. \tag{7.9}$$

When $\mathcal{D}_{j_1 i_1}^{(1,0)}$ acts on $P(D - 1, D)$, it acts either on a Green-function entry $G_{i_n j_n}$ or $G_{a_n b_n}$ or on $\mathcal{H}_{i_n j_n}$ or $\mathcal{H}_{a_n b_n}$, $n \in \llbracket 2, \dots, 2D \rrbracket$. In the former case we get by equation (6.14) a zero contribution, and in the latter case by equation (6.10) the number of free summation indices in $P(D - 1, D)$ gets reduced from $4(2D - 1)$ to $4(2D - 1) - 2$. Bearing this in mind, we expand the term in formula (7.9) using $\mathcal{H}_{a_1 b_1}$ to obtain, with $l \geq 10D$,

$$\begin{aligned} &\frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \mathbb{E} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right] \\ &= \frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \sum_{p_2+q_2=1}^l N^{-\frac{p_2+q_2+1}{2}} \mathbb{E} \left[\mathcal{D}_{b_1 a_1}^{(p_2, q_2)} \left[X^{i_1 j_1 a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right] \right] + O_{<}\left(\left(\frac{1}{\sqrt{N}}\right)^{2D}\right) \\ &= \frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \frac{1}{N} \mathbb{E} \left[X^{i_1 j_1 a_1 b_1} \mathcal{D}_{b_1 a_1}^{(1,0)} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right] \\ &\quad + \frac{1}{N^2} \sum_{i_1 j_1 a_1 b_1} \sum_{p_2+q_2=2}^l N^{-\frac{p_2+q_2+1}{2}} \mathbb{E} \left[\mathcal{D}_{b_1 a_1}^{(p_2, q_2)} \left[X^{i_1 j_1 a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D) \right] \right] \\ &\quad + O_{<}\left(\left(\frac{1}{\sqrt{N}}\right)^{2D}\right), \end{aligned} \tag{7.10}$$

where we used equation (6.14). For the first term on the right side in equation (7.10), the number of free summation indices in $\mathcal{D}_{b_1 a_1}^{(1,0)} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D)$ is $4(2D - 1) - 4$ by equations (6.10) and (6.14). Or, put differently, there are $2(2D - 1) - 2$ factors of \mathcal{H} s left that we can use in cumulant expansions.

For the second term on the right side of equation (7.10), either we get a zero contribution when $\mathcal{D}_{j_1 i_1}^{(1,0)}$ acts on a Green-function entry of $P(D - 1, D)$ or the number of free summation indices gets reduced by two if $\mathcal{D}_{j_1 i_1}^{(1,0)}$ acts on a factor of \mathcal{H} . For the higher derivative terms in $\mathcal{D}_{b_1 a_1}^{(p_2, q_2)}$, with $p_2 + q_2 \geq 2$, acting on $X^{i_1 j_1 a_1 b_1} \mathcal{D}_{j_1 i_1}^{(1,0)} P(D - 1, D)$, either the number of Green-function entries is increased by one for each derivative hitting a Green-function entry or the number of free summation indices is reduced by two for each derivative hitting a factor \mathcal{H} . We have now expanded the first term on the right of equation (7.8) in $\mathcal{H}_{i_1 j_1}$ and $\mathcal{H}_{a_1 b_1}$. Before we go on and expand the remaining \mathcal{H} s in $P(D - 1, D)$, we return to second term on the right of equation (7.8).

Second term on the right of equation (7.8).

Consider now the term

$$\frac{1}{N} \sum_{p_1+q_1=2}^l \sum_{i_1 j_1 a_1 b_1} N^{-\frac{p_1+q_1+1}{2}} \mathbb{E} \left[\mathcal{D}_{j_1 i_1}^{(p_1, q_1)} \left[X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} P(D - 1, D) \right] \right]. \tag{7.11}$$

Since $p_1 + q_1 \geq 2$, we do not have further cancellations from equation (6.14) in the expression

$$\mathbb{E} \left[\mathcal{D}_{j_1 i_1}^{(p_1, q_1)} X^{i_1 j_1 a_1 b_1} \mathcal{H}_{a_1 b_1} P(D - 1, D) \right].$$

If one of the derivatives in $\mathcal{D}_{j_1 i_1}^{(p_1, q_1)}$ acts on $\mathcal{H}_{a_1 b_1}$, the number of free summation indices is reduced by two; if none of the derivatives acts on $\mathcal{H}_{a_1 b_1}$, we use a cumulant expansion in $\mathcal{H}_{a_1 b_1}$ stopped at order $l \geq 10D$. The leading term containing $\mathcal{D}_{b_1 a_1}^{(1,0)}$ will then either give a zero contribution if it acts on any Green-function entry (by equation (6.14)) or reduce the number of free summation indices by two. For the terms containing $\mathcal{D}_{b_1 a_1}^{(p_2, q_2)}$, $p_2 + q_2 \geq 2$, we have no cancellation due to equation (6.14) but the number of free summation indices gets reduced by two for each derivative acting on a factor \mathcal{H} .

Classification of the expanded terms from equation (7.8).

We have now expanded all the terms on the right side of equation (7.8) using cumulant expansions and performed all the derivatives by Leibniz’s rule. The resulting terms can be classified by the number of collapses M of two free summation indices when $\mathcal{H}_{i_1 j_1}$ or $\mathcal{H}_{a_1 b_1}$ act on some other \mathcal{H} s (except their own twin), and the number of cumulant expansions L in total; the number of cumulant expansions L_1 starting from order 1 – that is, with $p_n + q_n \geq 1$; and the number of cumulant expansions L_2 starting from order 2 – that is, with $p_n + q_n \geq 2$. For the moment, either $L = 1$ or $L = 2$, with $L_1 + L_2 = L$. Because of the bounds $|G_{ij}(z)| < 1$, $\|G(z)\| \leq |\eta|^{-1} \leq N$ and Lemma 3.3(3), we may ignore the number of Green-function entries in the power counting, and do not keep track of them.

Iteration and continued expansion.

We have now fully expanded equation (7.8) in terms of $\mathcal{H}_{i_1 j_1}$ and $\mathcal{H}_{a_1 b_1}$. We will continue expanding in the remaining \mathcal{H} s while keeping track of the numbers M , L_1 and L_2 just introduced.

Pick now one of the resulting terms from before; if that term contains $\mathcal{H}_{i_2 j_2}$ and its twin $\mathcal{H}_{a_2 b_2}$, we expand first in $\mathcal{H}_{i_2 j_2}$. When $\mathcal{D}_{b_2 a_2}^{(0,1)}$ acts on $\mathcal{H}_{a_2 b_2}$, we get the cancellation with \varkappa from equation (7.4), so that we are left with a cumulant expansion with $p_2 + q_2 \geq 2$ only. In case the twin $\mathcal{H}_{a_2 b_2}$ is missing, we note that the number of free summation indices has already been reduced by two. If we pick a term that does not contain $\mathcal{H}_{i_2 j_2}$, we go on and expand in the next \mathcal{H} , $\mathcal{H}_{a_2 b_2}$ or, if missing, the next available \mathcal{H} . In this way we successively expand all factors \mathcal{H} , except those appearing in the error term of a cumulant expansion cut at order $l \geq 10D$.

Estimation of fully expanded terms.

A resulting fully expanded term containing no more \mathcal{H} s is then classified by the total number of collapses M of free summation indices, resulting from equation (6.10). The number of free summation indices in such a term is $8D - 2M$, whereas the number of total cumulant expansions L in that term is $4D - M$. As before, let L_1 be the number of cumulant expansions with $p_n + q_n = 1$ and let L_2 be the number of cumulant expansions with $p_n + q_n \geq 2$. Note that $L_2 = 4D - M - L_1$.

Hence a fully expanded term with given M, L_1 and L_2 gives a contribution to equation (7.6) bounded by

$$\begin{aligned} & \frac{1}{N^{2D}} N^{8D-2M} \left[\sum_{p_n+q_n=1}^l \left(\frac{1}{\sqrt{N}} \right)^{p_n+q_n+1} \right]^{L_1} \left[\sum_{p_n+q_n=2}^l \left(\frac{1}{\sqrt{N}} \right)^{p_n+q_n+1} \right]^{L_2} \\ & < N^{6-2M} N^{-L_1} N^{-\frac{3}{2}L_2} \\ & = \left(\frac{1}{\sqrt{N}} \right)^{-M+L_1}, \end{aligned} \tag{7.12}$$

where we used the facts that $|G_{ij}| < 1, \|G(z)\| \leq \frac{1}{|\eta|} \leq N$ with probability 1 and there are no more \mathcal{H} s in a fully expanded term, so that by Lemma 3.3 we get the first line. To obtain the second line we used the fact that D and $l \geq 10D$ are fixed numbers, and for the third line we used $L_2 = 4D - M - L_1$. Summarising, so far we have expanded equation (7.6) in all the factors \mathcal{H} and shown that each resulting fully expanded term with given M, L_1 and L_2 is bounded by equation (7.12).

We next claim that $L_1 \leq M$ for any fully expanded term. Indeed, if for some pair of indices $i_n j_n$ or $a_n b_n$ there is no collapse – meaning that the derivatives in $\mathcal{D}_{j_n i_n}^{(p_n, q_n)}$ (or $\mathcal{D}_{b_n a_n}^{(p_n, q_n)}$) exclusively act on Green-function entries – then we have due to equation (6.14) that $p_n + q_n \geq 2$ in order to get a nonzero contribution.

Thus we reach the maximum for $M = L_1$ in equation (7.12), and the term is stochastically dominated by 1 – that is, each fully expanded term is stochastically bounded by 1. The number of generated terms in the expansion is bounded by $(CD)^{cD}$ if we choose l to be proportional to D .

It follows that

$$\mathbb{E}P(D, D) = \mathbb{E}|\mathcal{X}(z, z') - \kappa(z, z')|^{2D} < 1 \tag{7.13}$$

for any D , and hence by Markov’s inequality we have

$$|\mathcal{X}(z, z') - \kappa(z, z')| < 1, \tag{7.14}$$

which was to be proven for fixed $z, z' \in \mathcal{E}$.

It remains to extend this bound to a uniform bound for all $z, z' \in \mathcal{E}$. Let $\mathcal{L} \subset \mathcal{E} \times \mathcal{E}$ be a lattice such that $|\mathcal{L}| = O(N^{10})$ and for any $(z, z') \in \mathcal{E} \times \mathcal{E}$ there is a $(z_0, z'_0) \in \mathcal{L}$ such that $|(z, z') - (z_0, z'_0)| = O(N^{-10})$. Since $\langle \mathcal{H}G(z)\mathcal{H}G(z') \rangle$ is Lipschitz continuous in (z, z') with constant bounded by $\eta_0^{-4} \leq N^4, \eta_0 = \min\{|\text{Im}z|, |\text{Im}z'|\}$, as follows from equation (6.12), the uniform estimate follows from a union bound over \mathcal{L} and formula (7.14). This concludes the proof of Proposition 7.1, modulo the assumption that $\mathbb{E}h_{i,ij}^2 = 0$. This condition can easily be removed, as we will show in section 9. \square

Remark 7.2. We can strengthen estimate (7.1) to

$$\begin{aligned} \mathcal{X}(z, z') &= m(z)m(z') + O_{<} \left(\Psi^2(z, z') \right) + O_{<} \left(\frac{1}{\sqrt{N}} \right) \\ &= m_{sc}(z)m_{sc}(z') + O_{<} \left(\Psi^2(z, z') \right) + O_{<} \left(\frac{1}{\sqrt{N}} \right). \end{aligned}$$

To establish this, one needs to count the number of off-diagonal Green-function entries generated along the expansion procedure and then use $|G_{ij}| < \Psi + \delta_{ij}$.

8. Real symmetric case

In this section, we outline how our results for the complex Hermitian setup carry over to the real symmetric one. We start with the analogue to Assumption 3.1.

Assumption 8.1. Fix an integer $k \geq 2$. We assume that $H_\iota := (h_{\iota,ij})$ are k independent real symmetric Wigner matrices of size $N \times N$ – that is, we assume that their entries are independent centered random variables, up to the symmetry constraints $h_{\iota,ij} = h_{\iota,ji}$, satisfying

$$\mathbb{E}h_{\iota,ij}^2 = \frac{1 + \delta_{ij}}{N}, \quad 1 \leq i, j \leq N, \quad \iota = 1, \dots, k, \tag{8.1}$$

and the families of random variables $\{h_{\iota,ij}\}$ have finite moments to all order – that is, they satisfy formula (3.2).

We then have the following result for the real symmetric case:

Theorem 8.2. Let H be given by definition (3.4) and assume that $H_\iota, \iota = 1, \dots, k$, satisfy Assumption 8.1 and that $\sigma_\iota, \iota = 1, \dots, k$, satisfy equation (3.3). Then

$$|w_\alpha^* H_\iota w_\beta - \sigma_\iota \lambda_\alpha \delta_{\alpha\beta}| < \frac{1}{\sqrt{N}}, \tag{8.2}$$

for all $\alpha, \beta \in \llbracket 1, N \rrbracket$ and $\iota \in \llbracket 1, k \rrbracket$.

Proof. In the following we sketch the proof of the theorem for $k = 2$. First we define the cumulants $\kappa_{\iota,ij}^{(p)} = \kappa_{\iota,ji}^{(p)}$ for the real random variables $h_{\iota,ij}$ as

$$\kappa_{\iota,ij}^{(p)} := (-i)^p \frac{\partial^p}{\partial s^p} \log \mathbb{E}e^{ish_{\iota,ij}} \Big|_{s=0} \tag{8.3}$$

and note that they satisfy estimate (6.8).

Second, we introduce the real symmetric analogue to $\mathcal{D}_{ji}^{(p,q)}$ by setting

$$\mathcal{D}_{ji}^{(p)} := \frac{1}{p!} N^{\frac{p+1}{2}} \left(\sigma_2 \kappa_{1,ji}^{(p+1)} \partial_{1,ji}^p - \sigma_1 \kappa_{2,ji}^{(p+1)} \partial_{2,ji}^p \right), \quad p \in \mathbb{N}. \tag{8.4}$$

With these definitions we obtain the following cumulant expansion formula for the real symmetric case: let F be a monomial in the Green-function entries and entries of \mathcal{H} as in Corollary 6.3; then for any $l \in \mathbb{N}$,

$$\mathbb{E}\mathcal{H}_{ij} F = \sum_{p=1}^l N^{-\frac{p+1}{2}} \mathbb{E}\mathcal{D}_{ji}^{(p)} F + \mathbb{E}\Omega_{l+1}(F), \tag{8.5}$$

where the error term satisfies the bound

$$|\mathbb{E}\Omega_{l+1}(F)| < N^{-(l+2)/2}. \tag{8.6}$$

Third, we recall the basic differentiation rule for the real symmetric setup:

$$\partial_{\iota,ji} G(z) = -G(z) \sigma_\iota \Delta^{ji} G(z) - G(z) \sigma_\iota \Delta^{ij} G(z), \quad \iota = 1, 2. \tag{8.7}$$

It is then easy to check that we have the computational rules

$$\mathcal{D}_{ji}^{(1)}G(z) = 0 \tag{8.8}$$

and

$$\mathcal{D}_{ji}^{(1)}\mathcal{H}_{ab} = \delta_{ja}\delta_{ib} + \delta_{ia}\delta_{jb}, \tag{8.9}$$

where $\mathcal{H}_{ab} = \sigma_2 H_1 - \sigma_1 H_2$ and where we used equations (3.3) and (8.1).

Armed with these definitions and rules, we turn to the computation of $\mathbb{E}\langle \mathcal{H}G(z)\mathcal{H}G(z') \rangle$. We follow the computation in section 6 up to equation (6.20), which now becomes

$$\begin{aligned} \frac{1}{N^2} \sum_{ijab} \mathbb{E} \mathcal{D}_{ji}^{(1)} \left[\text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{H}_{ab} \right] &= \frac{1}{N^2} \sum_{ijab} \mathbb{E} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) \mathcal{D}_{ji}^{(1)} \mathcal{H}_{ab} \\ &= \frac{1}{N^2} \sum_{ijab} \mathbb{E} \text{Tr} \left(\Delta^{ij} G \Delta^{ab} G' \right) (\delta_{ja}\delta_{ib} + \delta_{ia}\delta_{jb}) \\ &= \mathbb{E}m(z)m(z') + \frac{1}{N^2} \sum_{ij} \mathbb{E}G_{ij}G'_{ji} \\ &= \mathbb{E}m(z)m(z') + O_{<} \left(\Psi(z, z')^2 \right), \end{aligned} \tag{8.10}$$

where we used the local law for the Green function in formula (4.7) to get the last line, and the fact that Theorem 4.1 holds for real symmetric Wigner matrices as well. The only change was the addition of the error term $O_{<}(\Psi(z, z')^2)$ in equation (8.10). Following the computation in section 6 further, we conclude that Lemma 6.1 holds in the real symmetric setup too.

We move on to bound the higher moments of $\langle \mathcal{H}G\mathcal{H}G' \rangle$ following the arguments in section 7. Due to the modified rule (8.9) in the real setup, we redefine $\kappa(z, z')$ from equation (7.4) as

$$\kappa(z, z') := m(z)m(z') + \frac{1}{N^2} \sum_{ij} G_{ij}(z)G_{ji}(z'), \tag{8.11}$$

so that

$$\frac{1}{N^2} \sum_{ijab} \mathcal{D}_{ji}^{(1)} X^{ijab} \mathcal{H}_{ab} = \kappa(z, z') \tag{8.12}$$

holds with the adapted notation X^{ijab} given in equation (7.3). This modification of κ ensures that $\mathcal{X}(z, z') - \kappa(z, z')$ is a self-normalising quantity – that is, in the computation of $\mathbb{E}P(D, D)$, with P from definition (7.5), when some $\mathcal{H}_{i_n j_n}$ acts on its twin $\mathcal{H}_{a_n b_n}$ we get a zero contribution to $\mathbb{E}P(D, D)$ as in the complex Hermitian computation.

Yet if some $\mathcal{H}_{i_n j_n}$ acts on another \mathcal{H} which is not its own twin, then we get an additional contribution from the second term on the right side of equation (8.9) which is absent in the complex case. However, when this happens the number of free summation indices is reduced by two, and we continue to expand the resulting term in the same way as in the complex case. Thus the modified rule (8.9) produces more terms in the expansion of $\mathbb{E}P(D, D)$, but after all terms are fully expanded in the \mathcal{H} s, the sizes of the terms are estimated by the same power counting as in the complex Hermitian case. In this way one obtains

$$\langle \mathcal{H} \text{Im}G(z) \mathcal{H} \text{Im}G(z') \rangle = \text{Imm}(z)\text{Imm}(z') + O_{<}(1), \tag{8.13}$$

uniformly in $z, z' \in \mathcal{E}$, similar to Proposition 7.1. The proof of Theorem 8.2 is then concluded in the same way as in section 5. □

9. Complex case revisited

In this last section, we return to the complex Hermitian case. In the proof of Proposition 5.1 in section 7, we assumed for simplicity that $\mathbb{E}h_{i,j}^2 = \kappa_{i,j}^{(0,2)} = 0$. In this section we explain how this assumption can be removed. Even if $\kappa_{i,j}^{(0,2)} \neq 0$ and hence $\mathcal{D}_{ji}^{(0,1)} \neq 0$, we have

$$\mathcal{D}_{ji}^{(1,0)} G(z) = \mathcal{D}_{ji}^{(0,1)} G(z) = 0, \tag{9.1}$$

similar to equation (6.14). Further, equation (6.10) is modified as

$$\mathcal{D}_{ji}^{(0,1)} \mathcal{H}_{ab} = \delta_{ib} \delta_{ja}, \quad \mathcal{D}_{ji}^{(1,0)} \mathcal{H}_{ab} = \sigma_2^2 \kappa_{1,ji}^{(0,2)} \delta_{ia} \delta_{jb} + \sigma_1^2 \kappa_{2,ji}^{(0,2)} \delta_{ia} \delta_{bj}. \tag{9.2}$$

Since the cumulant expansions of Corollary 4.2 remain valid, it is straightforward to check that Proposition 7.1 holds true also when $\kappa_{i,j}^{(0,2)}$ do not necessarily vanish, after modifying the definition of \varkappa similarly to the real symmetric case to obtain self-normalising quantities in the moment bounds of $\mathcal{X} - \varkappa$. More precisely, redefining

$$\varkappa(z, z') := m(z)m(z') + \frac{1}{N} \sum_{ij} \left(\sigma_2^2 \kappa_{1,ji}^{(0,2)} G_{ij}(z)G_{ji}(z') + \sigma_1^2 \kappa_{2,ji}^{(0,2)} G_{ij}(z)G_{ji}(z') \right), \tag{9.3}$$

we find that

$$\frac{1}{N^2} \sum_{ijab} \left(\mathcal{D}_{ji}^{(0,1)} + \mathcal{D}_{ji}^{(1,0)} \right) X^{ijab} \mathcal{H}_{ab} = \varkappa(z, z'), \tag{9.4}$$

with X^{ijab} given in equation (7.3). We leave the further details aside. Finally, the proof of Theorem 3.4 from Proposition 7.1 remains unaffected by this modification.

Appendix A. Proof of Lemma 6.2

Proof. Fix the indices i and j . We write $F \equiv F(h_{1,ij}, h_{1,ji}, h_{2,ij}, h_{2,ji})$ to emphasise the explicit dependencies. From Lemma 4.2 and the definition of $\mathcal{D}_{ji}^{(p,q)}$ in definition (6.13) we directly obtain equation (6.15), where $\Omega_{l+1}(F)$ is the sum of two error terms $\Omega_{1,l+1}$ and $\Omega_{1,l+2}$, the first coming from cumulant expansion with respect to $h_{1,ij}$ and the second from expanding with respect to $h_{2,ij}$ in \mathcal{H}_{ij} . To bound the error term $\Omega_{1,l+1}$, we choose $M = N^{-1/4}$ in formula (4.11). Then together with the moment bounds in formula (3.2), for any (large) $D > 0$, we have

$$\begin{aligned} |\Omega_{1,l+1}| &\leq \frac{C_l}{N^{\frac{l+2}{2}}} \max_{p+q=l+1} \sup_{w \in \mathbb{C}, |w| \leq N^{-1/4}} \left| \partial_w^p \partial_{\bar{w}}^q F(w, \bar{w}, h_{2,ij}, h_{2,ji}) \right| \\ &\quad + \frac{C_{l,D}}{N^{D/2}} \left(\mathbb{E}_{ij} \left[\max_{p+q=l+1} \sup_{|w| \leq |h_{1,ij}|} \left| \partial_w^p \partial_{\bar{w}}^q F(w, \bar{w}, h_{2,ij}, h_{2,ji}) \right|^2 \right] \right)^{1/2} \end{aligned} \tag{A.1}$$

for N sufficiently large, where we used Hölder’s inequality and the moment assumption (3.2) to conclude that for any l and D ,

$$\mathbb{E}_{ij} |h_{1,ij}|^{l+2} \leq \frac{C_l}{N^{\frac{l+2}{2}}}, \quad \mathbb{E}_{ij} \left[|h_{1,ij}|^{2l+4} 1_{|h_{1,ij}| > N^{-1/4}} \right] \leq \frac{C_{l,D}}{N^{D/2}} \tag{A.2}$$

for N sufficiently large.

Consider next the Green-function entry $G_{ab} \equiv G_{ab}(z)$ for some fixed $z \in \mathcal{E}$ and some choice of indices a, b . We write $G_{ab} = G_{ab}(h_{1,ij}, h_{1,ji})$ in the following. Recall from the local law in formula (4.7) that $|G_{ab}(h_{1,ij}, h_{1,ji})| < \delta_{ab} + \Psi(z)$. Hence, using a Neumann expansion of the resolvent, we get

$$G_{ab}(w, \bar{w}) \tag{A.3}$$

$$= G_{ab}(h_{1,ij}, h_{1,ji}) - \text{Tr} \left[\Delta^{ba} G(h_{1,ij}, h_{1,ji}) \sigma_1 \left((w - h_{1,ij}) \Delta^{ij} + (\bar{w} - h_{1,ji}) \Delta^{ji} \right) G(w, \bar{w}) \right]. \tag{A.4}$$

Thus, denoting

$$\hat{\Lambda}_o := \sup_{w \in \mathbb{C}, |w| \leq N^{-1/4}} \max_{a,b} |G_{ab}(w, \bar{w})|,$$

we get from equation (A.3) and $|h_{1,ij}| < \frac{1}{\sqrt{N}}$ that

$$\hat{\Lambda}_o < 1 + \frac{1}{N^{1/4}} \hat{\Lambda}_o, \tag{A.5}$$

and hence $\hat{\Lambda}_o < 1$. Next observe that $\partial_{ij}^p \partial_{ji}^q F$ is a polynomial in the Green-function entries and the matrix entries of \mathcal{H} of degree at most $d + d' + l + 2$. Since F was a monomial, the number of monomial summands in $\partial_{ij}^p \partial_{ji}^q F$ depends on p, q and $d + d' + l + 2$, but is independent of N . Using the bounds $|\mathcal{H}_{ab}| < 1$ and $\hat{\Lambda}_o < 1$, we conclude that $\sup_{w \in \mathbb{C}, |w| \leq N^{-1/4}} \left| \partial_{ij}^p \partial_{ji}^q F(w, \bar{w}) \right| < 1$, for all p, q with $p + q \leq l + 1$. It follows that the first term on the right side of formula (A.1) is of order $O_{<}(N^{-(l+2)/2})$.

To control the second term on the right side of formula (A.1), we use once more that $\partial_{ij}^p \partial_{ji}^q F$ is a finite linear combination of monomials in the Green-function entries and the matrix entries of \mathcal{H} . The maximal number of Green-function entries occurring is $d + l + 2$, and estimating each factor by $\|G\|_\infty \leq \frac{1}{\eta_0} \leq N$, we get a contribution of order N^{d+l+2} from the Green-function entries. From the factors of \mathcal{H}_{ab} we use $|\mathcal{H}_{ab}| < 1$ and the moment bounds in formula (3.2) to conclude that

$$\left(\mathbb{E}_{ij} \left[\max_{p+q=l+1} \sup_{|w| \leq |h_{1,ij}|} \left| \partial_{ij}^p \partial_{ji}^q F(w, \bar{w}, h_{2,ij}, h_{2,ji}) \right|^2 \right] \right)^{1/2} < N^{d+l+2}. \tag{A.6}$$

Hence choosing D sufficiently large, the second term on the right side in formula (A.1) is bounded by $O_{<}(N^{-(l+2)/2})$. In sum, we have $|\Omega_{1,l+1}(F)| < N^{-(l+2)/2}$. We derive the corresponding bound on $\Omega_{2,l+1}(F)$ in the same way. □

Proof of Corollary 6.3. Corollary 6.3 follows from Lemma 6.2, together with an application of Lemma 3.3(3) using the estimates $|G_{ij}(z)| < 1, \|G(z)\| \leq |\eta|^{-1} \leq N$ and the moment assumptions in Assumption 3.1, combined with Hölder’s inequality. □

Acknowledgements. The first author is supported in part by Hong Kong RGC Grant GRF 16301519 and NSFC 11871425. The second author is supported in part by ERC Advanced Grant RANMAT 338804. The third author is supported in part by Swedish Research Council Grant VR-2017-05195 and the Knut and Alice Wallenberg Foundation.

Conflict of Interest: None.

References

[1] L. Benigni, ‘Fermionic eigenvector moment flow’, *Probab. Theory Related Fields* **179**(3) (2021), 733–775.
 [2] L. Benigni and P. Lopatto, ‘Optimal delocalization for generalized Wigner matrices’, Preprint, 2020, arXiv:2007.09585.
 [3] L. Benigni and P. Lopatto, ‘Fluctuations in local quantum unique ergodicity for generalized Wigner matrices’, Preprint, 2021, arXiv:2103.12013.

- [4] P. Białas, J. Spiechowicz and J. Luczka, ‘Quantum analogue of energy equipartition theorem’, *J. Phys. A* **52**(15) (2019), 15LT01.
- [5] P. Bourgade and H.-T. Yau, ‘The eigenvector moment flow and local quantum unique ergodicity’, *Comm. Math. Phys.* **350**(1) (2017), 231–278.
- [6] G. Cipolloni, L. Erdős and D. Schröder, ‘Eigenstate thermalization hypothesis for Wigner matrices’, Preprint, 2020, arXiv:2012.13215.
- [7] G. Cipolloni, L. Erdős and D. Schröder, ‘Normal fluctuation in quantum ergodicity for Wigner matrices’, Preprint, 2021, arXiv:2103.0673.
- [8] L. Erdős, A. Knowles, H.-T. Yau and J. Yin, ‘The local semicircle law for a general class of random matrices’, *Electron. J. Probab.* **18**(59) (2013), 1–58.
- [9] L. Erdős and H.-T. Yau, ‘A dynamical approach to random matrix theory’, *Courant Lecture Notes in Mathematics*, 28, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, (2017).
- [10] L. Erdős, H.-T. Yau and J. Yin, ‘Rigidity of eigenvalues of generalized Wigner matrices’, *Adv. Math.* **229**(3) (2012), 1435–1515.
- [11] Y. He and A. Knowles, ‘Mesoscopic eigenvalue statistics of Wigner matrices’, *Ann. Appl. Probab.* **27**(3) (2017), 1510–1550.
- [12] Y. He, A. Knowles and R. Rosenthal, ‘Isotropic self-consistent equations for mean-field random matrices’, *Probab. Theory Related Fields* **171**(1-2) (2018), 203–249.
- [13] A. Khorunzhy, B. Khoruzhenko and L. Pastur, ‘Asymptotic properties of large random matrices with independent entries’, *J. Math. Phys.* **37**(10) (1996), 5033–5060.
- [14] A. Knowles and J. Yin, ‘Eigenvector distribution of Wigner matrices’, *Probab. Theory Related Fields* **155**(3-4) (2013), 543–582.
- [15] A. Lytova and L. Pastur, ‘Central limit theorem for linear eigenvalue statistics of random matrices with independent entries’, *Ann. Probab.* **37** (2009), 1778–1840.
- [16] J. Marcinek and H.-T. Yau, ‘High dimensional normality of noisy eigenvectors’, Preprint, 2020, arXiv:2005.08425.
- [17] T. Tao and V. Vu, ‘Random matrices: Universal properties of eigenvectors’, *Random Matrices Theory Appl.* **1**(1) (2012), 1150001.