Bulk universality for generalized Wigner matrices

Bulk universality for generalized Wigner matrices

László Erdős[1]*, Horng-Tzer Yau[2]† and Jun Yin[2] ‡

Institute of Mathematics, University of Munich,
Theresienstr. 39, D-80333 Munich, Germany
lerdos@math.lmu.de [1]

Department of Mathematics, Harvard University
Cambridge MA 02138, USA
htyau@math.harvard.edu, jyin@math.harvard.edu [2]

Aug 14, 2010

Abstract

Consider $N \times N$ Hermitian or symmetric random matrices H where the distribution of the (i, j) matrix element is given by a probability measure ν_{ij} with a subexponential decay. Let σ_{ij}^2 be the variance for the probability measure ν_{ij} with the normalization property that $\sum_i \sigma_{ij}^2 = 1$ for all j. Under essentially the only condition that $c \le N\sigma_{ij}^2 \le c^{-1}$ for some constant $c > 0$, we prove that, in the limit $N \to \infty$, the eigenvalue spacing statistics of H in the bulk of the spectrum coincide with those of the Gaussian unitary or orthogonal ensemble (GUE or GOE). We also show that for band matrices with bandwidth M the local semicircle law holds to the energy scale M^{-1}.

AMS Subject Classification (2010): 15B52, 82B44

Keywords: Random band matrix, Local semicircle law, sine kernel.

*Partially supported by SFB-TR 12 Grant of the German Research Council
†Partially supported by NSF grants DMS-0602038, 0757425, 0804279
‡Partially supported by NSF grants DMS-100165

1 Introduction

One key universal quantity for random matrices is the eigenvalue gap distribution. Although the density of eigenvalues may depend on the specific model, the gap distribution or the short distance correlation function are believed to depend only on the symmetry class of the ensembles but are otherwise independent of the details of the distributions. There are two types of universality: the edge universality and the bulk universality. In this paper, we will focus on the bulk universality concerning the interior of the spectrum. The bulk universality was proved for very general classes of invariant ensembles (see, e.g. [3, 6, 7, 8, 9, 25, 26, 27] and references therein). For non-invariant ensembles, in particular for matrices with i.i.d. entries (Wigner matrices), the bulk universality was difficult to establish due to the lack of an explicit expression for the joint distribution of the eigenvalues.

The first rigorous partial result for bulk universality in the non-unitary case was given by Johansson [23] (see also Ben Arous and Péché [2] and the recent improvement [24]) stating that the bulk universality holds for Gaussian divisible *Hermitian* ensembles, i.e., Hermitian ensembles of the form

$$\widehat{H} + sV, \tag{1.1}$$

where $\widehat{H}$ is a Wigner matrix, V is an independent standard GUE matrix and s is a positive constant of order one. The restriction on Gaussian divisibility turned out to be very difficult to remove. In a series of papers [12, 13, 14, 17], we developed a new approach to prove the universality. The first step was to derive the local semicircle law, an estimate of the local eigenvalue density, down to energy scales containing around $\log N$ eigenvalues. Once such a strong form of the local semicircle law was obtained, the result of [23, 2] can be extended to a Gaussian convolution with variance only $s^2 \asymp N^{-1+\varepsilon}$. This tiny Gaussian component can then be removed via a reverse heat flow argument and this proves [17] the bulk universality for Hermitian ensembles provided that the distributions of the matrix elements are sufficiently differentiable.

The bulk universality for Hermitian ensembles was also proved later on by Tao and Vu [31] under the condition that the first four moments of the matrix elements match those of GUE, but without the differentiability assumption. The condition on the fourth moment was already removed in [31] by using the result for Gaussian divisible ensembles of [23, 2]; the third moment condition was then removed in [18] by using the result of [17].

The four moment theorem [31] is also valid for the symmetric ensembles, but the restriction on the matching of the first four moments cannot be weakened for the following reason. The key input to remove the fourth moment matching condition for the Hermitian case, the universality of the Gaussian divisible ensembles [23, 2], relied entirely on the asymptotic analysis of an *explicit* formula, closely related to a formula in Brézin-Hikami [5, 23], for the correlation functions of the eigenvalues for the *Hermitian ensembles* $\widehat{H} + sV$. Since similar formulas for symmetric matrices are very complicated, the corresponding result is not available and thus the matching of the fourth moment cannot be removed in this way. Although there is a proof [16] of universality for $s^2 \geq N^{-3/4}$ without using this formula, the main ingredient of that proof, establishing the uniqueness of the local equilibria of the Dyson Brownian motion, still heavily used explicit formulas related to GUE.

In [15] a completely different strategy was introduced based on a *local relaxation flow*, which locally behaves like a Dyson Brownian motion, but has a faster decay to equilibrium. This approach entirely eliminates explicit formulas and it gives a unified proof for the universality of *symmetric* and Hermitian Wigner matrices [15]. It was further generalized [19] to quaternion self-dual Wigner matrices and sample covariance matrices. The method not only applies to all these specific ensembles, but it also gives a conceptual interpretation that the occurrence of the universality is due to the relaxation to local equilibrium of the DBM. We remark that very recently the results of [31] were also extended to sample covariance matrices [33].

The main input of all these methods [17, 15, 19] and [31, 33] is an estimate of the local density of eigenvalues, the local semicircle law. This has been developed in the previous work on Wigner matrices [12, 13, 14], where the matrix elements were i.i.d. random variables. In this paper, we extend this method to random matrices with independent, but not necessarily identically distributed entries. If we denote the variance of the (i, j) entry of the matrix by σ_{ij}^2, our main interest is the case that σ_{ij} are not a constant but they satisfy the normalization condition $\sum_i \sigma_{ij}^2 = 1$ for all j. We will call such matrix ensembles *universal Wigner matrices*. For these ensembles Guionnet [21] and Anderson-Zeitouni [1] proved that the density of the eigenvalues converges to the Wigner semi-circle law. The simplest case is that of *generalized Wigner matrices*, where $N\sigma_{ij}^2$ is uniformly bounded from above and below by two fixed positive numbers. In this case, we prove the local semicircle law down to essentially the smallest possible energy scale N^{-1} (modulo $\log N$ factors). A much more difficult case is the *Wigner band matrices* where, roughly speaking, $\sigma_{ij}^2 = 0$ if $|i - j| > M$ for some $M < N$. In this case, we obtain the local semicircle law to the energy scale M^{-1}. We note that a certain three-dimensional version of Gaussian band matrices was considered by Disertori, Pinson and Spencer [10] using the supersymmetric method. They proved that the expectation of the density of eigenvalues is smooth and it coincides with the Wigner semicircle law.

With the local semicircle law proved up to the almost optimal scale, applying the method of [15, 19] leads to the identification of the correlation functions and the gap distribution for generalized Wigner matrices provided that the distribution of the matrix elements is continuous and satisfies the logarithmic Sobolev inequality. These additional assumptions can be removed if one can extend the Tao-Vu theorem [31] to generalized Wigner matrices. In Section 8, we will introduce an approach based on a Green's function comparison theorem, which states that the joint distributions of Green's functions of two ensembles at different energies with imaginary parts of order $1/N$ are identical provided that the first three moments of the two ensembles coincide and the fourth moments are close. Since local correlation functions and the gap distribution of the eigenvalues can be identified from Green's functions, it follows that the local correlation functions of these two ensembles are identical at the scale $1/N$. We can thus use this theorem to remove all continuity and logarithmic Sobolev inequality restrictions in our approach. In particular, this leads to the bulk universality for generalized Wigner matrices with the subexponential decay being essentially the only assumption on the probability law. We note that one major technical difficulty in [31], the level repulsion estimate, is not needed in the proof of the Green's function comparison theorem. It will be clear in Section 8 that, once the local semicircle law is established, the Green's function comparison theorem is a simple consequence of the standard resolvent perturbation theory.

2 Main results

We now state the main results of this paper. Since all our results hold for both Hermitian and symmetric ensembles, we will state the results for Hermitian matrices only. The modifications to the symmetric case are straightforward and they will be omitted. Let $H = (h_{ij})_{i,j=1}^N$ be an $N \times N$ Hermitian matrix where the matrix elements $h_{ij} = \overline{h}_{ji}$, $i \le j$, are independent random variables given by a probability measure ν_{ij} with mean zero and variance σ_{ij}^2. The variance of h_{ij} for $i > j$ is $\sigma_{ij}^2 = \mathbb{E}\,|h_{ij}|^2 = \sigma_{ji}^2$. For simplicity of the presentation, we assume that for any fixed $1 \le i < j \le N$, $\operatorname{Re} h_{ij}$ and $\operatorname{Im} h_{ij}$ are i.i.d. with distribution ω_{ij} i.e., $\nu_{ij} = \omega_{ij} \otimes \omega_{ij}$ in the sense that $\nu_{ij}(\mathrm{d}h) = \omega_{ij}(\mathrm{dRe}\, h)\omega_{ij}(\mathrm{dIm}\, h)$, but this assumption is not essential for the result. The distribution ν_{ij} and its variance σ_{ij}^2 may depend on N, but we suppress this in the notation. We assume that, for any j fixed,

$$\sum_i \sigma_{ij}^2 = 1\,. \tag{2.1}$$

Matrices with independent, zero mean entries and with the normalization condition (2.1) will be called *universal Wigner matrices*. For a forthcoming review on this matrix class, see [29], where the terminology of *random band matrices* was used.

Define C_{inf} and C_{sup} by

$$C_{inf} := \inf_{N,i,j}\{N\sigma_{ij}^2\} \le \sup_{N,i,j}\{N\sigma_{ij}^2\} =: C_{sup}. \tag{2.2}$$

Note that $C_{inf} = C_{sup}$ corresponds to the standard Wigner matrices and the condition $0 < C_{inf} \le C_{sup} < \infty$ defines more general Wigner matrices with comparable variances.

We will also consider an even more general case when σ_{ij} for different (i,j) indices are not comparable. The basic parameter of such matrices is the quantity

$$M := \frac{1}{\max_{ij}\sigma_{ij}^2}. \tag{2.3}$$

A special case is the band matrix, where $\sigma_{ij} = 0$ for $|i-j| > W$ with some parameter W. In this case, M and W are related by $M \le CW$.

Denote by $B := \{\sigma_{ij}^2\}_{i,j=1}^N$ the matrix of variances which is symmetric and doubly stochastic by (2.1), in particular it satisfies $-1 \le B \le 1$. Let the spectrum of B be supported in

$$\mathrm{Spec}(B) \subset [-1+\delta_-, 1-\delta_+] \cup \{1\} \tag{2.4}$$

with some nonnegative constants $\delta_\pm$. We will always have the following spectral assumption

$$\textit{1 is a simple eigenvalue of } B \textit{ and } \delta_- \textit{ is a positive constant, independent of } N. \tag{2.5}$$

The local semicircle law will be proven under this general condition, but the precision of the estimate near the spectral edge will also depend on δ_+ in an explicit way. For the orientation of the reader, we mention two special cases of universal Wigner matrices that provided the main motivation for our work.

Example 1. Generalized Wigner matrix. In this case we have

$$0 < C_{inf} \le C_{sup} < \infty, \tag{2.6}$$

and one can easily prove that 1 is a simple eigenvalue of B and (2.4) holds with

$$\delta_\pm \ge C_{inf}, \tag{2.7}$$

i.e., both δ_- and δ_+ are positive constants independent of N.

Example 2. Band matrix. The variances are given by

$$\sigma_{ij}^2 = W^{-1} f\Big(\frac{[i-j]_N}{W}\Big), \tag{2.8}$$

where $W \ge 1$, $f : \mathbb{R} \to \mathbb{R}_+$ is a bounded nonnegative symmetric function with $\int f = 1$ and we defined $[i-j]_N \in \mathbb{Z}$ by the property that $[i-j]_N \equiv i-j \mod N$ and $-\frac{1}{2}N < [i-j]_N \le \frac{1}{2}N$. Note that the relation (2.1) holds only asymptotically as $W \to \infty$ but this can be remedied by an irrelevant rescaling. If the bandwidth is comparable with N, then we also have to assume that $f(x)$ is supported in $|x| \le N/(2W)$.

The quantity M defined in (2.3) satisfies $M \le W/\|f\|_\infty$. In Appendix A we will show that (2.5) is satisfied for the choice of (2.8) if W is large enough.

The Stieltjes transform of the empirical eigenvalue distribution of H is given by

$$m(z) \equiv m_N(z) = \frac{1}{N}\mathrm{Tr}\ \frac{1}{H-z}\,, \quad z = E + i\eta. \tag{2.9}$$

We define the density of the semicircle law

$$\varrho_{sc}(x) \ := \ \frac{1}{2\pi}\sqrt{[4-x^2]_+}\,, \tag{2.10}$$

and, for $\mathrm{Im}\, z > 0$, its Stieltjes transform

$$m_{sc}(z) \ := \ \int_{\mathbb{R}} \frac{\varrho_{sc}(x)}{x-z}\,\mathrm{d}x\,. \tag{2.11}$$

The Stieltjes transform $m_{sc}(z) \equiv m_{sc}$ may also be characterized as the unique solution of

$$m_{sc} + \frac{1}{z + m_{sc}} \ = \ 0 \tag{2.12}$$

satisfying $\mathrm{Im}\, m_{sc}(z) > 0$ for $\mathrm{Im}\, z > 0$, i.e.,

$$m_{sc}(z) = \frac{-z + \sqrt{z^2-4}}{2}\,. \tag{2.13}$$

Here the square root function is chosen with a branch cut along the positive real axis. This guarantees that the imaginary part of m_{sc} is non-negative. The Wigner semicircle law states that $m_N(z) \to m_{sc}(z)$ for any fixed z provided that $\eta = \mathrm{Im}\, z > 0$ is independent of N. The local version of this result for universal Wigner matrices is the content of the following Theorem.

Theorem 2.1 (Local semicircle law) *Let $H = (h_{ij})$ be a Hermitian $N \times N$ random matrix where the matrix elements $h_{ij} = \overline{h}_{ji}$, $i \le j$, are independent random variables with $\mathbb{E}\, h_{ij} = 0$, $1 \le i, j \le N$, and assume that the variances $\sigma_{ij}^2 = \mathbb{E}|h_{ij}|^2$ satisfy (2.1), (2.4) and (2.5). Suppose that the distributions of the matrix elements have a uniformly subexponential decay in the sense that there exist constants α, $\beta > 0$, independent of N, such that for any $x > 0$ we have*

$$\mathbb{P}(|h_{ij}| \ge x^\alpha |\sigma_{ij}|) \le \beta e^{-x}. \tag{2.14}$$

Then there exist constants C_1, C_2, C and $c > 0$, depending only on α, β and δ_- in (2.5), such that for any $z = E + i\eta$ with $\eta = Im\, z > 0$, $|z| \le 10$ and

$$\frac{1}{\sqrt{M\eta}} \le \frac{\kappa^2}{(\log N)^{C_1}}, \tag{2.15}$$

where $\kappa := \big|\, |E| - 2\big|$, the Stieltjes transform of the empirical eigenvalue distribution of H satisfies

$$\mathbb{P}\left(|m_N(z) - m_{sc}(z)| \ge (\log N)^{C_2}\frac{1}{\sqrt{M\eta}\,\kappa}\right) \le CN^{-c(\log\log N)} \tag{2.16}$$

for sufficiently large N. In fact, the same result holds for the individual matrix elements of the Green's function $G_{ii}(z) = (H-z)^{-1}(i,i)$:

$$\mathbb{P}\left(\max_i |G_{ii}(z) - m_{sc}(z)| \geq (\log N)^{C_2} \frac{1}{\sqrt{M\eta}\,\kappa}\right) \leq C N^{-c(\log\log N)}. \tag{2.17}$$

We remark that once a local semicircle law is obtained on a scale essentially M^{-1}, it is straightforward to show that eigenvectors are delocalized on a scale at least of order M. The precise statement will be formulated in Corollary 3.2. We will prove Theorem 2.1 in Sections 3–5 by extending the approach of [12, 13, 14]. The main ingredients of this approach consist of i) a derivation of a self-consistent equation for the Green's function and ii) an induction on the scale of the imaginary part of the energy. The key novelty in this paper is that the self-consistent equation is formulated for the array of the diagonal elements of the Green's function $(G_{11}, G_{22}, \ldots, G_{NN})$ instead of the Stieltjes transform $m = \frac{1}{N}\mathrm{Tr}\, G = \frac{1}{N}\sum_i G_{ii}$ itself as in [14]. This yields for the first time a strong pointwise control on the diagonal elements G_{ii}, see (2.17).

The subexponential decay condition (2.14) can be weakened if we are not aiming at error estimates faster than any power law of N. This can be easily carried out and we will not pursue it in this paper.

Denote the eigenvalues of H by $\lambda_1, \ldots, \lambda_N$ and let $p_N(x_1, \ldots, x_N)$ be their (symmetric) probability density. For any $k = 1, 2, \ldots, N$, the k-point correlation function of the eigenvalues is defined by

$$p_N^{(k)}(x_1, x_2, \ldots, x_k) := \int_{\mathbb{R}^{N-k}} p_N(x_1, x_2, \ldots, x_N) \mathrm{d}x_{k+1} \ldots \mathrm{d}x_N. \tag{2.18}$$

We now state our main result concerning these correlation functions.

Theorem 2.2 (Universality for generalized Wigner matrices) *We consider a generalized hermitian Wigner matrix such that* (2.6) *holds. Assume that the distributions ν_{ij} of the (i,j) matrix elements have a uniformly subexponential decay in the sense of* (2.14)*. Suppose that the real and imaginary parts of h_{ij} are i.i.d., distributed according to ω_{ij}, i.e., $\nu_{ij}(\mathrm{d}h) = \omega_{ij}(\mathrm{d}Im\, h)\omega_{ij}(\mathrm{d}Reh)$. Let $m_k(i,j) = \int x^k \mathrm{d}\omega_{ij}(x)$, $1 \leq k \leq 4$, denote the k-th moment of ω_{ij} ($m_1 = 0$). Suppose that*

$$\inf_N \min_{1\leq i,j\leq N} \left\{ \frac{m_4(i,j)}{(m_2(i,j))^2} - \frac{(m_3(i,j))^2}{(m_2(i,j))^3} \right\} > 1, \tag{2.19}$$

then, for any $k \geq 1$ and for any compactly supported continuous test function $O : \mathbb{R}^k \to \mathbb{R}$, we have

$$\begin{aligned} \lim_{b\to 0} \lim_{N\to\infty} \frac{1}{2b} \int_{E-b}^{E+b} \mathrm{d}E' \int_{\mathbb{R}^k} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_k\, O(\alpha_1, \ldots, \alpha_k) \\ \times \frac{1}{\varrho_{sc}(E)^k} \Big(p_N^{(k)} - p_{GUE,N}^{(k)}\Big)\Big(E' + \frac{\alpha_1}{N\varrho_{sc}(E)}, \ldots, E' + \frac{\alpha_k}{N\varrho_{sc}(E)}\Big) = 0, \end{aligned} \tag{2.20}$$

where $p_{GUE,N}^{(k)}$ is the k-point correlation function of the GUE ensemble. The same statement holds for generalized symmetric Wigner matrices, with GOE replacing the GUE ensemble.

The limiting correlation functions of the GUE ensemble are given by the sine kernel

$$\frac{1}{\varrho_{sc}(E)^k} p_{GUE,N}^{(k)}\Big(E + \frac{\alpha_1}{N\varrho_{sc}(E)}, \ldots, E + \frac{\alpha_k}{N\varrho_{sc}(E)}\Big) \to \det\{K(\alpha_i - \alpha_j)\}_{i,j=1}^k, \qquad K(x) = \frac{\sin \pi x}{\pi x},$$

and similar universal formula is available for the limiting gap distribution.

Remark: The quantity in the bracket in (2.19) is always greater or equal to 1 for any real distribution with mean zero, which can be obtained by

$$m_3^2 = \big[\int x^3 \mathrm{d}\omega\big]^2 = \big[\int x(x^2-m_2)\mathrm{d}\omega\big]^2 \le \big[\int x^2\mathrm{d}\omega\big]\big[\int (x^2-m_2)^2\mathrm{d}\omega\big] = m_2(m_4-m_2^2)$$

and it is exactly 1 if the distribution is supported on two points. For example, if ω_{ij} is a rescaling of a fixed distribution $\widetilde{\omega}$ with variance $\frac{1}{2}$, i.e. $\omega_{ij}(x)\mathrm{d}x = \sigma_{ij}^{-1}\widetilde{\omega}(x/\sigma_{ij})\mathrm{d}x$, then condition (2.19) is satisfied under (2.6), as long as the support of $\widetilde{\omega}$ consists of at least three points. The case of a Bernoulli-type distribution supported on two points require a separate argument and it will be treated in the forthcoming paper [20].

We now state our main comparison theorem for matrix elements of Green's functions of two Wigner ensembles. As in the paper [31], we assume conditions on four moments. It will lead quickly to Theorem 6.4 stating that the correlation functions of eigenvalues of two matrix ensembles are identical up to scale $1/N$ provided that the first four moments of all matrix elements of these two ensembles are almost identical. Here we do not assume that the real and imaginary parts are i.i.d., hence the k-th moment of h_{ij} is understood as the collection of numbers $\int \bar{h}^s h^{k-s}\nu_{ij}(\mathrm{d}h)$, $s = 0, 1, 2, \ldots, k$. The main result in [31] compares the joint distribution of individual eigenvalues — which is not covered by our Theorem 2.3 — but it does not address directly the matrix elements of Green's functions. The key input for both theorems is the local semicircle law on the almost optimal scale $N^{-1+\varepsilon}$. The eigenvalue perturbation used in [31] requires certain estimates on the eigenvalue level repulsion; the proof of Theorem 2.3 is a straightforward resolvent perturbation theory.

Theorem 2.3 (Green's function comparison) *Suppose that we have two generalized $N \times N$ Wigner matrices, $H^{(v)}$ and $H^{(w)}$, with matrix elements h_{ij} given by the random variables $N^{-1/2}v_{ij}$ and $N^{-1/2}w_{ij}$, respectively, with v_{ij} and w_{ij} satisfying the uniform subexponential decay condition*

$$\mathbb{P}\big(|v_{ij}| \ge x^\alpha\big) \le \beta e^{-x}, \qquad \mathbb{P}\big(|w_{ij}| \ge x^\alpha\big) \le \beta e^{-x},$$

with some $\alpha, \beta > 0$. Fix a bijective ordering map on the index set of the independent matrix elements,

$$\phi : \{(i,j) : 1 \le i \le j \le N\} \to \Big\{1, \ldots, \gamma(N)\Big\}, \qquad \gamma(N) := \frac{N(N+1)}{2},$$

and denote by H_γ the generalized Wigner matrix whose matrix elements h_{ij} follow the v-distribution if $\phi(i,j) \le \gamma$ and they follow the w-distribution otherwise; in particular $H^{(v)} = H_0$ and $H^{(w)} = H_{\gamma(N)}$. Let $\kappa > 0$ be arbitrary and suppose that, for any small parameter $\tau > 0$ and for any $y \ge N^{-1+\tau}$, we have the following estimate on the diagonal elements of the resolvent

$$\mathbb{P}\left(\max_{0\le\gamma\le\gamma(N)} \max_{1\le k\le N} \max_{|E|\le 2-\kappa} \left|\left(\frac{1}{H_\gamma - E - iy}\right)_{kk}\right| \le N^{2\tau}\right) \ge 1 - CN^{-c\log\log N} \tag{2.21}$$

with some constants C, c depending only on τ, κ. Moreover, we assume that the first three moments of v_{ij} and w_{ij} are the same, i.e.

$$\mathbb{E}\bar{v}_{ij}^s v_{ij}^u = \mathbb{E}\bar{w}_{ij}^s w_{ij}^u, \qquad 0 \le s+u \le 3,$$

and the difference between the fourth moments of v_{ij} and w_{ij} is much less than 1, say

$$\big|\mathbb{E}\bar{v}_{ij}^s v_{ij}^{4-s} - \mathbb{E}\bar{w}_{ij}^s w_{ij}^{4-s}\big| \le N^{-\delta}, \qquad s = 0, 1, 2, 3, 4, \tag{2.22}$$

for some given $\delta > 0$. Let $\varepsilon > 0$ be arbitrary and choose an η with $N^{-1-\varepsilon} \le \eta \le N^{-1}$. For any sequence of positive integers $k_1, \dots, k_n$, set complex parameters $z_j^m = E_j^m \pm i\eta$, $j = 1, \dots k_m$, $m = 1, \dots, n$, with $|E_j^m| \le 2 - 2\kappa$ and with an arbitrary choice of the $\pm$ signs. Let $G^{(v)}(z) = (H^{(v)} - z)^{-1}$ denote the resolvent and let $F(x_1, \dots, x_n)$ be a function such that for any multi-index $\alpha = (\alpha_1, \dots, \alpha_n)$ with $1 \le |\alpha| \le 5$ and for any $\varepsilon' > 0$ sufficiently small, we have

$$\max\left\{ |\partial^\alpha F(x_1, \dots, x_n)| : \max_j |x_j| \le N^{\varepsilon'} \right\} \le N^{C_0 \varepsilon'} \tag{2.23}$$

and

$$\max\left\{ |\partial^\alpha F(x_1, \dots, x_n)| : \max_j |x_j| \le N^{2} \right\} \le N^{C_0} \tag{2.24}$$

for some constant C_0.

Then, there is a constant C_1, depending on α, β, $\sum_m k_m$ and C_0 such that for any η with $N^{-1-\varepsilon} \le \eta \le N^{-1}$ and for any choices of the signs in the imaginary part of z_j^m, we have

$$\begin{aligned} \left| \mathbb{E} F\left(\frac{1}{N^{k_1}} Tr\left[\prod_{j=1}^{k_1} G^{(v)}(z_j^1) \right], \dots, \frac{1}{N^{k_n}} Tr\left[\prod_{j=1}^{k_n} G^{(v)}(z_j^n) \right] \right) - \mathbb{E} F\left(G^{(v)} \to G^{(w)} \right) \right| \\ \le C_1 N^{-1/2 + C_1\varepsilon} + C_1 N^{-\delta + C_1 \varepsilon}, \end{aligned} \tag{2.25}$$

where the arguments of F in the second term are changed from the Green's functions of $H^{(v)}$ to $H^{(w)}$ and all other parameters remain unchanged.

Remark 1: We formulated Theorem 2.3 for functions of traces of monomials of the Green's function because this is the form we need in the application. However, the result (and the proof we are going to present) holds directly for matrix elements of monomials of Green's functions as well, namely, for any choice of $\ell_1, \dots, \ell_{2n}$, we have

$$\begin{aligned} \left| \mathbb{E} F\left(\frac{1}{N^{k_1 - 1}} \left[\prod_{j=1}^{k_1} G^{(v)}(z_j^1) \right]_{\ell_1, \ell_2}, \dots, \frac{1}{N^{k_n - 1}} \left[\prod_{j=1}^{k_n} G^{(v)}(z_j^n) \right]_{\ell_{2n-1}, \ell_{2n}} \right) - \mathbb{E} F\left(G^{(v)} \to G^{(w)} \right) \right| \\ \le C_1 N^{-1/2 + C_1\varepsilon} + C_1 N^{-\delta + C_1 \varepsilon}. \end{aligned} \tag{2.26}$$

We also remark that Theorem 2.3 holds for generalized Wigner matrices since $C_{sup} < \infty$ in (2.2). The positive lower bound on the variances, $C_{inf} > 0$, is not necessary for this theorem.

Remark 2: Although we state Theorem 2.3 for Hermitian and symmetric ensembles, similar results hold for real and complex sample covariance ensembles; the modification of the proof, to be given in Section 8, is obvious and we omit the details.

To summarize, our approach to prove the universality is based on the following three steps; a detailed outline will be given in Section 6. *Step 1.* Local semicircle law, i.e., Theorem 2.1. This will be proved in Sections 3–5. *Step 2.* Universality for ensembles with smooth distributions satisfying the logarithmic Sobolev inequality (LSI), Theorem 6.3. The key input is the general theorem, Theorem 6.2, concerning the universality for the local relaxation flow. In Section 7, by using the local semicircle law and the LSI, we

verify the assumptions for this theorem. *Step 3.* Green's function comparison theorem, Theorem 2.3. This removes the restriction on the smoothness and the LSI, and it will be proved in Section 8.

Convention. We will frequently use the notation C, c for generic positive constants whose exact values are irrelevant and may change from line to line. For two positive quantities A, B we also introduce the notation $A \asymp B$ to indicate that there exists a universal constant C such that $C^{-1} \leq A/B \leq C$.

3 Proof of local semicircle law

Proof of Theorem 2.1 Recall that $G_{ij} = G_{ij}(z)$ denotes the matrix element

$$G_{ij} = \left(\frac{1}{H-z}\right)_{ij} \tag{3.1}$$

and

$$m(z) = m_N(z) = \sum_{i=1}^{N} N^{-1} G_{ii}(z).$$

We will prove the following more detailed stronger result.

Theorem 3.1 *Assume the $N \times N$ random matrix H satisfies* (2.1), (2.4), (2.5) *and* (2.14), $\mathbb{E}\, h_{ij} = 0$, *for any $1 \leq i, j \leq N$. Let $z = E + i\eta$ $(\eta > 0)$ and let $g(z)$ be the real valued function defined by*

$$g(z) \equiv \min\left\{\sqrt{\kappa+\eta},\ \max\left\{\delta_+,\ \left|\mathrm{Re}\,[m_{sc}(z)^2] - 1\right|\right\}\right\}, \tag{3.2}$$

where $\kappa \equiv ||E| - 2|$ and δ_+ is given in (2.4). *Then for all $z = E + i\eta$ and*

$$\frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\; g^{\,2}(z)} \leq (\log N)^{-13-6\alpha}, \quad |z| \leq 10, \tag{3.3}$$

we have

$$\mathbb{P}\left\{\max_i |G_{ii}(z) - m_{sc}(z)| \geq (\log N)^{11+6\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\; g(z)}\right\} \leq C N^{-c(\log\log N)} \tag{3.4}$$

for sufficiently large N, with positive c and $C > 0$ depending only α and β in (2.14) *and δ_- in* (2.4) *and* (2.5).

Remark: The condition (3.3) is effectively a lower bound on η. The control function $g(z)$ can be estimated by

$$g(z) \asymp \begin{cases} \min\left\{\sqrt{\kappa+\eta},\ \max\left\{\delta_+,\ \eta/\sqrt{\kappa},\ \kappa\right\}\right\}, & |E| \leq 2 \text{ and } \kappa \geq \eta, \\ \sqrt{\kappa+\eta}, & \text{otherwise}, \end{cases} \tag{3.5}$$

up to some factor of order one. Note that the precise formula (3.2) for $g(z)$ is not important, only its asymptotic behaviour for small κ, η and δ_+ is relevant. The theorem remains valid if $g(z)$ is replaced by $\widetilde{g}(z)$ with $\widetilde{g}(z) \leq C g(z)$. In particular, $g(z)$ can be chosen to be order one when E is not near the edges of the spectrum. If we are only concerned with the case of generalized Wigner matrices, (2.6), we can choose

$g(z) = O(\sqrt{\kappa+\eta})$ for any $z = E + i\eta$ ($\eta > 0$). Note that Theorem 2.1 was obtained by replacing $g(z)$ with the lower bound $\kappa \le g(z)$ in Theorem 3.1.

Once the local semicircle law is established on scale $\eta \asymp 1/M$ (modulo logarithmic factors), we obtain the following supremum bound on the eigenvectors that can be interpreted as a lower bound of order $1/M$ on the localization length. The proof of this result now is simpler than in [12, 13], since we have a pointwise control on the diagonal elements of the Green's function. Let $\mathbf{u}_\alpha$ denote the normalized eigenvector of H belonging to the eigenvalue λ_α, $\alpha = 1, 2, \ldots, N$, i.e., $H\mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha$ and $\|\mathbf{u}_\alpha\| = 1$.

Corollary 3.2 *Let H be as in Theorem 3.1, for any fixed $\kappa > 0$, there exists C_κ that*

$$\mathbb{P}\left\{\exists\, \lambda_\alpha \in [-2+\kappa, 2-\kappa],\ H\mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha,\ \|\mathbf{u}_\alpha\| = 1,\ \ \|\mathbf{u}_\alpha\|_\infty \ge C_\kappa \frac{(\log N)^{13+6\alpha}}{M^{1/2}}\right\} \le CN^{-c\log\log N}. \quad (3.6)$$

For the case of generalized Wigner matrices, (2.6)*, we have the following more precise bound*

$$\mathbb{P}\left\{\exists\, \lambda_\alpha,\ H\mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha,\ \|\mathbf{u}_\alpha\| = 1,\ \ \|\mathbf{u}_\alpha\|_\infty \ge \frac{C(\log N)^{13+6\alpha}}{N^{1/2}\big[\big||\lambda_\alpha| - 2\big| + N^{-1}\big]^{1/2}}\right\} \le CN^{-c\log\log N}. \quad (3.7)$$

Proof of Corollary 3.2. Let $\eta = C_\kappa M^{-1}(\log N)^{26+12\alpha}$; C_κ can be chosen large enough so that (3.3) is satisfied for all $|\kappa'| \le \kappa$, making use of (3.5) . Choose $\{E_m\}$ as a grid of points in $[-2+\kappa, 2-\kappa]$ such that the distance between any two neighbors is of order η. Then with (3.4), we have

$$\mathbb{P}\left\{\max_j \max_m \operatorname{Im}|G_{jj}(E_m + i\eta)| \ge \operatorname{Im} m_{sc}(E_m + i\eta) + 1\right\} \le CN^{-c(\log\log N)}, \quad (3.8)$$

where we used $g(z) \le \sqrt{\kappa+\eta} \le C$ from (3.5). Then, with $|m_{sc}(z)| \le C$ (see (2.13)) and

$$\operatorname{Im} G_{jj}(E_m + i\eta) = \sum_\alpha \frac{\eta |u_\alpha(j)|^2}{|E_m - \lambda_\alpha|^2 + \eta^2}, \quad (3.9)$$

where $\mathbf{u}_\alpha = (u_\alpha(1),\ \ u_\alpha(2)\ldots u_\alpha(N))$, we have

$$\mathbb{P}\left(\max_j \max_m \sum_\alpha \frac{\eta |u_\alpha(j)|^2}{|E_m - \lambda_\alpha|^2 + \eta^2} \ge C\right) \le CN^{-c(\log\log N)}. \quad (3.10)$$

By the definition of E_m, for any $\lambda_\alpha \in [-2+\kappa, 2-\kappa]$, there exists m' such that $|E_{m'} - \lambda_\alpha|$ is of the order of η. Together with (3.10), we obtain (3.6).

In case of the generalized Wigner matrix (2.6), we have $g(z) = \sqrt{\kappa+\eta}$ and $M \asymp N$. Let η be the solution to $\eta = N^{-1}(\log N)^{26+12\alpha}(\kappa+\eta)^{-3/2}$, then $N^{-1} \le \eta \le CN^{-1}(\kappa + N^{-1})^{-3/2}(\log N)^{26+12\alpha}$. With this choice of η, (3.3) is satisfied, and $\max_i |G_{ii} - m_{sc}| \le C(\log N)^{-2}(\kappa+\eta)^{1/2}$ holds with an overwhelming probability by (3.4). Since $|\operatorname{Im} m_{sc}(z)| \le C\sqrt{\kappa+\eta}$, so $\max_i \operatorname{Im} G_{ii} \le C(\kappa+\eta)^{1/2}$. By the argument above, we obtain that $\|\mathbf{u}_\alpha\|_\infty^2 \le C\eta(\kappa+\eta)^{1/2}$ on this event. This proves (3.7). □

To prove that $G_{ii}(z)$ is very close to $m_{sc}(z)$ in the sense of (3.4), we will also need to control the off-diagonal elements. In fact we will show that all G_{ij} ($i \ne j$) are bounded by $O((M\eta)^{-1})$ up to some factor $(\log N)^C$. To state the result precisely, we first define some events in the probability space.

Recall that λ_α, $\alpha = 1, 2, \ldots, N$, denote the eigenvalues of $H = (h_{ij})$. Denote by Ω^0 the subset of the probability space such that

$$\max_\alpha |\lambda_\alpha| \leq 3\,. \tag{3.11}$$

Let $\widehat{\Omega}_z^d$ (here the superscript d means diagonal) be the subset of Ω^0 where the following inequality on the diagonal terms hold for any $1 \leq i \leq N$

$$|G_{ii}(z) - m_{sc}(z)| \leq (\log N)^{11+6\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\; g(z)} \tag{3.12}$$

(recall that $m_{sc}(z)$ was defined in (2.13)). Similarly, let $\widehat{\Omega}_z^o$ (here the superscript o means off-diagonal) be the subset of Ω^0 where the following inequality on the off-diagonal terms hold for any $1 \leq i \neq j \leq N$

$$|G_{ij}(z)| \leq (\log N)^{5+4\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}}. \tag{3.13}$$

Finally, denote by Ω_z^d the set

$$\Omega_z^d = \bigcap_{k=0}^{10N^5} \widehat{\Omega}_{z+ik/N^5}^d\,, \tag{3.14}$$

and similarly define Ω_z^o. These sets depend on N but we suppress this from the notations.

Proof of Theorem 3.1. The following proposition immediately implies Theorem 3.1.

□

Proposition 3.3 *Suppose that the assumptions of Theorem 3.1 hold. Then, for sufficiently large N, we have*

$$\mathbb{P}(\Omega_z^d \cap \Omega_z^o) \geq 1 - CN^{-c \log\log N} \tag{3.15}$$

for some positive constants c and C.

Following the work of [12], we will use a continuity argument. In Section 4 we will derive a self-consistent equation of the form

$$G_{ii} + \frac{1}{z + \sum_j \sigma_{ij}^2 G_{jj} + \Upsilon_i(z)} = 0, \qquad i = 1, 2 \ldots, N. \tag{3.16}$$

Later we will give an explicit formula for $\Upsilon_i(z)$, but for now we take (3.16) as the definition of Υ_i. Let $\widehat{\Omega}_z^\Upsilon(N) = \widehat{\Omega}_z^\Upsilon$ be the subset of Ω^0 where the following inequality holds

$$\Upsilon = \Upsilon(z) := \max_i |\Upsilon_i(z)| \leq (\log N)^{9+6\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}}. \tag{3.17}$$

We will use the following Lemmas that will be proved later in Section 4 and 5.

Lemma 3.4 *Let $z = E + i\eta$ be a fixed complex number satisfying* (3.3)*. Then there are constants C and c such that for $N \geq N_0$, with N_0 sufficiently large independent of E and η, the following estimates hold.*

(1) Suppose $3 \leq \eta \leq 10$. Then

$$\mathbb{P}(\widehat{\Omega}_z^o) \geq 1 - CN^{-c\log\log N}, \tag{3.18}$$

$$\mathbb{P}\left(\widehat{\Omega}_z^o \cap \widehat{\Omega}_z^\Upsilon\right) \geq 1 - CN^{-c\log\log N}, \tag{3.19}$$

and for any $1 \leq i \leq N$,

$$|\mathbb{E}\mathbf{1}(\widehat{\Omega}_z^o)\Upsilon_i(z)| \leq (\log N)^{10+8\alpha}\frac{(\kappa+\eta)^{1/2}}{M\eta} + CN^{-c(\log\log N)}. \tag{3.20}$$

(2) Suppose that $\eta \leq 3$. Setting $z' = z + iN^{-5}$, we have

$$\mathbb{P}(\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o) \geq \mathbb{P}(\Omega_{z'}^d \cap \Omega_{z'}^o) - CN^{-c\log\log N}, \tag{3.21}$$

$$\mathbb{P}\left(\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o \cap \widehat{\Omega}_z^\Upsilon\right) \geq \mathbb{P}\left(\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o\right) - CN^{-c\log\log N}, \tag{3.22}$$

and for any $1 \leq i \leq N$,

$$|\mathbb{E}\mathbf{1}(\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o)\Upsilon_i(z)| \leq (\log N)^{10+8\alpha}\frac{(\kappa+\eta)^{1/2}}{M\eta} + CN^{-c(\log\log N)}. \tag{3.23}$$

Lemma 3.5 *Suppose we are on the event $\widehat{\Omega}_z^\Upsilon$ for some fixed $z = E + i\eta$ satisfying* (3.3)*. Suppose either $3 \leq \eta \leq 10$ or the following inequality hold:*

$$\max_i |G_{ii} - m_{sc}(z)| \leq 2(\log N)^{-2}g(z). \tag{3.24}$$

Then, for sufficiently large N, we have

$$\max_i |G_{ii} - m_{sc}(z)| \leq \frac{(\log N)^2}{g(z)}\Upsilon(z). \tag{3.25}$$

Proof of Proposition 3.3. Recall that $\widehat{\Omega}_z^d$ is the subset of Ω^0 where (3.12) holds. Since on $\widehat{\Omega}_z^\Upsilon$ (3.12) follows from (3.17), the case $3 \leq \eta = \operatorname{Im} z \leq 10$ follows from Lemma 3.4 and Lemma 3.5 by taking a union bound for $0 \leq k \leq 10N^5$.

Now we prove (3.15) for the case $\eta \leq 3$ assuming that $z = E + i\eta$ satisfies (3.3). We have shown that (3.15) holds for $\eta = 3$, now we will successively decrease η by N^{-5} in each step, and we continue this inductive procedure as long as (3.3) is still satisfied for the reduced η. More precisely, let $z' = z + iN^{-5}$ and assume that (3.15) holds for z'. Our goal is to prove that

$$\mathbb{P}(\Omega_z^d \cap \Omega_z^o) \geq \mathbb{P}(\Omega_{z'}^d \cap \Omega_{z'}^o) - CN^{-c\log\log N}. \tag{3.26}$$

The number of steps we will be taking is of order N^5. Since $N^{-c\log\log N} \ll N^{-5}$, this proves (3.15) provided that we can establish (3.26).

From (3.21), the difference between the probabilities of the sets $\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_z^o$, and $\Omega_{z'}^d \cap \Omega_{z'}^o$, is negligible. With the definition of Ω_z^d and Ω_z^o in (3.14), we have

$$\Omega_z^d \cap \Omega_z^o \supset \widehat{\Omega}_z^d \cap \widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o. \tag{3.27}$$

Then, to prove (3.26), it remains to prove

$$\mathbb{P}(\widehat{\Omega}_z^d \cap \widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o) \geq \mathbb{P}(\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o) - CN^{-c \log\log N}, \tag{3.28}$$

i.e., we need to estimate the probability of the complement of $\widehat{\Omega}_z^d$ on the set $\widehat{\Omega}_z^o \cap \Omega_{z'}^d \cap \Omega_{z'}^o$. On this set, using (3.22), we can assume that the estimate (3.17) holds with a very high probability. We will show below that (3.24) holds on $\Omega_{z'}^d$. Then (3.25) together with (3.17) imply (3.12), the defining relation of $\widehat{\Omega}_z^d$. This will conclude (3.28) and complete the proof of Proposition 3.3. Therefore, we only have to verify (3.24).

Now we show that (3.24) holds on $\Omega_{z'}^d$. Recall $z' = z + iN^{-5}$ and we have the trivial estimate

$$|G_{ii}(z) - m_{sc}(z)| \leq |G_{ii}(z) - G_{ii}(z')| + |m_{sc}(z) - m_{sc}(z')| + |G_{ii}(z') - m_{sc}(z')|. \tag{3.29}$$

In the set $\Omega_{z'}^d$, we have

$$|G_{ii}(z') - m_{sc}(z')| \leq (\log N)^{11+6\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\, g(z')} \leq (\log N)^{-2} g(z'), \tag{3.30}$$

where in the second inequality we used (3.3). By the definition of $g(z)$ from (3.2), we have $g(z) \leq \sqrt{\kappa+\eta}$. Thus, if (3.3) holds, then, in particular,

$$\eta \geq CM^{-1}(\log N)^{26+12\alpha}. \tag{3.31}$$

This sets a lower bound on η. Together with $|z - z'| = 1/N^5$, we have the trivial continuity bound

$$|G_{ii}(z) - G_{ii}(z')| + |m_{sc}(z) - m_{sc}(z')| \leq N^{-2},$$

using $|\partial_z m_{sc}(z)| \leq |\mathrm{Im}\, z|^{-2}$, $|\partial_z G_{ii}(z)| = |[(H-z)^{-2}]_{ii}| \leq \|(H-z)^{-2}\| \leq |\mathrm{Im}\ z|^{-2}$ and $\eta > N^{-1}$ from (3.31). Thus

$$|G_{ii}(z) - m_{sc}(z)| \leq N^{-2} + (\log N)^{-2} g(z'). \tag{3.32}$$

Using $|g(z)| \geq C\eta \geq CN^{-1}$ and $|g'(z)| \leq C\eta^{-1} \leq CN$ for $\eta \leq 3$, we have the following estimate

$$|G_{ii}(z) - m_{sc}(z)| \leq 2(\log N)^{-2} g(z) \tag{3.33}$$

in the set $\Omega_{z'}^d$. Thus the assumption (3.24) holds in the set $\Omega_{z'}^d$. □

Under the assumptions of Theorem 3.1, with (3.15), (3.20), (3.23) and the definitions in (3.14), all these Ω's are sets of almost full probability, i.e.,

$$\mathbb{P}(\widehat{\Omega}_z^o),\ \ \mathbb{P}(\widehat{\Omega}_z^d),\ \ \mathbb{P}(\Omega_z^o),\ \ \mathbb{P}(\Omega_z^d),\ \ \mathbb{P}(\widehat{\Omega}_z^\Upsilon) \geq 1 - CN^{c \log\log N} \tag{3.34}$$

for some c, $C > 0$.

4 Self-consistent equation for Green's function

First, we introduce some notations.

Definition 4.1 *For any collection of s different numbers, $k_1, k_2, \dots k_s \in \{1, 2, \dots, N\}$, let $H^{(k_1,k_2,\dots,k_s)}$ denote the $N-s$ by $N-s$ submatrix of H after removing the k_i-th $(1 \le i \le s)$ rows and columns. Sometimes we use the notation $H^{(\mathbf{T})}$ where $\mathbf{T}$ denote the unordered set $\{k_1, k_2, \dots k_s\}$. Similarly, we define $\mathbf{a}^{(\ell;\, \mathbf{T})}$ to be the ℓ-th column of H with k_i-th $(1 \le i \le s)$ elements removed. Sometimes, we just use the short notation $\mathbf{a}^{\ell}$=$\mathbf{a}^{(\ell;\, \mathbf{T})}$.*

For $\mathbf{T} = \{k_1, k_2, \dots k_s\}$, we define

$$\begin{aligned} G_{ij}^{(\mathbf{T})} &:= [H^{(\mathbf{T})} - z]^{-1}(i,j), \\ Z_{ij}^{(\mathbf{T})} &:= \mathbf{a}^i \cdot [H^{(\mathbf{T})} - z]^{-1} \mathbf{a}^j = \sum_{k,l \notin \mathbf{T}} \overline{\mathbf{a}_k^i} G_{k,l}^{(\mathbf{T})} \mathbf{a}_l^j, \\ K_{ij}^{(\mathbf{T})} &:= h_{ij} - z\delta_{ij} - Z_{ij}^{(\mathbf{T})}. \end{aligned}$$

These quantities depend on z, but we mostly neglect this dependence in the notation.

We start the proof with deriving some identities between the matrix elements of $G = (H - z)^{-1}$ and $G^{(k_1,k_2,\dots,k_s)}$ using the following well known result in linear algebra that we quote without proof.

Lemma 4.1 *Let A , B, C be $n \times n$, $m \times n$ and $m \times m$ matrices. We define $(m+n) \times (m+n)$ matrix D as*

$$D = \begin{pmatrix} A & B^* \\ B & C \end{pmatrix} \tag{4.1}$$

and $n \times n$ matrix $\widehat{D}$ as

$$\widehat{D} = A - B^* C^{-1} B. \tag{4.2}$$

Then for any $1 \le i, j \le n$, we have

$$(D^{-1})_{ij} = (\widehat{D}^{-1})_{ij}$$

for the corresponding matrix elements.

Furthermore, let $\mathbf{T}$ denote the unordered set $\{k_1, k_2, \dots k_s\}$ and $1 \le k_i \le n$, $1 \le i \le s$. We define $D^{(\mathbf{T})}$ to be the $n+m-s$ by $n+m-s$ submatrix of D after removing the k_i-th $(1 \le i \le s)$ rows and columns and define $\widehat{D}^{(\mathbf{T})}$ to be the $n-s$ by $n-s$ submatrix of $\widehat{D}$ after removing the k_i-th $(1 \le i \le s)$ rows and columns. Then for any $1 \le i, j \le n$ and $i, j \notin \mathbf{T}$, we have

$$\left((D^{(\mathbf{T})})^{-1}\right)_{ij} = \left((\widehat{D}^{(\mathbf{T})})^{-1}\right)_{ij}$$

for the corresponding matrix elements.

Using Lemma 4.1 and Definition 4.1, for $1 \le i \ne j \le N$, we have

$$G_{ii} = (K_{ii}^{(i)})^{-1} = \frac{K_{jj}^{(ij)}}{K_{jj}^{(ij)} K_{ii}^{(ij)} - K_{ij}^{(ij)} K_{ji}^{(ij)}}. \tag{4.3}$$

For the off diagonal matrix elements G_{ij}, $(i \neq j)$, we have

$$G_{ij} = -\frac{K_{ij}^{(ij)}}{K_{jj}^{(ij)} K_{ii}^{(ij)} - K_{ij}^{(ij)} K_{ji}^{(ij)}} = -G_{ii} \frac{K_{ij}^{(ij)}}{K_{jj}^{(ij)}} = -G_{ii} G_{jj}^{(i)} K_{ij}^{(ij)}. \tag{4.4}$$

Similarly, we have the following result

Lemma 4.2 *Let* $\mathbf{T}$ *be an unordered set* $\{k_1, k_2, \ldots, k_s\}$ *with* $1 \leq k_t \leq N$ *for* $(1 \leq t \leq s)$ *or* $\mathbf{T} = \emptyset$. *For simplicity, we use the notation* $(i\,\mathbf{T})$ *for* $\{i\} \cup \mathbf{T}$ *and* $(ij\,\mathbf{T})$ *for* $\{i, j\} \cup \mathbf{T}$. *Then we have the following identities:*

1. *For any* $i \notin \mathbf{T}$

$$G_{ii}^{(\mathbf{T})} = (K_{ii}^{(i\,\mathbf{T})})^{-1}. \tag{4.5}$$

2. *For* $i \neq j$ *and* $i, j \notin \mathbf{T}$

$$G_{ij}^{(\mathbf{T})} = -G_{jj}^{(\mathbf{T})} G_{ii}^{(j\,\mathbf{T})} K_{ij}^{(ij\ \mathbf{T})} = -G_{ii}^{(\mathbf{T})} G_{jj}^{(i\,\mathbf{T})} K_{ij}^{(ij\ \mathbf{T})}. \tag{4.6}$$

3. *For* $i \neq j$ *and* $i, j \notin \mathbf{T}$

$$G_{ii}^{(\mathbf{T})} - G_{ii}^{(j\ \mathbf{T})} = G_{ij}^{(\mathbf{T})} G_{ji}^{(\mathbf{T})} (G_{jj}^{(\mathbf{T})})^{-1}. \tag{4.7}$$

4. *For any indices* i, j *and* k *that are different and* $i, j, k \notin \mathbf{T}$

$$G_{ij}^{(\mathbf{T})} - G_{ij}^{(k\ \mathbf{T})} = G_{ik}^{(\mathbf{T})} G_{kj}^{(\mathbf{T})} (G_{kk}^{(\mathbf{T})})^{-1}. \tag{4.8}$$

Proof of Lemma 4.2. The first two identities (4.5) and (4.6) are obvious extensions of (4.3) and (4.4). To prove (4.7), without loss of generality, we may assume that $i = 1$, $j = 2$ and $\mathbf{T} = \emptyset$. Let $D = H - z$ and $\widehat{D}$ defined as in (4.2) with $n = 2$ and $m = N - 2$. With Lemma 4.1, we have that for $i, j = 1$ or 2.

$$G_{ij} = (\widehat{D}^{-1})_{ij} \tag{4.9}$$

and

$$G_{ii}^{(j)} = \left((\widehat{D}^{(j)})^{-1} \right)_{ii}. \tag{4.10}$$

Since $\widehat{D}$ is just a 2×2 matrix, one can easily check that (4.7) holds. With the same method, one can obtain (4.8) . □

Lemma 4.3 *The diagonal matrix elements of the resolvent satisfy the following self-consistent equation.*

$$G_{ii} = \left(-z - \sum_j \sigma_{ij}^2 G_{jj} + \Upsilon_i \right)^{-1} \tag{4.11}$$

where $\Upsilon_i(z)$ *is given by*

$$\Upsilon_i(z) := \sigma_{ii}^2 G_{ii} + \sum_{j \neq i} \sigma_{ij}^2 G_{ij} G_{ji} [G_{ii}]^{-1} + \left(K_{ii}^{(i)} - \mathbb{E}_{\mathbf{a}^i} K_{ii}^{(i)} \right), \tag{4.12}$$

and $\mathbb{E}_{\mathbf{a}^i}$ is the expectation over $\mathbf{a}^i$. Let $\mathbf{T}$ denote the set $\{k_1, k_2, \dots k_m\}$, which also could be the empty set, then

$$|K_{ii}^{(i\mathbf{T})} - \mathbb{E}_{\mathbf{a}^i} K_{ii}^{(i\mathbf{T})}| \leq (\log N)^{3+2\alpha} \sqrt{M^{-1} + M^{-1} \max_k |G_{kk}^{(i\mathbf{T})}|^2 + \max_{k\neq l} |G_{kl}^{(i\mathbf{T})}|^2} \tag{4.13}$$

and for $i \neq j$

$$|K_{ij}^{(ij\mathbf{T})}| \leq (\log N)^{4+4\alpha} \sqrt{M^{-1} + (M\eta)^{-1} \max_l \left\{ Im\, G_{ll}^{(ij\mathbf{T})} \right\}}, \quad i \neq j, \tag{4.14}$$

hold with a probability larger than $1 - CN^{-c(\log\log N)}$ for sufficiently large N.

Proof of Lemma 4.3. We can write G_{11} as follows,

$$G_{11} = (K_{11}^{(1)})^{-1} = \frac{1}{\mathbb{E}_{\mathbf{a}^1} K_{11}^{(1)} + K_{11}^{(1)} - \mathbb{E}_{\mathbf{a}^1} K_{11}^{(1)}}. \tag{4.15}$$

Using the fact $G^{(1)} = (H^{(1)} - z)^{-1}$ is independent of $\mathbf{a}^1$ and $\mathbb{E}_{\mathbf{a}^1} \overline{\mathbf{a}^1(i)} \mathbf{a}^1(j) = \delta_{ij}\sigma_{1j}^2$, we obtain $\mathbb{E}_{\mathbf{a}^1} K_{11}^{(1)} = -z - \sum_{j\neq 1} \sigma_{1j}^2 G_{jj}^{(1)}$, and thus

$$G_{11} = \frac{1}{-z - \sum_{j\neq 1} \sigma_{1j}^2 G_{jj}^{(1)} + \left(K_{11}^{(1)} - \mathbb{E}_{\mathbf{a}_1} K_{11}^{(1)} \right)}. \tag{4.16}$$

Combining this identity with (4.7), we have

$$G_{11} = \left(-z - \sum_{j\neq 1} \sigma_{1j}^2 \left(G_{jj} - G_{1j} G_{j1} G_{11}^{-1} \right) + \left(K_{11}^{(1)} - \mathbb{E}_{\mathbf{a}^1} K_{11}^{(1)} \right) \right)^{-1}. \tag{4.17}$$

Clearly G_{11} can be replaced with any G_{ii} and this proves (4.11) with the definition (4.12).

Now we prove (4.13) and (4.14). Define

$$v_{ij} \equiv h_{ij}/\sigma_{ij}, \tag{4.18}$$

hence $\mathbb{E}v_{ij} = 0$ and $\mathbb{E}|v_{ij}|^2 = 1$. If $\sigma_{ij} = 0$, i.e., $h_{ij} = 0$ almost surely, then we set $v_{ij} = 0$. By the definition of $K_{ii}^{(i\mathbf{T})}$, we write

$$K_{ii}^{(i\mathbf{T})} = h_{ii} - z - \sum_{k,l\notin(i\mathbf{T})} \overline{\mathbf{a}_k^i} G_{kl}^{(i\mathbf{T})} \mathbf{a}_l^i = h_{ii} - z - \sum_{k,l\notin(i\mathbf{T})} \overline{v_{ik}} \sigma_{ik} G_{kl}^{(i\mathbf{T})} \sigma_{li} v_{li} \tag{4.19}$$

and

$$\mathbb{E}_{\mathbf{a}^i} K_{ii}^{(i\mathbf{T})} = -z - \sum_{k\notin(i\mathbf{T})} \sigma_{ik} G_{kk}^{(i\mathbf{T})} \sigma_{ki}. \tag{4.20}$$

We note h_{ii}, v_{ij} and $G_{kl}^{(i\mathbf{T})}$ are independent for $k, l \notin (i\mathbf{T})$. With the sub-exponential decay (2.14) and $\sigma_{ij}^2 \leq 1/M$, we have for any i, j

$$\mathbb{P}\left\{ |h_{ij}| \leq (\log N)^{3+2\alpha} M^{-1/2} \right\} \geq 1 - CN^{-c(\log\log N)}. \tag{4.21}$$

In Corollary B.3 of Appendix B we will prove a general large deviations result. Applying (B.15) to the last term in (4.19), with the choice

$$B_{kl} = \sigma_{ik} G^{(i\mathbf{T})}_{kl} \sigma_{li} \tag{4.22}$$

and with $\sum_j \sigma_{ij}^2 = 1$ and $\sigma_{ii}^2 \le 1/M$, we obtain that

$$\left| \sum_{k,l \notin (i\mathbf{T})} \overline{v_{ik}} \sigma_{ik} G^{(i\mathbf{T})}_{kl} \sigma_{li} v_{li} - \sum_{k \notin (i\mathbf{T})} \sigma_{ik} G^{(i\mathbf{T})}_{kk} \sigma_{ki} \right| \le (\log N)^{3+2\alpha} \sqrt{M^{-1} \max_k |G^{(i\mathbf{T})}_{kk}|^2 + \max_{k \ne l} |G^{(i\mathbf{T})}_{kl}|^2} \tag{4.23}$$

holds with a probability larger than $1 - CN^{-c(\log\log N)}$. Together with (4.21), we obtain that (4.13) holds with a probability larger than $1 - CN^{-c(\log\log N)}$ for sufficiently large N.

Next we prove (4.14). By the definition of $K^{(ij\mathbf{T})}_{ij}$, $i \ne j$, we can write

$$K^{(ij\mathbf{T})}_{ij} = h_{ij} - \sum_{k,l \notin (ij\mathbf{T})} \overline{v_{ik}} \sigma_{ik} G^{(ij\mathbf{T})}_{kl} \sigma_{lj} v_{lj}. \tag{4.24}$$

Applying (B.16), (4.21) and $\sigma_{ij}^2 \le 1/M$, we obtain that

$$|K^{(ij\mathbf{T})}_{ij}| \le (\log N)^{4+4\alpha} \sqrt{M^{-1} + \sum_{k,l \notin (ij\mathbf{T})} \left| \sigma_{ik} G^{(ij\mathbf{T})}_{kl} \sigma_{lj} \right|^2} \tag{4.25}$$

holds with a probability larger than $1 - CN^{-c(\log\log N)}$ for sufficiently large N. With Schwarz's inequality, for any i, j,

$$\sum_{kl} \left| \sigma_{ik} G^{(ij\mathbf{T})}_{kl} \sigma_{lj} \right|^2 \le \left(\sum_{kl} |\sigma_{ik}|^4 |G^{(ij\mathbf{T})}_{kl}|^2 \right)^{1/2} \left(\sum_{kl} |G^{(ij\mathbf{T})}_{kl}|^2 |\sigma_{jl}|^4 \right)^{1/2}. \tag{4.26}$$

Denote $u^{(ij\mathbf{T})}_\alpha$ and $\lambda^{(ij\mathbf{T})}_\alpha$ ($\alpha = 1, 2, \ldots, N - |\mathbf{T}| - 2$) the l^2-normalized eigenvectors and eigenvalues of $H^{(ij\mathbf{T})}$. Let $u^{(ij\mathbf{T})}_\alpha(l)$ denote the l-th coordinate of $u^{(ij\mathbf{T})}_\alpha$, then for any l

$$\sum_k |G^{(i\,j\,\mathbf{T})}_{kl}|^2 = \left(|G^{(i\,j\,\mathbf{T})}|^2 \right)_{ll} = \sum_\alpha \frac{|u^{(i\,j\,\mathbf{T})}_\alpha(l)|^2}{|\lambda^{(i\,j\,\mathbf{T})}_\alpha - z|^2} = \frac{\mathrm{Im}\ G^{(i\,j\,\mathbf{T})}_{ll}(z)}{\eta}. \tag{4.27}$$

Here we defined $|A|^2 := A^* A$ for any matrix A. Inserting (4.27) into (4.25) and using the definition of M in (2.3), we obtain that (4.14) holds with a probability larger than $1 - CN^{-c(\log\log N)}$ for sufficiently large N. □

Proof of Lemma 3.4. We first prove (3.18) in the range $3 \le \eta \le 10$. Recall that Ω^0 is the subset of the entire probability space where $\|H\| \le 3$ see (3.11). By (7.11) from Lemma 7.2 (using that $M \ge (\log N)^9$ is implied by (3.3)), we have $\mathbb{P}(\Omega^0) \ge 1 - N^{-c\log\log N}$. Denote λ_α and $\mathbf{u}_\alpha$ the eigenvalues and eigenvectors of $H = (h_{ij})$. From the identity

$$G_{ii} = \sum_\alpha \frac{|u_\alpha(i)|^2}{\lambda_\alpha - z} \tag{4.28}$$

and $\max_\alpha |\lambda_\alpha| \le 3$, we have that

$$\eta^{-1} \ge |G_{ii}| \ge |\mathrm{Im}\, G_{ii}| \ge \frac{\eta}{(|E|+3)^2+\eta^2} \tag{4.29}$$

holds in Ω^0. Together with $3 \le \eta \le 10$ and $|E| \le 10$, we obtain

$$c \le |G_{ii}| \le C \tag{4.30}$$

with some positive constants. From the interlacing property of the eigenvalues of the matrix and its submatrices, we find that not only $\|H\| \le 3$ but also $\|H^{(\mathbf{T})}\| \le 3$ holds on the set Ω^0. Thus for any j, k such that i, j and k are all different, the bounds

$$c \le |G_{ii}|, \;\; |G^{(j)}_{ii}|, \;\; |G^{(jk)}_{ii}| \le C. \tag{4.31}$$

hold in Ω^0 by a similar argument that led to (4.30). Thus (4.6) implies

$$\mathbf{1}(\Omega^0)|G^{(ij)}_{ij}| \le C^2 \mathbf{1}(\Omega^0)|K^{(ij)}_{ij}| \tag{4.32}$$

and (3.18) follows make use of (4.14) and $\eta > 3$.

Now we prove (3.19). Recall that the self consistent equation (3.16) with the error term $\Upsilon_i(z)$ is given by (4.12), i.e.,

$$\Upsilon_i(z) = \sigma^2_{ii} G_{ii} + \sum_{j \ne i} \sigma^2_{ij} G_{ij} G_{ji} G^{-1}_{ii} + \left(K^{(i)}_{ii} - \mathbb{E}_{\mathbf{a}^i} K^{(i)}_{ii} \right). \tag{4.33}$$

Now we bound $\Upsilon_i(z)$ in $\widehat{\Omega}^o_z$. Since $\sigma^2_{ii} \le M^{-1}$, with (4.31), the first term of the r.h.s. of (4.33) is less than $O(M^{-1})$. Then with (2.1), and using the bound on G_{ij} $(i \ne j)$ from (3.13) and the one on G_{ii} from (4.31), we obtain that the second term of the r.h.s. of (4.33) is less than $C(\log N)^{10+8\alpha}(M\eta)^{-1}$ (and with (3.31), we know it is much less than 1), i.e., in $\widehat{\Omega}^o_z$

$$\left| \sigma^2_{ii} G_{ii} + \sum_{j \ne i} \sigma^2_{ij} G_{ij} G_{ji} G^{-1}_{ii} \right| \le C(\log N)^{10+8\alpha}(M\eta)^{-1}. \tag{4.34}$$

The last term of the r.h.s. of (4.33) can be bounded, using (4.13) with $\mathbf{T} = \emptyset$, with a very large probability. Using (4.8) and (4.31), the $G^{(i)}_{kl}$'s in (4.13) can be bounded as

$$|G^{(i)}_{kl}| \le |G_{kl}| + C|G_{ki}||G_{il}|. \tag{4.35}$$

Therefore, again with the bound on G_{ij} $(i \ne j)$ in (3.13) and the one on G_{ii} from (4.31), we see that

$$|K^{(i)}_{ii} - \mathbb{E}_{\mathbf{a}^i} K^{(i)}_{ii}| \le 2(\log N)^{8+6\alpha}(M\eta)^{-1/2} \tag{4.36}$$

holds in $\widehat{\Omega}^o_z$ with a probability larger than $\mathbb{P}(\widehat{\Omega}^o_z) - CN^{-c(\log\log N)}$ for sufficiently large N. Inserting (4.34) and (4.36) into (4.33) and together $\eta \ge 3$, we have proved (3.19).

Now we prove (3.20) for $\eta \geq 3$. By the definition of Υ_i in (4.33), we have

$$\left|\mathbb{E}\left[\mathbf{1}(\widehat{\Omega}_z^o)\Upsilon_i(z)\right]\right| \leq \mathbb{E}\mathbf{1}(\widehat{\Omega}_z^o)\left|\sigma_{ii}^2 G_{ii} + \sum_{j\neq i}\sigma_{ij}^2 G_{ij}G_{ji}G_{ii}^{-1}\right| + \left|\mathbb{E}\,\mathbf{1}([\widehat{\Omega}_z^o]^c)\left(K_{ii}^{(i)} - \mathbb{E}_{\mathbf{a}^i}K_{ii}^{(i)}\right)\right|, \tag{4.37}$$

since $\mathbb{E}\left(K_{ii}^{(i)} - \mathbb{E}_{\mathbf{a}^i}K_{ii}^{(i)}\right) = 0$. Using (4.31), in Ω^0 we have that $|\sigma_{ii}^2 G_{ii} + \sum_{j\neq i}\sigma_{ij}^2 G_{ij}G_{ji}G_{ii}^{-1}|$ is always less than a constant C for some $C > 0$. Inserting this and (4.34) into (4.37), we obtain that

$$\left|\mathbb{E}\left[\mathbf{1}(\widehat{\Omega}_z^o)\Upsilon_i(z)\right]\right| \leq C(\log N)^{10+8\alpha}(M\eta)^{-1} + \left|\mathbb{E}\left[\mathbf{1}([\widehat{\Omega}_z^o]^c)\left|K_{ii}^{(i)} - \mathbb{E}_{\mathbf{a}^i}K_{ii}^{(i)}\right|\right]\right| + CN^{-c\log\log N}. \tag{4.38}$$

We now claim that for some large enough $C > 0$ there exists $c > 0$ such that

$$\mathbb{P}(|Z_{ii}^{(i)}| \geq N^C) \leq e^{-N^c} \qquad \text{and} \qquad \mathbb{P}(|K_{ii}^{(i)}| \geq N^C) \leq e^{-N^c}. \tag{4.39}$$

The first estimate follows from the definition of $Z_{ii}^{(i)}$ given in Definition 4.1 by using the sub-exponential decay of the matrix elements and by using the trivial bound $|G_{kl}^{(i)}| \leq \eta^{-1} \neq N$. The second estimate is a trivial consequence of the first one and the definition of $K_{ii}^{(i)}$. Together with (4.38), we obtain (3.20) in the case that $3 \leq \eta \leq 10$.

We now prove (3.21) and (3.22) for the case $\eta \leq 3$ satisfying (3.3). We will work in the event $\Omega_{z'}^d \cap \Omega_{z'}^d$ where $z' = z + iN^{-5}$. Similarly as we proved (3.33), from the bound below (3.31) and the Lipschitz continuity of $g(z)$, we obtain that

$$|G_{ii}(z) - m_{sc}(z)| \leq 2(\log N)^{11+6\alpha}\frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\; g(z)} \tag{4.40}$$

and

$$|G_{ij}(z)| \leq 2(\log N)^{5+4\alpha}\frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}} \tag{4.41}$$

hold in $\Omega_{z'}^d \cap \Omega_{z'}^d$. We note the r.h.s of these inequalities are much less than $(\log N)^{-1}$ by (3.3). From the explicit formula (2.13) we obtain that $c \leq |m_{sc}(z)| \leq C$ for any $|z| \leq 10$ with some positive constants. Using this observation and the fact that the r.h.s. of (4.40) is much less than $(\log N)^{-1}$, we have

$$c \leq |G_{ii}(z)| \leq C\,.$$

Hence, using (3.31), (4.7), (4.8) and the lower bound of $|G_{ii}|$, one can easily obtain that

$$|G_{ii}(z) - m_{sc}(z)|,\;\; |G_{ii}^{(j)}(z) - m_{sc}(z)|,\;\; |G_{ii}^{(jk)}(z) - m_{sc}(z)| \leq C(\log N)^{11+6\alpha}\frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}\; g(z)} \tag{4.42}$$

hold in $\Omega_{z'}^o \cap \Omega_{z'}^d$, (for the third term in l.h.s., we have also used the lower bounds of $G_{ii}^{(j)}$'s as above). Then we also have

$$c \leq |G_{ii}(z)|,\;\; |G_{ii}^{(j)}(z)|,\;\; |G_{ii}^{(jk)}(z)| \leq C \tag{4.43}$$

with some positive constants.

The definition of $m_{sc}(z)$ implies Im $m_{sc}(z) \leq C\sqrt{\kappa+\eta}$. Then with (4.42), (3.3) and $g(z) \leq \sqrt{\kappa+\eta}$, we have that

$$\operatorname{Im} G_{ii}^{(jk)}(z) \leq C\sqrt{\kappa+\eta} \tag{4.44}$$

holds in $\Omega_{z'}^o \cap \Omega_{z'}^d$ for some constant $C > 0$. Inserting it into (4.14), we obtain that

$$|K_{ij}^{(ij)}(z)| \le C(\log N)^{4+4\alpha} \frac{(\kappa+\eta)^{1/4}}{\sqrt{M\eta}} \tag{4.45}$$

hold in $\Omega_{z'}^o \cap \Omega_{z'}^d$ with a probability larger than $\mathbb{P}(\Omega_{z'}^o \cap \Omega_{z'}^d) - CN^{-c(\log\log N)}$ for sufficiently large N. Again, with (4.6) and (4.43), we obtain (3.21) for sufficiently large N.

Then, as we proved in (4.34) and (4.36), we get that

$$|\sigma_{ii}^2 G_{ii}(z) + \sum_{j\neq i} \sigma_{ij}^2 G_{ij} G_{ji} G_{ii}^{-1}(z)| \le C(\log N)^{10+8\alpha}(M\eta)^{-1} \tag{4.46}$$

and

$$|K_{jj}^{(j)}(z) - \mathbb{E}_{\mathbf{a}^j} K_{jj}^{(j)}(z)| \le C(\log N)^{8+6\alpha}(M\eta)^{-1/2} \tag{4.47}$$

hold in $\widehat{\Omega}_z^o \cap \Omega_{z'}^o \cap \Omega_{z'}^d$ with a probability larger than $\mathbb{P}(\widehat{\Omega}_z^o \cap \Omega_{z'}^o \cap \Omega_{z'}^d) - CN^{-c(\log\log N)}$, which implies (3.22).

Finally, similarly as using (4.37)- (4.38) to prove (3.20), we can obtain (3.23) in the case that $\eta < 3$. □

5 Stability of the self-consistent equation: proof of Lemma 3.5

In this section, we prove Lemma 3.5, i.e., we will prove the stability of the self-consistent equation with a precise error estimate given in (3.25). We set $m_{sc} = m_{sc}(z)$ and $\Upsilon = \max_i |\Upsilon_i(z)|$ for simplicity of notation and we will omit all z dependences in all the symbols. With the definition of $m_{sc}(z)$ in (2.12) and (2.13), the following properties of $m_{sc}(z)$ can be easily established:

Lemma 5.1 *Let $z = E + i\eta$ with $\eta > 0$ and $|z| \le 20$. Then we have*

$$|z + m_{sc}|^{-2} = |m_{sc}|^2 \le 1 \tag{5.1}$$

and

$$\left|(z + m_{sc}(z))^{-2} - 1\right| \ge C\sqrt{\kappa+\eta} \tag{5.2}$$

for some constant C. Furthermore, suppose that either $2 \le |E| \le 10$ or $\kappa \le \eta$. Then

$$|z + m_{sc}|^{-2} = |m_{sc}|^2 \le 1 - C\sqrt{\kappa+\eta}. \tag{5.3}$$

For small values of $|z^2 - 4| \asymp \kappa + \eta$, $m_{sc}(z)$ has the asymptotic expansion

$$m_{sc} = \mp 1 + \frac{1}{2}\sqrt{z^2-4} + O(|z^2-4|), \quad \textit{near} \quad z \asymp \pm 2. \tag{5.4}$$

□

We first prove (3.25) for the case that $3 \le \eta \le 10$. In this case, we can easily check that $g(z) = \sqrt{\kappa+\eta}$. Denote the difference between G_{ii} and m_{sc} by

$$v_i = G_{ii} - m_{sc}, \qquad 1 \le i \le N.$$

By the self consistent equation (3.16), (2.1) and (2.12), we have

$$v_i = \frac{\sum_i \sigma_{ij}^2 v_j + \Upsilon_i}{(z + m_{sc} + \sum_j \sigma_{ij}^2 v_j + \Upsilon_i)(z + m_{sc})}, \ \ 1 \le i \le N. \tag{5.5}$$

For $\eta \ge 3$, $|z + m_{sc}(z)| > 2$ by (2.13). Using $|G_{ii}| \le \eta^{-1}$ and $|m_{sc}| \le \eta^{-1}$, we obtain

$$|v_i| \le 2/\eta \le 2/3, \qquad 1 \le i \le N. \tag{5.6}$$

From the assumption (3.17) and (3.3), we have $\Upsilon = \max_i |\Upsilon_i| \ll 1$ in this region. Together with $|z+m_{sc}(z)| > 2$ and (5.6), we obtain that the absolute value of the r.h.s. of (5.5) is less than

$$\frac{\sup_i |v_i|}{|z + m_{sc}(z)| - \sup_i |v_i|} + O(\Upsilon). \tag{5.7}$$

Taking the absolute value of (5.5) and maximizing over n, we have

$$\sup_n |v_n| \le \frac{\sup_i |v_i|}{|z + m_{sc}| - \sup_i |v_i|} + O(\Upsilon). \tag{5.8}$$

The denominator satisfies $|z + m_{sc}(z)| - \sup_i |v_i| \ge 2 - 2/3 = 4/3$, therefore we obtain $\sup |v_i| = \sup_i |G_{ii} - m_{sc}(z)| \le O(\Upsilon)$, which shows (3.25) for $3 \le \eta \le 10$.

Next, we prove (3.25) in the case that $\eta \le 3$ with η satisfying (3.3) and under the condition (3.24). Define

$$m = m(z) := \frac{1}{N} \sum_i G_{ii} \quad \text{and} \quad u_i := G_{ii} - m. \tag{5.9}$$

Combining (3.17), (3.3), (3.24) with the fact that $g(z) \le C$, we can see that

$$\Upsilon \le (\log N)^{-4} g^2(z) \le C(\log N)^{-4}, \ \ |G_{ii} - m_{sc}(z)| \le C(\log N)^{-2}. \tag{5.10}$$

Together with (5.1), we have

$$|z + m_{sc}(z)| - |G_{ii} - m_{sc}(z)| - |\Upsilon| \ge C$$

for some $C > 0$. Furthermore (3.24) implies

$$|m(z) - m_{sc}| \le 2(\log N)^{-2} g(z) \tag{5.11}$$

thus there exists $C > 0$ such that

$$|z + m(z)| - |G_{ii} - m(z)| - |\Upsilon| \ge C. \tag{5.12}$$

Therefore, expanding the self consistent equation (3.16) around $z + m(z)$, we obtain that

$$0 = G_{ii} + \frac{1}{z + \sum_j \sigma_{ij}^2 G_{jj} + \Upsilon_i} = G_{ii} + \frac{1}{z + m(z)} + \Omega_i \tag{5.13}$$

where Ω_i is defined by the second equality and it satisfies

$$\Omega_i = -\frac{\sum_j \sigma_{ij}^2 u_j}{(z + m(z))^2} + O(\|\mathbf{u}\|_\infty^2) + O(\Upsilon) \tag{5.14}$$

with error bounds uniform in i. Here $\|\mathbf{u}\|_\infty = \max_i |u_i|$. Taking the average of the r.h.s of (5.13) with respect to i, we obtain that

$$m(z) + \frac{1}{z + m(z)} = -\Omega \tag{5.15}$$

where

$$\Omega := \frac{1}{N} \sum_i \Omega_i, \tag{5.16}$$

and it satisfies

$$|\Omega| \leq O(\|\mathbf{u}\|_\infty^2) + O(\Upsilon). \tag{5.17}$$

Here we used $\sum_i \sum_j \sigma_{ij}^2 u_j = \sum_j u_j = 0$. The bound (3.24) , (5.11) and $g(z) \leq \sqrt{\kappa + \eta}$ (from (3.2)) implies that

$$\|\mathbf{u}\|_\infty \leq 4(\log N)^{-2} g(z) \leq C(\log N)^{-2} \sqrt{\kappa + \eta}. \tag{5.18}$$

Together with (5.10) and (5.17), we obtain

$$|\Omega| \leq C(\log N)^{-4}(\kappa + \eta).$$

To bound $m(z)$, we use the following lemma.

Lemma 5.2 *Let $z = E + i\eta \in \mathbb{C}$, $|z| \leq 10$ and let $\delta > 0$ be a sufficiently small constant. Let $t \in \mathbb{C}$ such that*

$$|t| \leq \delta(\kappa + \eta). \tag{5.19}$$

Suppose there is a function $s_z(t) \in \mathbb{C}$ that solves the equation

$$s_z(t) + \frac{1}{z + s_z(t)} = t, \tag{5.20}$$

with $Im\, s_z(t) > 0$ and the estimate

$$|s_z(t) - m_{sc}(z)| \leq \delta \sqrt{\kappa + \eta} \tag{5.21}$$

holds. Then

$$|s_z(t) - m_{sc}(z)| \leq C \frac{|t|}{\sqrt{\kappa + \eta}} \tag{5.22}$$

for some constant $C > 0$.

Proof of Lemma 5.2. It follows from (5.20) that

$$s_z(t) = t + \frac{-z - t}{2} \pm \frac{\sqrt{(z + t)^2 - 4}}{2}. \tag{5.23}$$

We denote by $s_z^1(t)$ and $s_z^2(t)$ the two solutions of this equation, which are continuous with respect to t locally in the neighborhood (5.19). When $t = 0$, one of them is equal to $m_{sc}(z)$, we choose $s_z^1(0) = m_{sc}(z)$. From (5.23), we have

$$|s^1 - s^2| = |(z + t)^2 - 4|^{1/2}. \tag{5.24}$$

Then, for small enough δ, if $|t| \le \delta(\kappa+\eta)$, then $|s^1 - s^2| \ge \frac{1}{2}\min\{|z-2|, |z+2|\}$ by using (5.24) and that $\kappa + \eta \asymp \min\{|z-2|, |z+2|\}$. We thus see that only one out of s^1 and s^2 can satisfy (5.21). With the assumption that $s_z^1(0) = m_{sc}(z)$, it is s^1 that satisfies (5.21). Then

$$s_z(t) - m_{sc}(z) = s_z^1(t) - s_z^1(0) \tag{5.25}$$

and (5.22) follows from the fact that

$$|\partial_t s_z(t)| \le O\left(\frac{1}{|\sqrt{(z+t)^2-4}|}\right) \le O\left(\frac{1}{\sqrt{\kappa+\eta}}\right), \tag{5.26}$$

where for the second inequality, we used $|t| \le \delta(\kappa+\eta)$. □

Using Lemma 5.2, for $s_z(t) = m(z)$ and $t = -\Omega$, we have

$$|m(z) - m_{sc}(z)| \le \frac{C|\Omega|}{\sqrt{\kappa+\eta}} \le C(\log N)^{-2}\|\mathbf{u}\|_\infty + \frac{C\Upsilon}{\sqrt{\kappa+\eta}}, \tag{5.27}$$

where in the second inequality we used (5.17) and (5.18). Subtracting (5.17) from (5.13), we have the equation for u_i

$$u_i = G_{ii} - m(z) = \frac{\sum_j \sigma_{ij}^2 u_j}{(z+m(z))^2} + \Omega + O(\|\mathbf{u}\|_\infty^2) + O(\Upsilon) = w_i + \frac{\sum_j \sigma_{ij}^2 u_j}{(z+m_{sc}(z))^2}, \tag{5.28}$$

where w_i is defined as $u_i - (\sum_j \sigma_{ij}^2 u_j)(z+m_{sc})^{-2}$. By (5.17), it is bounded by

$$\|\mathbf{w}\|_\infty = O(\|\mathbf{u}\|_\infty^2) + O\left(\|\mathbf{u}\|_\infty |(z+m)^{-2} - (z+m_{sc})^{-2}|\right) + O(\Upsilon). \tag{5.29}$$

Then, using (5.11) and (5.1), we obtain that

$$|(z+m)^{-2} - (z+m_{sc})^{-2}| \le C|m(z) - m_{sc}(z)|. \tag{5.30}$$

Inserting this into (5.29), using the bounds on $\|\mathbf{u}\|_\infty$ in (5.18) and (5.27), we have

$$\|\mathbf{w}\|_\infty = O(\|\mathbf{u}\|_\infty^2) + O(\Upsilon). \tag{5.31}$$

From (5.3) in Lemma 5.1, whenever $|E| \ge 2$ or $\kappa \le \eta$, in which case $g(z) \asymp \sqrt{\kappa+\eta}$, we have

$$|z + m_{sc}|^{-2} \le 1 - C\sqrt{\kappa+\eta}, \tag{5.32}$$

for some $C > 0$. Therefore (5.28) imply in this region that

$$\|\mathbf{u}\|_\infty \le C(\kappa+\eta)^{-1/2}\|\mathbf{w}\|_\infty. \tag{5.33}$$

Using (5.31), we get

$$\|\mathbf{u}\|_\infty \le \frac{C}{\sqrt{\kappa+\eta}}\|\mathbf{u}\|_\infty^2 + \frac{C}{\sqrt{\kappa+\eta}}\Upsilon, \tag{5.34}$$

and using $\|\mathbf{u}\|_\infty \ll \sqrt{\kappa+\eta}$ from (5.18), we conclude that

$$\|\mathbf{u}\|_\infty \le O\Big(\frac{\Upsilon}{\sqrt{\kappa+\eta}}\Big). \tag{5.35}$$

Combining this with the bound on $m - m_{sc}$ (5.27) and $G_{ii} - m_{sc} = m - m_{sc} + u_i$, we obtain (3.25).

Finally, we consider the main interesting regime: $|E| \le 2$ and $\kappa \ge \eta$. We claim that the following inequality about $m_{sc}(z)$ holds.

Lemma 5.3 *Let $1 > \delta_- > 0$ be a given constant. Then there exist small real numbers $\tau \ge 0$ and $c_1 > 0$, depending only on δ_-, such that we have*

$$\max_{x\in[-1+\delta_-,1-\delta_+]} \left\{ \left| \tau + x\, m_{sc}^2 \right|^2 \right\} \le (1 - c_1\, \widehat{g}(z))\,(1+\tau)^2 \tag{5.36}$$

with

$$\widehat{g}(z) = \max\{\delta_+,\ |1 - \mathrm{Re}\, m_{sc}^2(z)|\} \tag{5.37}$$

for any positive number δ_+ such that $-1 + \delta_- \le 1 - \delta_+$.

We postpone the proof of this lemma to the end of this subsection and we first complete the main argument. Recall that $B = \{\sigma_{ij}^2\}_{i,j=1}^N$ is the matrix of variances which is symmetric. We also recall $\delta_\pm$ from (2.4) and we will apply Lemma 5.3 with these δ_- and δ_+. Fix z, set $\zeta := m_{sc}^2(z) = (m_{sc}(z) + z)^{-2}$ and rewrite (5.28) as

$$\mathbf{u} = (I - \zeta B)^{-1}\mathbf{w} = \frac{1}{1+\tau}\Big[I - \frac{\zeta B + \tau}{1+\tau}\Big]^{-1}\mathbf{w} \tag{5.38}$$

with τ given in Lemma 5.3. Define $Q := I - |\mathbf{e}\rangle\langle\mathbf{e}|$ to be the projection onto the orthogonal complement of the normalized eigenvector $\mathbf{e} = N^{-1/2}(1, 1, \dots, 1)$ belonging to the simple eigenvalue 1 of B. Note that B and Q commute and that the spectrum of BQ lies in $[-1+\delta_-, 1-\delta_+]$. Denote by $\|A\|$ the usual $\ell^2 \to \ell^2$ norm of a matrix A. Since

$$\Big\| \frac{\zeta B + \tau}{1+\tau} Q \Big\| \le \sup_{x\in[-1+\delta_-,1-\delta_+]} \Big| \frac{\zeta x + \tau}{1+\tau} \Big| \le (1 - c_1\ \widehat{g}(z))^{1/2} < 1$$

by the Lemma 5.3 and $\mathbf{w} \perp (1, ..., 1)$, the Neumann expansion of (5.38) converges on span $((1, ..., 1))^\perp$ and

$$\mathbf{u} = (I - \zeta B)^{-1}\mathbf{w} = \frac{1}{1+\tau} \sum_{n=0}^{\infty} \Big(\frac{\zeta B + \tau}{1+\tau} \Big)^n \mathbf{w}. \tag{5.39}$$

We will compute the $\ell^\infty \to \ell^\infty$ norm of this matrix. First note that

$$\Big\| \frac{\zeta B + \tau}{1+\tau} \Big\|_{\infty\to\infty} = \max_i \sum_j \Big| \Big(\frac{\zeta B + \tau}{1+\tau} \Big)_{ij} \Big| \le \frac{1}{1+\tau} \max_i \sum_j |\zeta B_{ij} + \tau \delta_{ij}| \le \frac{|\zeta| + \tau}{1+\tau} \le 1, \tag{5.40}$$

since $|\zeta| = |m_{sc}|^2 \le 1$ and $\sum_j |B_{ij}| = \sum_j B_{ij} = \sum_j \sigma_{ij}^2 = 1$. Then we have

$$\Big\| \Big(\frac{\zeta B + \tau}{1+\tau} \Big)^n \mathbf{u} \Big\| = \Big\| \Big(\frac{\zeta B + \tau}{1+\tau} \Big)^n Q\mathbf{u} \Big\| \le \sup_{x\in[-1+\delta_-,1-\delta_+]} \Big| \frac{\zeta x + \tau}{1+\tau} \Big|^n \|\mathbf{u}\| \le (1 - c_1\ \widehat{g}(z))^{n/2} \|\mathbf{u}\|$$

by Lemma 5.3. Since for any $N \times N$ matrix we have

$$\|A\|_{\infty\to\infty} \le \sqrt{N} \|A\|,$$

we obtain

$$\left\| \left(\frac{\zeta B + \tau}{1 + \tau} \right)^n Q \right\|_{\infty \to \infty} \leq \sqrt{N} (1 - c_1 \, \widehat{g}(z))^{n/2}. \tag{5.41}$$

Thus, estimating the first $n \leq n_0 := (\log N)(c_1 \widehat{g}(z))^{-1}$ terms in (5.39) by (5.40), and the rest by (5.41), we get

$$\|\mathbf{u}\|_\infty \leq \left(\frac{\log N}{c_1 \widehat{g}(z)} + \sum_{n=n_0}^{\infty} \sqrt{N} (1 - c_1 \, \widehat{g}(z))^{n/2} \right) \|\mathbf{w}\|_\infty \leq C \frac{\log N}{\widehat{g}(z)} \|\mathbf{w}\|_\infty.$$

Using the bound (5.31) on $\|\mathbf{w}\|_\infty$ and the bound (5.18) on $\|\mathbf{u}\|_\infty$, we have

$$\|\mathbf{u}\|_\infty \leq (\log N)^{-1} \|\mathbf{u}\|_\infty + C \frac{\log N}{\widehat{g}(z)} \Upsilon \tag{5.42}$$

which implies

$$\|\mathbf{u}\|_\infty \leq C \frac{\log N}{\widehat{g}(z)} \Upsilon \tag{5.43}$$

for some $C > 0$. Combining this with (5.27), we find

$$|G_{ii} - m_{sc}| \leq |m(z) - m_{sc}(z)| + |u_i| \leq C \left[\|\mathbf{u}\|_\infty + \frac{\Upsilon}{\sqrt{\kappa + \eta}} \right] \leq C \left[\frac{\log N}{\hat{g}(z)} + \frac{1}{\sqrt{\kappa + \eta}} \right] \Upsilon$$

which implies (3.25), since $g(z) = \min\{\sqrt{\kappa + \eta}, \widehat{g}(z)\}$. □

Proof of Lemma 5.3. First, if $\widehat{g}(z) = \delta_+$, then we choose $\tau = 0$. With $|m_{sc}| \leq 1$ in (5.1), one can see that (5.36) holds.

In the case of $\widehat{g}(z) = |1 - \mathrm{Re}\, m_{sc}^2|$, we have $\widehat{g} \leq 2$ by using $|m_{sc}| \leq 1$. We choose $\tau = \delta_-/10$, then

$$\max_{x \in [-1+\delta_-, 1-\delta_+]} \left| \tau + x m_{sc}^2 \right|^2 \leq \max \left\{ \left| \frac{\delta_-}{10} + m_{sc}^2 \right|^2, \left| \frac{\delta_-}{10} - (1 - \delta_-) m_{sc}^2 \right|^2 \right\}, \tag{5.44}$$

and

$$\left| \frac{\delta_-}{10} - (1 - \delta_-) m_{sc}^2 \right|^2 \leq \left| 1 - \frac{9\delta_-}{10} \right|^2 \leq (1 - \tau \widehat{g}(z))(1 + \delta_-/10)^2 \,. \tag{5.45}$$

For the other term in r.h.s. of (5.44), we have

$$\left| \frac{\delta_-}{10} + m_{sc}^2 \right|^2 = |m_{sc}|^4 + 0.2\delta_- \mathrm{Re}\,(m_{sc}^2) + (\delta_-/10)^2 \,. \tag{5.46}$$

With $|m_{sc}| \leq 1$ in (5.1) and $\widehat{g}(z) = |1 - \mathrm{Re}\,(m_{sc}^2)|$ in this case, (5.46) is bounded as

$$\left| \frac{\delta_-}{10} + m_{sc}^2 \right|^2 \leq \left(1 + \frac{\delta_-}{10} \right)^2 - 0.2\delta_- \widehat{g}(z) \leq \left(1 + \frac{\delta_-}{10} \right)^2 (1 - C\widehat{g}(z)) \tag{5.47}$$

for some C depending on δ_-. At last, we complete the proof by combining (5.47) and (5.45). □

6 Proof of the universality of local statistics

We now outline the main steps to prove Theorem 2.2.

Step 1. Local relaxation flow. Following [19], we first prove that the local eigenvalue statistics of Dyson Brownian motion (DBM) at a fixed time t are the same as those of GUE if $t \asymp N^{-\varepsilon_0}$ for some ε_0. The DBM is generated by the flow

$$H_t = e^{-t/2} H_0 + (1 - e^{-t})^{1/2} V, \tag{6.1}$$

where H_0 is the initial matrix and V is an independent GUE matrix whose matrix elements are centered Gaussian random variables with variance $1/N$. Strictly speaking, for each matrix element we have used the Ornstein-Uhlenbeck (OU) process on $\mathbb{C}$ instead of the Brownian motion which was used in the original definition of DBM in [19]. It is easy to check that the eigenvalues of H_t follow a process, very similar to the original DBM in [19], but with a drift. With a slight abuse of terminology, we will still call this process DBM. More precisely, let

$$\mu = \mu_N(\mathrm{d}\mathbf{x}) = \frac{e^{-\mathcal{H}(\mathbf{x})}}{Z_\beta} \mathrm{d}\mathbf{x}, \qquad \mathcal{H}(\mathbf{x}) = N \left[\beta \sum_{i=1}^N \frac{x_i^2}{4} - \frac{\beta}{N} \sum_{i<j} \log |x_j - x_i| \right] \tag{6.2}$$

($\beta = 2$ for GUE) be the probability measure of the eigenvalues of the general β ensemble, $\beta \geq 1$ (in this section, we often use the notation x_j for the eigenvalues to follow the notations of [19]). In this paper we consider the $\beta = 2$ case for simplicity, but we stress that our proof applies to the case of symmetric matrices as well. Denote the distribution of the eigenvalues at time t by $f_t(\mathbf{x})\mu(\mathrm{d}\mathbf{x})$. Then f_t satisfies

$$\partial_t f_t = \mathscr{L} f_t. \tag{6.3}$$

where (see (2.2) in [19])

$$\mathscr{L} = \sum_{i=1}^N \frac{1}{2N} \partial_i^2 + \sum_{i=1}^N \left(-\frac{\beta}{4} x_i + \frac{\beta}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right) \partial_i. \tag{6.4}$$

Theorem 6.1 *Suppose that the probability law for the initial matrix H_0 satisfies the assumptions of Theorem 2.2. Then there exists $\varepsilon_0 > 0$ such that for any*

$$t \geq N^{-\varepsilon_0}, \tag{6.5}$$

the probability law for the eigenvalues of H_t satisfies (2.20)*, i.e., for any $k \geq 1$ and for any compactly supported continuous test function $O : \mathbb{R}^k \to \mathbb{R}$, we have*

$$\lim_{\kappa \to 0} \lim_{N \to \infty} \frac{1}{2\kappa} \int_{E-\kappa}^{E+\kappa} \mathrm{d}E' \int_{\mathbb{R}^k} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_k \, O(\alpha_1, \ldots, \alpha_k) \times \frac{1}{\varrho_{sc}(E)^k} \Big(p_N^{(k)} - p_{GUE,N}^{(k)} \Big) \Big(E' + \frac{\alpha_1}{N \varrho_{sc}(E)}, \ldots, E' + \frac{\alpha_k}{N \varrho_{sc}(E)} \Big) = 0,$$

where $p_{GUE,N}^{(k)}$ is the k-point correlation function of the GUE ensemble.

Proof of Theorem 6.1. We first recall the following general theorem concerning the Dyson Brownian motion from [19] that asserts that under four general assumptions, the local eigenvalue statistics of the time evolved matrix H_t coincide with GUE. The first assumption (called Assumption I in [19]) is a convexity bound on $\mathcal{H}$ which is automatically satisfied in our case and we only have to verify the following three assumptions.

Assumption II. There exists a continuous, compactly supported density function $\varrho(x) \geq 0$, $\int_{\mathbb{R}} \varrho = 1$, on the real line, independent of N, such that for any fixed $a, b \in \mathbb{R}$

$$\lim_{N\to\infty} \sup_{t\geq 0} \left| \int \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(x_j \in [a,b]) f_t(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) - \int_a^b \varrho(x) \mathrm{d}x \right| = 0. \tag{6.6}$$

For the next assumption, we introduce a notation. Let $\gamma_j = \gamma_{j,N}$ denote the location of the j-th point under the limiting density, i.e., γ_j is defined by

$$N \int_{-\infty}^{\gamma_j} \varrho(x) \mathrm{d}x = j, \qquad 1 \leq j \leq N, \quad \gamma_j \in \mathrm{supp}\varrho. \tag{6.7}$$

We will call γ_j the *classical location* of the j-th point.

Assumption III. There exists an $\varepsilon > 0$ such that

$$\sup_{t\geq 0} \int \frac{1}{N} \sum_{j=1}^{N} (x_j - \gamma_j)^2 f_t(\mathrm{d}\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) \leq C N^{-1-2\varepsilon} \tag{6.8}$$

with a constant C uniformly in N.

The final assumption is an upper bound on the local density. For any $I \in \mathbb{R}$, let

$$\mathcal{N}_I := \sum_{i=1}^{N} \mathbf{1}(x_i \in I)$$

denote the number of points in I.

Assumption IV. For any compact subinterval $I_0 \subset \{E \,:\, \varrho(E) > 0\}$ independent of N, and for any $\delta > 0$, $\sigma > 0$ and $r > 0$, there are constants c depending on I_0, δ, σ and r such that for any interval $I \subset I_0$ with $|I| \geq N^{-1+\sigma}$, we have

$$\sup_{\tau \geq N^{-2\varepsilon+\delta}} \int \mathbf{1}\big\{\mathcal{N}_I \geq KN|I|\big\} f_\tau \mathrm{d}\mu \leq N^{-c\log\log N}, \quad K = N^r \tag{6.9}$$

where ε is the exponent from Assumption III.

Theorem 6.2 *[19, Theorem 2.1] Let $\varepsilon > 0$ be the exponent from Assumption III. Suppose that there is a time $\tau < N^{-2\varepsilon}$ such that the following entropy bound holds*

$$S_\mu(f_\tau) := \int f_\tau \log f_\tau \mathrm{d}\mu \leq C N^m \tag{6.10}$$

for some fixed exponent m. Suppose that the Assumptions II, III and IV hold for the solution f_t of the forward equation (6.3) for all time $t \geq \tau$. Let $E \in \mathbb{R}$ be a point where $\varrho(E) > 0$. Then for any $k \geq 1$ and for any compactly supported continuous test function $O : \mathbb{R}^k \to \mathbb{R}$, we have

$$\lim_{b\to 0}\lim_{N\to\infty}\sup_{t\geq N^{-2\varepsilon+\delta}} \frac{1}{2b}\int_{E-b}^{E+b}\mathrm{d}E'\int_{\mathbb{R}^k}\mathrm{d}\alpha_1\ldots\mathrm{d}\alpha_k\, O(\alpha_1,\ldots,\alpha_k) \times \frac{1}{\varrho(E)^k}\Big(p^{(k)}_{t,N}-p^{(k)}_{\mu,N}\Big)\Big(E'+\frac{\alpha_1}{N\varrho(E)},\ldots,E'+\frac{\alpha_k}{N\varrho(E)}\Big)=0\,. \tag{6.11}$$

Theorem 6.2 was exactly Theorem 2.1 of [19] except that the assumption (6.10) on the entropy in [19] was stated for the initial probability density f_0. Clearly, we can start the flow (6.4) from a fixed time $\tau \ll N^{-2\varepsilon+\delta}$ since the statement of Theorem 6.2 concerns only the time $t \geq N^{-2\varepsilon+\delta}$. In the case that the flow (6.4) is generated from the matrix evolution (6.1), the entropy assumption (6.10) is satisfied automatically. To see this, let ν_t^{ij} denote the probability measure of the ij-th element of the matrix H_t, $i \leq j$, and $\bar{\nu}_t$ the probability measure of the matrix H_t. Let $\bar{\mu}$ denote the probability measure of the GUE and μ^{ij} the probability measure of its ij-th element which is a Gaussian measure with mean zero and variance $1/N$. Since the dynamics of matrix elements are independent (subject to the Hermitian condition), we have the identity

$$\int \log\left(\frac{\mathrm{d}\bar{\nu}_t}{\mathrm{d}\bar{\mu}}\right)\mathrm{d}\bar{\nu}_t = \sum_{ij}\int \log\left(\frac{\mathrm{d}\nu_t^{ij}}{\mathrm{d}\mu^{ij}}\right)\mathrm{d}\nu_t^{ij}. \tag{6.12}$$

The process $t \to \nu_t^{ij}$ is an Ornstein-Uhlenbeck process and each entropy term on the right hand side of the last equation is bounded by CN provided that $t \geq 1/N$ and ν_0^{ij} has a subexponential decay. It is easy to check from the explicit OU kernel. Since the entropy of the marginal distribution on the eigenvalues is bounded by the entropy of the total measure on the matrix, we have proved that

$$\int f_{1/N}\log f_{1/N}\mathrm{d}\mu \leq CN^3, \tag{6.13}$$

and this verifies (6.10). Therefore, in order to apply Theorem 6.2, we only have to verify the Assumptions II, III and IV. Clearly, Assumption II follows from Theorem 2.1 (note that in the case of generalized Wigner matrix, $M \asymp N$ and $g(z) \asymp \sqrt{\kappa+\eta}$). Assumption IV also follows from Theorem 2.1 by noting that $\mathcal{N}_I \leq C\,\mathrm{Im}(E+i\eta)$ if I is an interval of length η about E. We also note that Assumption IV in [19] was stated in a slightly stronger form, requiring a large deviation bound (6.9) for all $K \geq 1$, but inspecting the proof of Theorem 2.1 of [19] reveals that Assumption IV is used only for K larger than some positive power of N and smaller than N (the main observation is that the upper limit of the summation in (7.16) of [19] is effectively N and not ∞).

Having verified all other assumptions, it remains to prove (6.8), which we state as the next theorem.

Theorem 6.3 *Suppose H satisfies the assumptions of Theorem 2.2, in particular, it is a generalized Wigner matrix with positive constants C_{inf}, C_{sup} in (2.6). Let $\widetilde{\nu}_{ij}(x)\mathrm{d}x := \sigma_{ij}\nu_{ij}(\sigma_{ij}x)\mathrm{d}x$ be the rescaling of the distributions ν_{ij} of the matrix elements and suppose that they satisfy the logarithmic Sobolev inequality (LSI) with a constant C_S independent of N, i, j, i.e.,*

$$\int u\log u\mathrm{d}\widetilde{\nu}_{ij} \leq C_S\int|\nabla\sqrt{u}|^2\mathrm{d}\widetilde{\nu}_{ij} \tag{6.14}$$

holds for any smooth probability density u, $\int u \mathrm{d}\widetilde{\nu}_{ij} = 1$. Denote λ_i the i-th eigenvalue of H in increasing order, $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$. Then there exists $\varepsilon > 0$ depending on α, β in (2.14) but independent of C_{inf}, C_{sup} and C_S such that

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(\lambda_i - \gamma_i)^2 \le CN^{-1-2\varepsilon}, \tag{6.15}$$

if N is sufficiently large (depending on C_{inf}, C_{sup}, C_S, α and β).

The proof of Theorem 6.3 will be given in Section 7. It is easy to check that if an initial matrix $H = H_0$ satisfies the conditions of Theorem 6.3, then its evolution H_t under the Ornstein-Uhlenbeck flow will also satisfy these conditions with constants changed at most by a factor two. The main condition to check is that the logarithmic Sobolev inequality (6.14) holds for $0 \le t \ll 1$. But this was proved in the argument following Lemma 5.3 of [15] using an estimate on the logarithmic Sobolev constant for convolution of two measures, i.e., Lemma B.1 of [15]. Therefore Theorem 6.3 guarantees (6.15) for all positive times $t > 0$ and this proves Assumption III provided that the initial distribution satisfies the LSI (6.14). We have thus proved Theorem 2.2 for matrix ensembles of the form

$$h_{ij} = e^{-t/2}\widehat{h}_{ij} + (1 - e^{-t})^{1/2} N^{-1/2} \xi_{ij}^G, \quad t \ge N^{-2\varepsilon+\delta} \tag{6.16}$$

where ξ_{ij}^G are i.i.d. complex random variables with Gaussian distribution with mean 0 and variance 1, and $\widehat{h}_{ij}$'s are independent random variables such that the rescaled variables $\widehat{\zeta}_{ij} = \widehat{h}_{ij}/\sigma_{ij}$ satisfy th LSI assumption (6.14). In (6.16) $\delta > 0$ is arbitrary and ε is fixed in Theorem 6.3. In particular, with the choice $\delta = \varepsilon$ and $t \asymp N^{-\varepsilon}$, we have proved Theorem 2.2 for matrix ensembles $h_{ij} = \sigma_{ij}\zeta_{ij}$ if ζ_{ij} is of the form

$$\zeta_{ij} = (1-\gamma)^{1/2}\widehat{\zeta}_{ij} + \gamma^{1/2}\xi_{ij}^G, \quad \gamma \asymp N^{-\varepsilon}, \qquad \text{distribution of } \widehat{\zeta}_{ij} \text{ satisfies (6.14).} \tag{6.17}$$

Step 2. Eigenvalue correlation function comparison theorem.

The next step is to prove that the correlation functions of eigenvalues for two matrix ensembles are identical up to scale $1/N$ provided that the first four moments of all matrix elements of these two ensembles are almost identical. This theorem is a corollary of Theorem 2.3 and we state it as the following correlation function comparison theorem. The proof will be given in Section 8. Note that the assumption (2.21) in Theorem 2.3 is satisfied by Theorem 3.1; in case of generalized Wigner matrix we have $g(z) = \sqrt{\kappa + \eta}$ and $M \asymp N$, so in the regime where $|E|$ is separated away from 2, we have from (3.4), that $G_{ii}(z)$ is uniformly bounded (modulo logarithmic factors).

Theorem 6.4 *Suppose the assumptions of Theorem 2.3 hold. Let $p_{v,N}^{(k)}$ and $p_{w,N}^{(k)}$ be the $k-$point functions of the eigenvalues w.r.t. the probability law of the matrix $H^{(v)}$ and $H^{(w)}$, respectively. Then for any $|E| < 2$, any $k \ge 1$ and any compactly supported continuous test function $O : \mathbb{R}^k \to \mathbb{R}$ we have*

$$\int_{\mathbb{R}^k} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_k \, O(\alpha_1, \ldots, \alpha_k)\Big(p_{v,N}^{(k)} - p_{w,N}^{(k)}\Big)\Big(E + \frac{\alpha_1}{N}, \ldots, E + \frac{\alpha_k}{N}\Big) = 0. \tag{6.18}$$

Step 3. Approximation of a measure by Ornstein-Uhlenbeck process for small time.

Summarizing, we have proved Theorem 2.2 in Step 1 for matrix ensembles whose probability distributions of the normalized matrix elements ζ_{ij} are of the form (6.17). Using the Green's function comparison theorem,

i.e. Theorem 6.4, we extended the class of distributions to all random variables whose first four moments can almost be matched (more precisely, match the first three moments and almost match the fourth moments in the sense of (2.22)) by random variables in the class (6.17). In order to complete the proof of Theorem 2.2, it remains to prove that for all measures in the class given by the assumptions of Theorem 2.2, i.e., measures satisfying the subexponential decay condition, the uniformly bounded-variance condition (2.6) and the moment restriction (2.19) for the real and imaginary parts, we can find random variables in the class (6.17) to almost match the first four moments. Since the real and imaginary parts are i.i.d., it is sufficient to match them individually, i.e., we can work with real random variables normalized to variance one. This is the content of the following Lemma 6.5. Notice that the uniformity in the conditions (2.19) and (2.14) guarantees that the bounds (6.19) hold with uniform constants C_1, C_2. This implies the uniformity of the LSI constants, needed in Theorem 6.3, for the random variables constructed in Lemma 6.5. The proof of this Lemma will be given in Appendix C. We have thus proved Theorem 2.2. □

Lemma 6.5 *Let m_3 and m_4 be two real numbers such that*

$$m_4 - m_3^2 - 1 \geq C_1, \;\; m_4 \leq C_2 \tag{6.19}$$

for some positive constants C_1 and C_2. Then for any sufficient small $\gamma > 0$ (depending on C_1 and C_2), there exists a real random variable ξ_γ whose distribution satisfies LSI and the first 4 moments of

$$\xi' = (1-\gamma)^{1/2}\xi_\gamma + \gamma^{1/2}\xi^G \tag{6.20}$$

are 0, 1, $m_3(\xi') = m_3$ and $m_4(\xi')$, and

$$|m_4(\xi') - m_4| \leq C\gamma \tag{6.21}$$

for some C depending on C_1 and C_2, where ξ^G is real Gaussian random variable with mean 0 and variance 1, independent of ξ_γ. The LSI constant of ξ_γ (and thus ξ') is bounded from above by a function of C_1 and C_2.

7 Proof of Theorem 6.3

Theorem 6.3 states that the eigenvalues are at a distance $N^{-1/2-\varepsilon}$ from their classical locations in a quadratic average sense. We will deduce this conclusion from the information on the closeness of the local density to the semicircle law. We note the constants appearing in this section may also depend on α and β in (2.14), but we will not mention the dependence in the proof.

First we reformulate a result, which we have proved in [19], in a somewhat more general setup. It states that random points, λ_j, are close to a fixed set of locations, γ_j, if the local fluctuation is controlled, if the averaged counting function is close to the counting function of the γ_j's in L^1-sense and if some tightness holds. For simplicity, the result is stated for the case when γ_j's are the classical locations given by the semicircle law $\varrho = \varrho_{sc}$, (6.7), but the statement (and its proof) holds for any density function with support being a compact interval and with square root singularity at the edges. In particular, we applied this result in [19] for the Marchenko-Pastur (MP) distribution instead of the semicircle law. The counting function of γ_j can be replaced by its continuous version, i.e., by the distribution function of the semicircle law which defined by

$$n_{sc}(E) := \int_{-\infty}^{E} \varrho_{sc}(x)\mathrm{d}x. \tag{7.1}$$

Lemma 7.1 *Let $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$ be an ordered collection of random points in $\mathbb{R}$. Denote the averaged counting function of λ_j's*

$$n^\lambda(E) = \frac{1}{N}\mathbb{E}\#[\lambda_j \le E]. \tag{7.2}$$

Suppose the following four assumptions hold.

1. *[Tightness at the edge] There exist $m < 7$ and $\varepsilon > 0$ such that*

$$n^\lambda(-2 - N^{-1/m}) \le Ce^{-N^\varepsilon} \quad \text{and} \quad n^\lambda(2 + N^{-1/m}) \ge 1 - Ce^{-N^\varepsilon} \tag{7.3}$$

   and for any $K \ge 3$,

$$n^\lambda(K) \ge 1 - e^{-N^\varepsilon \log K} \quad \text{and} \quad n^\lambda(-K) \le e^{-N^\varepsilon \log K}. \tag{7.4}$$

2. *[L^1-closeness of the counting functions]*

$$\int_{-\infty}^{\infty} \left|n^\lambda(E) - n_{sc}(E)\right| \mathrm{d}E \le CN^{-6/7}. \tag{7.5}$$

3. *[Fluctuation of moving averages] For any small $\delta > 0$ there is a constant C such that for any $j, K \in \mathbb{N}$ with $j + K \le N + 1$, the local averages $\lambda_{j,K} := K^{-1}\sum_{i=0}^{K-1} \lambda_{j+i}$ satisfy*

$$\mathbb{P}\left(|\lambda_{j,K} - \mathbb{E}\,\lambda_{j,K}| \ge N^{-1/2+\delta}K^{-1/2}\right) \le Ce^{-N^{\delta/2}}. \tag{7.6}$$

4. *[Positivity of the bulk density] There exists a small enough $\delta > 0$ such that: for any interval I with $|I| = N^{-5/8}$ and $I \subset [-2 + N^{-\delta}, 2 - N^{-\delta}]$, the number of the λ's in I is bounded from below as follows*

$$\mathbb{P}\left(\#\{\lambda_j \in I\} \ge N^{-\delta}N|I|\right) \ge 1 - CN^{c\log\log N}. \tag{7.7}$$

Then there exists $\varepsilon > 0$ (independent of the constants in these four assumptions) such that

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(\lambda_i - \gamma_i)^2 \le CN^{-1-\varepsilon} \tag{7.8}$$

when N is large enough (depending on the constants in these four assumptions).

Proof of Lemma 7.1. In Theorem 9.1 of [19] we have proved the analogous result on the singular values of the covariance matrix, where the role of the semicircle law was played by the MP law and the spectral edges, ± 2, were replaced by $\lambda_\pm$, the two edges of the support of the MP distribution. In that paper we first proved the analogues of these four assumptions, then we presented the proof of (7.8) via a general argument that used only these assumptions. Inspecting the proofs of Lemma 9.5, 9.6 and 9.7 in [19], leading to (7.8), we observe that only equations (9.6), (9.8), (9.9) and (9.13) from [19] were used, in addition to the lower bound on the density of the points in the scale $N^{-5/8}$, which is used below (9.51) of [19]. The lower bound on the density is granted by the last assumption (7.7) (even with a better control on the probability than we required in [19]). Repeating the argument from [19], for the proof of Lemma 7.1 it is sufficient to check that the first three assumptions in Lemma 7.1 imply equations (9.6), (9.8), (9.9) and (9.13) in [19]. We now explain how to obtain these necessary bounds from our assumptions.

The first condition (7.3) corresponds to the input for Lemma 9.2 in [19], in particular, the analogue of (9.6) of [19],

$$-2 - N^{-1/m} \leq \mathbb{E}\lambda_j \leq 2 + N^{1/m},$$

follows immediately from (7.3) and (7.4). We note that (9.6) in [19] contains a threshold $N^{-1/5}$ but actually in the proof we only needed it to be much less than $N^{-1/7}$ (see (9.36)–(9.37) of [19] for the application of (9.6)).

The second condition (7.5) corresponds to Eq. (9.8) in [19]. As we showed in the proof of Lemma 9.3 of [19], Eq. (9.9) directly follows from (9.8). Here the analogous bound

$$\sup_E \left|n^\lambda(E) - n_{sc}(E)\right| \leq CN^{3/7}$$

follows directly from (7.5) in the same way.

Finally, the third condition (7.6) is exactly the same as (9.13) in [19]. Simply repeating now the proof of Theorem 9.1 from [19], we proved Lemma 7.1. □

Theorem 6.3 will now follow from Lemma 7.1 if we prove that the four conditions in the Lemma 7.1 hold in the case of generalized Wigner matrices (2.6). The last condition (7.7) follows from the local semicircle law (Theorem 2.1) and from the fact that $\varrho_{sc}(x) \geq c\sqrt{\kappa}$ for $x \in (-2+\kappa, 2-\kappa)$. Here we list the first three conditions as three separate lemmas that will be proven in the next three subsections. This will complete the proof of Theorem 6.3. □

Lemma 7.2 *(1) Let H be a generalized Wigner matrix with subexponential decay, in fact it is sufficient to assume that* (2.14) *and the upper bound $C_{sup} < \infty$ in* (2.6) *hold. Define $n^\lambda(E)$ as in* (7.2)*. Then*

$$n^\lambda(-2 - N^{-1/6+\varepsilon}) \leq Ce^{-N^{\varepsilon'}} \text{ and } n^\lambda(2 + N^{-1/6+\varepsilon}) \geq 1 - Ce^{-N^{\varepsilon'}} \tag{7.9}$$

for any small $\varepsilon > 0$ with an $\varepsilon' > 0$ depending on ε. Furthermore, for $K \geq 3$,

$$n^\lambda(-K) \leq e^{-N^\varepsilon \log K} \text{ and } n^\lambda(K) \geq 1 - e^{-N^\varepsilon \log K} \tag{7.10}$$

for some $\varepsilon > 0$.

(2) In fact, the last tightness bound holds in a more general situation, namely, let the universal Wigner matrix H satisfy (2.1), (2.14) *and $M \geq (\log N)^9$ where M is defined in* (2.3)*. Then we have*

$$n^\lambda(-3) \leq CN^{-c\log\log N} \text{ and } n^\lambda(3) \geq 1 - CN^{-c\log\log N}. \tag{7.11}$$

Lemma 7.3 *Let H satisfy the conditions of Theorem 6.3. Then for any $\varepsilon > 0$ we have*

$$\int_{-\infty}^{\infty} \left|n^\lambda(E) - n_{sc}^\lambda(E)\right| \mathrm{d}E \leq CN^{-1+\varepsilon}. \tag{7.12}$$

Lemma 7.4 *Let H satisfy the conditions in Theorem 6.3, in particular, let the distribution of the matrix elements satisfy the uniform LSI* (6.14)*. For $j, K \in \mathbb{N}$, $j + K \leq N + 1$, define $\lambda_{j,K} = K^{-1}\sum_{i=0}^{K-1} \lambda_{j+i}$. Then for any $\delta > 0$ small enough,*

$$\mathbb{P}\left(|\lambda_{j,K} - \mathbb{E}(\lambda_{j,K})| \geq N^{-1/2+\delta}K^{-1/2}\right) \leq Ce^{-N^\delta}, \tag{7.13}$$

with C depending on C_{sup} in (2.6) *and C_S in* (6.14)*.*

7.1 Proof of Lemma 7.2.

Extreme eigenvalues are typically controlled by the moment method, evaluating $\mathbb{E}\,\mathrm{Tr}\,H^k$ for large k using some graphical representation. Our proof follows the standard path, but since we were unable to find a reference that would apply precisely to our case, we include the proof for completeness. The main technical estimate (7.18) is borrowed from [34]. We remark that if we use the strongest result in [34], one can improve the exponent 1/6 to 1/4 in (7.9).

We start with the proof of (7.9) and (7.10) in the case of generalized Wigner matrices (see (2.6)). First we truncate the random variables. With the assumption of subexponential decay of h_{ij}, for any small $\delta > 0$, one can find a $\widehat{h}_{ij}$ such that

$$\mathbb{P}(\widehat{h}_{ij} = h_{ij}) \geq 1 - e^{-N^{\varepsilon'}} \tag{7.14}$$

and

$$|\widehat{h}_{ij}| \leq N^{-1/2+\delta}, \quad \mathbb{E}(\widehat{h}_{ij}) = 0, \quad \mathbb{E}(|\widehat{h}_{ij}|^2) \leq \mathbb{E}(|h_{ij}|^2) \tag{7.15}$$

for some small number ε', depending on δ. Then we only need to bound the spectral norm of the new matrix $\widehat{H} = (\widehat{h}_{ij})$. To prove (7.9), it only remains to prove that, for some small $\varepsilon' > 0$,

$$\mathbb{P}(\|\widehat{H}\| \geq 2 + N^{-1/6+\varepsilon}) \leq e^{-N^{\varepsilon'}}. \tag{7.16}$$

With

$$\mathbb{P}(\|\widehat{H}\| \geq 2 + N^{-1/6+\varepsilon}) \leq \frac{\mathbb{E}\mathrm{Tr}\,\widehat{H}^k}{(2 + N^{-1/6+\varepsilon})^k},$$

for any even k, (7.16) follows from

$$\mathbb{E}\,\mathrm{Tr}\,\widehat{H}^{k_0} \leq 2^{k_0 + O(\log N)}, \tag{7.17}$$

with the choice of $k_0 = N^{1/6-\delta/3}$ and $\delta = 3\varepsilon/2$, since $\|\widehat{H}\|^k \leq \mathrm{Tr}\,\widehat{H}^k$ for even powers. The proof of (7.10) is analogous.

To estimate $\mathbb{E}(\mathrm{Tr}\,\widehat{H}^k)$ for $k \in \mathbb{N}$, we start with introducing some notations and concepts on graphs.

Let p and k be given integers. We define the concept of *ordered closed walk* of k edges on an abstract ordered set $A_p := \{a_1, a_2, \ldots, a_p\}$ of p elements with the natural ordering $a_1 < a_2 < \ldots < a_p$. An ordered closed walk on p vertices with k edge is determined by a sequence $\underline{w} = (w_1, w_2, \ldots, w_k)$ of the elements of A_p with the following properties:

i) Along the walk, the fresh vertices from A_p are adjoined in increasing order, i.e., $\max_{j \leq m} w_j \leq \max_{j \leq m-1} w_j + 1$.

ii) $\{w_1, w_2, \ldots w_k\} = A_p$, i.e., all points of A_p are visited.

iii) Let $\Gamma(\underline{w})$ denote the undirected graph associated with $\underline{w}$, i.e., the vertex set of $\Gamma(\underline{w})$ is A_p, the edges are given by $(w_1, w_2), (w_2, w_3), \ldots (w_k, w_1)$; with multiple edges as well as self-loops ($w_i = w_{i+1}$ for some i) allowed. Then every edge of Γ appears at least twice.

Let $\mathcal{W}(k, p)$ denote the set of ordered closed walks on p vertices with k edges. Their number was estimated in Lemma 2.1 of [34]

$$W(k, p) := |\mathcal{W}(k, p)| \leq \binom{k}{2p-2} p^{2(k-2p+2)} 2^{2p-2}. \tag{7.18}$$

This bound will be sufficient for the proof of (7.9) with exponent $1/6+\varepsilon$. We remark that Lemma 4.1 of [34] gives a different bound on (7.18) that is better by essentially a factor $[(k-2p)/p]^{k-2p}$. Applying this bound, one could improve the exponent in (7.9) to $1/4+\varepsilon$ but we will not pursue this improvement here.

We also need the concept of *labelling* the elements of A_p by the set $\{1,2,\dots,N\}$. A labelling is given by a function $\ell : A_p \to \{1,2,\dots,N\}$ and we require that ℓ be injective. The set of such labelling functions is denoted by $\mathcal{L}(p,N)$.

With these notations, we have the formula

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\,\widehat{H}^k &= \sum_{i_1,i_2,\dots i_k=1}^{N} \mathbb{E}\widehat{h}_{i_1 i_2}\widehat{h}_{i_2 i_3}\dots\widehat{h}_{i_k i_1} \\ &= \sum_{p=1}^{k/2+1} \sum_{\underline{w}\in\mathcal{W}(k,p)} \sum_{\ell\in\mathcal{L}(p,N)} \mathbb{E}\widehat{h}_{\ell(w_1)\ell(w_2)}\widehat{h}_{\ell(w_2)\ell(w_3)}\dots\widehat{h}_{\ell(w_k)\ell(w_1)}. \end{aligned} \tag{7.19}$$

To verify this formula, for any given sequence $i_1, i_2, \dots, i_k$ on the l.h.s., let p denote the number of different elements in this sequence and let the set A_p be identified with these different elements in the order of their appearance (i.e. for any m we let $a_m := i_s$ for some s if $i_s \neq i_t$, $t<s$, and i_s is the m-th freshest element among $i_1, i_2, \dots, i_s$, i.e., $|\{i_1, i_2, \dots, i_{s-1}\}| = m-1$). Let $w_1, w_2, \dots, w_k$ encode the sequence $i_1, i_2, \dots, i_k$ with the new labels $a_1, a_2, \dots a_p$. One may think of the walk, $w_1, w_2, \dots, w_k$, as the topological structure of the sequence $(i_1, i_2, \dots, i_k)$ where the original labels from the set $\{1,2,\dots,N\}$ have been replaced by abstract labels, defined intrinsically from the repetition structure of $(i_1, i_2, \dots, i_k)$. Formula (7.19) is a resummation of all sequences $(i_1, i_2, \dots, i_k)$ in terms of topological walks (first and second sum) and then reintroducing the original labelling with $\{1,2,\dots,N\}$ (third sum). Since the first moment of $\widehat{h}_{ij}$ vanishes and different matrix elements are independent, all terms on the right hand side have zero expectation in which at least one factor $\widehat{h}_{ij}$ appears only once. This justifies the requirement iii) in the definition of the ordered closed walks. The restriction $p \le k/2+1$ in the summation then comes from iii). This proves (7.19).

To compute the expectation on the r.h.s. of the (7.19), we need to introduce the concept of the *skeleton of the walk*. Given $\underline{w} \in \mathcal{W}(k,p)$, its skeleton $S(\underline{w})$ is the undirected graph on A_p that is obtained from $\Gamma(\underline{w})$ after replacing each multiple (parallel) edge by a single undirected edge. Here $S(\underline{w})$ allows self-loops (as long as every edge has multiplicity 1). Thus the edge set $E(S(\underline{w}))$ of the skeleton coincides with the edge set $E(\Gamma(\underline{w}))$ after neglecting multiplicity and direction. The skeleton is a subgraph of the complete graph on A_p. We will also define the *tree of the walk*, $T(\underline{w})$, which is just a spanning tree of the skeleton $S(\underline{w})$ built up successively along the walk by a greedy algorithm: include an edge to the $T(\underline{w})$ if it does not create a loop together with the previously adjoined edges. Since $\Gamma(\underline{w})$ is connected, and then so is $S(\underline{w})$, thus $T(\underline{w})$ is indeed a tree on p vertices, in particular the number of its edges is

$$|E(T(\underline{w}))| = p-1. \tag{7.20}$$

and $S(\underline{w}) \setminus T(\underline{w})$ has total edge multiplicity less than $k-2(p-1)$.

For any edge $e \in E(S(\underline{w}))$ of the skeleton, let $\nu(e)$ denote the multiplicity of e in $\Gamma(\underline{w})$ (edges with both orientations are taken into account). Clearly

$$\sum_{e\in E(S)} \nu(e) = k \tag{7.21}$$

for any skeleton graph $S = S(\underline{w})$ for $w \in \mathcal{W}(k,p)$. Finally, for a given edge $e = (a_\alpha, a_\beta)$ in a subgraph of A_p and for any labelling $\ell \in \mathcal{L}(p,N)$, we define the induced labelling of the edge e by $\ell(e) = (\ell(a_\alpha), \ell(a_\beta))$.

With these notations we have

$$\left|\mathbb{E}\widehat{h}_{\ell(w_1)\ell(w_2)}\widehat{h}_{\ell(w_2)\ell(w_3)}\dots\widehat{h}_{\ell(w_k)\ell(w_1)}\right| \le \prod_{e\in E(S(\underline{w}))} \mathbb{E}\,|\widehat{h}_{\ell(e)}|^{\nu(e)}.$$

Note that $|\widehat{h}_{ij}| = |\widehat{h}_{ji}|$, therefore there is no ambiguity in the notation $|\widehat{h}_{\ell(e)}|$. Since $|\widehat{h}| \le N^{-1/2+\delta}$ and $\nu(e) \ge 2$, we have

$$\mathbb{E}\,|\widehat{h}_{\ell(e)}|^{\nu(e)} \le N^{(-1/2+\delta)(\nu(e)-2)}\sigma^2_{\ell(e)}, \tag{7.22}$$

or, alternatively,

$$\mathbb{E}\,|\widehat{h}_{\ell(e)}|^{\nu(e)} \le N^{(-1/2+\delta)\nu(e)}. \tag{7.23}$$

We will use (7.22) for the edges of the tree, $e \in E(T(\underline{w}))$, and we use (7.23) for the remaining edges $e \in E(S(\underline{w})) \setminus E(T(\underline{w}))$. We can now estimate (7.19) using (7.21) and (7.20):

$$\begin{aligned}
|\mathbb{E}\,\mathrm{Tr}\,\widehat{H}^k| &\le \sum_{p=1}^{k/2+1} \sum_{\underline{w}\in\mathcal{W}(k,p)} \sum_{\ell\in\mathcal{L}(p,N)} \prod_{e\in E(S(\underline{w}))} \mathbb{E}\,|\widehat{h}_{\ell(e)}|^{\nu(e)} \\
&\le \sum_{p=1}^{k/2+1} \sum_{\underline{w}\in\mathcal{W}(k,p)} N^{(-1/2+\delta)(k-2(p-1))} \sum_{\ell\in\mathcal{L}(p,N)} \prod_{e\in E(T(\underline{w}))} \sigma^2_{\ell(e)} \\
&\le \sum_{p=1}^{k/2+1} \sum_{\underline{w}\in\mathcal{W}(k,p)} N^{1+(-1/2+\delta)(k-2(p-1))}.
\end{aligned} \tag{7.24}$$

In the last step we used that

$$\sum_{\ell\in\mathcal{L}(p,N)} \prod_{e\in E(T)} \sigma^2_{\ell(e)} \le N.$$

holds for any tree T. This identity follows from successively summing up the labels for vertices with degree one in T by using the identity $\sum_i \sigma^2_{ij} = 1$.

Using (7.18), we obtain the bound

$$|\mathbb{E}\,\mathrm{Tr}\,\widehat{H}^k| \le \sum_{p=1}^{k/2+1} S(k,p), \tag{7.25}$$

with

$$S(k,p) := \binom{k}{2p-2} p^{2(k-2p+2)} 2^{2p-2} N^{1+(-1/2+\delta)(k-2(p-1))}. \tag{7.26}$$

It is easy to show that

$$S(k,p-1) \le \frac{N^{2\delta}k^6}{4N} S(k,p). \tag{7.27}$$

Choosing $k = N^{1/6-\delta/3}$, we have $S(k,p-1) \le S(k,p)$. Inserting this into (7.25), we obtain (7.17) and complete the proof.

Now we prove (7.11) with the same method. Similarly, with the assumption on the distribution of h_{ij}, one can find a $\widehat{h}_{ij}$ such that

$$\mathbb{P}(\widehat{h}_{ij} = h_{ij}) \ge 1 - CN^{-c\log\log N} \tag{7.28}$$

and

$$|\widehat{h}_{ij}| \le M^{-1/2}n, \quad \mathbb{E}(\widehat{h}_{ij}) = 0, \quad \mathbb{E}(|\widehat{h}_{ij}|^2) \le \mathbb{E}(|h_{ij}|^2) + N^{-c\log\log N} \tag{7.29}$$

for $n = (\log N)(\log\log N)$. Here $\widehat{h}_{ij}$ can be obtained by considering the cutoff random variables $h_{ij}\mathbf{1}\big(|h_{ij}| \le M^{-1/2}(\log N)(\log\log N)\big)$ and then slightly modifying them to recover their zero expectation value.

We can again bound $|\mathbb{E}\ \mathrm{Tr}\,\widehat{H}^k|$ as in (7.25) but with a slightly different $S(k,p)$; instead of the factor $N^{1+(-1/2+\delta)(k-2(p-1))}$ we will have $N\cdot M^{(-1/2+\delta)(k-2(p-1))}$ in the definition (7.26). These modified $S(k,p)$ numbers satisfy

$$S(k,p-1) \le \frac{n^2k^6}{4M}S(k,p) \tag{7.30}$$

and

$$S(k,k/2+1) = 2^k\cdot N. \tag{7.31}$$

Choosing $k=n$, we have $\frac{n^2k^6}{4M} < 1$. Thus we obtain

$$|\mathbb{E}\ \mathrm{Tr}\,\widehat{H}^k| \le 2^k\cdot 2nN, \tag{7.32}$$

which implies (7.11) . □

7.2 Proof of Lemma 7.3.

First we show that the estimate on the expectation of $m-m_{sc}$ is better than the estimate (2.16) on $m-m_{sc}$ itself.

Lemma 7.5 *Assume that the $N\times N$ generalized Wigner matrix H (see (2.6)) satisfies (2.1), (2.4), (2.5) and (2.14), $\mathbb{E}\,h_{ij}=0$, for any $1\le i,j\le N$ (i.e. the assumptions of Theorem 3.1 apart from (3.3) hold). Then we have, with some $C>0$,*

$$|\mathbb{E}m(z) - m_{sc}(z)| \le \frac{(\log N)^C}{(N\eta)(\kappa+\eta)} \tag{7.33}$$

for any $z=E+i\eta$, $\eta>0$.

As a preparation to the proof, we need the following technical lemma that we state under more general conditions so that it is applicable for universal Wigner matrices.

Lemma 7.6 *With the assumption of Theorem 3.1, suppose (3.3) holds, we have the estimate*

$$\left|\mathbb{E}m(z) + \frac{1}{\mathbb{E}m(z)+z}\right| \le \frac{(\log N)^{C_0}\sqrt{\kappa+\eta}}{(M\eta)g^2(z)} \tag{7.34}$$

for some sufficiently large positive constants C_0 (depending on α,β in (2.14)).

Proof of Lemma 7.6. Recall the definitions of Ω_z^o, Ω_z^d and $\widehat{\Omega}_z^\Upsilon$ in (3.17), (3.12), (3.13) and (3.14) and we define

$$\Omega_z \equiv \Omega_z^o \cap \Omega_z^d \cap \widehat{\Omega}_z^\Upsilon. \tag{7.35}$$

With (3.34), we have

$$\mathbb{P}(\Omega_z) \ge 1 - CN^{-c\log\log N}. \tag{7.36}$$

The r.h.s. of (7.34) is larger than N^{-2}. Then with (7.36) and $|m(z)| \le \eta^{-1} \le M$ (see (3.31)), we only need to prove

$$\left| \mathbb{E}\mathbf{1}(\Omega_z) m(z) - \frac{1}{\mathbb{E}\mathbf{1}(\Omega_z) m(z) + z} \right| \le \frac{C(\log N)^C \sqrt{\kappa + \eta}}{(M\eta) g(z)^2}. \tag{7.37}$$

Taking the expectation of the self consistent equation (4.11) with (4.12), we obtain that

$$\mathbb{E}\big[\mathbf{1}(\Omega_z) \cdot G_{ii}\big] + \mathbb{E}\Big[\mathbf{1}(\Omega_z)\Big(z + \sum_j \sigma_{ij}^2 G_{jj} + \Upsilon_i\Big)^{-1}\Big] = 0. \tag{7.38}$$

For simplicity, we define

$$A_i := \mathbb{E}\big[\mathbf{1}(\Omega_z) \cdot G_{ii}\big], \qquad A := \sum_i A_i / N\,.$$

Together with (7.35) and (7.36), we have

$$|A - m_{sc}|,\ |A_j - A| \ll 1.$$

Then, similarly to (5.12), on the event Ω_z we have

$$|z + A| - |\sum_j \sigma_{ij}^2 A_j - A| - |\Upsilon| > C,$$

by using $|z + m_{sc}| \ge 1$ and that on the set Ω_z, Υ is small. Therefore, we can expand (7.38) as

$$\begin{aligned} 0 = \quad & A_i + \frac{1}{z + A} - \frac{\sum_j \sigma_{ij}^2 A_j - A}{(z + A)^2} - \frac{\mathbb{E}\mathbf{1}(\Omega_z)\Upsilon_i}{(z + A)^2} \\ & + O\left(\frac{\mathbb{E}\Big[\mathbf{1}(\Omega_z) \Big|\sum_j \sigma_{ij}^2 G_{jj} - A\Big|^2\Big]}{(z + A)^3} \right) + O\left(\frac{\mathbb{E}\big[\mathbf{1}(\Omega_z) |\Upsilon_i|^2\big]}{(z + A)^3} \right). \end{aligned} \tag{7.39}$$

Then summing up $1 \le i \le N$, we obtain that

$$\left| A + \frac{1}{z + A} \right| \le C \max_i \Big|\mathbb{E}\big[\mathbf{1}(\Omega_z)\Upsilon_i\big]\Big| + C \max_i \mathbb{E}\Big[\mathbf{1}(\Omega_z)\Big| \sum_j \sigma_{ij}^2 G_{jj} - A\Big|^2\Big] + C\mathbb{E}\big[\mathbf{1}(\Omega_z) |\Upsilon|^2\big]. \tag{7.40}$$

Applying (3.12) and the definition of Ω_z, we can bound the second and third terms in the r.h.s. of (7.40) with some constant C as follows,

$$\left| A + \frac{1}{z + A} \right| \le C \max_i \Big|\mathbb{E}\big[\mathbf{1}(\Omega_z)\Upsilon_i\big]\Big| + \frac{(\log N)^C \sqrt{\kappa + \eta}}{M\eta g(z)^2}. \tag{7.41}$$

If $\eta > 3$, we estimate $\mathbb{E}\big[\mathbf{1}(\Omega_z)\Upsilon_i\big]$ as

$$\Big|\mathbb{E}\big[\mathbf{1}(\Omega_z)\Upsilon_i\big]\Big| \le \Big|\mathbb{E}\big[\mathbf{1}(\widehat{\Omega}_z^o)\Upsilon_i\big]\Big| + \mathbb{E}\Big[\mathbf{1}([\widehat{\Omega}_z^o]^c)\mathbf{1}(\widehat{\Omega}_z)|\Upsilon_i|\Big]. \tag{7.42}$$

With (4.7), we have

$$|G_{ij}G_{ji}/G_{ii}| \le 2/\eta. \tag{7.43}$$

Then, with the definition of Υ_i (4.12) and (4.39), we have

$$\mathbb{P}(|\max \Upsilon_i| \geq N^C) \leq e^{-N^c} \tag{7.44}$$

for some positive constants c and C. Inserting this and (3.20) into (7.42), we have

$$\left|A + \frac{1}{z+A}\right| \leq \frac{(\log N)^C \sqrt{\kappa+\eta}}{M\eta g^2(z)} \tag{7.45}$$

in the case of $\eta > 3$. If $\eta < 3$, similarly, with (3.23) we have the same result. This proves (7.37) and thus completes the proof of Lemma 7.6. □

Proof of Lemma 7.5. First we will prove the result for large η, more precisely we show (7.33) under the additional assumption that

$$N\eta(\kappa+\eta)^{3/2} \geq (\log N)^{C_1}, \tag{7.46}$$

with a sufficiently large constant C_1.

In the case of the generalized Wigner matrix, (2.6), we have $M \geq (C_{sup})^{-1}N$ and $\delta_+ \geq C_{inf}$ (2.7), then

$$g(z) \asymp \sqrt{\kappa+\eta}$$

up to an $O(1)$ factor. Note that with a sufficiently large C_1, (7.46) implies (3.3) and thus combining Lemma 7.6 with Lemma 5.2 we obtain (7.33) under the condition that η satisfies (7.46).

To prove (7.33) for any $\eta > 0$, it remains to consider the case when (7.46) does not hold. For a fixed E, let $\eta^* = \eta^*(E) > 0$ be the (unique) solution of $N\eta(\kappa+\eta)^{3/2} = (\log N)^{C_1}$, i.e. when (7.46) becomes an equality. In particular, we know that

$$|\mathbb{E}m(z^*) - m_{sc}(z^*)| \leq \frac{(\log N)^C}{(N\eta^*)(\kappa+\eta^*)}. \tag{7.47}$$

Consider $\eta < \eta^*$, set $z = E + i\eta$, $z^* = E + i\eta^*$ and estimate

$$|\mathbb{E}m(z) - m_{sc}(z)| \leq |\mathbb{E}m(z^*) - m_{sc}(z^*)| + \int_\eta^{\eta^*} \left|\partial_y\big(\mathbb{E}m(E+iy) - m_{sc}(E+iy)\big)\right| \mathrm{d}y. \tag{7.48}$$

Note that

$$|\partial_y m(E+iy)| = \Big|\frac{1}{N}\sum_j \partial_y G_{jj}(E+iy)\Big| \tag{7.49}$$

$$\leq \frac{1}{N}\sum_{jk}|G_{jk}(E+iy)|^2 = \frac{1}{Ny}\sum_j \mathrm{Im}G_{jj}(E+iy) = \frac{1}{y}\mathrm{Im}m(E+iy), \tag{7.50}$$

and similarly

$$|\partial_y m_{sc}(E+iy)| = \Big|\int \frac{\varrho_{sc}(x)}{(x-E-iy)^2}\mathrm{d}x\Big| \leq \int \frac{\varrho_{sc}(x)}{|x-E-iy|^2}\mathrm{d}x = \frac{1}{y}\mathrm{Im}m_{sc}(E+iy).$$

Now we use the fact that the functions $y \to y\mathrm{Im}m(E+iy)$ and $y \to y\mathrm{Im}m_{sc}(E+iy)$ are monotone increasing for any $y > 0$ since both are Stieltjes transforms of a positive measure. Therefore the integral in (7.48) can be bounded by

$$\int_\eta^{\eta^*} \frac{\mathrm{d}y}{y}\big[\mathrm{Im}\mathbb{E}m(E+iy) + \mathrm{Im}m_{sc}(E+iy)\big] \le \eta^*\big[\mathrm{Im}\mathbb{E}m(E+i\eta^*) + \mathrm{Im}m_{sc}(E+i\eta^*)\big]\int_\eta^{\eta^*} \frac{\mathrm{d}y}{y^2} \tag{7.51}$$

By the choice of η^* and using that $\mathrm{Im}\, m_{sc}(z^*) \le C\sqrt{\kappa+\eta^*}$, we have

$$\mathrm{Im}\, m_{sc}(z^*) \le \frac{(\log N)^C}{(N\eta^*)(\kappa+\eta^*)}. \tag{7.52}$$

and then $\mathrm{Im}\,\mathbb{E}m(z^*)$ can be estimated from (7.47). Inserting these estimates into (7.48) and (7.51), and using (7.47), we get

$$|\mathbb{E}m(z) - m_{sc}(z)| \le |\mathbb{E}m(z^*) - m_{sc}(z^*)| + \frac{2(\log N)^C}{N\eta^*(\kappa+\eta^*)}\frac{\eta^*}{\eta} \le \frac{(\log N)^C}{N\eta(\kappa+\eta)}$$

with a possible larger C in the r.h.s. This completes the proof of Lemma 7.5. □

With Lemma 7.5, it follows that for any E and $\eta > 0$,

$$\big|n^\lambda(E+\eta) - n^\lambda(E-\eta)\big| + \big|n^\lambda_{sc}(E+\eta) - n^\lambda_{sc}(E-\eta)\big| \le \eta(\log N)^C\left(1 + \frac{1}{N\eta(|E-2|+\eta)}\right). \tag{7.53}$$

Now we return to the main argument to prove (7.12) in Lemma 7.3. Given (7.10), we only need to prove

$$\int_{-3}^{3} \big|n^\lambda(E) - n^\lambda_{sc}(E)\big|\,\mathrm{d}E \le CN^{-1+\varepsilon}. \tag{7.54}$$

This inequality follows from the next lemma by choosing the signed measure

$$\varrho^\Delta(\mathrm{d}x) = \varrho_{sc}(\mathrm{d}x) - \frac{\mathrm{d}n^\lambda(E)}{\mathrm{d}E}, \tag{7.55}$$

whose Stieltjes transform is given by

$$m^\Delta(z) = m_{sc}(z) - \mathbb{E}m(z) \tag{7.56}$$

and the conditions (7.58) and (7.59) are provided by (7.33) and (7.53). This will complete the proof of Lemma 7.3.

Lemma 7.7 *Let $\varrho^\Delta(\mathrm{d}x)$ be a finite signed measure with support in $[-K, K]$ for some $K > 0$. Let*

$$m^\Delta(z) := \int_{\mathbb{R}} \frac{\varrho^\Delta(\mathrm{d}x)}{x-z}, \qquad n^\Delta(E) := \int_{-\infty}^{E} \varrho^\Delta(\mathrm{d}x) \tag{7.57}$$

be the Stieltjes transform and the distribution function of $\varrho^\Delta(\mathrm{d}x)$, respectively. Let κ_x, κ_E denote $||x|-2|$ and $||E|-2|$. We assume that m^Δ satisfies the following bound with some constant C:

$$\big|m^\Delta(x+iy)\big| \le \frac{(\log N)^C}{(Ny)(\kappa_x+y)} \quad \textit{for} \quad y > 0,\ \ |x| \le K+1, \tag{7.58}$$

and for any $a > 0$

$$\int_{E-a}^{E+a} |\varrho^{\Delta}|(\mathrm{d}x) \le a(\log N)^C \left(1 + \frac{1}{Na(\kappa_E + a)}\right). \tag{7.59}$$

Then

$$\int_{-K}^{K} \mathrm{d}E \left| n^{\Delta}(E) \right| \le CN^{-1}(\log N)^{C'} \tag{7.60}$$

for some constant $C' > 0$ *when* N *is sufficiently large.*

This lemma is similar to Lemma B.1 in [16], but with different assumptions. Since the assumptions here are stronger than (B.3) and (B.4) in [16], we actually obtain a better bound (7.60) than in [16], where the l.h.s. of (7.60) was bounded by $N^{-6/7}$.

Proof of Lemma 7.7. For simplicity, we omit the Δ superscript in the proof. For a fixed $E \in [-K, K]$, $\eta > 0$, define a function $f = f_{E,\eta}$: $\mathbb{R} \to \mathbb{R}$: such that $f(x) = 1$ for $x \in [-K, E - \eta]$, $f(x)$ vanishes for $x \in (-\infty, -K - 1) \cap [E + \eta, \infty)$, moreover $|f'(x)| \le C\eta^{-1}$ and $|f''(x)| \le C\eta^{-2}$. Then

$$\left| n(E) - \int_{\mathbb{R}} f_{E,\eta}(\lambda)\varrho(\lambda)\mathrm{d}\lambda \right| \le \int_{E-\eta}^{E+\eta} |\varrho|(\mathrm{d}x) \le \eta(\log N)^C \left(1 + \frac{1}{N\eta(\kappa_E + \eta)}\right). \tag{7.61}$$

We will choose $\eta = N^{-1}$ and set $f_E := f_{E,\eta}$ with $\eta = 1/N$. Then to prove (7.60), we only need to prove that

$$\left| \int_{|E| \le K+1} \int_{\mathbb{R}} f_E(\lambda)\varrho(\lambda)\mathrm{d}\lambda \mathrm{d}E \right| \le N^{-1}(\log N)^{C'} \tag{7.62}$$

for some $C' > 0$.

To express $f_E(\lambda)$ in terms of the Stieltjes transform, we use the Helffer-Sjöstrand functional calculus, as (B.12) in [16]. We formulate this result in a more general form.

Lemma 7.8 *Let* $f_{E,\eta}$ *be given as above with some* $E \in [-K, K]$, $K \ge 3$, *and* $0 < \eta \le 1/2$. *Suppose that the Stieltjes transform* m *of the signed measure* ϱ *satisfies*

$$|m(x + iy)| \le \frac{L}{(Ny)^{\tau}(\kappa_x + y)^{\sigma}} \quad \textit{for} \quad y > 0, \ \ |x| \le K + 1, \tag{7.63}$$

with some exponents $0 \le \tau, \sigma \le 1$ *and some constant* L. *Then*

$$\left| \int f_E(\lambda)\varrho(\lambda)\mathrm{d}\lambda \right| \le \frac{CL|\log \eta|}{N^{\tau}(\kappa_E + \eta)^{\sigma}}, \tag{7.64}$$

with some constant C *depending on* K.

The condition of this lemma with $\tau = \sigma = 1$ and $L = (\log N)^C$ coincides with (7.58), therefore, after integrating in E and using $\eta = 1/N$, we obtain (7.62) which completes the proof of Lemma 7.7. □

Proof of Lemma 7.8. Analogously to (B.13), (B.14) and (B.15) in [16] we obtain that

$$\begin{aligned}\left|\int f_E(\lambda)\varrho(\lambda)\mathrm{d}\lambda\right| \leq \quad & C\int_{\mathbb{R}^2}(|f_E(x)|+|y||f_E'(x)|)|\chi'(y)||m(x+iy)|\mathrm{d}x\mathrm{d}y \\ & +C\left|\int_{|y|\leq\eta}\int y f_E''(x)\chi(y)\mathrm{Im}\, m(x+iy)\mathrm{d}x\mathrm{d}y\right| \\ & +C\left|\int_{|y|\geq\eta}\int_{\mathbb{R}} y f_E''(x)\chi(y)\mathrm{Im}\, m(x+iy)\mathrm{d}x\mathrm{d}y\right|,\end{aligned} \tag{7.65}$$

where $\chi(y)$ is a smooth cutoff function with support in $[-1,1]$, with $\chi(y)=1$ for $|y|\leq 1/2$ and with bounded derivatives. The first term is estimated by

$$\int_{\mathbb{R}^2}(|f_E(x)|+|y||f_E'(x)|)|\chi'(y)||m(x+iy)|\mathrm{d}x\mathrm{d}y \leq \frac{CL}{N^\tau}, \tag{7.66}$$

using (7.63) and the support of χ'.

With (7.63) and $|f_E''|\leq C\eta^{-2}$ and

$$\mathrm{supp} f_E'(x)\subset\{|x-E|\leq\eta\},$$

the second term in r.h.s. of (7.65) is bounded by

$$\begin{aligned}CL\left|\int_{0\leq y\leq\eta}\int_{|x-E|\leq\eta}\frac{y|f_E''(x)|}{(Ny)^\tau(\kappa_x+y)^\sigma}\mathrm{d}x\mathrm{d}y\right| \quad & \leq \frac{CL}{N^\tau\eta^2}\left|\int_{0\leq y\leq\eta}\int_{|x-E|\leq\eta}\frac{y^{1-\tau}}{(\kappa_x+y)^\sigma}\mathrm{d}x\mathrm{d}y\right| \\ & \leq \frac{CL\eta^{1-\tau}|\log\eta|}{N^\tau(\kappa_E+\eta)^\sigma}.\end{aligned} \tag{7.67}$$

Here we used that for $y\leq 1/2$ we have

$$\int_{|x-E|\leq\eta}\frac{1}{(\kappa_x+y)^\sigma}\mathrm{d}x \leq \frac{C\eta|\log y|}{(\kappa_E+\eta)^\sigma}.$$

As the (B.17) and (B.19) in [16], we integrate the third term in (7.65) by parts first in x, then in y. Then bound it with absolute value by

$$C\int_{|x|\leq K+1}\eta|f_E'(x)||\mathrm{Re}\, m(x+i\eta)|\mathrm{d}x+C\int_{\mathbb{R}^2}|f_E'(x)\chi'(y)\mathrm{Re}\, m(x+iy)|+\frac{C}{\eta}\int_{\eta\leq y\leq 1}\int_{|x-E|\leq\eta}|\mathrm{Re}\, m(x+iy)|\mathrm{d}x\mathrm{d}y. \tag{7.68}$$

The middle term is bounded as (7.66). With (7.63) again, we have

$$\begin{aligned}(7.68)\leq \quad & \frac{CL}{(N\eta)^\tau}\int_{|x-E|\leq\eta}\frac{1}{(\kappa_x+\eta)^\sigma}\mathrm{d}x+\frac{CL}{(N\eta)^\tau}+\frac{CL}{(N\eta)^\tau}\int_{\eta\leq y\leq 1}\int_{|x-E|\leq\eta}\frac{1}{(\kappa_x+y)^\sigma}\mathrm{d}x\mathrm{d}y \\ & \leq \frac{CL\eta^{1-\tau}|\log\eta|}{N^\tau(\kappa_E+\eta)^\sigma}.\end{aligned} \tag{7.69}$$

Then combining (7.65), (7.66), (7.67), (7.68) and (7.69) we obtain (7.64) and complete the proof of Lemma 7.8. □

7.3 Proof of Lemma 7.4

Define the variables v_{ij} as

$$h_{ij} = \sigma_{ij} v_{ij}. \tag{7.70}$$

Denote by u_α and λ_α the eigenvectors and eigenvalues of H. For any collection of real numbers, $C_\alpha \in \mathbb{R}$, we have

$$\sum_{ij} \left| \sum_\alpha C_\alpha \frac{\partial \lambda_\alpha}{\partial v_{ij}} \right|^2 = \sum_{ij} \left| \sum_\alpha C_\alpha \sigma_{ij} \bar{u}_\alpha(i) u_\alpha(j) \right|^2 = \sum_{ij} \sigma_{ij}^2 \left| \sum_\alpha C_\alpha \bar{u}_\alpha(i) u_\alpha(j) \right|^2 \le C_{sup} N^{-1} \sum_\alpha |C_\alpha|^2. \tag{7.71}$$

With the choice $C_\alpha = K^{-1}$, $\alpha = j, j+1, \ldots, j+K-1$, and $C_\alpha = 0$ otherwise, we get $|\nabla \lambda_{j,K}|^2 \le C_{sup}(NK)^{-1}$. Using the Bobkov-Götze concentration inequality [4] and the uniform bound on the LSI constant (6.14), we get

$$\mathbb{P}\big(|\lambda_{j,K} - \mathbb{E}\,\lambda_{j,K}| \ge \gamma\big) \le e^{-\gamma T} \mathbb{E}\, e^{C_S T^2 |\nabla \lambda_{j,K}|^2} \le e^{-\gamma T + C_S C_{sup} T^2/(NK)}$$

for any T and γ. Choosing $\gamma = N^{-1/2+\delta} K^{-1/2}$ and $T = (NK)^{1/2}$, we obtain (7.13). □

8 Proof of the Green's function comparison theorem

Proof of Theorem 2.3. From the trivial bound

$$\mathrm{Im}\left(\frac{1}{H - E - i\eta}\right)_{jj} \le \left(\frac{y}{\eta}\right) \mathrm{Im}\left(\frac{1}{H - E - iy}\right)_{jj}, \qquad \eta \le y,$$

and from (2.21) we have the following a priori bound

$$\mathbb{P}\left(\max_{0\le\gamma\le\gamma(N)} \max_{1\le k\le N} \max_{|E|\le 2-\kappa} \sup_{\eta \ge N^{-1-\varepsilon}} \left| \mathrm{Im}\left(\frac{1}{H_\gamma - E \pm i\eta}\right)_{kk} \right| \le N^{3\tau+\varepsilon} \right) \ge 1 - CN^{-c\log\log N}. \tag{8.1}$$

Note that the supremum over η can be included by establishing the estimate first for a fine grid of η's with spacing N^{-10} and then extend the bound for all η by using that the Green's functions are Lipschitz continuous in η with a Lipschitz constant η^{-2}.

Let λ_m and u_m denote the eigenvalues and eigenvectors of H_γ, then by the definition of the Green's function, we have

$$\left| \left(\frac{1}{H_\gamma - z}\right)_{jk} \right| \le \sum_{m=1}^N \frac{|u_m(j)||u_m(k)|}{|\lambda_m - z|} \le \left[\sum_{m=1}^N \frac{|u_m(j)|^2}{|\lambda_m - z|} \right]^{1/2} \left[\sum_{m=1}^N \frac{|u_m(k)|^2}{|\lambda_m - z|} \right]^{1/2}.$$

Define a dyadic decomposition

$$U_n = \{m : 2^{n-1}\eta \le |\lambda_m - E| < 2^n \eta\}, \qquad n = 1, 2, \ldots, n_0 := C \log N, \tag{8.2}$$

$$U_0 = \{m : |\lambda_m - E| < \eta\}, \qquad U_\infty := \{m : 2^{n_0}\eta \le |\lambda_m - E|\},$$

and divide the summation over m into $\cup_n U_n$

$$\sum_{m=1}^N \frac{|u_m(j)|^2}{|\lambda_m - z|} = \sum_n \sum_{m\in U_n} \frac{|u_m(j)|^2}{|\lambda_m - z|} \le C \sum_n \sum_{m\in U_n} \mathrm{Im}\, \frac{|u_m(j)|^2}{\lambda_m - E - i2^n\eta} \le C \sum_n \mathrm{Im}\left(\frac{1}{H_\gamma - E - i2^n\eta}\right)_{jj}.$$

Using the estimate (2.21) for $n = 0, 1, \ldots, n_0$ and a trivial bound of $O(1)$ for $n = \infty$, we have proved that

$$\mathbb{P}\left(\sup_{0\le\gamma\le\gamma(N)} \sup_{1\le k,\ell\le N} \max_{|E|\le 2-\kappa} \sup_{\eta\ge N^{-1-\varepsilon}} \left| \left(\frac{1}{H_\gamma - E \pm i\eta} \right)_{k\ell} \right| \le N^{4\tau+\varepsilon} \right) \ge 1 - CN^{-c\log\log N}. \tag{8.3}$$

For simplicity, we will consider the case when the test function F has only $n = 1$ variable and $k_1 = 1$, i.e., we consider the trace of a first order monomial; the general case follows analogously. Consider the telescoping sum of differences of expectations

$$\begin{aligned} \mathbb{E}\, F\left(\frac{1}{N}\mathrm{Tr}\, \frac{1}{H^{(v)} - z} \right) - \mathbb{E}\, F\left(\frac{1}{N}\mathrm{Tr}\, \frac{1}{H^{(w)} - z} \right) \\ = \sum_{\gamma=1}^{\gamma(N)} \left[\mathbb{E}\, F\left(\frac{1}{N}\mathrm{Tr}\, \frac{1}{H_\gamma - z} \right) - \mathbb{E}\, F\left(\frac{1}{N}\mathrm{Tr}\, \frac{1}{H_{\gamma-1} - z} \right) \right]. \end{aligned} \tag{8.4}$$

Let $E^{(ij)}$ denote the matrix whose matrix elements are zero everywhere except at the (i, j) position, where it is 1, i.e., $E^{(ij)}_{k\ell} = \delta_{ik}\delta_{j\ell}$. Fix an $\gamma \ge 1$ and let (i, j) be determined by $\phi(i, j) = \gamma$. We will compare $H_{\gamma-1}$ with H_γ. Note that these two matrices differ only in the (i, j) and (j, i) matrix elements and they can be written as

$$H_{\gamma-1} = Q + \frac{1}{\sqrt{N}} V, \qquad V := v_{ij}E^{(ij)} + v_{ji}E^{(ji)}$$

$$H_\gamma = Q + \frac{1}{\sqrt{N}} W, \qquad W := w_{ij}E^{(ij)} + w_{ji}E^{(ji)},$$

with a matrix Q that has zero matrix element at the (i, j) and (j, i) positions and where we set $v_{ji} := \overline{v}_{ij}$ for $i < j$ and similarly for w. Define the Green's functions

$$R = \frac{1}{Q - z}, \qquad S = \frac{1}{H_\gamma - z}.$$

We first claim that the estimate (8.3) holds for the Green's function R as well. To see this, we have, from the resolvent expansion,

$$R = S + N^{-1/2}SVS + \ldots + N^{-9/5}(SV)^9 S + N^{-5}(SV)^{10}R.$$

Since V has only at most two nonzero element, when computing the (k, ℓ) matrix element of this matrix identity, each term is a finite sum involving matrix elements of S or R and v_{ij}, e.g. $(SVS)_{k\ell} = S_{ki}v_{ij}S_{j\ell} + S_{kj}v_{ji}S_{i\ell}$. Using the bound (8.3) for the S matrix elements, the subexponential decay for v_{ij} and the trivial bound $|R_{ij}| \le \eta^{-1}$, we obtain that the estimate (8.3) holds for R.

We can now start proving the main result. By the resolvent expansion,

$$S = R - N^{-1/2}RVR + N^{-1}(RV)^2R - N^{-3/2}(RV)^3R + N^{-2}(RV)^4R - N^{-5/2}(RV)^5S,$$

so we can write

$$\frac{1}{N}\mathrm{Tr}\, S = \widehat{R} + \xi, \qquad \xi = \sum_{m=1}^{4} N^{-m/2}\widehat{R}^{(m)} + N^{-5/2}\Omega$$

with

$$\widehat{R} = \frac{1}{N}\mathrm{Tr}\, R, \qquad \widehat{R}^{(m)} = (-1)^m \frac{1}{N}\mathrm{Tr}\,(RV)^m R, \qquad \Omega = -\frac{1}{N}\mathrm{Tr}\,(RV)^5 S.$$

For each diagonal element in the computation of these traces, the contribution to $\widehat{R}$, $\widehat{R}^{(m)}$ and Ω is a sum of a few terms. E.g.

$$\widehat{R}^{(2)} = \frac{1}{N}\sum_k \Big[R_{ki}v_{ij}R_{jj}v_{ji}R_{ik} + R_{ki}v_{ij}R_{ji}v_{ij}R_{jk} + R_{kj}v_{ji}R_{ii}v_{ij}R_{jk} + R_{kj}v_{ji}R_{ij}v_{ji}R_{ik}\Big]$$

and similar formulas hold for the other terms.

Then we have

$$\begin{aligned}\mathbb{E}F\left(\frac{1}{N}\mathrm{Tr}\,\frac{1}{H_\gamma - z}\right) =& \mathbb{E}F\left(\widehat{R}+\xi\right) \\ =& \mathbb{E}\left[F(\widehat{R}) + F'(\widehat{R})\xi + F''(\widehat{R})\xi^2 + \ldots + F^{(5)}(\widehat{R}+\xi')\xi^5\right] \\ =& \sum_{m=0}^{5} N^{-m/2}\mathbb{E}A^{(m)},\end{aligned} \tag{8.5}$$

where ξ' is a number between 0 and ξ and it depends on $\widehat{R}$ and ξ; the $A^{(m)}$'s are defined as

$$A^{(0)} = F(\widehat{R}), \quad A^{(1)} = F'(\widehat{R})\widehat{R}^{(1)}, \quad A^{(2)} = F''(\widehat{R})(\widehat{R}^{(1)})^2 + F'(\widehat{R})\widehat{R}^{(2)},$$

and similarly for $A^{(3)}$ and $A^{(4)}$. Finally,

$$A^{(5)} = F'(\widehat{R})\Omega + F^{(5)}(\widehat{R}+\xi')(\widehat{R}^{(1)})^5 + \ldots.$$

The expectation values of the terms $A^{(m)}$, $m \le 4$, with respect to v_{ij} are determined by the first four moments of v_{ij}, for example

$$\mathbb{E}\,A^{(2)} = F'(\widehat{R})\Big[\frac{1}{N}\sum_k R_{ki}R_{jj}R_{ik} + \ldots\Big]\mathbb{E}\,|v_{ij}|^2 + F''(\widehat{R})\Big[\frac{1}{N^2}\sum_{k,\ell} R_{ki}R_{j\ell}R_{\ell j}R_{ik} + \ldots\Big]\mathbb{E}\,|v_{ij}|^2$$

$$+F'(\widehat{R})\Big[\frac{1}{N}\sum_k R_{ki}R_{ji}R_{jk} + \ldots\Big]\mathbb{E}\,v_{ij}^2 + F''(\widehat{R})\Big[\frac{1}{N^2}\sum_{k,\ell} R_{ki}R_{j\ell}R_{\ell i}R_{jk} + \ldots\Big]\mathbb{E}\,v_{ij}^2.$$

Note that the coefficients involve up to four derivatives of F and normalized sums of matrix elements of R. Using the estimate (8.3) for R and the derivative bounds (2.23) for the typical values of $\widehat{R}$, we see that all these coefficients are bounded by $N^{C(\tau+\varepsilon)}$ with a very large probability, where C is an explicit constant. We use the bound (2.24) for the extreme values of $\widehat{R}$ but this event has a very small probability by (8.3). Therefore, the coefficients of the moments $\mathbb{E}\,\bar{v}_{ij}^s v_{ij}^u$, $u+s \le 4$, in the quantities $A^{(0)}, \ldots, A^{(4)}$ are essentially bounded, modulo a factor $N^{C(\tau+\varepsilon)}$. Notice that the fourth moment of v_{ij} appears only in the $m=4$ term that already has a prefactor N^{-2} in (8.5). Therefore, to compute the $m \le 4$ terms in (8.5) up to a precision $o(N^{-2})$, it is sufficient to know the first three moments of v_{ij} exactly and the fourth moment only with a precision $N^{-\delta}$; if τ and ε are chosen such that $C(\tau+\varepsilon) < \delta$, then the discrepancy in the fourth moment is irrelevant.

Finally, we have to estimate the error term $A^{(5)}$. All terms without Ω can be dealt with as before; after estimating the derivatives of F by $N^{C(\tau+\varepsilon)}$, one can perform the expectation with respect to v_{ij} that is independent of $\widehat{R}^{(m)}$. For the terms involving Ω one can argue similarly, by appealing to the fact that the matrix elements of S are also essentially bounded by $N^{C(\tau+\varepsilon)}$, see (8.3), and that v_{ij} has subexponential decay. Alternatively, one can use Hölder inequality to decouple S from the rest and use (8.3) directly, for example:

$$\mathbb{E}|F'(\widehat{R})\Omega| = \frac{1}{N}\mathbb{E}|F'(\widehat{R})\mathrm{Tr}\,(RV)^5 S| \le \frac{1}{N}\left[\mathbb{E}(F'(\widehat{R}))^2\mathrm{Tr}\,S^2\right]^{1/2}\left[\mathbb{E}\,\mathrm{Tr}\,(RV)^5(VR^*)^5\right]^{1/2} \le CN^{C(\tau+\varepsilon)}.$$

Note that exactly the same perturbation expansion holds for the resolvent of $H_{\gamma-1}$, just v_{ij} is replaced with w_{ij} everywhere. By the moment matching condition, the expectation values $\mathbb{E}A^{(m)}$ of terms for $m \le 3$ in (8.5) are identical and the $m = 4$ term differs by $N^{-\delta+C(\tau+\varepsilon)}$. Choosing $\tau = \varepsilon$, we have

$$\mathbb{E}\,F\left(\frac{1}{N}\mathrm{Tr}\,\frac{1}{H_\gamma - z}\right) - \mathbb{E}\,F\left(\frac{1}{N}\mathrm{Tr}\,\frac{1}{H_{\gamma-1} - z}\right) \le CN^{-5/2+C\varepsilon} + CN^{-2-\delta+C\varepsilon}.$$

After summing up in (8.4) we have thus proved that

$$\mathbb{E}\,F\left(\frac{1}{N}\mathrm{Tr}\,\frac{1}{H^{(v)} - z}\right) - \mathbb{E}\,F\left(\frac{1}{N}\mathrm{Tr}\,\frac{1}{H^{(w)} - z}\right) \le CN^{-1/2+C\varepsilon} + CN^{-\delta+C\varepsilon}.$$

The proof can be easily generalized to functions of several variables. This concludes the proof of Theorem 2.3. □

Proof of Theorem 6.4. Define an approximate delta function (times π) at the scale η by

$$\theta_\eta(x) = \mathrm{Im}\,\frac{1}{x - i\eta}.$$

For notational simplicity, we will prove only the case of three point correlation functions; the proof is analogous for the general case. By definition of the correlation function, for any fixed E, $\alpha_1, \alpha_2, \alpha_3$,

$$\begin{aligned}\mathbb{E}_{\mathbf{w}}\frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}\theta_\eta\Big(\lambda_i - E - \frac{\alpha_1}{N}\Big)\theta_\eta\Big(\lambda_j - E - \frac{\alpha_2}{N}\Big)\theta_\eta\Big(\lambda_k - E - \frac{\alpha_3}{N}\Big)\\ = \int \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 p^{(3)}_{w,N}(x_1,x_2,x_3)\theta_\eta(x_1 - E_1)\theta_\eta(x_2 - E_2)\theta_\eta(x_3 - E_3), \qquad E_j := E + \frac{\alpha_j}{N}. \end{aligned} \tag{8.6}$$

By the exclusion-inclusion principle,

$$\mathbb{E}_{\mathbf{w}}\frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}\theta_\eta(x_1 - E_1)\theta_\eta(x_2 - E_2)\theta_\eta(x_3 - E_3) = \mathbb{E}_{\mathbf{w}}A_1 + \mathbb{E}_{\mathbf{w}}A_2 + \mathbb{E}_{\mathbf{w}}A_3, \tag{8.7}$$

where

$$A_1 := \frac{1}{N(N-1)(N-2)}\prod_{j=1}^{3}\left[\frac{1}{N}\sum_i \theta_\eta(\lambda_i - E_j)\right],$$

$$A_3 := \frac{2}{N(N-1)(N-2)}\sum_i \theta_\eta(\lambda_i - E_1)\theta_\eta(\lambda_i - E_2)\theta_\eta(\lambda_i - E_3) + \dots$$

and

$$A_2 := B_1 + B_2 + B_3, \quad \text{with} \quad B_3 = -\frac{1}{N(N-1)(N-2)} \sum_i \theta_\eta(\lambda_i - E_1)\theta_\eta(\lambda_i - E_2) \sum_k \theta_\eta(\lambda_k - E_3),$$

and similarly, B_1 consists of terms with $j = k$, while B_2 consists of terms with $i = k$.

Notice that, modulo a trivial change in the prefactor, $\mathbb{E}_{\mathbf{w}} A_1$ can be approximated by

$$\mathbb{E}_{\mathbf{w}} F\left(\frac{1}{N}\mathrm{Im}\,\mathrm{Tr}\,\frac{1}{H^{(v)} - z_1}, \ldots, \frac{1}{N}\mathrm{Im}\,\mathrm{Tr}\,\frac{1}{H^{(v)} - z_3}\right),$$

where the function F is chosen to be $F(x_1, x_2, x_3) := x_1 x_2 x_3$ if $\max_j |x_j| \le N^{\varepsilon}$ and it is smoothly cutoff to go to zero in the regime $\max_j |x_j| \ge N^{2\varepsilon}$. The difference between the expectation of F and A_1 is negligible, since it comes from the regime where $N^{\varepsilon} \le \max_j \frac{1}{N}|\mathrm{Im}\,\mathrm{Tr}\,(H^{(v)} - z_j)^{-1}| \le N^2$, which has an exponentially small probability by (8.3) (the upper bound on the Green's function always holds since $\eta \ge N^{-2}$). Here the arguments of F are imaginary parts of the trace of the Green's function, but this type of function is allowed when applying Theorem 2.3, since

$$\mathrm{Im}\,\mathrm{Tr}\,G(z) = \frac{1}{2}\big[\mathrm{Tr}\,G(z) - \mathrm{Tr}\,G(\bar{z})\big].$$

We remark that the main assumption (2.21) for Theorem 2.3 is satisfied by using (2.17) of Theorem 2.1 with the choice of $M \asymp N$.

Similarly, we can approximate $\mathbb{E}_{\mathbf{w}} B_3$ by

$$\mathbb{E}_{\mathbf{w}} G\left(\frac{1}{N^2}\mathrm{Tr}\left\{\mathrm{Im}\,\frac{1}{H^{(v)} - z_1}\mathrm{Im}\,\frac{1}{H^{(v)} - z_2}\right\}, \frac{1}{N}\mathrm{Im}\,\mathrm{Tr}\,\frac{1}{H^{(v)} - z_3}\right),$$

where $G(x_1, x_2) = x_1 x_2$ with an appropriate cutoff for large arguments, and there are similar expressions for B_1, B_2 and also for A_3, the latter involving the trace of the product of three resolvents. By Theorem 2.3, these expectations w.r.t. $\mathbf{w}$ in the approximations of $\mathbb{E}_{\mathbf{w}} A_i$ can be replaced by expectations w.r.t. $\mathbf{v}$ with only negligible errors provided that $\eta \ge N^{-1-\varepsilon}$. We have thus proved that

$$\lim_{N\to\infty} \int \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \big[p^{(3)}_{w,N}(x_1, x_2, x_3) - p^{(3)}_{v,N}(x_1, x_2, x_3)\big]\theta_\eta(x_1 - E_1)\theta_\eta(x_2 - E_2)\theta_\eta(x_3 - E_3) = 0. \tag{8.8}$$

Set $\eta = N^{-1-\varepsilon}$ for the rest of the proof. We now show that the validity of (8.8) for any choice of E, $\alpha_1, \alpha_2, \alpha_3$ (recall $E_j = E + \alpha_j/N$) implies that the rescaled correlation functions, $p^{(3)}_{w,N}(E + \beta_1/N, \ldots, E + \beta_3/N)$ and $p^{(3)}_{v,N}(E + \beta_1/N, \ldots, E + \beta_3/N)$, as functions of the variables $\beta_1, \beta_2, \beta_3$, have the same weak limit.

Let O be a smooth, compactly supported test function and let

$$O_\eta(\beta_1, \beta_2, \beta_3) := \frac{1}{(\pi N)^3} \int_{\mathbb{R}^3} \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \mathrm{d}\alpha_3 O(\alpha_1, \alpha_2, \alpha_3)\theta_\eta\left(\frac{\beta_1 - \alpha_1}{N}\right) \ldots \theta_\eta\left(\frac{\beta_3 - \alpha_3}{N}\right)$$

be its smoothing on scale $N\eta$. Then we can write

$$\begin{aligned}
&\int_{\mathbb{R}^3} \mathrm{d}\beta_1 \mathrm{d}\beta_2 \mathrm{d}\beta_3\; O(\beta_1, \beta_2, \beta_3) p^{(3)}_{w,N}\left(E + \frac{\beta_1}{N}, \ldots, E + \frac{\beta_3}{N}\right) \\
&\qquad = \int_{\mathbb{R}^3} \mathrm{d}\beta_1 \mathrm{d}\beta_2 \mathrm{d}\beta_3\; O_\eta(\beta_1, \beta_2, \beta_3) p^{(3)}_{w,N}\left(E + \frac{\beta_1}{N}, \ldots, E + \frac{\beta_3}{N}\right) \\
&\qquad\quad + \int_{\mathbb{R}^3} \mathrm{d}\beta_1 \mathrm{d}\beta_2 \mathrm{d}\beta_3\; (O - O_\eta)(\beta_1, \beta_2, \beta_3) p^{(3)}_{w,N}\left(E + \frac{\beta_1}{N}, \ldots, E + \frac{\beta_3}{N}\right).
\end{aligned} \tag{8.9}$$

The first term on the right side, after the change of variables $x_j = E + \beta_j/N$, is equal to

$$\int_{\mathbb{R}^3} \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \mathrm{d}\alpha_3 \; O(\alpha_1, \alpha_2, \alpha_3) \int_{\mathbb{R}^3} \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 p^{(3)}_{w,N}(x_1, x_2, x_3) \theta_\eta(x_1 - E_1)\theta_\eta(x_2 - E_2)\theta_\eta(x_3 - E_3), \quad (8.10)$$

i.e., it can be written as an integral of expressions of the form (8.8) for which limits with $p_{w,N}$ and $p_{v,N}$ coincide.

Finally, the second term on the right hand side of (8.9) is negligible. To see this, notice that for any test function Q, we have

$$\begin{aligned}
&\int_{\mathbb{R}^3} \mathrm{d}\beta_1 \mathrm{d}\beta_2 \mathrm{d}\beta_3 \; Q(\beta_1, \beta_2, \beta_3) p^{(3)}_{w,N}\left(E + \frac{\beta_1}{N}, \dots, E + \frac{\beta_3}{N}\right) \\
&\qquad = N^3 \int_{\mathbb{R}^3} \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \; Q\big(N(x_1 - E), N(x_2 - E), N(x_3 - E)\big) p^{(3)}_{w,N}(x_1, x_2, x_3) \\
&\qquad = \Big(1 - \frac{1}{N}\Big)\Big(1 - \frac{2}{N}\Big) \mathbb{E}_{\mathbf{w}} \sum_{i \neq j \neq k} Q\big(N(\lambda_i - E), N(\lambda_j - E), N(\lambda_k - E)\big). \qquad (8.11)
\end{aligned}$$

If the test function Q were supported on a ball of size $N^{\varepsilon'}$, $\varepsilon' > 0$, then this last term were bounded by

$$\|Q\|_\infty \mathbb{E}_{\mathbf{w}} \mathcal{N}^3_{CN^{-1+\varepsilon'}}(E) \leq C \|Q\|_\infty N^{4\varepsilon'}. \quad (8.12)$$

Here $\mathcal{N}_\tau(E)$ denotes the number of eigenvalues in the interval $[E - \tau, E + \tau]$ and in the estimate we used the local semicircle law on intervals of size $\tau \geq N^{-1+\varepsilon'}$.

Set now $Q := O - O_\eta$. From the definition of O_η, it is easy to see that the function

$$Q_1(\beta_1, \beta_2, \beta_3) = O(\beta_1, \beta_2, \beta_3) - O_\eta(\beta_1, \beta_2, \beta_3) \prod_{j=1}^3 1(|\beta_j| \leq N^{\varepsilon'})$$

satisfies the bound $\|Q_1\|_\infty \leq \|Q\|_\infty = \|O - O_\eta\|_\infty \leq CN\eta = CN^{-\varepsilon}$. So choosing $\varepsilon' < \varepsilon/4$, the contribution of Q_1 is negligible. Finally, $Q_2 = Q - Q_1$ is given by

$$Q_2(\beta_1, \beta_2, \beta_3) = -O_\eta(\beta_1, \beta_2, \beta_3) \left[1 - \prod_{j=1}^3 1(|\beta_j| \leq N^{\varepsilon'})\right]$$

and

$$\begin{aligned}
|Q_2| \leq & C \left[\frac{1}{1 + \beta_1^2}\right] \left[\frac{1}{1 + \beta_2^2}\right] \left[\frac{1}{1 + \beta_3^2}\right] \Big\{1(|\beta_1| \geq N^{\varepsilon'}) + \dots\Big\} \\
\leq & C \left\{N^{-\varepsilon'} \left[\frac{N^{\varepsilon'}}{N^{2\varepsilon'} + \beta_1^2}\right] \left[\frac{1}{1 + \beta_2^2}\right] \left[\frac{1}{1 + \beta_3^2}\right] + \dots\right\}. \qquad (8.13)
\end{aligned}$$

Hence the contribution of Q_2 in the last term of (8.11) is bounded by

$$CN^{-3-\varepsilon'} \mathbb{E}_{\mathbf{w}} \sum_{i,j,k} \left\{\left[\frac{N^{-1+\varepsilon'}}{N^{-2+2\varepsilon'} + (\lambda_i - E)^2}\right] \left[\frac{N^{-1}}{N^{-2} + (\lambda_j - E)^2}\right] \left[\frac{N^{-1}}{N^{-2} + (\lambda_k - E)^2}\right] + \dots\right\}$$

From Theorem 2.1, the last term is bounded by $N^{-\varepsilon'}$ up to some logarithmic factor. This completes the proof of Theorem 6.4. □

A Spectral condition for band matrices

Lemma A.1 *Let $B = (\sigma_{ij})$ satisfying (2.1) and (2.8) with $W \geq 1$ and with f being a nonnegative symmetric function with $\int f = 1$ and $f \in L^\infty(\mathbb{R})$. Then we have*

$$B \geq -1 + \delta \tag{A.1}$$

for some $\delta > 0$ and W large enough, depending on f.

Proof. Recall that the discrete Fourier transform in $d = 1$ dimensions is defined as follows. Let $\varepsilon := 1/N$ and

$$\Lambda_\varepsilon := \Lambda = \varepsilon\mathbb{Z}/\mathbb{Z}$$

be the periodic one dimensional lattice (torus) of size 1 and spacing ε with its dual lattice being

$$\Lambda^*_\varepsilon := \Lambda^* := \Big(2\pi\mathbb{Z}\Big)/\Big(\frac{2\pi}{\varepsilon}\mathbb{Z}\Big).$$

Let ψ be a function on Λ. Then its Fourier transform $\mathcal{F}_N\psi$ is a function on Λ^* defined as

$$\mathcal{F}_N\psi(p) = \varepsilon \sum_{x\in\Lambda} \psi(x) e^{-ip\cdot x}$$

and it is an isometry

$$\varepsilon \sum_{x\in\Lambda} \overline{\psi(x)}\phi(x) = \sum_{p\in\Lambda^*} \overline{\mathcal{F}_N\psi(p)}\mathcal{F}_N\phi(p).$$

In our case, $x = k/N$ and define

$$F_W(x) := NW^{-1} f(xN/W).$$

Then, for $p \in \Lambda^*$, we have

$$(\mathcal{F}_N F_W)(p) = \sum_{x\in\Lambda} F_W(x) e^{-ip\cdot x} = \sum_{k=1}^{N} W^{-1} f(k/W) e^{-i(Wp/N)\cdot(k/W)} = \widehat{f}(q) + o(1), \quad q = Wp/N,$$

where the error term vanishes as $W \to \infty$ and $\widehat{f}$ denotes the usual Fourier transform in $L^1(\mathbb{R})$

$$\widehat{f}(q) = \int f(y) e^{-iqy} \mathrm{d}y.$$

With this formula, and with the notation $\psi_N(j) := \psi(j/N)$ for any ψ defined on Λ, we have

$$(B\psi_N)(k) = \varepsilon \sum_{\ell=1}^{N} NW^{-1} f((k-\ell)/W)\psi_N(\ell) = \varepsilon \sum_{y\in\Lambda} F_W\Big(\frac{k}{N} - y\Big)\psi(y) = \sum_{p\in\Lambda^*} e^{ipk/N} \mathcal{F}_N F_W(p) \mathcal{F}_N\psi(p)$$

and

$$\sum_{p\in\Lambda^*} |\mathcal{F}_N\psi(p)|^2 = \varepsilon \sum_{j=1}^{N} |\psi_N(j)|^2$$

which is normalized to be 1. Hence B on the Fourier side acts as a multiplication by the function $\mathcal{F}_N F_W$, so

$$\mathrm{Spec}B = \mathrm{Range}\, \mathcal{F}_N F_W \subset \mathrm{supp}\, \widehat{f} + o(1).$$

Since f is nonnegative, symmetric function and $\int f = 1$, we have $\widehat{f}$ is real and

$$\inf \widehat{f} > -1 + \delta$$

for some $\delta > 0$, which completes the proof.

B Large deviation estimates

In this Appendix we prove two large deviations results. They are weaker than the corresponding results of Hanson and Wright [22], used in [14], but they require only independent, not necessarily identically distributed random variables, moreover the proofs are much simpler.

Lemma B.1 *Let a_i $(1 \le i \le N)$ be N independent complex random variables with mean zero, variance σ^2 and uniform subexponential decay, i.e., there exist α, $\beta > 0$ that for any $x > 0$*

$$\mathbb{P}(|a_i| \ge x^\alpha) \le \beta e^{-x}. \tag{B.1}$$

Then for any $A_i \in \mathbb{C}$ $(1 \le i \le N)$ and $D \ge 1$ we have,

$$\mathbb{P}\left\{ \left| \sum_i a_i A_i \right| \ge D\sigma \Big(\sum_i |A_i|^2 \Big)^{1/2} \right\} \le C \exp\big(- cD^{\frac{2}{2+\alpha}} \big) \tag{B.2}$$

for some positive constants C and c depending on α and β in (B.1).

Proof of Lemma B.1. Without loss of generality, we may assume that $\sigma = 1$. The assumption (B.1) implies that the $k-$th moment of a_i is bounded by:

$$\mathbb{E}|a_i|^k \le (Ck)^{\alpha k} \tag{B.3}$$

for some $C > 0$ depending on α and β.

First, for $p \in \mathbb{N}$, we estimate

$$\mathbb{E} \left| \sum_{i=1}^N a_i A_i \right|^p . \tag{B.4}$$

With the Marcinkiewicz–Zygmund inequality, for an integer $p \ge 2$, we have

$$\mathbb{E} \left| \sum_i a_i A_i \right|^p \le (Cp)^{p/2} \mathbb{E} \left[\left(\sum_i |a_i A_i|^2 \right)^{p/2} \right] \tag{B.5}$$

(for the estimate of the constant, see e.g. Exercise 2.2.30 of [30]). Using (B.3), we have $\mathbb{E}|a_{i_1} a_{i_2} \cdots a_{i_{p/2}}|^2 \le (Cp)^{\alpha p}$. Inserting it into (B.5), we obtain

$$\mathbb{E} \left| \sum_i a_i A_i \right|^p \le (Cp^{\frac{1}{2}+\alpha})^p \left(\sum_i |A_i|^2 \right)^{p/2} , \tag{B.6}$$

which implies (B.2) by choosing an even integer p of the order $(D/Ce)^{\frac{2}{2+\alpha}}$ and applying a high moment Markov inequality. □

Lemma B.2 *Let a_i $(1 \le i \le N)$ be N independent random complex variables with mean zero, variance σ^2 and having the uniform subexponential decay* (B.1). *Let $B_{ij} \in \mathbb{C}$ $(1 \le i, j \le N)$. Then we have that*

$$\mathbb{P}\left\{\left|\sum_{i=1}^{N} \overline{a}_i B_{ii} a_i - \sum_{i=1}^{N} \sigma^2 B_{ii}\right| \ge D\sigma^2\Big(\sum_{i=1}^{N} |B_{ii}|^2\Big)^{1/2}\right\} \le C\exp\big(-cD^{\frac{1}{1+\alpha}}\big) \tag{B.7}$$

and

$$\mathbb{P}\left\{\left|\sum_{i\neq j} \overline{a}_i B_{ij} a_j\right| \ge D\sigma^2\Big(\sum_{i\neq j} |B_{ij}|^2\Big)^{1/2}\right\} \le C\exp\big(-cD^{\frac{1}{2(1+\alpha)}}\big) \tag{B.8}$$

for some positive constants C and c depending on α and β in (B.1).

Proof of Lemma B.2. Without loss of generality, we may again assume that $\sigma = 1$. First, we prove (B.7). Notice that $|a_i|^2 - 1$ $(1 \le i \le N)$ are independent random variables with mean 0 and variance less than some constants C. Furthermore, the k-th moment of $|a_i|^2 - 1$ is bounded as

$$\mathbb{E}(|a_i|^2 - 1)^k \le (Ck)^{2\alpha k}. \tag{B.9}$$

Then following the proof of the Lemma B.1 with $|a_i|^2 - 1$ replacing a_i, we obtain (B.7).

Next, we prove (B.8). For any $p \in \mathbb{N}$, $p \ge 2$, we estimate

$$\mathbb{E}\left|\sum_i \overline{a}_i \xi_i\right|^p \equiv \mathbb{E}\left|\sum_{i>j} \overline{a}_i B_{ij} a_j\right|^p \tag{B.10}$$

where $\xi_i := \sum_{j<i} B_{ij} a_j$. Note that a_i and ξ_i are independent for any fixed i. By the definition,

$$X_n \equiv \sum_{i=1}^{n} \overline{a}_i \xi_i \tag{B.11}$$

is martingale. Using the Burkholder inequality, we have that

$$\mathbb{E}\left|\sum_i \overline{a}_i \xi_i\right|^p \le (Cp)^{3p/2}\mathbb{E}\left[\Big(\sum_i |\overline{a}_i \xi_i|^2\Big)^{p/2}\right] \tag{B.12}$$

(for the constant, see Section VII.3 of [28]). By the generalized Minkowski inequality, by the independence of a_i and ξ_i and using (B.3), we have

$$\left[\mathbb{E}\Big(\sum_i |\overline{a}_i \xi_i|^2\Big)^{p/2}\right]^{2/p} \le \sum_i \left[\mathbb{E}|\overline{a}_i \xi_i|^p\right]^{2/p} = \sum_i \left[\mathbb{E}(|\overline{a}_i|^p)\mathbb{E}(|\xi_i|^p)\right]^{2/p} \le (Cp)^{2\alpha}\sum_i \left[\mathbb{E}(|\xi_i|^p)\right]^{2/p}.$$

Using (B.6), we have

$$\mathbb{E}(|\xi_i|^p) \le (Cp^{\frac{1}{2}+\alpha})^p \Big(\sum_j |B_{ij}|^2 \Big)^{p/2}.$$

Combining this with (B.12) we obtain

$$\mathbb{E} \left| \sum_i \overline{a}_i \xi_i \right|^p \le (Cp)^{2p(1+\alpha)} \Big(\sum_i \sum_j |B_{ij}|^2 \Big)^{p/2}. \tag{B.13}$$

Then choosing $(D/Ce)^{\frac{1}{2(1+\alpha)}}$ and applying Markov inequality, we obtain (B.8) . □

In our applications we will need these two lemmas when D is a power of $\log N$. For simplicity, we do not want to keep track of the precise powers in the estimate and we are interested only in error bounds that decay faster than any fixed power of N, say $CN^{-\log\log N}$. Therefore, in this paper we will use the following weaker form of these two lemmas, the stronger form will be useful in future applications.

Corollary B.3 *Let a_i $(1 \le i \le N)$ be N independent random complex variables with mean zero, variance σ^2 and having the uniform subexponential decay* (B.1)*. Let A_i, $B_{ij} \in \mathbb{C}$ $(1 \le i, j \le N)$. Then we have that*

$$\mathbb{P}\left\{ \left| \sum_{i=1}^N a_i A_i \right| \ge (\log N)^{\frac{3}{2}+\alpha} \sigma \Big(\sum_i |A_i|^2 \Big)^{1/2} \right\} \le CN^{-\log\log N}, \tag{B.14}$$

$$\mathbb{P}\left\{ \left| \sum_{i=1}^N \overline{a}_i B_{ii} a_i - \sum_{i=1}^N \sigma^2 B_{ii} \right| \ge (\log N)^{\frac{3}{2}+2\alpha} \sigma^2 \Big(\sum_{i=1}^N |B_{ii}|^2 \Big)^{1/2} \right\} \le CN^{-\log\log N}, \tag{B.15}$$

$$\mathbb{P}\left\{ \left| \sum_{i\ne j} \overline{a}_i B_{ij} a_j \right| \ge (\log N)^{3+2\alpha} \sigma^2 \Big(\sum_{i\ne j} |B_{ij}|^2 \Big)^{1/2} \right\} \le CN^{-\log\log N}, \tag{B.16}$$

for some constants C depending on α and β in (B.1)*.*

C Proof of Lemma 6.5

We first prove a version of this lemma when the fourth moment exactly matches, i.e., $\gamma = 0$, then we explain how to deal with the approximation. More precisely, we first show the following:

Lemma C.1 *Under the condition*

$$m_4 - m_3^2 - 1 \ge C_1, \;\; m_4 \le C_2 \tag{C.1}$$

for some positive constants C_1 and C_2, there exists a real random variable ξ such that the first four moments of ξ are 0, 1, m_3 and m_4 and the distribution ν of ξ satisfies logarithmic Sobolev inequality and the LSI constant is bounded from above by a function of C_1 and C_2. Moreover, ν can be chosen to be absolutely continuous with a smooth positive density, $\nu(\mathrm{d}x) = e^{-U(x)}\mathrm{d}x$, such that the derivatives of U satisfy

$$|U^{(k)}(x)| \le C_k(1+x^2)^{Ck} \tag{C.2}$$

with some fixed constant C and k-dependent constants C_k.

Remark. The last statement about the smoothness of U will not be needed in this paper, but we state it for further reference.

Proof. We start with the case $|m_3| > \delta$, where δ is small enough number to depend only on C_1, see below. Let ξ be the sum of two Gaussians, with density function of the form

$$f_\xi(x) = \frac{b}{(a+b)}\frac{1}{\sqrt{2\pi\sigma}}e^{-(x-a)^2/(2\sigma)} + \frac{a}{(a+b)}\frac{1}{\sqrt{2\pi\sigma}}e^{-(x+b)^2/(2\sigma)} \tag{C.3}$$

with some parameters $a > 0$, $b > 0$, $\sigma > 0$. If the first 4 moments of $f_\xi(x)$ are 0, 1, m_3 and m_4, then we have the relations

$$m_4 = 1 + \frac{m_3^2}{1-\sigma} + 4\sigma - 2\sigma^2, \tag{C.4}$$

$$ab = 1 - \sigma \ \text{ and } \ a - b = \frac{m_3}{1-\sigma}. \tag{C.5}$$

With m_3, m_4 in (C.1) and $|m_3| \geq \delta$, one can always find a solution of (C.4) such that $0 < \sigma < 1$. Actually, one can see that $c < \sigma < C$, where c and C only depend on C_1, C_2 in (6.19) and δ.

Once σ is found, it is easy to check that one can always find real solutions a, b for (C.5) as long as m_3, m_4 satisfy (C.1) and $|m_3| \geq \delta$. Since the solutions a, b, σ are continuous with respect to m_3 and m_4, then they are uniformly bounded. Distributions of the form (C.3) satisfy the LSI, since the are log concave away from a compact set. Since the parameters a, b, σ are in a compact set, the LSI constant will remain uniformly bounded with a bound depending on C_1, C_2 and δ. It is clear that the density function (C.3) is positive and its logarithm satisfies (C.2).

Now we consider the case that $|m_3| < \delta$ with a small $\delta = \frac{1}{100}\min\{1, C_1\}$, where C_1 is the constant in (C.1). Without loss of generality, we may assume $m_3 > 0$. We consider the following three parameter family of probability densities

$$f_{d,\beta,\varepsilon}(x) = (1-\varepsilon)g_{d,\beta}(x) + \varepsilon h(x)$$

with

$$g_{d,\beta}(x) = \frac{\beta+1}{2d^{\beta+1}} \cdot |x|^\beta \cdot \mathbf{1}(|x| \leq d), \qquad h(x) = \frac{b}{(a+b)}\frac{1}{\sqrt{2\pi}}e^{-(x-a)^2/2} + \frac{a}{(a+b)}\frac{1}{\sqrt{2\pi}}e^{-(x+b)^2/2},$$

where the parameters are in the range $-1 < \beta < \infty$, $0 < d < \infty$, $0 \leq \varepsilon \ll 1$ and a, b will be chosen explicitly. Simple calculation shows that the moments of $f_{d,\beta,\varepsilon}$ are $m_1 = 0$,

$$m_2 = (1-\varepsilon)\frac{\beta+1}{\beta+3}d^2 + \varepsilon(1+ab), \tag{C.6}$$

$$m_3 = \varepsilon ab(a-b), \tag{C.7}$$

$$m_4 = (1-\varepsilon)\frac{\beta+1}{\beta+5}d^4 + \varepsilon\Big[3 + ab(6 + a^2 + b^2 - ab)\Big]. \tag{C.8}$$

Choosing, say, $a = 2$, $b = 1$, and setting $m_2 = 1$, we obtain $d^2 = \frac{1-3\varepsilon}{1-\varepsilon}\frac{\beta+3}{\beta+1}$ from the first equation, $\varepsilon = m_3/2$ from the second equation and finally the last equation becomes

$$m_4 = \frac{(1-3m_3/2)^2}{1-m_3/2}\frac{(\beta+3)^2}{(\beta+1)(\beta+5)} + \frac{23}{2}m_3. \tag{C.9}$$

Recall that we are in the regime where $|m_3| \le \delta \le C_1/100$. For any fixed $0 \le m_3 \le \delta$, the right hand side of (C.9) is a monotonically decreasing function in $\beta \in (-1, \infty)$ whose value goes down from ∞ to $\frac{(1-3m_3/2)^2}{1-m_3/2} + \frac{23}{2}m_3 \le 1 + 20\delta$. But we know from (C.1) that $C_2 \ge m_4 \ge 1 + 100\delta$, thus there is a value β such that (C.9) holds, moreover, β is in a compact subinterval of $(-1, \infty)$ that depends only on δ and C_2. It is then easy to check that the support and the supremum norm of the density $g_{d,\beta}$ also remains in a compact set, depending only on δ. Therefore we constructed a probability measure with the given moments, that is a linear combination of two Gaussians plus a compactly supported piece with a nonnegative bounded density. To ensure smoothness, we replace $g_{d,\beta}$ with $\widetilde{g}_{d,\beta,\tau} := \vartheta_\tau * g_{d,\beta}$, where $\vartheta_\tau(x) = \tau^{-1}\vartheta(x/\tau)$ and ϑ is a compactly supported nonnegative smooth symmetric function with $\int \vartheta = 1$. The first moment m_1 is unchanged and the formulas (C.6) for the higher moments will get modified by an error term of order τ. Let τ be much smaller than all other parameters in this proof. It is easy to see that, by a simple calculation treating τ as a small perturbation, one can still choose a, b, ε and β in the previous argument to match $m_2 = 1, m_3$ and m_4.

Finally, note that the sum of two Gaussians satisfy the LSI, as well as its compact perturbation and the new LSI constant depends only on the supremum norm of the density of the perturbation. Since all these parameters remain uniformly controlled by C_1 and C_2, we proved Lemma C.1, i.e., Lemma 6.5 for $\gamma = 0$. □

Now consider the case $\gamma > 0$. For any real random variable ζ, independent of ξ^G, and with the first 4 moments being 0, 1, $m_3(\zeta)$ and $m_4(\zeta) < \infty$, the first 4 moments of

$$\zeta' = (1-\gamma)^{1/2}\zeta + \gamma^{1/2}\xi^G \tag{C.10}$$

are 0, 1,

$$m_3(\zeta') = (1-\gamma)^{3/2} m_3(\zeta) \tag{C.11}$$

and

$$m_4(\zeta') = (1-\gamma)^2 m_4(\zeta) + 6\gamma - 3\gamma^2. \tag{C.12}$$

Given m_3 and m_4, satisfying (C.1) and using Lemma C.1, we obtain that for any γ small enough, there exists a real random variable ξ_γ such that the first four moments are 0, 1,

$$m_3(\xi_\gamma) = (1-\gamma)^{-3/2} m_3 \tag{C.13}$$

and

$$m_4(\xi_\gamma) = m_3(\xi_\gamma)^2 + (m_4 - m_3^2).$$

With $m_4 \le C_2$, we have $m_3^2 \le C_2$, thus

$$|m_4(\xi_\gamma) - m_4| \le C\gamma \tag{C.14}$$

for some C depending on C_2.

Hence with (C.11) and (C.12), we obtain that $\xi' = (1-\gamma)^{1/2}\xi_\gamma + \gamma^{1/2}\xi^G$ satisfies $m_3(\xi') = m_3$ and (6.21). With Lemma C.1, we obtain that the LSI constant of ξ_γ is bounded by a constant only depends on C_1 and C_2, which completes the proof of Lemma 6.5. □

References

[1] Anderson, G.; Zeitouni, O. : A CLT for a band matrix model. Probab. Theory Related Fields **134** (2006), no. 2, 283–338.

[2] Ben Arous, G., Péché, S.: Universality of local eigenvalue statistics for some sample covariance matrices. *Comm. Pure Appl. Math.* **LVIII.** (2005), 1–42.

[3] Bleher, P., Its, A.: Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. *Ann. of Math.* **150** (1999): 185–266.

[4] Bobkov, S. G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999), no. 1, 1–28.

[5] Brézin, E., Hikami, S.: Correlations of nearby levels induced by a random potential. *Nucl. Phys. B* **479** (1996), 697–706, and Spectral form factor in a random matrix theory. *Phys. Rev. E* **55** (1997), 4067–4083.

[6] Deift, P.: Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. *Courant Lecture Notes in Mathematics* **3**, American Mathematical Society, Providence, RI, 1999

[7] Deift, P., Gioev, D.: Random Matrix Theory: Invariant Ensembles and Universality. *Courant Lecture Notes in Mathematics* **18**, American Mathematical Society, Providence, RI, 2009

[8] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.* **52** (1999):1335–1425.

[9] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Strong asymptotics of orthogonal polynomials with respect to exponential weights. *Comm. Pure Appl. Math.* **52** (1999): 1491–1552.

[10] Disertori, M., Pinson, H., Spencer, T.: Density of states for random band matrices. Commun. Math. Phys. **232**, 83–124 (2002)

[11] Dyson, F.J.: Correlations between eigenvalues of a random matrix. *Commun. Math. Phys.* **19**, 235-250 (1970).

[12] Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37**, No. 3, 815–852 (2008)

[13] Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.* **287**, 641–655 (2009)

[14] Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. Int Math Res Notices **2010** (3): 436-479 (2010)

[15] Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. To appear in *Invent. Math.* arxiv.org/abs/0907.5605

[16] Erdős, L., Ramirez, J., Schlein, B., Yau, H.-T.: *Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation.* Electr. J. Prob. **15**, Paper 18, 526–604 (2010)

[17] Erdős, L., Péché, G., Ramírez, J., Schlein, B., and Yau, H.-T., Bulk universality for Wigner matrices. Comm. Pure Appl. Math. **63**, No. 7, 895-925 (2010)

[18] Erdős, L., Ramírez, J., Schlein, B., Tao, T., Vu, V. and Yau, H.-T., Bulk universality for Wigner hermitian matrices with subexponential decay. *Math. Res. Lett.* **17** (4), 667–674 (2010).

[19] Erdős, L., Schlein, B., Yau, H.-T., Yin, J.: The local relaxation flow approach to universality of the local statistics for random matrices. To appear in Annales Inst. H. Poincaré, Prob. and Stat. Preprint arXiv:0911.3687

[20] Erdős, L., Yau, H.-T., Yin, J.: Universality for generalized Wigner matrices with Bernoulli distribution. To appear in Journal of Combinatorics. Preprint arXiv:1003.3813

[21] Guionnet, A.: Large deviation upper bounds and central limit theorems for band matrices, *Ann. Inst. H. Poincaré Probab. Statist* **38** , (2002), pp. 341-384.

[22] Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Math. Stat.* **42** (1971), no.3, 1079-1083.

[23] Johansson, K.: Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* **215** (2001), no.3. 683–705.

[24] Johansson, K.: Universality for certain hermitian Wigner matrices under weak moment conditions. Preprint arxiv.org/abs/0910.4467

[25] Mehta, M.L.: Random Matrices. Academic Press, New York, 1991.

[26] Mehta, M.L., Gaudin, M.: On the density of eigenvalues of a random matrix. *Nuclear Phys.* **18**, 420-427 (1960).

[27] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian matrix models. *J. Stat. Phys.* **130** (2008), no.2., 205-250.

[28] Shiryayev, A. N.: Probability. Graduate Text in Mathematics. **54**. Springer, 1984.

[29] Spencer, T.: *Random banded and sparse matrices (Chapter 23)* to appear in "Oxford Handbook of Random Matrix Theory", edited by G. Akemann, J. Baik and P. Di Francesco.

[30] Stroock, D.W.: Probability theory, an analytic view. Cambridge University Press, 1993.

[31] Tao, T. and Vu, V.: Random matrices: Universality of the local eigenvalue statistics. Preprint arXiv:0906.0510.

[32] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics up to the edge. Preprint. arXiv:0908.1982

[33] Tao, T. and Vu, V.: Random covariance matrices: Universality of local statistics of eigenvalues. Preprint. arXiv:0912.0966

[34] Vu, V.: Spectral norm of random matrices. *Combinatorica*, **27** (6) (2007), 721-736.