Lectures on the local semicircle law for Wigner matrices

Florent Benaych-Georges   Antti Knowles

June 12, 2019

These notes provide an introduction to the local semicircle law from random matrix theory, as well as some of its applications. We focus on *Wigner matrices*, Hermitian random matrices with independent upper-triangular entries with zero expectation and constant variance. We state and prove the local semicircle law, which says that the eigenvalue distribution of a Wigner matrix is close to Wigner's semicircle distribution, down to spectral scales containing slightly more than one eigenvalue. This local semicircle law is formulated using the Green function, whose individual entries are controlled by large deviation bounds.

We then discuss three applications of the local semicircle law: first, complete delocalization of the eigenvectors, stating that with high probability the eigenvectors are approximately flat; second, rigidity of the eigenvalues, giving large deviation bounds on the locations of the individual eigenvalues; third, a comparison argument for the local eigenvalue statistics in the bulk spectrum, showing that the local eigenvalue statistics of two Wigner matrices coincide provided the first four moments of their entries coincide. We also sketch further applications to eigenvalues near the spectral edge, and to the distribution of eigenvectors.

Contents

1. Introduction

These notes are based on lectures given by Antti Knowles at the conference *États de la recherche en matrices aléatoires* at the Institut Henri Poincaré in December 2014. In them, we state and prove the local semicircle law and give several applications to the distribution of eigenvalues and eigenvectors. We favour simplicity and transparency of arguments over generality of results. In particular, we focus on one of the simplest ensembles of random matrix theory: Wigner matrices.

A Wigner matrix is a Hermitian random matrix whose entries are independent up to the symmetry constraint, and have zero expectation and constant variance. This definition goes back to the seminal work of Wigner [79], where he also proved the semicircle law, which states that the asymptotic eigenvalue distribution of a Wigner matrix is given with high probability by the semicircle distribution. Wigner matrices occupy a central place in random matrix theory, as a simple yet nontrivial ensemble on which many of the fundamental features of random matrices, such as universality of the local eigenvalue statistics and eigenvector delocalization, may be analysed. Moreover, many of the techniques developed for Wigner matrices, such as the ones presented in these notes, extend to other random matrix ensembles or serve as a starting point for more sophisticated methods.

The local semicircle law is a far-reaching generalization of Wigner's original semicircle law, and constitutes the key tool for analysing the local distribution of eigenvalues and eigenvectors of Wigner matrices, and in particular in the study of the universality of Wigner matrices. Roughly, it states that the eigenvalue distribution is well approximated by the semicircle distribution down to scales containing only slightly more than one eigenvalue. The first instance of the local semicircle law down to optimal spectral scales was obtained by Erdős, Schlein, and Yau in [40], following several previous results on larger spectral scales [39, 52]. Since then, the local semicircle law has been improved and generalized in a series of works [41, 38, 43, 44, 45, 34, 37, 25]. The proof presented in these notes is modelled on the argument of Erdős, Knowles, Yau, and Yin from [37], which builds on the works [39, 40, 41, 43, 44, 45] of Erdős, Schlein, Yau, and Yin.

In order to keep these notes focused, for the applications we restrict ourselves to relatively simple consequences of the semicircle law: eigenvalue rigidity, eigenvector delocalization, and a four-moment comparison theorem for the local eigenvalue statistics. Further topics and applications, such as distribution of eigenvalues near the spectral edges, Dyson Brownian motion and its local relaxation time, distribution of eigenvectors, and spectral statistics of deformed matrix ensembles, are not covered here. For further reading on Wigner matrices, we recommend the books [60, 3]. Additional applications of the local semicircle law, in particular in the analysis of the local relaxation time of Dyson Brownian motion, are given in the survey [30]. In Section 11.4, we briefly outline applications of the local semicircle law to the analysis of eigenvectors and eigenvalues near the spectral edge.

Outline. In Section 2 we define Wigner matrices and state the local semicircle law, or *local law* for short. We also give two simple consequences of the local law: *eigenvalue rigidity* and *complete eigenvector delocalization*. Sections 3–7 are devoted to the proof of the local law. Section 3 collects some basic tools from linear algebra and probability theory that are used throughout the proof. Section 4 gives a detailed outline of the proof. In Section 5 we perform the first of two major steps of the proof: the weak local law, which yields control down to optimal scales but with non-optimal error bounds. In Section 6 we perform the second major step of the proof, which yields optimal error bounds and concludes the proof of the local law. The key estimate used to obtain the optimal

error bounds is a fluctuation averaging result, whose proof is given in Section 7.

Having concluded the proof of the local law, in Sections 8–10 we draw some simple consequences. In Section 8 we prove the semicircle law on small scales, which provides large deviation bounds on the number of eigenvalues in small intervals. In Section 9 we prove eigenvalue rigidity, which provides large deviation bounds on the locations of individual eigenvalues. In Section 10 we extend the estimates from the local law to arbitrary scales and distances from the spectrum.

In Section 11 we illustrate how to use the local law to obtain a four-moment comparison theorem for the local eigenvalue statistics, using the Green function comparison method from [43]. We also sketch how to extend such comparison methods to the edge of the spectrum and to eigenvectors. Finally, in Section 12 we discuss further generalizations of the local law, and also other random matrix models for which local laws have been established.

The appendices contain some basic tools from linear algebra, spectral theory, and probability that are used throughout the notes, as well as some standard results on the semicircle distribution and the norm of Wigner matrices.

Pedagogical aspirations aside, in these notes we also give a coherent summary of the different guises of the local semicircle law that have proved useful in random matrix theory. They have all appeared, at least implicitly, previously in the literature; we take this opportunity to summarize them explicitly in Sections 2 and 10.

Conventions. We use C to denote a generic large positive constant, which may depend on some fixed parameters and whose value may change from one expression to the next. Similarly, we use c to denote a generic small positive constant. For two positive quantities A_N and B_N depending on N we use the notation $A_N \asymp B_N$ to mean $C^{-1} A_N \leqslant B_N \leqslant C A_N$ for some positive constant C. In statements of assumptions, we use the special constant $\tau > 0$, which should be thought of as a fixed arbitrarily small number. A smaller τ always leads to a weaker assumption. We always use the Euclidean norm $|\mathbf{v}|$ for vectors $\mathbf{v}$, and denote by $\|A\|$ the associated operator norm of a matrix A. Finally, in some heuristic discussions we use the symbols $\ll$, $\gg$, and $\approx$ to mean "much less than", "much greater than", and "approximately equal to", respectively.

2. The local law

Let $H = H^* = (H_{ij} : 1 \leqslant i, j \leqslant N) \in \mathbb{C}^{N \times N}$ be an $N \times N$ random Hermitian matrix. We normalize H so that its eigenvalues are typically of order one, $\|H\| \asymp 1$. A central goal of random matrix theory is to understand the distribution of the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N$ and the normalized associated eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$ of H.

Since most quantities that we are interested in depend on N, we shall almost always omit the explicit argument N from our notation. Hence, every quantity that is not explicitly a constant is in fact a sequence indexed by $N \in \mathbb{N}$.

2.1. The Green function. The main tool in the study of the eigenvalues and eigenvectors of H is the *Green function* or *resolvent*

$$G(z) := (H - z)^{-1},$$

where $z \in \mathbb{C} \setminus \{\lambda_1, \ldots, \lambda_N\}$ is the *spectral parameter*. Here, and throughout the following, we identify the matrix zI with the scalar z.

Since we are ultimately interested in the eigenvalues and eigenvectors, it might seem that the Green function is an unnecessary distraction. In fact, however, the Green function is a much simpler and more stable object than the eigenvalues and eigenvectors. Using the Green function as the central object instead of eigenvalues and eigenvectors allows one to establish results whose direct proof (i.e. working directly on the eigenvalues and eigenvectors) is not feasible or at least much more complicated. We give an informal summary of the key features of the Green function that make it such a powerful tool.

(i) The Green function contains the complete information about the eigenvalues and eigenvectors. Indeed, by spectral decomposition we have

$$G(z) = \sum_{i=1}^{N} \frac{\mathbf{u}_i \mathbf{u}_i^*}{\lambda_i - z}. \tag{2.1}$$

Hence, G is a matrix-valued meromorphic function in the complex plane, whose poles are the eigenvalues λ_i of H and whose residues the spectral projections $\mathbf{u}_i \mathbf{u}_i^*$ of H. We may therefore use complex analysis to extract the full information about λ_i and $\mathbf{u}_i$ from $G(z)$. Note that the spectral projection $\mathbf{u}_i \mathbf{u}_i^*$ is a more natural object than the eigenvector $\mathbf{u}_i$, since it discards the ambiguous phase of $\mathbf{u}_i$. More details are given in Section 11.

(ii) Green functions form a basis of a powerful *functional calculus*: every "reasonable" function $f(H)$ of H may be expressed as a superposition of Green functions. For instance, if f is holomorphic in an open domain containing the spectrum of H, we have

$$f(H) = -\frac{1}{2\pi\mathrm{i}} \oint_\Gamma f(z) G(z) \, \mathrm{d}z, \tag{2.2}$$

where Γ is a contour encircling the spectrum of H.

A more general and powerful functional calculus is provided by the *Helffer-Sjöstrand functional calculus* given in Appendix C.

(iii) The Green function is trivial to differentiate in H. For instance,

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

and higher order derivatives have analogous simple expressions. This is in stark contrast to the eigenvalues and eigenvectors, whose derivatives in the entries of H, while explicit, are notoriously unwieldy in higher orders.

A related observation is that the perturbation theory of Green functions is trivial. Let $\widetilde{H} = H + \Delta$ be a perturbation of H, and write $\widetilde{G}(z) := (\widetilde{H} - z)^{-1}$. Then by iterating the resolvent identity $\widetilde{G} = G - G\Delta\widetilde{G}$ we get the *resolvent expansion*

$$\widetilde{G}(z) = \sum_{k=0}^{n-1} G(z)(-\Delta G(z))^k + \widetilde{G}(z)(-\Delta G(z))^n. \tag{2.3}$$

Analogous perturbation series for the eigenvalues and eigenvectors are thoroughly unpleasant.

5

(iv) The Green function satisfies a host of very useful identities. One example is the resolvent expansion from (2.3). More sophisticated identities can be derived from Schur's complement formula; see Lemma 3.5 below.

(v) The Green function is analytically stable. More precisely, it is better suited for doing estimates and analysing the eigenvalue density on small scales than other families of functions. Common alternative families are the moments H^k, $k \in \mathbb{N}$, and the unitary one-parameter group $\mathrm{e}^{\mathrm{i}tH}$, $t \in \mathbb{R}$. Each of these families serves as a basis of a functional calculus. For example, we have the Fourier representation $f(H) = \int \mathrm{d}t\, \hat{f}(t)\, \mathrm{e}^{-\mathrm{i}tH}$, where $\hat{f}$ is the Fourier transform of f. The advantage of Green functions is that integrals of the form (2.2) are far less oscillatory than e.g. the Fourier representation. For many arguments, this means that reaching small spectral scales using Green functions is much easier than using moments or the unitary one-parameter group. In fact, in these notes we show how to obtain control on the optimal scales using Green functions. Control of analogous precision using moments or the unitary one-parameter group has until now not been achieved.

We conclude this subsection with a cultural discussion on the term *Green function* (sometimes also called *Green's function*). Many mathematical problems can be formulated in terms of a linear operator $L : f \to Lf$ acting on functions $f = f(x)$ defined on some set $\mathcal{X}$. A prominent example is the Dirichlet problem on an open domain $\mathcal{X} \subset \mathbb{R}^d$, where the linear operator $L = -\Delta$ is the Laplacian with Dirichlet boundary conditions. For such a problem, the *Green function* $G(x, y)$ is a function of two variables, $x, y \in \mathcal{X}$, defined as the integral kernel of the resolvent $(L - z)^{-1}$ with respect to some natural measure on $\mathcal{X}$, often Lebesgue measure if $\mathcal{X}$ is an open subset of $\mathbb{R}^d$ or the counting measure if $\mathcal{X}$ is discrete. Here the spectral parameter z is often taken to be 0. We discuss a few examples.

(i) In the above example of the Dirichlet problem with $\mathcal{X} \subset \mathbb{R}^d$ open and $L = -\Delta$ with Dirichlet boundary conditions, the Green function $L^{-1}(x, y)$ is the integral kernel of L^{-1} with respect to Lebesgue measure. It gives the solution of the problem $Lf = \rho$ via the integral formula $f(x) = \int_{\mathcal{X}} \mathrm{d}y\, G(x, y)\rho(y)$.

(ii) For a discrete-time random walk on $\mathbb{Z}^d$, L is the transition matrix of the walk, and the Green function $G(x, y) = (1 - L)^{-1}(x, y)$ is the integral kernel of $(1 - L)^{-1}$ with respect to the counting measure on $\mathbb{Z}^d$. The quantity $G(x, y)$ has the probabilistic interpretation of the expected time spent at y starting from x.

(iii) For a continuous-time $\mathbb{R}^d$-valued Markov process $(X_t)_{t \geqslant 0}$ with generator L, the Green function $G(x, y) = (-L)^{-1}(x, y)$ is the integral kernel of $(-L)^{-1}$ with respect to Lebesgue measure. The quantity $G(x, y)\, \mathrm{d}y$ has the probabilistic interpretation of the expected time spent in $\mathrm{d}y$ starting from x. Indeed, for a test function f we find $\int \mathrm{d}y\, G(x, y)f(y) = ((-L)^{-1}f)(x) = \int_0^\infty \mathrm{d}t\, (\mathrm{e}^{Lt}f)(x) = \int_0^\infty \mathrm{d}t\, \mathbb{E}^x[f(X_t)] = \mathbb{E}^x\big[\int_0^\infty \mathrm{d}t\, f(X_t)\big]$, where in the third step we used that the semigroup $S(t)$ defined by $S(t)f(x) := \mathbb{E}^x[f(X_t)]$ is given as $S(t) = \mathrm{e}^{Lt}$ in terms of the generator L.

(iv) A Green function can be associated with much more general differential operators than the Laplacian from (i), and can for instance also be used for parabolic and hyperbolic partial differential equations. For example, let $\mathcal{X} = \mathbb{R} \times \mathbb{R}^d$ and $L = \partial_0^2 - \partial_1^2 - \cdots - \partial_d^2$ be the d'Alambert operator. A Green function $G(x, y)$ gives a solution of the wave equation $Lf = \rho$

6

using the integral formula $f(x) = \int \mathrm{d}y\, G(x,y)\rho(y)$. Note that, since the homogeneous wave equation $Lf = 0$ has nontrivial solutions, L is not invertible and the Green function is not unique. Nevertheless, Green functions can be easily constructed.

(v) Let $\mathcal{X} = \{1, \ldots, d\}$, and L be a positive matrix on $\mathbb{R}^d$. Define the Gaussian probability measure $\mu(\mathrm{d}\xi) := \frac{1}{Z} \mathrm{e}^{-\frac{1}{2}\langle \xi, L\xi \rangle} \mathrm{d}\xi$. Then the *two-point correlation function* $G(x,y) := \int \mu(\mathrm{d}\xi)\, \xi_x\, \xi_y$ is simply equal to the matrix entry $(L^{-1})(x,y)$, i.e. the integral kernel of L^{-1} with respect to the counting measure on $\mathcal{X}$. Generally, in statistical mechanics correlation functions of free fields are often referred to as Green functions. This convention is natural for classical fields, where a free field is a Gaussian measure, but it is also commonly used in quantum field theory.

(vi) Finally, in the context of random matrix theory we take $\mathcal{X} = \{1, \ldots, N\}$, $L = H$, and write G_{ij} instead of $G(x,y)$. The spectral parameter z plays a central role and is therefore often kept in the notation. Moreover, we do not distinguish the resolvent $(H-z)^{-1}$ from its entries $G_{ij} = (H-z)^{-1}_{ij}$, and refer to both as the Green function.

2.2. Local and global laws. We define the *empirical eigenvalue distribution*

$$\mu \ := \ \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}\,. \tag{2.4}$$

For many matrix models H one has $\mu \to \varrho$ as $N \to \infty$ (in some sense to be made precise), where ϱ is a deterministic probability measure that does not depend on N. (For instance, for the Wigner matrices defined in Definition 2.2 below, ϱ is the semicircle distribution (2.10).) This convergence is best formulated in terms of *Stieltjes transforms*.

We define the Stieltjes transforms

$$s(z) \ := \ \int \frac{\mu(\mathrm{d}x)}{x-z} \ = \ \frac{1}{N}\operatorname{Tr} G(z) \tag{2.5}$$

and

$$m(z) \ := \ \int \frac{\varrho(\mathrm{d}x)}{x-z} \tag{2.6}$$

of the probability measures μ and ϱ. Note that $s(z)$ is random while $m(z)$ is deterministic.

From now on, we use the notation

$$z \ = \ E + \mathrm{i}\eta$$

for the real and imaginary parts of z. We always assume that z lies in the open upper-half plane $\mathbb{C}_+$, i.e. that $\eta > 0$.

Note that

$$\frac{1}{\pi}\operatorname{Im} s(z) \ = \ \frac{1}{N}\sum_{i=1}^{N} \frac{\eta/\pi}{(\lambda_i - E)^2 + \eta^2} \ = \ (\mu * \theta_\eta)(E)\,, \tag{2.7}$$

where we defined

$$\theta_\eta(x) \ := \ \frac{1}{\pi}\operatorname{Im}\frac{1}{x-\mathrm{i}\eta} \ = \ \frac{\eta/\pi}{x^2 + \eta^2}\,. \tag{2.8}$$

7

The function θ_η is an approximate delta function as $\eta \downarrow 0$, with breadth η. We conclude that control of $\pi^{-1} \operatorname{Im} s(z)$ is tantamount to control of the empirical distribution μ smoothed out on the scale η. Hence, η is called the *spectral resolution*.

Stieltjes transforms in particular provide a convenient way of studying convergence of (random) probability measures in distribution. The following lemma is proved in Appendix E.

Lemma 2.1. *Let $\mu \equiv \mu_N$ be a random probability measure, and ϱ a deterministic probability measure. Let $s \equiv s_N$ and m denote the Stieltjes transforms of μ and ϱ respectively. Then*

$$\mu \xrightarrow{\mathrm{d}} \varrho \qquad \Longleftrightarrow \qquad s(z) \longrightarrow m(z) \text{ for all fixed } z \in \mathbb{C}_+ \,, \tag{2.9}$$

where both convergences are as $N \to \infty$ and with respect to convergence in probability in the randomness of μ or s, and $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution of probability measures[1].

Note that in Lemma 2.1 the spectral parameter z does not depend on N. Hence, recalling the discussion following (2.7), we see that the convergence $s(z) \to m(z)$ has a spectral resolution of order one. A result of the form

$$s(z) \longrightarrow m(z) \text{ for all } z \in \mathbb{C}_+$$

(for instance in probability) is therefore called a *global law*.

A *local law* is a result that controls the error $s(z) - m(z)$ for all z satisfying $\eta \gg N^{-1}$. In other words, a local law admits arguments z that depend on N, so that the spectral resolution η may be much smaller than the global scale 1. We always require the spectral resolution to be bigger than N^{-1}. This restriction is necessary and its origin easy to understand. Since we assumed that $\|H\| \asymp 1$ and H has N eigenvalues, the typical separation of eigenvalues is of order N^{-1}. Individual eigenvalues are expected to fluctuate about their mean locations, so that the random quantity $s(z)$ can only be expected to be close to the deterministic quantity $m(z)$ if some averaging mechanism ensures that $s(z)$ depends strongly on a large number of eigenvalues. This means that the spectral resolution η has to be larger than the typical eigenvalue spacing N^{-1}. See Figure 2.1 for an illustration of the spectral scale η and the approximation from (2.7).

For most applications to the distribution of the eigenvalues and eigenvectors, control of just $s(z) = N^{-1} \operatorname{Tr} G(z)$ is not enough, and one needs control of the Green function $G(z)$ regarded as a matrix. The need for such a stronger control is obvious if we are interested in the eigenvectors of H (recall the spectral decomposition from (2.1)). However, perhaps surprisingly, to understand the distribution of individual eigenvalues we also need to control the individual entries of G; see Section 11 for more details.

We shall see that, at least for Wigner matrices, the matrix $G(z)$ is close (in some sense to be made precise) to $m(z)$, a multiple of the identity matrix.

2.3. The local law for Wigner matrices. From now we focus on the case where H is a Wigner matrix.

Definition 2.2 (Wigner matrix). *A* Wigner matrix, *or* Wigner ensemble, *is a Hermitian $N \times N$ matrix $H = H^*$ whose entries H_{ij} satisfy the following conditions.*

---

[1]Explicitly, "$\mu \xrightarrow{\mathrm{d}} \varrho$ in probability" means: for all continuous bounded functions $f : \mathbb{R} \to \mathbb{R}$ and all $\varepsilon > 0$ we have $\mathbb{P}(|\int f \, \mathrm{d}(\mu - \varrho)| > \varepsilon) \to 0$ as $N \to \infty$.
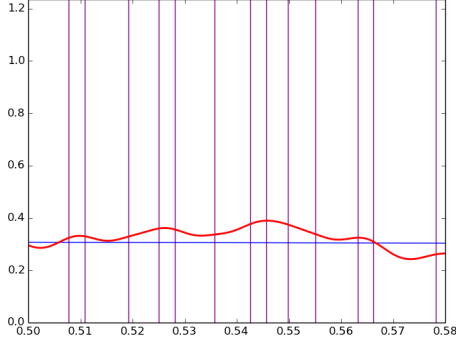
(a) Spectral resolution $\eta = N^{-0.8}$.
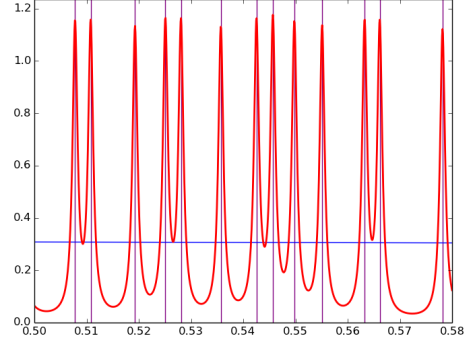
(b) Spectral resolution $\eta = N^{-1.2}$.

Figure 2.1. *Role of the spectral resolution η.* In both plots we draw the eigenvalues of a 500×500 GUE matrix (vertical lines), $\operatorname{Im} s(z)/\pi$ as a function of E for a fixed η (red curve), and $\operatorname{Im} m(z)/\pi$ as a function of E for the same fixed η (blue curve). Both plots are drawn using the same realization of the underlying matrix and the same x- and y-ranges.

(i) *The upper-triangular entries $(H_{ij} : 1 \leqslant i \leqslant j \leqslant N)$ are independent.*

(ii) *For all i, j we have $\mathbb{E}H_{ij} = 0$ and $\mathbb{E}|H_{ij}|^2 = N^{-1}(1 + O(\delta_{ij}))$.*

(iii) *The random variables $\sqrt{N} H_{ij}$ are bounded in any L^p space, uniformly in N, i, j.*

The conditions (i) and (ii) are standard. The factor $1 + O(\delta_{ij})$ in (ii) admits more general variances for the diagonal entries of H; as we shall see, many properties of the model are insensitive to the variances of the diagonal entries. Note that we do not assume the entries of H to be identically distributed, and impose no constraint on $\mathbb{E}(H_{ij})^2$ associated with some symmetry class.

The condition (iii) is technical and made for convenience. Explicitly, it states that for each $p \in \mathbb{N}$ there exists a constant C_p such that $\|\sqrt{N} H_{ij}\|_p \leqslant C_p$ for all N, i, j, where we use the notation

$$\|X\|_p := \left(\mathbb{E}|X|^p\right)^{1/p}.$$

In the literature, various other conditions on the tails of the entries $\sqrt{N} H_{ij}$ have been used, ranging from sub-Gaussian tails to merely the existence of the second moment in (ii). The assumption (iii) can be relaxed (for instance by truncation), but we shall not pursue this direction in these notes.

This choice of the normalization N^{-1} in (ii) ensures that $\|H\| \asymp 1$ as required above. A simple heuristic way to convince ourselves that this is the right normalization is to compute the average square distance of an eigenvalue from the origin:

$$\mathbb{E}\frac{1}{N}\sum_{i=1}^N \lambda_i^2 = \frac{1}{N}\mathbb{E}\operatorname{Tr} H^2 = \frac{1}{N}\sum_{i,j=1}^N \mathbb{E}|H_{ij}|^2 = 1 + O(N^{-1}),$$

as desired.

Example 2.3 (Gaussian ensembles). *Let X be an $N \times N$ matrix with i.i.d. real standard normal entries. Then the* Gaussian orthogonal ensemble *(GOE) is defined as*

$$\frac{X + X^*}{\sqrt{2N}}\,.$$

Similarly, let Y be an $N \times N$ matrix with i.i.d. complex standard normal entries. (This means that $\operatorname{Re} Y_{ij}$ and $\operatorname{Im} Y_{ij}$ are independent standard normals.) Then the Gaussian unitary ensemble *(GUE) is defined as*

$$\frac{Y + Y^*}{\sqrt{4N}}\,.$$

The entries of the GOE and the GUE are centred Gaussian random variables. Moreover, it is easy to check that if H is the GOE then $N\mathbb{E}|H_{ij}|^2 = 1 + \delta_{ij}$, and if H is the GUE then $N\mathbb{E}|H_{ij}|^2 = 1$. The terminology orthogonal/unitary stems from the fact that the GOE is invariant under conjugation $H \mapsto OHO^*$ by an arbitrary deterministic orthogonal matrix O, and the GUE is invariant under conjugation $H \mapsto UHU^*$ by an arbitrary deterministic unitary matrix U.

Next, we define the *semicircle distribution*

$$\varrho(\mathrm{d}x) \;:=\; \frac{1}{2\pi}\sqrt{(4 - x^2)_+}\,\mathrm{d}x\,. \tag{2.10}$$

It is not hard to check (see Lemma B.2) that the Stieltjes transform (2.6) of the semicircle distribution ϱ is

$$m(z) \;=\; \int \frac{\varrho(\mathrm{d}x)}{x - z} \;=\; \frac{-z + \sqrt{z^2 - 4}}{2}\,, \tag{2.11}$$

where the square root $\sqrt{z^4 - 4}$ is chosen with a branch cut in the segment $[-2, 2]$ so that $\sqrt{z^2 - 4} \sim z$ as $z \to \infty$. This ensures that $m(z) \in \mathbb{C}_+$ for $z \in \mathbb{C}_+$.

Clearly, we have

$$m(z) + \frac{1}{m(z)} + z \;=\; 0\,. \tag{2.12}$$

In fact, it is easy to show, for $z \in \mathbb{C}_+$, that $m(z)$ is characterized as the unique solution of (2.12) in the upper half-plane $\mathbb{C}_+$. This latter characterization is often more convenient than the form (2.11).

The following global law, illustrated by Figure 2.2, is well known (see e.g. [4, 3]). (It will also follow from the local law, Theorem 2.6 below.)

Theorem 2.4 (Global law). *Let H be a Wigner matrix and ϱ defined by (2.10). Then for all fixed $z \in \mathbb{C}_+$ we have $s(z) \to m(z)$ in probability as $N \to \infty$.*

In order to state the local law, we use the following notion of high-probability bounds that was introduced in [33]. It provides a simple way of systematizing and making precise statements of the form "X is bounded with high probability by Y up to small powers of N". As we shall see, it is a very convenient way of wrapping various details of convergence in high probability, such as the low-probability exceptional events, into an object that very rarely needs to be unwrapped.
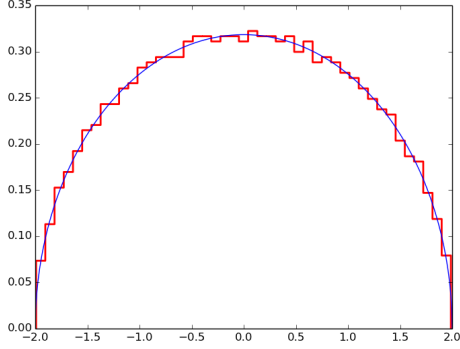
(a) Histogram of the spectrum (in red) and density of the semicircle distribution (in blue).
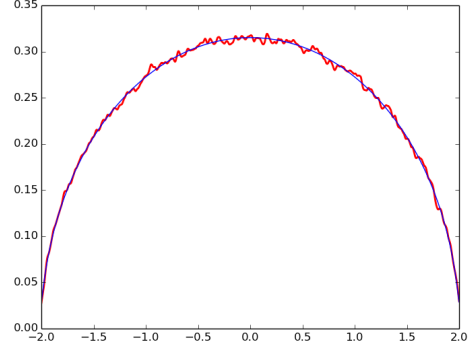
(b) As functions of E, for fixed spectral resolution $\eta = 0.02$: $\operatorname{Im} s(z)/\pi$ (in red) and $\operatorname{Im} m(z)/\pi$ (in blue).

Figure 2.2. *Global semicircle law.* Both plots, made with the same realization of a GUE matrix with $N = 2 \cdot 10^3$, give slightly different (yet equivalent, by Lemma 2.1) illustrations of the global law from Theorem 2.4. As observed in (2.7), $\operatorname{Im} m(z)/\pi$ is close to the density of the semicircle distribution when η is close (but not too close; see Figure 2.1) to zero.

Definition 2.5 (Stochastic domination). *Let*

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

be two families of nonnegative random variables, where $U^{(N)}$ is a possibly N-dependent parameter set.

We say that X is stochastically dominated by Y, *uniformly in u, if for all (small) $\varepsilon > 0$ and (large) $D > 0$ we have*

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^\varepsilon Y^{(N)}(u)\right] \leqslant N^{-D}$$

for large enough $N \geqslant N_0(\varepsilon, D)$. The stochastic domination is always uniform in all parameters (such as matrix indices and spectral parameters z) that are not explicitly fixed.

If X is stochastically dominated by Y, uniformly in u, we use the notation $X \prec Y$. Moreover, if for some complex family X we have $|X| \prec Y$ we also write $X = O_\prec(Y)$.

Finally, we say that an event $\Xi = \Xi^{(N)}$ holds with high probability *if $1 - \mathbf{1}(\Xi) \prec 0$, i.e. if for any $D > 0$ there is $N_0(D)$ such that for all $N \geqslant N_0(D)$ we have $\mathbb{P}(\Xi^{(N)}) \geqslant 1 - N^{-D}$.*

For example, it is easy to check that $|H_{ij}| \prec N^{-1/2}$. Note that this notation implicitly means that $\prec$ is uniform in the indices i, j.

We may now state the main result of these notes.

Theorem 2.6 (Local law for Wigner matrices). *Let H be a Wigner matrix. Fix $\tau > 0$ and define the domain*

$$\mathbf{S} \equiv \mathbf{S}_N(\tau) := \left\{E + \mathrm{i}\eta : |E| \leqslant \tau^{-1}, N^{-1+\tau} \leqslant \eta \leqslant \tau^{-1}\right\}. \tag{2.13}$$

Then we have

$$s(z) \; = \; m(z) + O_{\prec}\left(\frac{1}{N\eta}\right) \tag{2.14}$$

and

$$G_{ij}(z) \; = \; m(z)\delta_{ij} + O_{\prec}(\Psi(z)) \tag{2.15}$$

uniformly for $i, j = 1, \dots, N$ and $z \in \mathbf{S}$, where we defined the deterministic error parameter

$$\Psi(z) \; := \; \sqrt{\frac{\operatorname{Im} m(z)}{N\eta}} + \frac{1}{N\eta} \,. \tag{2.16}$$

For $|E| \leqslant 2$, both estimates (2.14) and (2.15), as encapsulated by the respective error parameters $(N\eta)^{-1}$ and Ψ, are optimal up to the details in the definition of $\prec$. For $|E| > 2$ the optimal bounds are better than those of Theorem 2.6. Remarkably, these optimal bounds for $|E| > 2$ are essentially a consequence of the weaker ones from Theorem 2.6; see Theorems 10.1 and 10.3 below for the precise statements.

To help the interpretation of the error parameter Ψ, we consider the cases where E is in the *bulk* $(-2, 2)$ of the spectrum and at the *edge* $\{-2, 2\}$ of the spectrum. Suppose first that $E \in (-2, 2)$ is fixed in the bulk. Then we easily find $\operatorname{Im} m(z) \asymp 1$ for $\eta \in [N^{-1}, 1]$. Hence, the first term of (2.16) dominates and we have $\Psi(z) \asymp (N\eta)^{-1/2}$, which is much smaller than $\operatorname{Im} m(z)$ for $\eta \gg N^{-1}$. Note that the scale N^{-1} is the typical separation of the eigenvalues in the bulk, as can be seen for instance by choosing $i \in [cN, (1 - c)N]$ in (2.19) below, for some constant $c > 0$.

On the other hand, if $E = 2$ is at the edge, we find $\operatorname{Im} m(z) \asymp \sqrt{\eta}$. We conclude that the first term of (2.16) dominates over the second if $\eta \gg N^{-2/3}$ and the second over the first if $\eta \ll N^{-2/3}$. Note that the threshold $N^{-2/3}$ is precisely the typical separation of the eigenvalues near the edge, as can be seen for instance by choosing $i \leqslant C$ in (2.19) below, for some constant $C > 0$. Hence, we conclude that at the edge $\Psi(z)$ is much smaller than $\operatorname{Im} m(z)$ provided that $\eta \gg N^{-2/3}$. See Figure 2.3.

Remark 2.7. The estimate (2.15) holds uniformly in $i, j = 1, \dots, N$ and $z \in \mathbf{S}$. In general, such an estimate does of course not imply a simultaneous high-probability bound for all $z \in \mathbf{S}$. However, thanks to the Lipschitz-continuity of all quantities in (2.15), this implication does hold. More precisely, we have

$$\sup_{z \in \mathbf{S}} \max_{i,j} \frac{|G_{ij}(z) - m(z)\delta_{ij}|}{\Psi(z)} \; \prec \; 1 \,. \tag{2.17}$$

To obtain (2.17) from (2.15), we define the N^{-3}-net $\widehat{\mathbf{S}} := (N^{-3}\mathbb{Z}^2) \cap \mathbf{S}$. Hence, $|\widehat{\mathbf{S}}| \leqslant CN^6$ and for any $z \in \mathbf{S}$ there exists a $w \in \widehat{\mathbf{S}}$ such that $|s - w| \leqslant 2N^{-3}$. By a simple union bound, we find from (2.15) that (2.17) holds with $\mathbf{S}$ replaced with $\widehat{\mathbf{S}}$. Using that $G_{ij}(z)$, $m(z)$, and $\Psi(z)$ are all N^2-Lipschitz on $\mathbf{S}$ (as follows immediately from their definitions) and that $\Psi(z) \geqslant 1/N$ on $\mathbf{S}$, we easily deduce (2.17).

2.4. Applications of Theorem 2.6. We now state some important consequences of Theorem 2.6. The first one gives the semicircle law on *small scales*. In Lemma 2.1 we saw that control of $s(z) - m(z)$ for fixed z yields control of $\mu - \varrho$ on large scales. Using the strong local control on $s(z) - m(z)$ from Theorem 2.6, we may correspondingly obtain strong bounds on the local convergence of μ to ϱ.
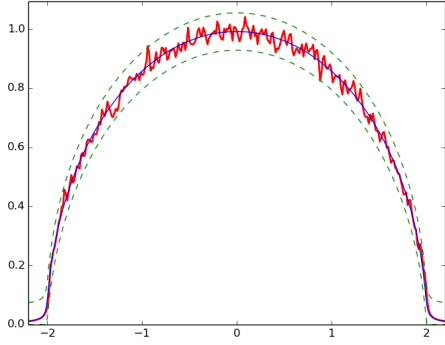
(a) $\mathrm{Im}\, s(z)$ (in red), $\mathrm{Im}\, m(z)$ (in blue) and the error bounds $\mathrm{Im}\, m(z) \pm 1/(N\eta)$ (green dashed lines).
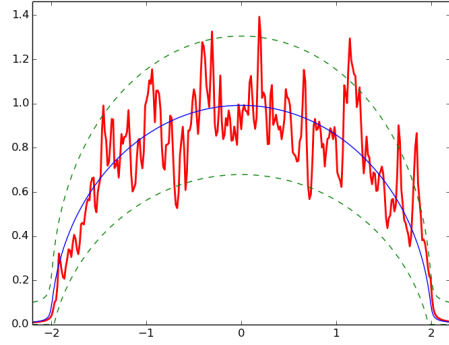
(b) $\mathrm{Im}\, G_{11}(z)$ (in red), $\mathrm{Im}\, m(z)$ (in blue) and the error bounds $\mathrm{Im}\, m(z) \pm \Psi(z)$ (green dashed lines).

(c) $\mathrm{Im}\, G_{12}(z)$ (in red) and the error bounds $\pm \Psi(z)$ (green dashed lines).

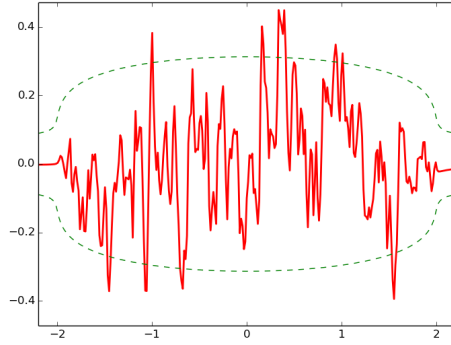**Figure 2.3.** *Local semicircle law.* All plots were obtained from the same realization of a GUE matrix with $N = 10^3$. We draw several functions of E at a fixed spectral resolution $\eta = N^{-0.6}$. As expected from Theorem 2.6, the fluctuations and the error bounds are larger for the matrix entries than for the normalized trace of the Green function.

Theorem 2.8 (Semicircle law on small scales). *For a Wigner matrix we have*

$$\mu(I) \;=\; \varrho(I) + O_{\prec}(N^{-1})$$

uniformly for all intervals $I \subset \mathbb{R}$.

We postpone the proof of Theorem 2.8 to Section 8. The second consequence of Theorem 2.6 is an *eigenvalue rigidity* estimate, which gives large deviation bounds on the locations of the eigenvalues. Eigenvalue rigidity for Wigner matrices was first established in [45], using the optimal error bound from (2.14) that was first obtained there.

For $i = 1, \ldots, N$ we define the *typical location of λ_i* as the quantile γ_i satisfying

$$N \int_{\gamma_i}^{2} \varrho(\mathrm{d}x) \;=\; i - 1/2 \,. \tag{2.18}$$

13

(a) $N = 50$, GUE matrix

(b) $N = 10^3$, GUE matrix (zoomed in)

(c) $N = 50$, i.i.d. variables with semicircle distribution

(d) $N = 10^3$, i.i.d. variables with semicircle distribution (zoomed in)
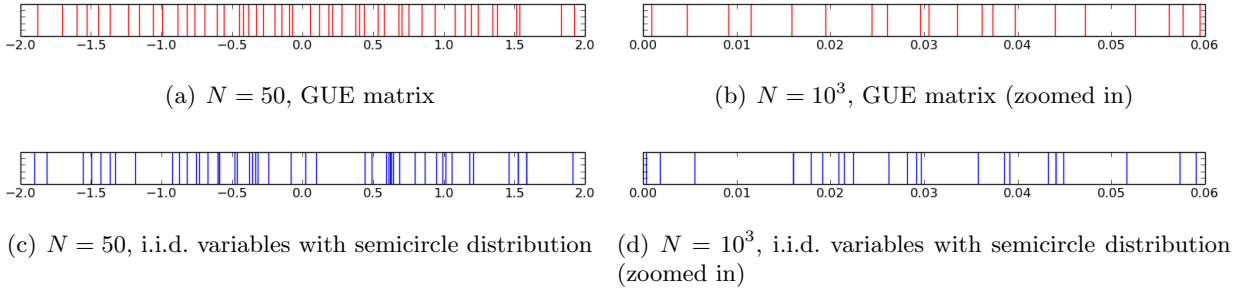
Figure 2.4. *Local eigenvalue density.* We compare the eigenvalues of an $N \times N$ GUE matrix (abscissas of the red vertical lines) with a sample of N i.i.d. variables with semicircle distribution (abscissas of the blue vertical lines). The GUE eigenvalue spacings are far more regular than those of the i.i.d. variables; the former are typically of order N^{-1} (by Theorem 2.8 or Theorem 2.9), while the latter are typically of order $N^{-1/2}$.

It is easy to see that the typical eigenvalue locations γ_i satisfy

$$2 - \gamma_i \; \asymp \; \left(\frac{i}{N}\right)^{2/3} \qquad (i \leqslant N/2), \tag{2.19}$$

and by symmetry a similar estimate holds for $i \geqslant N/2$. Note that the right-hand side of (2.19) is characteristic of the *square root decay* of the density of ϱ near the boundary of its support.

Theorem 2.9 (Eigenvalue rigidity). *For a Wigner matrix we have*

$$|\lambda_i - \gamma_i| \; \prec \; N^{-2/3}\big(i \wedge (N + 1 - i)\big)^{-1/3}$$

uniformly for $i = 1, \ldots, N$.

In particular, the extreme eigenvalues λ_1 and λ_N are with high probability located at a distance of order at most $N^{-2/3}$ from their typical locations. Moreover, the bulk eigenvalues, λ_i satisfying $|i| \asymp |N - i| \asymp N$, are with high probability located at a distance of order at most N^{-1} from their typical locations. The eigenvalue rigidity is a manifestation of a strong repulsion between the eigenvalues, which tend to form a rather rigid "jelly" where neighbouring eigenvalues avoid getting too close and are therefore pinned down by the influence of their neighbours. In contrast, if $\lambda_1, \ldots, \lambda_N$ were i.i.d. random variables distributed according to the semicircle distribution, the global semicircle law for $s - m$ would remain true, but a standard exercise in order statistics shows that in this case λ_i would typically fluctuate on the scale $N^{-1/2}$. This is much larger than the scale N^{-1} from Theorem 2.9. Correspondingly, for this i.i.d. model we would expect to find gaps of order $N^{-1/2}$, and the error term in Theorem 2.8 would have to be replaced with $O_{\prec}(N^{-1/2})$. See Figures 2.4 and 2.5 for an illustration of this rigidity.

We postpone the proof of Theorem 2.9 to Section 9. A very easy corollary of Theorems 2.6 and 2.9 is the *complete delocalization of eigenvectors*, illustrated in Figure 2.6. The complete delocalization of eigenvectors of Wigner matrices was first obtained in [39] as a corollary of the local law on optimal scales proved there.

We use the notation $\mathbf{u}_i = (u_i(k))_{k=1}^N$ for the components of the eigenvectors.
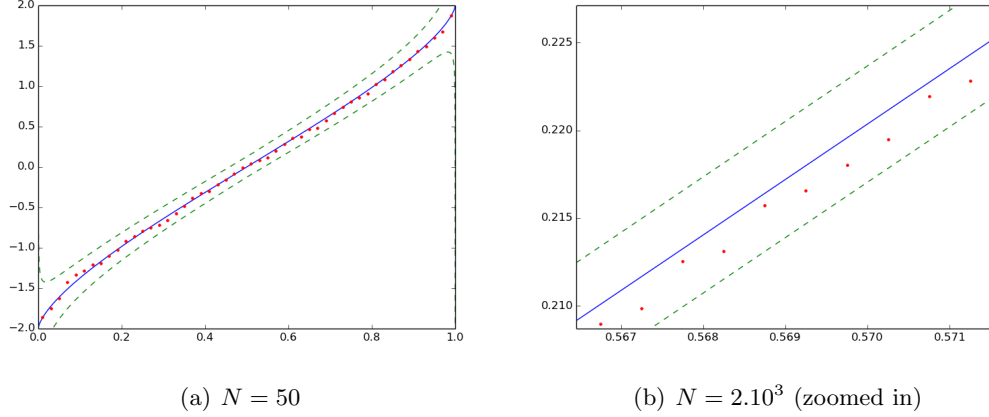
(a) $N = 50$ (b) $N = 2.10^3$ (zoomed in)

Figure 2.5. *Eigenvalue rigidity.* The quantile function $x \in [0,1] \mapsto Q(x)$ of the semicircle distribution (in blue), the error bounds $x \in [0,1] \mapsto Q(x) \pm 5N^{-1}(x \wedge (1-x))^{-1/3}$ (green dashed lines), and a scatter plot of the function $(i-1/2)/N \mapsto \lambda_i$ where λ_i is the i-th eigenvalue of an $N \times N$ GUE matrix (red points).

Theorem 2.10 (Complete eigenvector delocalization). *For a Wigner matrix we have*

$$|u_i(k)|^2 \prec \frac{1}{N}.$$

uniformly for $i, k = 1, \ldots, N$.

Proof. By Theorem 2.9 we know that $|\lambda_i| \leqslant 3$ with high probability. We choose a random spectral parameter by choosing $z := \lambda_i + i\eta$ with $\eta := N^{-1+\tau}$. Hence for small enough $\tau > 0$ we have $z \in \mathbf{S}$ with high probability. Moreover, by Remark 2.7 we may invoke Theorem 2.6 with the random error parameter z. Since $|m(z)| \leqslant 1$ (as can be easily checked using (2.11)) and $\Psi(z) \leqslant C$, we get from (2.15) that

$$1 \succ \operatorname{Im} G_{kk}(z) = \sum_{j=1}^{N} \frac{\eta}{(\lambda_j - \lambda_i)^2 + \eta^2} |u_j(k)|^2 \geqslant \frac{1}{\eta} |u_i(k)|^2 = N^{1-\tau} |u_i(k)|^2.$$

Since $\tau > 0$ was arbitrary, the conclusion follows. $\qquad\square$

To understand the meaning of Theorem 2.10, observe that for each fixed i the map $k \mapsto p_i(k) := |u_i(k)|^2$ is a (random) probability measure on $\{1, \ldots, N\}$. A fundamental question about eigenvectors is whether p_i is mainly concentrated at a few points or whether it is evenly spread out over $\{1, \ldots, N\}$. Complete eigenvector delocalization states that, for each eigenvector $\mathbf{u}_i$ and with high probability, the latter possibility always holds. Note that if H is a GOE/GUE matrix, the law of H is invariant under orthogonal/unitary conjugations, and its eigenvector matrix therefore uniformly distributed on the orthogonal/unitary group. In particular, each eigenvector $\mathbf{u}_i$ is uniformly distributed on the unit sphere, and we easily find that $|u_i(k)|^2 \leqslant (\log N)^a N^{-1}$ for all i, k with high probability, provided that $a > 1/2$. Up to the technical details in the definition of $\prec$ (see Remark 2.11 below), this simple argument therefore recovers, for the special cases of GOE and GUE, the result of Theorem 2.10 (which holds of course for all Wigner matrices).
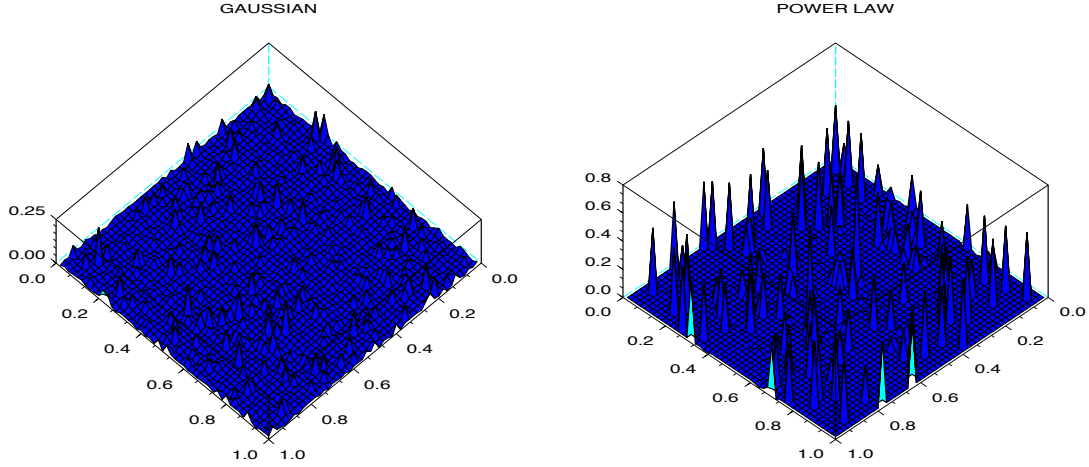
15

Figure 2.6. *Complete eigenvector delocalization.* We illustrate two plots of the function $(i/N, k/N) \mapsto ||u_i(k)|^2 - N^{-1}|$, where $u_i(k)$ denotes the k-th component of the i-th eigenvector of a Wigner matrix, and $N = 50$. The Wigner matrix of the left-hand plot has Gaussian entries (GOE), while that of the right-hand plot has heavy-tailed entries of the form $SU^{-(2+\varepsilon)}$ for S, U independent such that $S = \pm 1$ is symmetric, U is uniformly distributed on $[0, 1]$ and $\varepsilon = 10^{-3}$.

Remark 2.11. The notion of stochastic domination from Definition 2.5 greatly simplifies many statements and arguments. On the other hand, the specific notion of high-probability bounds that it yields is not as sharp as possible: the factors of N^ε can be improved to powers of $\log N$ and the polynomial error probabilities N^{-D} can be improved to exponential error probabilities. Such extensions may in fact be obtained by a routine but tedious extension of the arguments presented in these notes. We do not pursue this further.

The local law has many further applications, for instance to the universality of the local eigenvalue statistics and the distribution of eigenvectors; see Section 11 below for more details.

2.5. Sketch of the proof of Theorem 2.6. We conclude this section with a sketch of the proof of Theorem 2.6, which is the main content of these notes. A more detailed presentation of the main steps of the proof is given in Section 4.

(a) The starting point of the proof is *Schur's complement formula*

$$\frac{1}{G_{ii}} = H_{ii} - z - \sum_{k,l \neq i} H_{ik} G_{kl}^{(i)} H_{li}, \tag{2.20}$$

where $G^{(i)}$ denotes the Green function of the matrix obtained from H by removing the i-th row and column. The coefficients $G_{kl}^{(i)}$ are independent of the family $(H_{ik})_{k=1}^N$, so that we may apply large deviation estimates to the sum, finding that it is in fact close with high probability to its expectation with respect to the randomness in the i-th column, which is in turn close to $\frac{1}{N} \sum_k G_{kk} = s$.

(b) We split the right-hand side of (2.20) into a leading term and a random error term, which yields after multiplication of both sides by G_{ii}

$$1 = -zG_{ii} - sG_{ii} + (\text{error term}). \qquad (2.21)$$

Taking $\eta \geqslant 1$, averaging over i, and estimating the random error term, we get

$$1 + s^2 + sz = O_{\prec}(N^{-1/2}) \qquad (2.22)$$

for $\eta \geqslant 1$.

(c) As $m(z)$ is a solution of (2.22) with the right-hand side set to zero, a stability analysis of (2.22) yields an estimate of $s(z) - m(z)$ for $\eta \geqslant 1$. This is the global semicircle law. Going back to (2.20) and using further analogous formulas for the off-diagonal entries, we obtain weak control of the entries $G_{ij} - m\delta_{ij}$ on the global scale $\eta \geqslant 1$.

(d) The heart of the proof of the local law is a *boostrapping* from large scales to smaller scales. Roughly, we prove the local semicircle law on a certain scale η, and then use the Lipschitz continuity of $G_{ij} - m\delta_{ij}$ to deduce an a priori bound on the slightly smaller scale $\eta' = \eta - N^{-3}$. Using this a priori bound as input, we use the formula (2.20) and similar formulas for the off-diagonal entries to obtain an improved bound $G_{ij} - m\delta_{ij}$ on the same scale η'. Thanks to this self-improving estimation, we may iterate this procedure an order N^3 times to reach the optimal scale $\eta \approx N^{-1}$. At each step of this iteration, we lose some probability. This loss has to be carefully tracked to ensure that the combined losses in probability from N^3 iterations remain much smaller than 1. This procedure yields control of $G_{ij} - m\delta_{ij}$ for all $z \in \mathbf{S}$, albeit with an error bound that is larger than Ψ from (2.16).

(e) In order to obtain (2.15) with the optimal error bound Ψ, we use a *fluctuation averaging argument*. Roughly, it says that the random variables $1/G_{ii}$, when centred with respect to the randomness of the i-th column of H, are not too strongly dependent, and hence their average is typically much smaller than any one of them.

3. Proof of Theorem 2.6 (a): preliminaries

We now move on to the proof of Theorem 2.6. For the proof we often omit the spectral parameter $z \in \mathbf{S}$ from our notation, bearing in mind that we are typically dealing with random sequences (indexed by $N \in \mathbb{N}$) of functions (of $z \in \mathbf{S}$).

In this preliminary section we collect the main tools of the proof.

Definition 3.1 (Minors). *We consider general matrices whose indices lie in subsets of $\{1, \ldots, N\}$. For $T \subset \{1, \ldots, N\}$ we define $H^{(T)}$ as the $(N - |T|) \times (N - |T|)$ matrix*

$$H^{(T)} = (H_{ij})_{i,j \in \{1,\ldots,N\} \setminus T}.$$

Moreover, we define the Green function of $H^{(T)}$ through

$$G^{(T)}(z) := \left(H^{(T)} - z\right)^{-1}.$$

We also set

$$\sum_i^{(T)} := \sum_{i:i\notin T}.$$

When $T = \{a\}$, we abbreviate $(\{a\})$ by (a) in the above definitions; similarly, we write (ab) instead of $(\{a,b\})$.

Note that in the minor $H^{(T)}$ it is important to keep the original values of the matrix indices, and not to identify $\{1,\ldots,N\} \setminus T$ with $\{1,\ldots,N-|T|\}$.

Definition 3.2 (Conditional expectation). *Let $X \equiv X(H)$ be a random variable. For $i \in \{1,\ldots,N\}$ we define the operations P_i and Q_i through*

$$P_i X := \mathbb{E}(X|H^{(i)}), \qquad Q_i X := X - P_i X.$$

Note that P_i and Q_i form a family of commuting orthogonal projections on $L^2(\mathbb{P})$.

The following lemma collects basic bounds on m. In order to state it, we define the *distance to the spectral edge*

$$\kappa \equiv \kappa(E) := \big||E| - 2\big|. \tag{3.1}$$

Lemma 3.3 (Basic properties of m). *Fix $\tau > 0$. Then for $z \in \mathbf{S}$ we have*

$$|m(z)| \asymp 1 \tag{3.2}$$

and

$$\operatorname{Im} m(z) \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } |E| \leqslant 2 \\ \frac{\eta}{\sqrt{\kappa+\eta}} & \text{if } |E| \geqslant 2. \end{cases} \tag{3.3}$$

Proof. The proof is an elementary exercise using (2.11) and (2.12). $\qquad\square$

The following lemma collects properties of stochastic domination $\prec$. Roughly, it states that $\prec$ satisfies the usual arithmetic properties of order relations. We shall use it tacitly throughout the following.

Lemma 3.4 (Basic properties of $\prec$).   (i) *Suppose that $X(u,v) \prec Y(u,v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leqslant N^C$ for some constant C then*

$$\sum_{v\in V} X(u,v) \prec \sum_{v\in V} Y(u,v)$$

uniformly in u.

(ii) *Suppose that $X_1(u) \prec Y_1(u)$ uniformly in u and $X_2(u) \prec Y_2(u)$ uniformly in u. Then $X_1(u)X_2(u) \prec Y_1(u)Y_2(u)$ uniformly in u.*

Proof. This is a simple union bound. $\qquad\square$

The following resolvent identities form the backbone of all of our calculations. The idea behind them is that a Green function entry G_{ij} depends strongly on the i-th and j-th columns of H, but weakly on all other columns. The first identity determines how to make a Green function entry G_{ij} independent of the randomness in the k-th row, where $k \neq i, j$. The second identity expresses the dependence of a Green function entry G_{ij} on the matrix elements in the i-th or in the j-th column of H.

Lemma 3.5 (Resolvent identities). *For any Hermitian matrix H and $T \subset \{1, \ldots, N\}$ the following identities hold. If $i, j, k \notin T$ and $i, j \neq k$ then*

$$G_{ij}^{(T)} = G_{ij}^{(Tk)} + \frac{G_{ik}^{(T)} G_{kj}^{(T)}}{G_{kk}^{(T)}}. \tag{3.4}$$

If $i, j \notin T$ satisfy $i \neq j$ then

$$G_{ij}^{(T)} = -G_{ii}^{(T)} \sum_k^{(Ti)} H_{ik} G_{kj}^{(Ti)} = -G_{jj}^{(T)} \sum_k^{(Tj)} G_{ik}^{(Tj)} H_{kj}. \tag{3.5}$$

Proof. See Appendix A. $\qquad\square$

We also need the *Ward identity*

$$\sum_j |G_{ij}|^2 = \frac{\operatorname{Im} G_{ii}}{\eta}. \tag{3.6}$$

It may be proved by applying the resolvent identity to $G - G^*$, or, alternatively, by spectral decomposition:

$$\sum_j |G_{ij}|^2 = (GG^*)_{ii} = \sum_j \frac{|u_j(i)|^2}{|\lambda_j - z|^2} = \sum_j \frac{1}{\eta} \operatorname{Im} \frac{|u_j(i)|^2}{\lambda_j - z} = \frac{1}{\eta} \operatorname{Im} G_{ii}.$$

Assuming $\eta \gg \frac{1}{N}$, (3.6) shows that the squared ℓ^2-norm $\frac{1}{N} \sum_j |G_{ij}|^2$ is smaller by the factor $\frac{1}{N\eta} \ll 1$ than the diagonal element $|G_{ii}|$. This fact plays a crucial role in our proof: it is precisely this gain of $\eta^{-1} \ll N$ as compared to N that allows us to perform the bootstrapping underlying the proof. This identity is also the source of the ubiquitous factors of η in denominators. It was first used systematically in the proof the local semicircle law for random matrices in [43].

Finally, we record the following large deviation bounds.

Lemma 3.6 (Large deviation bounds). *Let $\left(X_i^{(N)}\right)$, $\left(Y_i^{(N)}\right)$, $\left(a_{ij}^{(N)}\right)$, and $\left(b_i^{(N)}\right)$ be independent families of random variables, where $N \in \mathbb{N}$ and $i, j = 1, \ldots, N$. Suppose that all entries $X_i^{(N)}$ and $Y_i^{(N)}$ are independent and satisfy the conditions*

$$\mathbb{E} X = 0, \qquad \|X\|_p \leqslant \mu_p \tag{3.7}$$

for all p with some constants μ_p.

(i) Suppose that $\left(\sum_i |b_i|^2\right)^{1/2} \prec \Psi$. Then $\sum_i b_i X_i \prec \Psi$.

(ii) Suppose that $\left(\sum_{i\neq j}|a_{ij}|^2\right)^{1/2} \prec \Psi$. Then $\sum_{i\neq j} a_{ij}X_iX_j \prec \Psi$.

(iii) Suppose that $\left(\sum_{i,j}|a_{ij}|^2\right)^{1/2} \prec \Psi$. Then $\sum_{i,j} a_{ij}X_iY_j \prec \Psi$.

If all of the above random variables depend on an index u and the hypotheses of (i) – (iii) are uniform in u, then so are the conclusions.

Proof. See Appendix D $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

4. Outline of the proof of Theorem 2.6

(a) The diagonal entries of G can be written using Schur's complement formula as

$$\frac{1}{G_{ii}} \;=\; H_{ii} - z - \sum_{k,l}^{(i)} H_{ik}G_{kl}^{(i)}H_{li}. \tag{4.1}$$

Since the coefficients $G_{kl}^{(i)}$ in (4.1) are independent of the entries $(H_{ik})_{k=1}^N$, we can condition on $H^{(i)}$ and apply Lemma 3.6 to find that the sum on the right-hand side of (4.1) is with high probability close to its P_i-expectation, $\frac{1}{N}\sum_k^{(i)} G_{kk}^{(i)}$. Using Lemma 3.5 we can get rid of the upper index (i) up to a small error term, and find that the sum on the right-hand side of (4.1) is with high probability close to s.

In order to quantify these errors, we introduce the random z-dependent error parameter

$$\Lambda \;:=\; \max_{i,j}|G_{ij} - m\delta_{ij}|.$$

Then using large deviation estimates as outlined above as well as the Ward identity (3.6), we find that if either $\eta \geqslant 1$ or $\Lambda \leqslant N^{-\tau/10}$, then (4.1) reads

$$\frac{1}{G_{ii}} = -z - s + O_\prec(\Psi_\Theta), \qquad \Psi_\Theta \;:=\; \sqrt{\frac{\operatorname{Im} m + |s - m|}{N\eta}}. \tag{4.2}$$

Combined with analogous arguments applied to the off-diagonal entries starting from (3.5), we obtain the estimate

$$\max_{i,j}|G_{ij} - \delta_{ij}s| \;\prec\; \Psi_\Theta \tag{4.3}$$

on the individual entries of G. Note that the diagonal entries are compared to the empirical quantity s instead of the deterministic quantity m.

(b) In the next step, we compare s to m. Rewriting (4.2) as

$$1 = -zG_{ii} - sG_{ii} + O_\prec(\Psi_\Theta) \tag{4.4}$$

and averaging over i, we obtain

$$1 + zs + s^2 \;=\; O_\prec(\Psi_\Theta). \tag{4.5}$$

Recall from (2.12) that m solves the same equation as (4.5) without the error term on the right-hand side. Thus, concluding estimating $s - m$ from (4.5) in terms of the error $O_\prec(\Psi_\Theta)$ involves a stability analysis of the quadratic equation (4.5).

20

(c) If $\eta \geqslant 1$ we have $\Psi_\Theta \leqslant CN^{-1/2}$, and it is then not hard to deduce that $|s - m| \prec N^{-1/2}$, so that (4.3) implies $\Lambda \prec N^{-1/2}$. Here the stability analysis of (4.5) is simple because the two solutions of (4.5) are well separated.

(d) In order to estimate $s - m$ for $\eta \leqslant 1$, we use a *bootstrapping* from large to small scales, a self-consistent multiscale induction where estimates on a given scale are used as a priori information for obtaining estimates on the subsequent, smaller, scale. This is the core of the proof. The bootstrapping is performed at fixed energy E along the sequence $z_k := E + \mathrm{i}(1 - kN^{-3})$, where $k = 0, 1, \ldots, N^3 - N^{2+\tau}$. At each step of the bootstrapping we establish the estimate $\Lambda(z_k) \prec (N\eta)^{-1/4}$.

The bootstrapping is started at z_0 for which the simple analysis of (c) applies.

To explain the bootstrapping, suppose that $\Lambda(z_k) \prec (N\eta)^{-1/4}$ holds at z_k. Then we use the trivial Lipschitz continuity of Λ with Lipschitz constant N^2 and the estimate $|z_{k+1} - z_k| = N^{-3}$ to deduce that with high probability $\Lambda \leqslant N^{-\tau/10}$ at z_{k+1}. This is the a priori assumption needed to obtain the estimates (4.2) and (4.3) at z_{k+1}. Hence, we also obtain (4.5) at z_{k+1}. In order to obtain a bound on $|s - m|$ from (4.5) and hence complete the induction step, we need to perform a stability analysis of (4.5). For small η and E near the spectral edge, this stability analysis requires some care, because the two solutions of (4.5) can be close. This completes the induction step.

(e) From (c)–(d) we obtain $\Lambda \prec (N\eta)^{-1/4}$ for all z_k at some fixed energy E. By a union bound, we obtain a simultaneous high-probability estimate for all z in the N^{-3}-net $\mathbf{S} \cap N^{-3}\mathbb{Z}^2 \subset \mathbf{S}$. By Lipschitz continuity of Λ with Lipschitz constant N^2, we extend this estimate to a simultaneous high-probability bound for all $z \in \mathbf{S}$.

(f) In (e) we obtained the *weak local law* $\Lambda \prec (N\eta)^{-1/4}$ for all $z \in \mathbf{S}$. Although the error bound $(N\eta)^{-1/4}$ is far from optimal, this is a highly nontrivial and very useful estimate, because it prove that Λ is small, and in particular bounded, for all $z \in \mathbf{S}$. This estimate serves as the key input in the proof of the local law, where the optimal error bound is obtained.

The key tool behind the proof of the optimal error bound is a *fluctuation averaging argument*, which states that the error term in (4.4) becomes much smaller after averaging over i. Thus, the optimal error in (9.2) is in fact much smaller than in (4.4). The main work behind this improvement is to estimate averages of the form $\sum_{i=1}^N X_i$ where $X_i := Q_i \frac{1}{G_{ii}}$ has (by definition) expectation zero. Clearly, if the variables X_i were all equal then the averaging would not change anything, and if they were all independent we would gain a factor $N^{-1/2}$ from the averaging. In fact, with the choice $X_i = Q_i \frac{1}{G_{ii}}$ neither is true, as the variables X_i are obviously not equal but they are nevertheless strongly correlated. As it turns out, the averaging yields a gain of order $(N\eta)^{-1/2}$, so that the variables X_i are almost uncorrelated for $\eta = 1$ and almost equal for $\eta \approx N^{-1}$.

5. Proof of Theorem 2.6 (b): weak local law

The conclusion of Theorem 2.6 may be regarded as consisting of two separate achievements on the quantities $G_{ij} - m\delta_{ij}$ and $s - m$: first, control for small values of η; second, optimal error bounds. These two achievements in fact entail separate difficulties in the proof, and we correspondingly

separate the proof into two parts. The weak local law presented in this section pertains to the first achievement.

The goal of the weak local law is to establish the smallness of $G_{ij} - m\delta_{ij}$ for all $z \in \mathbf{S}$, without however trying to get optimal error bounds. Once the weak law has been established, we will use it to improve the error estimates.

In this section we state and prove the weak law. The central quantity is the random z-dependent error parameter

$$\Lambda := \max_{i,j} |G_{ij} - m\delta_{ij}|.$$

Proposition 5.1 (Weak local law). *We have*

$$\Lambda \prec (N\eta)^{-1/4}$$

for $z \in \mathbf{S}$.

The rest of this section is devoted to the proof of Proposition 5.1. In addition to Λ, we shall need the additional error parameters

$$\Lambda_* := \max_{i \neq j} |G_{ij}|, \qquad \Theta := |s - m|, \qquad \Psi_\Theta := \sqrt{\frac{\operatorname{Im} m + \Theta}{N\eta}}.$$

Our starting point is Schur's complement formula (see Lemma A.1), which we write as

$$\frac{1}{G_{ii}} = H_{ii} - z - \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}. \tag{5.1}$$

We have the following simple identity.

Lemma 5.2. *We have*

$$\frac{1}{G_{ii}} = -z - s + Y_i,$$

where we defined

$$Y_i := H_{ii} + A_i - Z_i, \qquad A_i := \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}}, \qquad Z_i := Q_i \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}. \tag{5.2}$$

Proof. Using (3.4) we get

$$P_i \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} = \frac{1}{N} \sum_k^{(i)} G_{kk}^{(i)} = \frac{1}{N} \sum_k G_{kk} - \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}} = s - \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}}.$$

Hence,

$$\sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} = s - \frac{1}{N} \sum_k \frac{G_{ki} G_{ik}}{G_{ii}} + Q_i \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} = s - A_i + Z_i.$$

The claim now follows using (5.1). $\qquad \square$

The strategy behind the proof of Proposition 5.1 is a self-consistent multiscale approach, which involves a bootstrapping from the large scale $\eta \geqslant 1$ down to the small scale $\eta = N^{-1+\tau}$. The bootstrapping hypothesis will be the event $\phi = 1$, where we defined the z-dependent indicator function

$$\phi := \mathbf{1}\big(\Lambda \leqslant N^{-\tau/10}\big).\tag{5.3}$$

The following result gives trivial bounds on entries of G on the event $\phi = 1$.

Lemma 5.3. *Let $p \in \mathbb{N}$ be fixed. Then we have*

$$\phi|G_{ij}^{(T)}| + \phi\left|\frac{1}{G_{ii}^{(T)}}\right| \prec 1$$

uniformly for $i, j \notin T \subset \{1, \ldots, N\}$ satisfying $|T| \leqslant p$.

Proof. The proof follows using (3.2) and a repeated application of (3.4), whose details we leave to the reader. $\qquad\square$

The following lemma contains the key a priori estimates used to perform the bootstrapping. Note that the assumptions for the estimates are either $\eta \geqslant 1$ (start of bootstrapping) or $\phi = 1$ (bootstrapping assumption used for iteration).

Lemma 5.4 (Main estimates). *For $z \in \mathbf{S}$ we have*

$$\big(\phi + \mathbf{1}(\eta \geqslant 1)\big)\big(\Lambda_* + |A_i| + |Z_i|\big) \prec \Psi_\Theta \tag{5.4}$$

and

$$\big(\phi + \mathbf{1}(\eta \geqslant 1)\big)|G_{ii} - s| \prec \Psi_\Theta.\tag{5.5}$$

Proof. We begin with Z_i under the bootstrapping assumption $\phi = 1$. We split

$$\phi|Z_i| \leqslant \phi\left|\sum_k^{(i)}\left(|H_{ik}|^2 - \frac{1}{N}\right)G_{kk}^{(i)}\right| + \phi\left|\sum_{k\neq l}^{(i)} H_{ik}G_{kl}^{(i)}H_{li}\right|.\tag{5.6}$$

Using Lemmas 3.6 and 5.3 we estimate the first term of (5.6) as

$$\phi\left|\sum_k^{(i)}\left(|H_{ik}|^2 - \frac{1}{N}\right)G_{kk}^{(i)}\right| \prec \phi\left(\frac{1}{N^2}\sum_k^{(i)}|G_{kk}^{(i)}|^2\right)^{1/2} \prec \phi N^{-1/2} \leqslant C\phi\Psi_\Theta,$$

where in the last step we used (3.3) to estimate $\Psi_\Theta \geqslant cN^{-1/2}$.

Next, using Lemma 3.6 we estimate the second term of (5.6) as

$$\phi\left|\sum_{k\neq l}^{(i)} H_{ik}G_{kl}^{(i)}H_{li}\right| \prec \phi\left(\sum_{k,l}^{(i)}\frac{1}{N^2}|G_{kl}^{(i)}|^2\right)^{1/2} = \phi\left(\sum_k^{(i)}\frac{1}{N^2\eta}\operatorname{Im}G_{kk}^{(i)}\right)^{1/2}.$$

Here the second step follows from the Ward identity (3.6) applied to the matrix $H^{(i)}$. Using (3.4) and Lemma 5.3 we therefore conclude

$$\phi\left|\sum_{k\neq l}^{(i)} H_{ik}G_{kl}^{(i)}H_{li}\right| \prec \phi\left(\sum_k\frac{\operatorname{Im}G_{kk} + \Lambda_*^2}{N^2\eta}\right)^{1/2} \leqslant \phi\Psi_\Theta + \phi\Lambda_*.$$

23

We conclude from (5.6) that
$$\phi|Z_i| \;\prec\; \phi\Psi_\Theta + \phi\Lambda_* \,. \tag{5.7}$$

Next, from Lemma 5.3 we get the trivial bound
$$\phi|A_i| \;\prec\; \phi\Lambda_*^2 \;\leqslant\; \phi\Lambda_* \,. \tag{5.8}$$

In order to estimate Λ_*, we take $i \neq j$ and use (3.5) twice to get
$$G_{ij} \;=\; -G_{ii}\sum_k^{(i)} H_{ik}G_{kj}^{(i)} \;=\; -G_{ii}G_{jj}^{(i)}\left(H_{ij} - \sum_{k,l}^{(ij)} H_{ik}G_{kl}^{(ij)}H_{lj} \right). \tag{5.9}$$

We estimate the second term using Lemma 3.6 and the identity (3.6) (applied to $H^{(ij)}$) according to

$$\phi\left|\sum_{k,l}^{(ij)} H_{ik}G_{kl}^{(ij)}H_{lj}\right| \;\prec\; \phi\left(\frac{1}{N^2}\sum_{k,l}^{(ij)}|G_{kl}^{(ij)}|^2\right)^{1/2} \;=\; \phi\left(\frac{1}{N^2\eta}\sum_k^{(ij)}\operatorname{Im}G_{kk}^{(ij)}\right)^{1/2}$$
$$\leqslant\; \phi\left(\sum_k \frac{\operatorname{Im}G_{kk} + C\Lambda_*^2}{N^2\eta}\right)^{1/2} \;\leqslant\; \phi\Psi_\Theta + C\phi(N\eta)^{-1/2}\Lambda_* \,,$$

where the third step follows using a repeated application of (3.4). Using $|H_{ij}| \prec N^{-1/2}$, Lemma 5.3, and the estimate $(N\eta)^{-1/2} \leqslant CN^{-\tau/4}$ valid for $z \in \mathbf{S}$, we therefore conclude that
$$\phi\Lambda_* \;\prec\; \phi\Psi_\Theta + \phi N^{-\tau/4}\Lambda_* \,.$$

Using the definition of $\prec$, we deduce that
$$\phi\Lambda_* \;\prec\; \phi\Psi_\Theta \,. \tag{5.10}$$

Plugging (5.10) into (5.7) and (5.8) concludes the proof of the term ϕ in (5.4).

Moreover, the proof of the term $\mathbf{1}(\eta \geqslant 1)$ is almost the same, using the trivial bound $|G_{ij}^{(T)}| \leqslant 1$ instead of Lemma 5.3.

What remains is the estimate (5.5). From Lemma 5.2 and (5.4) we get
$$(\phi + \mathbf{1}(\eta \geqslant 1))\left|\frac{1}{G_{ii}} - \frac{1}{G_{jj}}\right| \;\leqslant\; (\phi + \mathbf{1}(\eta \geqslant 1))(|Y_i| + |Y_j|) \;\prec\; \phi\Psi_\Theta \,,$$

which implies, by Lemma 5.3 that
$$\big(\phi + \mathbf{1}(\eta \geqslant 1)\big)|G_{ii} - G_{jj}| \;\prec\; \Psi_\Theta \,.$$

Averaging over j yields (5.5). $\qquad\square$

We emphasize here the essential role played by the identity (3.6) in the proof. It is the source of the factors $(N\eta)^{-1}$ in our error bounds. Using it, we can estimate the sum over N Green function entries by η^{-1} times a Green function entry. Since $\eta^{-1} \ll N$, this is a nontrivial gain. This gain is responsible not only for the optimal error bounds for Λ but also for the self-improving mechanism in the estimation of Λ that allows us to bootstrap in η.

The following elementary lemma is used to estimate $s - m$. It may be regarded as a quantitative stability result for the equation (2.12).

24

Lemma 5.5. *Let $z \in \mathbf{S}$. The equation*

$$u^2 + zu + 1 = 0$$

has two solutions, $m \in \mathbb{C}_+$ and $\tilde{m} \in \mathbb{C}_-$. They satisfy

$$|m - \tilde{m}| \asymp \sqrt{\kappa + \eta}.$$

Moreover, if u solves

$$u^2 + zu + 1 + r = 0 \qquad (|r| \leqslant 1) \tag{5.11}$$

then

$$\min\{|u - m|, |u - \tilde{m}|\} \leqslant \frac{C|r|}{\sqrt{\kappa + \eta + |r|}} \tag{5.12}$$

for some constant C.

Proof. See Appendix B $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

The following lemma starts the bootstrapping at $\eta \geqslant 1$.

Lemma 5.6. *For $z \in \mathbf{S}$, we have*

$$\mathbf{1}(\eta \geqslant 1)\Lambda \prec N^{-1/2}$$

Proof. Let $\eta \geqslant 1$. From (5.4) and Lemma 3.3 we get $\Lambda_* \prec N^{-1/2}$. What remains therefore is the estimate of $G_{ii} - m$. By Lemmas 5.2 and 5.4, we have

$$1 + sG_{ii} + zG_{ii} = O_{\prec}(N^{-1/2}).$$

Averaging over i yields

$$1 + s^2 + sz = O_{\prec}(N^{-1/2}). \tag{5.13}$$

Lemma 5.5 therefore gives

$$\min\{|s - m|, |s - \tilde{m}|\} \prec N^{-1/2}.$$

We now need to make sure that s is close to m and not $\tilde{m}$. Since $\eta \geqslant 1$, it is easy to see that $\operatorname{Im} \tilde{m} \leqslant -c$ for some positive constant c. Since $\operatorname{Im} s > 0$, we therefore deduce that

$$\Theta = |s - m| \prec N^{-1/2}.$$

Using (5.5), we deduce that

$$|G_{ii} - m| \leqslant |G_{ii} - s| + \Theta \prec N^{-1/2},$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We may now come to the heart of the proof of Proposition 5.1: the self-consistent multiscale bootstrapping from large to small η. If all error estimates were deterministic, we could do a standard continuity argument by choosing a continuous path $\eta(t) = 1 - t$. However, owing to the stochastic nature of our estimates, the bootstrapping has to be done in a finite number of steps. At each step we lose some probability, which imposes an upper bound on the number of allowed steps. On the other hand, the steps have to be small enough to be able to carry over information from one step

to the next using continuity. At each step, the deterioration of the error probabilities has to be tracked carefully. Roughly, at each step we lose an additive factor N^{-D} in the error probabilities, which means that we can perform an order N^C steps for any fixed constant C. Using the Lipschitz continuity of all error parameters, it turns out to be enough to perform steps of size N^{-3}. Hence, for the *stochastic continuity argument* that underlies the self-consistent multiscale approach, we replace the purely topological notions of a continuous path and continuous error parameters with the quantitative notions of a lattice and Lipschitz continuous error parameters.

Thus, we define the discrete subset $\widehat{\mathbf{S}} := \mathbf{S} \cap (N^{-3}\mathbb{Z}^2)$. Since s, m, G_{ij}, and Ψ are all N^2-Lipschitz on $\mathbf{S}$, it is enough to prove Proposition 5.1 for $z \in \widehat{\mathbf{S}}$. This is the content of the following lemma.

Lemma 5.7. *For $z \in \widehat{\mathbf{S}}$ we have $\Lambda \prec (N\eta)^{-1/4}$.*

Proof. The case $\eta \geqslant 1$ follows from Lemma 5.6. For the other $z \in \widehat{\mathbf{S}}$ we construct a lattice of points $z_k \in \widehat{\mathbf{S}}$ as follows. Let $z_0 = E + \mathrm{i} \in \widehat{\mathbf{S}}$ define $z_k := E + \mathrm{i}\eta_k$ with $\eta_k := 1 - kN^{-3}$.

Fix $\varepsilon \in (0, \tau/16)$ and $D > 0$. Set $\delta_k := (N\eta_k)^{-1/2}$ and define the events

$$
\Omega_k := \left\{ \Theta(z_k) \leqslant \frac{N^\varepsilon \delta_k}{\sqrt{\kappa + \eta_k + \delta_k}} \right\}, \qquad \Xi_k := \left\{ \Lambda(z_k) \leqslant N^{2\varepsilon}\sqrt{\delta_k} \right\}.
$$

By Lemma 5.6 and $\Theta \leqslant \Lambda$,

$$
\mathbb{P}((\Omega_0 \cap \Xi_0)^c) \leqslant N^{-D}.
$$

In the bootstrapping step we prove that, conditioned on the event $\Omega_{k-1} \cap \Xi_{k-1}$, the event $\Omega_k \cap \Xi_k$ has high probability. The estimate on Θ is obtained from Lemma 5.5 by noting that s solves an equation of the form (5.11). Then we use the estimate (5.12). The key difficulty is in making sure that s is close to m and not $\tilde{m}$. Depending on the relative sizes of $|m - \tilde{m}|$ and $|r|$, we need to consider two cases. In the first case $|m - \tilde{m}|$ is large enough for us to deduce by continuity from the previous step that u cannot be close to $\tilde{m}$. In the second case $|m - \tilde{m}|$ is so small that it does not matter whether u is close to m or $\tilde{m}$. The crossover from the first to the second case happens at the index

$$
K := \min\{k : \delta_k \geqslant N^{-2\varepsilon}(\kappa + \eta_k)\}.
$$

Note that, as k increases, η_k decreases and δ_k increases. Hence, $\delta_k < N^{-2\varepsilon}(\kappa + \eta_k)$ for all $k < K$ and $\delta_k \geqslant N^{-2\varepsilon}(\kappa + \eta_k)$ for all $k \geqslant K$.

Case 1: $1 \leqslant k < K$. Since Λ is N^2-Lipschitz on $\mathbf{S}$, we have $\phi(z_k) = 1$ on Ξ_{k-1}. Using Lemma 5.4, we therefore deduce that

$$
\mathbf{1}(\Xi_{k-1})\big(|Y_i(z_k)| + \Lambda_*(z_k)\big) \prec \delta_k,
$$

where we also used that $\operatorname{Im} m + \phi\Theta \leqslant C$. Plugging this into Lemma 5.2, averaging over i, and invoking Lemma 5.5 yields

$$
\mathbb{P}\left(\Xi_{k-1} \cap \left(\min\big\{|s(z_k) - m(z_k)|, |s(z_k) - \tilde{m}(z_k)|\big\} > \frac{N^\varepsilon \delta_k}{\sqrt{\kappa + \eta_k + \delta_k}} \right) \right) \leqslant N^{-D}. \tag{5.14}
$$

On the other hand, by Lipschitz continuity, on Ω_{k-1} we have

$$
|s(z_k) - m(z_k)| \leqslant \frac{2N^\varepsilon \delta_k}{\sqrt{\kappa + \eta_k + \delta_k}}.
$$

26

Moreover, by Lemma 5.5, we have

$$|m(z_k) - \tilde{m}(z_k)| \asymp \sqrt{\kappa + \eta_k} \geqslant \frac{N^{2\varepsilon}\delta_k}{\sqrt{\kappa + \eta_k + \delta_k}},$$

where in the last step we used that $\delta_k \leqslant N^{-2\varepsilon}(\kappa + \eta_k)$. Therefore, on Ω_{k-1} we have

$$\min\{|s(z_k) - m(z_k)|, |s(z_k) - \tilde{m}(z_k)|\} = |s(z_k) - m(z_k)|.$$

We have therefore proved that

$$\mathbb{P}(\Omega_{k-1} \cap \Xi_{k-1} \cap \Omega_k^c) \leqslant N^{-D}. \tag{5.15}$$

Moreover, using Lemma 5.4 we find

$$\mathbf{1}(\Xi_{k-1} \cap \Omega_k)\Lambda(z_k) \prec \mathbf{1}(\Xi_{k-1} \cap \Omega_k)(\Psi_\Theta(z_k) + \Theta(z_k)) \leqslant CN^\varepsilon\sqrt{\delta_k},$$

where in the first step we used that $\phi(z_k) = 1$ on Ξ_{k-1} and the general estimate $\Lambda \leqslant \Lambda_* + \max_i |G_{ii} - s| + \Theta$, and in the second step we used the definitions of Ω_k and δ_k. We conclude that

$$\mathbb{P}(\Omega_{k-1} \cap \Xi_{k-1} \cap \Omega_k \cap \Xi_k^c) \leqslant N^{-D}. \tag{5.16}$$

Summarizing, defining $B_k := \Omega_k \cap \Xi_k$, we get from (5.15) and (5.16) that

$$\mathbb{P}(B_0^c) \leqslant N^{-D}, \qquad \mathbb{P}(B_{k-1} \cap B_k^c) \leqslant 2N^{-D} \qquad (k < K). \tag{5.17}$$

Therefore we get for $k < K$ that

$$\mathbb{P}(B_k^c) \leqslant \mathbb{P}(B_0^c) + \sum_{l=1}^{k} \mathbb{P}(B_l^c \cap B_{l-1}) \leqslant 2N^3 N^{-D}.$$

Case 2: $k \geqslant K$. In this case the argument is similar but somewhat easier, since there is no need to use the assumption Ω_{k-1}. Hence, we do not track Θ and we do not need the events Ω_k. If

$$\min\{|s(z_k) - m(z_k)|, |s(z_k) - \tilde{m}(z_k)|\} \leqslant \frac{N^\varepsilon\delta_k}{\sqrt{\kappa + \eta_k + \delta_k}}$$

then

$$\Theta(z_k) \leqslant N^\varepsilon\sqrt{\delta_k} + |m(z_k) - \tilde{m}(z_k)| \leqslant CN^\varepsilon\sqrt{\delta_k},$$

where we used Lemma 5.5. Using (5.14) and Lemma 5.4 we therefore deduce that

$$\mathbb{P}(\Xi_{k-1} \cap \Xi_k^c) \leqslant N^{-D}.$$

From (5.17) we therefore conclude that

$$\mathbb{P}(\Xi_k^c) \leqslant CN^{3-D}$$

for all k. Since $\varepsilon > 0$ and D were arbitrary, the claim follows. $\qquad\square$

6. Proof of Theorem 2.6 (c): optimal error bounds

In this section we complete the proof of Theorem 2.6 by improving the error bounds from Proposition 5.1 to optimal ones. The main observation is that the equation for s of the form (5.13) arises from an averaging over i. When deriving (5.13) (and its analogue for smaller η), we simply estimated the average $\frac{1}{N}\sum_i Y_i$ by $\max_i|Y_i|$. (Recall the definition (5.2).) In fact, the random variables Y_i have a small expectation, so that their average should be smaller than the typical size of each individual variable. If they were independent, we would gain a factor $N^{-1/2}$ by a trivial concentration result. However, they are not independent. In fact, for small η different Y_i are strongly correlated, and one typically has

$$\frac{1}{N}\operatorname{Var}(Y_i) \;\ll\; \operatorname{Var}\!\left(\frac{1}{N}\sum_j Y_j\right) \;\ll\; \operatorname{Var}(Y_i)\,.$$

The upper bound corresponds to fully correlated variables, and the lower bound to uncorrelated variables. For general η, the variables Y_i are in between these two extremes. The extreme cases are reached for $\eta \approx N^{-1}$ (almost fully correlated) and $\eta \approx 1$ (almost uncorrelated, in fact almost independent); see Remark 6.3 below for more details.

The key result that allows us to obtain optimal bounds on the average $\frac{1}{N}\sum_i Y_i$ is the following *fluctuation averaging* result. Recall the definitions of P_i and Q_i from Definition 3.2.

Proposition 6.1 (Fluctuation averaging). *Suppose that for $z \in \mathbf{S}$ we have*

$$\Lambda_* \;\prec\; \Phi, \tag{6.1}$$

where

$$N^{-1/2} \;\leqslant\; \Phi \;\leqslant\; N^{-c} \tag{6.2}$$

for some constant $c > 0$. Then for $z \in \mathbf{S}$ we have

$$\frac{1}{N}\sum_i Q_i \frac{1}{G_{ii}} \;=\; O_\prec(\Phi^2)\,.$$

The proof of Proposition 6.1 is postponed to next section. In this section we use Proposition 6.1 to complete the proof of Theorem 2.6. Throughout, we use that $1 - \phi \prec 0$ in $\mathbf{S}$, as follows from Proposition 5.1. This allows us for instance to drop the indicator functions on the left-hand sides of (5.4) and (5.5).

Lemma 6.2. *Under the assumptions of Proposition 6.1, the error terms Y_i defined in Lemma 5.2 satisfy*

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \;=\; O_\prec(\Phi^2)\,.$$

Proof. From Schur's complement formula (5.1) we get

$$Q_i \frac{1}{G_{ii}} \;=\; H_{ii} - Q_i \sum_{k,l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li} \;=\; H_{ii} - Z_i \;=\; Y_i - A_i\,,$$

so that $Y_i = A_i + Q_i 1/G_{ii}$. From the definition of A_i in Lemma 5.2 we obtain $A_i = O_\prec(\Phi^2)$, where we used Proposition 5.1 and (3.2). The claim now follows from Proposition 6.1. $\qquad\square$

28

Remark 6.3. Since $A_i = O_\prec(\Phi^2)$ even without averaging, we find that the important part of Y_i is $X_i := Q_i \frac{1}{G_{ii}}$. Suppose for simplicity that we are in the bulk, i.e. $E \in [-2+c, 2-c]$ for some constant $c > 0$. As evidenced by Theorem 2.6, X_i is typically of size $(N\eta)^{-1/2}$ in the bulk. Moreover, again by Theorem 2.6, we have the bound $\Lambda_* \prec \Phi$ with $\Phi := (N\eta)^{-1/2}$. Thus, we find that the effect of the averaging on X_i is to multiply it by a factor $(N\eta)^{-1/2}$. We conclude that for $\eta \approx 1$, the average $\frac{1}{N}\sum_i X_i$ is smaller than X_i by a factor $N^{-1/2}$, which corresponds to the behaviour of independent random variables $X_1, \ldots, X_N$. In contrast, for $\eta \approx N^{-1}$, the average $\frac{1}{N}\sum_i X_i$ is of the same size as X_i, which corresponds to the behaviour of fully correlated random variables $X_1 = \cdots = X_N$.

Now, suppose that $\Theta \prec (N\eta)^{-\sigma}$ in $\mathbf{S}$, where $\sigma \in [1/4, 1]$. Using Proposition 5.1 and Lemma 5.4, we get

$$\Lambda_* + |Y_i| + |G_{ii} - s| \prec \Phi_\sigma := \sqrt{\frac{\operatorname{Im} m + (N\eta)^{-\sigma}}{N\eta}}. \tag{6.3}$$

Using Lemma 5.2, we get

$$-s - z + Y_i = \frac{1}{G_{ii} - s + s} = \frac{1}{s} - \frac{1}{s^2}(G_{ii} - s) + \frac{1}{s^2 G_{ii}}(G_{ii} - s)^2.$$

Averaging over i and using that $|s^2 G_{ii}|^{-1} \prec 1$, we get

$$-s - z + \frac{1}{N}\sum_i Y_i = \frac{1}{s} + O_\prec(\Phi_\sigma^2).$$

Using Lemma 6.2, we get

$$s^2 + sz + 1 = O_\prec(\Phi_\sigma^2). \tag{6.4}$$

This is the desired perturbed quadratic equation for s.

Next, fix $\varepsilon, D > 0$ and define

$$\mathcal{E}_\sigma := \frac{N^\varepsilon \Phi_\sigma^2}{\sqrt{\kappa + \eta + \Phi_\sigma^2}}.$$

From (6.4) and Lemma 5.5 combined with the fact that s, m, and $\mathcal{E}_\sigma$ are N^2-Lipschitz, we find (using a net argument as in Remark 2.7) that there exists an event Ξ that does not depend on z such that $1 - \mathbf{1}(\Xi) \prec 0$ and

$$\mathbf{1}(\Xi) \min\{|s - m|, |s - \tilde{m}|\} \leqslant \mathcal{E}_\sigma \tag{6.5}$$

for all $z \in \mathbf{S}$.

As in the proof of Lemma 5.7, we now have to show that s is always close to m and not $\tilde{m}$. The argument is a continuity argument very similar to that from the proof of Lemma 5.7. In fact, since we do not have to worry about the deterioration of error probabilities, it is even easier. We do not even have to use a lattice, since on the high-probability event Ξ the estimate (6.5) holds simultaneously for all $z \in \mathbf{S}$. The case $\eta \geqslant 1$ is again trivial by the simple observation that $\operatorname{Im} \tilde{m} \leqslant -c$ for some positive constant c. Let us therefore focus on the case $\eta \leqslant 1$. For the following we fix a realization $H \in \Xi$, so that the argument is entirely deterministic.

Let $z = E + \mathrm{i} \in \mathbf{S}$ and define $z_t := z + \mathrm{i}\eta_t$ where $\eta_t := 1 - t$. The crossover time is defined by

$$T := \min\{t \geqslant 0 : \Phi_\sigma(z_t)^2 \geqslant N^{-2\varepsilon}(\kappa + \eta_t)\}.$$

Using the initial bound $|s(z_0) - m(z_0)| \leqslant \mathcal{E}_\sigma(z_0)$ and the estimates

$$\min\{|s(z_t) - m(z_t)|, |s(z_t) - \tilde{m}(z_t)|\} \leqslant \mathcal{E}_\sigma(z_t), \qquad |m(z_t) - \tilde{m}(z_t)| \geqslant c\sqrt{\kappa + \eta_t} \gg \mathcal{E}_\sigma(z_t)$$

valid for $t \in [0, T]$, we get using the continuity of s, m, and $\tilde{m}$, that $|s(z_t) - m(z_t)| \leqslant \mathcal{E}_\sigma(z_t)$ for $t \in [0, T]$.

Next, for $t > T$ we use the estimate

$$|s(z_t) - m(z_t)| \leqslant \mathcal{E}_\sigma(z_t) + |m(z_t) - \tilde{m}(z_t)| \leqslant 2N^\varepsilon \mathcal{E}_\sigma(z_t).$$

We conclude that $|s - m| \leqslant 2N^\varepsilon \mathcal{E}_\sigma$ on Ξ and for all $z \in \mathbf{S}$. Since ε and D were arbitrary, we conclude that if $\Theta \prec (N\eta)^{-\sigma}$ then

$$\Theta \prec \frac{\Phi_\sigma^2}{\sqrt{\kappa + \eta + \Phi_\sigma^2}} \leqslant \frac{\operatorname{Im} m}{N\eta\sqrt{\kappa + \eta}} + \sqrt{\frac{(N\eta)^{-\sigma}}{N\eta}} \leqslant \frac{C}{N\eta} + (N\eta)^{-1/2-\sigma/2}, \tag{6.6}$$

where in the last step we used Lemma 3.3. Note that the second estimate is in general wasteful. It turns out to be optimal for $E \in [-2, 2]$, but if E is away from the limiting spectrum $[-2, 2]$, it can be improved. We shall return to this point in Lemma 9.2 below.

In conclusion, we have shown for any $\sigma \in [1/4, 1]$ that

$$\Theta \prec (N\eta)^{-\sigma} \qquad \Longrightarrow \qquad \Theta \prec (N\eta)^{-1/2-\sigma/2}. \tag{6.7}$$

Starting from $\sigma = 1/4$ from Proposition 5.1 and iterating (6.7) a bounded number of times, we get $\Theta \prec (N\eta)^{-1}$. (Note that the number of iterations is independent of N; it depends only on the constant ε in the definition of $\prec$, which is arbitrary but fixed.) This concludes the proof of (2.14). Finally, (2.15) follows from (2.14) and the estimates (5.4) and (5.5). This concludes the proof of Theorem 2.6.

7. Fluctuation averaging: proof of Proposition 6.1

The first instance of the fluctuation averaging mechanism appeared in [44] for the Wigner case. A different proof (with a better bound on the constants) was given in [45]. A conceptually streamlined version of the original proof was extended to sparse matrices [34] and to sample covariance matrices [65]. Finally, an extensive analysis in [33] treated the fluctuation averaging of general polynomials of Green function entries and identified the order of cancellations depending on the algebraic structure of the polynomial. Moreover, in [33] an additional cancellation effect was found for the quantity $Q_i|G_{ij}|^2$. These improvements played a key role in obtaining the diffusion profile for the Green function of band matrices. The version we present here is based on the greatly simplified proof given in [37].

We start with a simple lemma which summarizes the key properties of $\prec$ when combined with expectation. Note that if X and Y are deterministic, $X \prec Y$ from Definition 2.5 simply means that for each $\varepsilon > 0$, we have for large enough N and all u that $X^{(N)}(u) \leqslant N^\varepsilon Y^{(N)}(u)$.

Lemma 7.1. *Suppose that the deterministic control parameter Φ satisfies $\Phi \geqslant N^{-C}$ for some constant $C > 0$, and that for all p there is a constant C_p such that the nonnegative random variable X satisfies $\mathbb{E}X^p \leqslant N^{C_p}$. Then we have the equivalence*

$$X \prec \Phi \qquad \Longleftrightarrow \qquad \mathbb{E}X^n \prec \Phi^n \text{ for any fixed } n \in \mathbb{N}. \tag{7.1}$$

Moreover, with P_i and Q_i as in Definition 3.2, we have

$$X \prec \Phi \qquad \Longrightarrow \qquad P_i X^n \prec \Phi^n \text{ and } Q_i X^n \prec \Phi^n \text{ for any fixed } n \in \mathbb{N}. \tag{7.2}$$

Finally, if $X \equiv X(u)$ and $\Phi \equiv \Phi(u)$ depend on some parameter u and any of the above hypothesis is uniform in u, then so is the corresponding conclusion.

Proof. If $\mathbb{E}X^n \prec \Phi^n$ then for any $\varepsilon > 0$ we get from Markov's inequality

$$\mathbb{P}(X > N^\varepsilon \Phi) \;\leqslant\; \frac{\mathbb{E}X^n}{N^{n\varepsilon}\Phi^n} \;\leqslant\; N^{-\varepsilon(n-1)}.$$

Choosing n large enough (depending on ε) proves the implication "$\Longleftarrow$" of (7.1). Conversely, if $X \prec \Phi$ then for any $D > 0$ we get

$$\mathbb{E}X \;=\; \mathbb{E}X\mathbf{1}(X \leqslant N^\varepsilon\Phi) + \mathbb{E}X\mathbf{1}(X > N^\varepsilon\Phi)$$
$$\leqslant\; N^\varepsilon\Phi + \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{P}(X > N^\varepsilon\Phi)} \;\leqslant\; N^\varepsilon\Phi + N^{C_2/2 - D/2},$$

Using $\Phi \geqslant N^{-C}$ and choosing D large enough, we obtain the implication "$\Longrightarrow$" of (7.1) for $n = 1$. The same implicitation for arbitrary n follows from the fact that $X \prec \Phi$ implies $X^n \prec \Phi^n$ for any fixed n. Moreover, (7.2) follows using (7.1) and Jensen's inequality for conditional expectations. The final claim about uniformity is trivial. $\qquad\square$

We shall apply Lemma 7.1 to the entries of G. In order to verify its assumptions, we record the following bounds.

Lemma 7.2. *Suppose that $\Lambda_* \prec \Phi$ for some deterministic control parameter Φ satisfying (6.2). Fix $\ell \in \mathbb{N}$. Then for any $i \neq j$ and $T \subset \{1, \dots, N\}$ satisfying $|T| \leqslant \ell$ and $i, j \notin T$ we have*

$$G_{ij}^{(T)} \;=\; O_\prec(\Phi), \qquad \frac{1}{G_{ii}^{(T)}} \;=\; O_\prec(1). \tag{7.3}$$

Moreover, we have the rough bounds $\left|G_{ij}^{(T)}\right| \leqslant N$ and

$$\mathbb{E}\left|\frac{1}{G_{ii}^{(T)}}\right|^n \;\leqslant\; N^\varepsilon \tag{7.4}$$

for any $\varepsilon > 0$ and $N \geqslant N_0(n, \varepsilon)$.

Proof. The bounds (7.3) follow easily by a repeated application of (3.4), the estimate $\Lambda \prec N^{-c}$ from Proposition 5.1, and the lower bound in (3.2). The deterministic bound $\left|G_{ij}^{(T)}\right| \leqslant N$ follows immediately from $\eta \geqslant N^{-1}$ by definition of a spectral domain.

In order to prove (7.4), we use Schur's complement formula (5.1) applied to $1/G_{ii}^{(T)}$, where the expectation is estimated using (iii) of Definition 2.2 and $\left|G_{ij}^{(T)}\right| \leqslant N$. This gives

$$\mathbb{E}\left|\frac{1}{G_{ii}^{(T)}}\right|^p \;\prec\; N^{C_p}$$

for all $p \in \mathbb{N}$. Since $1/\left|G_{ii}^{(T)}\right| \prec 1$, (7.4) therefore follows from (7.1). $\qquad\square$

Proof of Proposition 6.1. First we claim that, for any fixed $\ell \in \mathbb{N}$, we have

$$\left| Q_k \frac{1}{G_{kk}^{(T)}} \right| \prec \Phi \tag{7.5}$$

uniformly for $T \subset \{1, \ldots, N\}$, $|T| \leqslant \ell$, and $k \notin T$. To simplify notation, for the proof we set $T = \emptyset$; the proof for nonempty T is the same. From Schur's complement formula (5.1) we get $|Q_k(G_{kk})^{-1}| \leqslant |H_{kk}| + |Z_k|$. The first term is estimated by $|H_{kk}| \prec N^{-1/2} \leqslant \Phi$. The second term is estimated exactly as in (5.6):

$$|Z_k|^2 \prec \frac{1}{N^2} \sum_k^{(i)} |G_{kk}^{(i)}|^2 + \sum_{k,l}^{(i)} \frac{1}{N^2} |G_{kl}^{(i)}|^2 \prec N^{-1} + \Phi^2 \prec \Phi^2 \,,$$

where in the second step we used Lemma 7.2. This concludes the proof of (7.5).

Abbreviate $X_k := Q_k(G_{kk})^{-1}$. We shall estimate $N^{-1} \sum_k X_k$ in probability by estimating its p-th moment by Φ^{2p}, from which the claim will easily follow using Markov's inequality. Before embarking on the estimate for arbitrary p, we illustrate its idea by estimating the variance

$$\mathbb{E} \left| N^{-1} \sum_k X_k \right|^2 = N^{-2} \sum_{k,l} \mathbb{E} X_k \overline{X}_l = N^{-2} \sum_k \mathbb{E} X_k \overline{X}_k + N^{-2} \sum_{k \neq l} \mathbb{E} X_k \overline{X}_l \,. \tag{7.6}$$

Using Lemma 7.1, we find that the first term on the right-hand side of (7.6) is $O_{\prec}(N^{-1}\Phi^2) = O_{\prec}(\Phi^4)$, where we used the estimate (6.2). Let us therefore focus on the second term of (7.6). Using the fact that $k \neq l$, we apply (3.4) to X_k and X_l to get

$$\mathbb{E} X_k \overline{X}_l = \mathbb{E} Q_k \left(\frac{1}{G_{kk}} \right) Q_l \overline{\left(\frac{1}{G_{ll}} \right)} = \mathbb{E} Q_k \left(\frac{1}{G_{kk}^{(l)}} - \frac{G_{kl}G_{lk}}{G_{kk}G_{kk}^{(l)}G_{ll}} \right) Q_l \overline{\left(\frac{1}{G_{ll}^{(k)}} - \frac{G_{lk}G_{kl}}{G_{ll}G_{ll}^{(k)}G_{kk}} \right)}. \tag{7.7}$$

We multiply out the parentheses on the right-hand side. The crucial observation is that if the random variable Y is $H^{(i)}$-measurable then

$$\mathbb{E} Q_i(X)Y = \mathbb{E} Q_i(XY) = 0. \tag{7.8}$$

Hence out of the four terms obtained from the right-hand side of (7.7), the only nonvanishing one is

$$\mathbb{E} Q_k \left(\frac{G_{kl}G_{lk}}{G_{kk}G_{kk}^{(l)}G_{ll}} \right) Q_l \overline{\left(\frac{G_{lk}G_{kl}}{G_{ll}G_{ll}^{(k)}G_{kk}} \right)} \prec \Phi^4 \,.$$

This concludes the proof of $\mathbb{E} \left| N^{-1} \sum_k X_k \right|^2 \prec \Phi^4$.

After this pedagogical interlude we move on to the full proof. Fix some even integer p and write

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p = \frac{1}{N^p} \sum_{k_1, \ldots, k_p} \mathbb{E} X_{k_1} \cdots X_{k_{p/2}} \overline{X}_{k_{p/2+1}} \cdots \overline{X}_{k_p} \,.$$

Next, we regroup the terms in the sum over $\mathbf{k} := (k_1, \ldots, k_p)$ according to the partition of $\{1, \ldots, p\}$ generated by the indices $\mathbf{k}$. To that end, let $\mathfrak{P}_p$ denote the set of partitions of $\{1, \ldots, p\}$, and $\mathcal{P}(\mathbf{k})$

the element of $\mathfrak{P}_p$ defined by the equivalence relation $r \sim s$ if and only if $k_r = k_s$. In short, we reorganize the summation according to coincidences among the indices $\mathbf{k}$. Then we write

$$\mathbb{E}\left|\frac{1}{N}\sum_k X_k\right|^p = \sum_{P\in\mathfrak{P}_p}\frac{1}{N^p}\sum_{\mathbf{k}}\mathbf{1}(\mathcal{P}(\mathbf{k}) = P)V(\mathbf{k})\,, \tag{7.9}$$

where we defined

$$V(\mathbf{k}) := \mathbb{E}X_{k_1}\cdots X_{k_{p/2}}\overline{X}_{k_{p/2+1}}\cdots \overline{X}_{k_p}\,.$$

Fix $\mathbf{k}$ and set $P := \mathcal{P}(\mathbf{k})$ to be the partition induced by the coincidences in $\mathbf{k}$. For any $r \in \{1,\ldots,p\}$, we denote by $[r]$ the block of r in P. Let $L \equiv L(P) := \{r : [r] = \{r\}\} \subset \{1,\ldots,p\}$ be the set of "lone" labels. We denote by $\mathbf{k}_L := (k_r)_{r\in L}$ the summation indices associated with lone labels.

The Green function entry G_{kk} depends strongly on the randomness in the k-column of H, but only weakly on the randomness in the other columns. We conclude that if r is a lone label then all factors X_{k_s} with $s \neq r$ in $V(\mathbf{k})$ depend weakly on the randomness in the k_r-th column of H. Thus, the idea is to make all Green function entries inside the expectation of $V(\mathbf{k})$ as independent as possible of the randomness in the rows of H indexed by $\mathbf{k}_L$ (see Definition 3.2), using the identity (3.4). To that end, we say that a Green function entry $G_{xy}^{(T)}$ with $x, y \notin T$ is *maximally expanded* if $\mathbf{k}_L \subset T \cup \{x, y\}$. The motivation behind this definition is that using (3.4) we cannot add upper indices from the set $\mathbf{k}_L$ to a maximally expanded Green function entry. We shall apply (3.4) to all Green function entries in $V(\mathbf{k})$. In this manner we generate a sum of monomials consisting of off-diagonal Green function entries and inverses of diagonal Green function entries. We can now repeatedly apply (3.4) to each factor until either they are all maximally expanded or a sufficiently large number of off-diagonal Green function entries has been generated. The cap on the number of off-diagonal entries is introduced to ensure that this procedure terminates after a finite number of steps.

In order to define the precise algorithm, let $\mathcal{A}$ denote the set of monomials in the off-diagonal entries $G_{xy}^{(T)}$, with $T \subset \mathbf{k}_L$, $x \neq y$, and $x, y \in \mathbf{k} \setminus T$, as well as the inverse diagonal entries $1/G_{xx}^{(T)}$, with $T \subset \mathbf{k}_L$ and $x \in \mathbf{k} \setminus T$. Starting from $V(\mathbf{k})$, the algorithm will recursively generate sums of monomials in $\mathcal{A}$. Let $d(A)$ denote the number of off-diagonal entries in $A \in \mathcal{A}$. For $A \in \mathcal{A}$ we shall define $w_0(A), w_1(A) \in \mathcal{A}$ satisfying

$$A = w_0(A) + w_1(A)\,, \qquad d(w_0(A)) = d(A)\,, \qquad d(w_1(A)) \geqslant \max\{2, d(A) + 1\}\,. \tag{7.10}$$

The idea behind this splitting is to use (3.4) on one entry of A; the first term on the right-hand side of (3.4) gives rise to $w_0(A)$ and the second to $w_1(A)$. The precise definition of the algorithm applied to $A \in \mathcal{A}$ is as follows.

(1) If all factors of A are maximally expanded or $d(A) \geqslant p + 1$ then stop the expansion of A. In other words, the algorithm cannot be applied to A in the future.

(2) Otherwise choose some (arbitrary) factor of A that is not maximally expanded. If this entry is off-diagonal, $G_{xy}^{(T)}$, write

$$G_{xy}^{(T)} = G_{xy}^{(Tu)} + \frac{G_{xu}^{(T)}G_{uy}^{(T)}}{G_{uu}^{(T)}} \tag{7.11}$$

33

for the smallest $u \in \mathbf{k}_L \setminus (T \cup \{x, y\})$. If the chosen entry is diagonal, $1/G_{xx}^{(T)}$, write

$$\frac{1}{G_{xx}^{(T)}} = \frac{1}{G_{xx}^{(Tu)}} - \frac{G_{xu}^{(T)}G_{ux}^{(T)}}{G_{xx}^{(T)}G_{xx}^{(Tu)}G_{uu}^{(T)}} \tag{7.12}$$

for the smallest $u \in \mathbf{k}_L \setminus (T \cup \{x\})$. Then the splitting $A = w_0(A) + w_1(A)$ is defined by the splitting induced by (7.11) or (7.12), in the sense that we replace the factor $G_{xy}^{(T)}$ or $1/G_{xx}^{(T)}$ in the monomial A by the right-hand sides of (7.11) or (7.12).

(This algorithm contains some arbitrariness in the choice of the factor of A to be expanded. It may be removed for instance by first fixing some ordering of all Green function entries $G_{ij}^{(T)}$. Then in (2) we choose the first factor of A that is not maximally expanded.) Note that (7.11) and (7.12) follow from (3.4). It is clear that (7.10) holds with the algorithm just defined.

We now apply this algorithm recursively to each entry $A^r := 1/G_{k_r k_r}$ in the definition of $V(\mathbf{k})$. More precisely, we start with A^r and define $A_0^r := w_0(A^r)$ and $A_1^r := w_1(A^r)$. In the second step of the algorithm we define four monomials

$$A_{00}^r := w_0(A_0^r), \qquad A_{01}^r := w_0(A_1^r), \qquad A_{10}^r := w_1(A_0^r), \qquad A_{11}^r := w_1(A_1^r),$$

and so on, at each iteration performing the steps (1) and (2) on each new monomial independently of the others. Note that the lower indices are binary sequences that describe the recursive application of the operations w_0 and w_1. In this manner we generate a binary tree whose vertices are given by finite binary strings σ. The associated monomials satisfy $A_{\sigma i}^r := w_i(A_\sigma^r)$ for $i = 0, 1$, where σi denotes the binary string obtained by appending i to the right end of σ. See Figure 7.1 for an illustration of the tree.
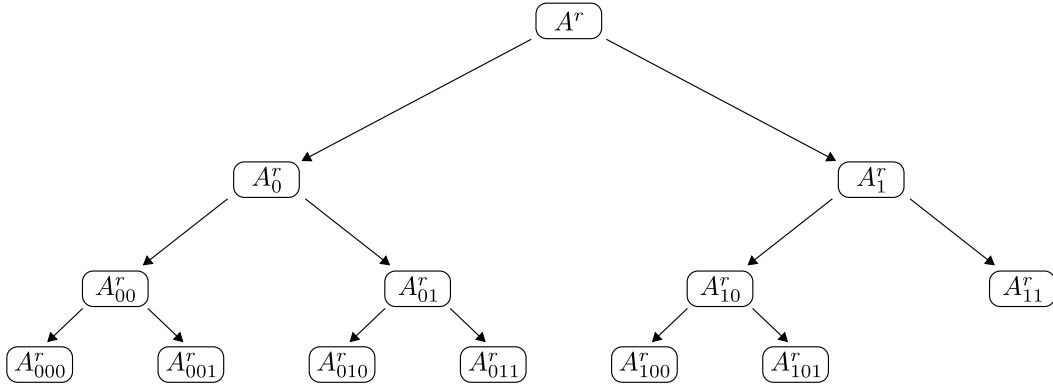
Figure 7.1. The binary tree generated by applying the algorithm (1)–(2) to a monomial A^r. Each vertex of the tree is indexed by a binary string σ, and encodes a monomial A_σ^r. An arrow towards the left represents the action of w_0 and an arrow towards the right the action of w_1. The monomial A_{11}^r satisfies the assumptions of step (1), and hence its expansion is stopped, so that the tree vertex 11 has no children.

We stop the recursion of a tree vertex whenever the associated monomial satisfies the stopping rule of step (1). In other words, the set of leaves of the tree is the set of binary strings σ such that

either all factors of A_σ^r are maximally expanded or $d(A_\sigma^r) \geqslant p + 1$. We claim that the resulting binary tree is finite, i.e. that the algorithm always reaches step (1) after a finite number of iterations. Indeed, by the stopping rule in (1), we have $d(A_\sigma^r) \leqslant p + 1$ for any vertex σ of the tree. Since each application of w_1 increases $d(\cdot)$ by at least one, and in the first step (i.e. when applied to A^r) by two, we conclude that the number of ones in any σ is at most p. Since each application of w_1 increases the number of Green function entries by at most four, and the application of w_0 does not change this number, we find that the number of Green function entries in A_σ^r is bounded by $4p + 1$. Hence the maximal number of upper indices in A_σ^r for any tree vertex σ is $(4p + 1)p$. Since each application of w_0 increases the total number of upper indices by one, we find that σ contains at most $(4p + 1)p$ zeros. We conclude that the maximal length of the string σ (i.e. the depth of the tree) is at most $(4p+1)p+p = 4p^2 + 2p$. A string σ encoding a tree vertex contains at most p ones. Denoting by k the number of ones in a string encoding a leaf of the tree, we find that the number of leaves is bounded by $\sum_{k=0}^p \binom{4p^2+2p}{k} \leqslant (Cp^2)^p$. Therefore, denoting by $\mathcal{L}_r$ the set of leaves of the binary tree generated from A^r, we have $|\mathcal{L}_r| \leqslant (Cp^2)^p$.

By definition of the tree and w_0 and w_1, we have the decomposition

$$X_{k_r} \;=\; Q_{k_r} \sum_{\sigma \in \mathcal{L}_r} A_\sigma^r \,. \tag{7.13}$$

Moreover, each monomial A_σ^r for $\sigma \in \mathcal{L}_r$ either consists entirely of maximally expanded Green function entries or satisfies $d(A_\sigma^r) = p + 1$. (This is an immediate consequence of the stopping rule in (1)).

Next, we observe that for any string σ we have

$$A_\sigma^k \;=\; O_\prec\!\big(\Phi^{b(\sigma)+1}\big)\,, \tag{7.14}$$

where $b(\sigma)$ is the number ones in the string σ. Indeed, if $b(\sigma) = 0$ then this follows from (7.5); if $b(\sigma) \geqslant 1$ this follows from the last statement in (7.10) and (7.3).

Using (7.9) and (7.13) we have the representation

$$V(\mathbf{k}) \;=\; \sum_{\sigma_1 \in \mathcal{L}_1} \cdots \sum_{\sigma_p \in \mathcal{L}_p} \mathbb{E}\big(Q_{k_1} A_{\sigma_1}^1\big) \cdots \big(Q_{k_p} \overline{A_{\sigma_p}^p}\big)\,. \tag{7.15}$$

We now claim that any nonzero term on the right-hand side of (7.15) satisfies

$$\big(Q_{k_1} A_{\sigma_1}^1\big) \cdots \big(Q_{k_p} \overline{A_{\sigma_p}^p}\big) \;=\; O_\prec\!\big(\Phi^{p+|L|}\big)\,. \tag{7.16}$$

Proof of (7.16). Before embarking on the proof, we explain its idea. By (7.14), the naive size of the left-hand side of (7.16) is Φ^p. The key observation is that each lone label $s \in L$ yields one extra factor Φ to the estimate. This is because by (7.8), the expectation in (7.15) would vanish if all other factors $\big(Q_{k_r} A_{\sigma_r}^r\big)$, $r \neq s$, were $H^{(k_s)}$-measurable. The expansion of the binary tree makes this dependence explicit by exhibiting k_s as a lower index. But this requires performing an operation w_1 with the choice $u = k_s$ in (7.11) or (7.12). However, w_1 increases the number of off-diagonal element by at least one. In other words, every index associated with a lone label must have a "partner" index in a different Green function entry which arose by application of w_1. Such a partner index may only be obtained through the creation of at least one off-diagonal Green function entry. The actual proof below shows that this effect applies *cumulatively* for all lone labels.

35

In order to prove (7.16), we consider two cases. Consider first the case where for some $r = 1, \ldots, p$ the monomial $A^r_{\sigma_r}$ on the left-hand side of (7.16) is not maximally expanded. Then $d(A^r_{\sigma_r}) = p+1$, so that (7.3) yields $A^r_{\sigma_r} \prec \Phi^{p+1}$. Therefore the observation that $A^s_{\sigma_s} \prec \Phi$ for all $s \neq r$, together with (7.2) implies that the left-hand side of (7.16) is $O_\prec(\Phi^{2p})$. Since $|L| \leqslant p$, (7.16) follows.

Consider now the case where $A^r_{\sigma_r}$ on the left-hand side of (7.16) is maximally expanded for all $r = 1, \ldots, p$. The key observation is the following claim about the left-hand side of (7.16) with a nonzero expectation.

(∗) For each $s \in L$ there exists $r = \tau(s) \in \{1, \ldots, p\} \setminus \{s\}$ such that the monomial $A^r_{\sigma_r}$ contains a Green function entry with lower index k_s.

In other words, after expansion, the lone label s has a "partner" label $r = \tau(s)$, such that the index k_s appears also in the expansion of A^r (note that there may be several such partner labels r). To prove (∗), suppose by contradiction that there exists an $s \in L$ such that for all $r \in \{1, \ldots, p\} \setminus \{s\}$ the lower index k_s does not appear in the monomial $A^r_{\sigma_r}$. To simplify notation, we assume that $s = 1$. Then, for all $r = 2, \ldots, p$, since $A^r_{\sigma_r}$ is maximally expanded, we find that $A^r_{\sigma_r}$ is $H^{(k_1)}$-measurable. Therefore we have

$$\mathbb{E}\big(Q_{k_1} A^1_{\sigma_1}\big)\big(Q_{k_2} A^2_{\sigma_2}\big) \cdots \big(Q_{k_p} \overline{A^p_{\sigma_p}}\big) \;=\; \mathbb{E}Q_{k_1}\Big(A^1_{\sigma_1}\big(Q_{k_2} A^2_{\sigma_2}\big) \cdots \big(Q_{k_p} \overline{A^p_{\sigma_p}}\big)\Big) \;=\; 0\,,$$

where in the last step we used (7.8). This concludes the proof of (∗).

For $r \in \{1, \ldots, p\}$ we define $l(r) := \sum_{s \in L} \mathbf{1}(\tau(s) = r)$, the number of times that the label r was chosen as a partner to some lone label s. We now claim that

$$A^r_{\sigma_r} \;=\; O_\prec\big(\Phi^{1+l(r)}\big)\,. \tag{7.17}$$

To prove (7.17), fix $r \in \{1, \ldots, p\}$. By definition, for each $s \in \tau^{-1}(\{r\})$ the index k_s appears as a lower index in the monomial $A^r_{\sigma_r}$. Since $s \in L$ is by definition a lone label and $s \neq r$, we know that k_s does not appear as an index in A^r. By definition of the monomials associated with the tree vertex σ_r, it follows that $b(\sigma_r)$, the number of ones in σ_r, is at least $|\tau^{-1}(\{r\})| = l(r)$ since each application of w_1 adds precisely one new (lower) index. Note that in this step it is crucial that $s \in \tau^{-1}(\{r\})$ was a lone label. Recalling (7.14), we therefore get (7.17).

Using (7.17) and Lemma 7.1 we find

$$\Big|\big(Q_{k_1} A^1_{\sigma_1}\big) \cdots \big(Q_{k_p} \overline{A^p_{\sigma_p}}\big)\Big| \;\prec\; \prod_{r=1}^{p} \Phi^{1+l(r)} \;=\; \Phi^{p+|L|}\,.$$

This concludes the proof of (7.16). $\qquad\square$

Summing over the binary trees in (7.15) and using Lemma 7.1, we get from (7.16)

$$V(\mathbf{k}) \;=\; O_\prec\big(\Phi^{p+|L|}\big)\,. \tag{7.18}$$

We now return to the sum (7.9). We perform the summation by first fixing $P \in \mathfrak{P}_p$, with associated lone labels $L = L(P)$. We find

$$\bigg|\frac{1}{N^p} \sum_{\mathbf{k}} \mathbf{1}(\mathcal{P}(\mathbf{k}) = P)\bigg| \;\leqslant\; (N^{-1})^{p-|P|} \;\leqslant\; (N^{-1/2})^{p-|L|}\,;$$

36

in the first step we used that the summation is performed over $|P|$ free indices, the remaining $p - |P|$ being estimated by N^{-1}; in the second step we used that each block of P that is not contained in L consists of at least two labels, so that $p - |P| \geqslant (p - |L|)/2$. From (7.9) and (7.18) we get

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p \prec \sum_{P \in \mathfrak{P}_p} (N^{-1/2})^{p - |L(P)|} \, \Phi^{p + |L(P)|} \leqslant C_p \Phi^{2p} \,,$$

where in the last step we used the lower bound from (6.2) and estimated the summation over $\mathfrak{P}_p$ with a constant C_p (which is bounded by $(Cp^2)^p$). Summarizing, we have proved that

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p \prec \Phi^{2p} \tag{7.19}$$

for any $p \in 2\mathbb{N}$.

We conclude the proof of Proposition 6.1 with a simple application of Markov's inequality. Fix $\varepsilon > 0$ and $D > 0$. Using (7.19) and Markov's inequality we find

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_k X_k \right| > N^\varepsilon \Phi^2 \right) \leqslant N \, N^{-\varepsilon p}$$

for large enough $N \geqslant N_0(\varepsilon, p)$. Choosing $p \geqslant \varepsilon^{-1}(1 + D)$ concludes the proof of Proposition 6.1. $\square$

We conclude this section with an alternative proof of Proposition 6.1. While the underlying argument remains similar, the following proof makes use of an additional decomposition of the space of random variables, which avoids the use of the stopping rule from Step (1) in the above proof of Proposition 6.1. This decomposition may be regarded as an abstract reformulation of the stopping rule.

Alternative proof of Proposition 6.1. As before, we set $X_k := Q_k(G_{kk})^{-1}$. The decomposition is defined using the operations P_i and Q_i, introduced in Definition 3.2. It is immediate that P_i and Q_i are projections, that $P_i + Q_i = 1$, and that all of these projections commute with each other (by Fubini's theorem). For a set $A \subset \{1, \ldots, N\}$ we use the notations $P_A := \prod_{i \in A} P_i$ and $Q_A := \prod_{i \in A} Q_i$.

Let p be even and introduce the shorthand $\widetilde{X}_{k_s} := X_{k_s}$ for $s \leqslant p/2$ and $\widetilde{X}_{k_s} := \overline{X}_{k_s}$ for $s > p/2$. Then we get

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p = \frac{1}{N^p} \sum_{k_1, \ldots, k_p} \mathbb{E} \prod_{s=1}^p \widetilde{X}_{k_s} = \frac{1}{N^p} \sum_{k_1, \ldots, k_p} \mathbb{E} \prod_{s=1}^p \left(\prod_{r=1}^p (P_{k_r} + Q_{k_r}) \widetilde{X}_{k_s} \right).$$

Introducing the notations $\mathbf{k} = (k_1, \ldots, k_p)$ and $[\mathbf{k}] = \{k_1, \ldots, k_p\}$, we therefore get by multiplying out the parentheses

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p = \frac{1}{N^p} \sum_{\mathbf{k}} \sum_{A_1, \ldots, A_p \subset [\mathbf{k}]} \mathbb{E} \prod_{s=1}^p \left(P_{A_s^c} Q_{A_s} \widetilde{X}_{k_s} \right). \tag{7.20}$$

Next, by definition of $\widetilde{X}_{k_s}$, we have that $\widetilde{X}_{k_s} = Q_{k_s} \widetilde{X}_{k_s}$, which implies that $P_{A_s^c} \widetilde{X}_{k_s} = 0$ if $k_s \notin A_s$. Hence may restrict the summation to A_s satisfying

$$k_s \in A_s \tag{7.21}$$

for all s. Moreover, we claim that the right-hand side of (7.20) vanishes unless

$$k_s \in \bigcup_{q \neq s} A_q \tag{7.22}$$

for all s. Indeed, suppose that $k_s \in \bigcap_{q \neq s} A_q^c$ for some s, say $s = 1$. In this case, for each $s = 2, \ldots, p$, the factor $P_{A_s^c} Q_{A_s} \widetilde{X}_{k_s}$ is $H^{(k_1)}$-measurable. Thus we get

$$\mathbb{E} \prod_{s=1}^{p} \left(P_{A_s^c} Q_{A_s} \widetilde{X}_{k_s} \right) = \mathbb{E} \left(P_{A_1^c} Q_{A_1} Q_{k_1} \widetilde{X}_{k_1} \right) \prod_{s=2}^{p} \left(P_{A_s^c} Q_{A_s} \widetilde{X}_{k_s} \right)$$

$$= \mathbb{E} Q_{k_1} \left(\left(P_{A_1^c} Q_{A_1} \widetilde{X}_{k_1} \right) \prod_{s=2}^{p} \left(P_{A_s^c} Q_{A_s} \widetilde{X}_{k_s} \right) \right) = 0,$$

where in the last step we used that $\mathbb{E} Q_i(X) = 0$ for any i and any random variable X.

We conclude that the summation on the right-hand side of (7.20) is restricted to indices satisfying (7.21) and (7.22). Under these two conditions we have

$$\sum_{s=1}^{p} |A_s| \geq 2\, |[\mathbf{k}]|\,, \tag{7.23}$$

since each index k_s must belong to at least two different sets A_q: to A_s (by (7.21)) as well as to some A_q with $q \neq s$ (by (7.22)).

Next, we claim that for $k \in A$ we have

$$|Q_A X_k| \prec \Phi^{|A|}. \tag{7.24}$$

Before proving (7.24), we show it may be used to complete the proof. Using (7.20), (7.24), and Lemma 7.1, we find

$$\mathbb{E} \left| \frac{1}{N} \sum_k X_k \right|^p \prec C_p \frac{1}{N^p} \sum_{\mathbf{k}} \Phi^{2|[k]|} = C_p \sum_{u=1}^{p} \Phi^{2u} \frac{1}{N^p} \sum_{\mathbf{k}} \mathbf{1}(|[\mathbf{k}]| = u)$$

$$\leq C_p \sum_{u=1}^{p} \Phi^{2u} N^{u-p} \leq C_p (\Phi + N^{-1/2})^{2p} \leq C_p \Phi^{2p},$$

where in the first step we estimated the summation over the sets $A_1, \ldots, A_p$ by a combinatorial factor C_p depending on p, in the fourth step we used the elementary inequality $a^n b^m \leq (a+b)^{n+m}$ for positive a, b, and in the last step we used (6.2). Thus we have proved (7.19), from which the claim follows exactly as in the first proof of Proposition 6.1.

What remains is the proof of (7.24). The case $|A| = 1$ (corresponding to $A = \{k\}$) follows from (7.5), exactly as in the first proof of Proposition 6.1. To simplify notation, for the case $|A| \geq 2$ we assume that $k = 1$ and $A = \{1, 2, \ldots, t\}$ with $t \geq 2$. It suffices to prove that

$$\left| Q_t \cdots Q_2 \frac{1}{G_{11}} \right| \prec \Phi^t. \tag{7.25}$$

We start by writing, using (3.4),

$$Q_2 \frac{1}{G_{11}} = Q_2 \frac{1}{G_{11}^{(2)}} - Q_2 \frac{G_{12}G_{21}}{G_{11}G_{11}^{(2)}G_{22}} = Q_2 \frac{G_{12}G_{21}}{G_{11}G_{11}^{(2)}G_{22}},$$

where the first term vanishes since $G_{11}^{(2)}$ is $H^{(2)}$-measurable. We now consider

$$Q_3 Q_2 \frac{1}{G_{11}} = Q_2 Q_3 \frac{G_{12}G_{21}}{G_{11}G_{11}^{(2)}G_{22}},$$

and apply (3.4) with $k = 3$ to each Green function entry on the right-hand side, and multiply everything out. The result is a sum of fractions of entries of G, whereby all entries in the numerator are off-diagonal and all entries in the denominator are diagonal. The leading order term vanishes,

$$Q_2 Q_3 \frac{G_{12}^{(3)}G_{21}^{(3)}}{G_{11}^{(3)}G_{11}^{(23)}G_{22}^{(3)}} = 0,$$

so that the surviving terms have at least three (off-diagonal) Green function entries in the numerator. We may now continue in this manner; at each step the number of (off-diagonal) Green function entries in the numerator increases by at least one.

More formally, we obtain a sequence $A_2, A_3, \ldots, A_t$, where $A_2 := Q_2 \frac{G_{12}G_{21}}{G_{11}G_{11}^{(2)}G_{22}}$ and A_i is obtained by applying (3.4) with $k = i$ to each entry of $Q_i A_{i-1}$, and keeping only the nonvanishing terms. The following properties are easy to check by induction.

(i) $A_i = Q_i A_{i-1}$.

(ii) A_i consists of the projection $Q_2 \cdots Q_i$ applied to a sum of fractions such that all entries in the numerator are off-diagonal and all entries in the denominator are diagonal.

(iii) The number of (off-diagonal) entries in the numerator of each term of A_i is at least i.

By Lemma 7.1 combined with (ii) and (iii) we conclude that $|A_i| \prec \Phi^i$. From (i) we therefore get

$$Q_t \cdots Q_2 \frac{1}{G_{11}} = A_t = O_\prec(\Phi^t).$$

This is (7.25). Hence the proof is complete. □

8. Semicircle law on small scales: proof of Theorem 2.8

We define the signed measure $\hat{\mu}$ and its Stieltjes transform $\hat{s}$ through

$$\hat{\mu} := \mu - \varrho, \qquad \hat{s}(z) := \int \frac{\hat{\mu}(\mathrm{d}x)}{x - z} = s(z) - m(z). \qquad (8.1)$$

The basic idea behind the proof of Theorem 2.8 is to estimate $\hat{\mu}(I)$ using the Helffer-Sjöstrand formula from Appendix C in terms of its Stieltjes transfrom, $\hat{s}$, which is controlled by Theorem 2.6.

To that end, fix $\varepsilon > 0$ and define $\eta := N^{-1+\varepsilon}$. Then for any interval $I \subset [-3, 3]$ we choose a smoothed indicator function $f \equiv f_{I,\eta} \in \mathcal{C}_c^\infty(\mathbb{R}; [0, 1])$ satisfying $f(x) = 1$ for $x \in I$, $f(x) = 0$ for $\mathrm{dist}(x, I) \geq \eta$, $\|f'\|_\infty \leq C\eta^{-1}$, and $\|f''\|_\infty \leq C\eta^{-2}$. Note that the supports of f' and f'' have Lebesgue measure at most 2η, which we tacitly use from now on. Next, choose a smooth, even, cutoff function $\chi \in \mathcal{C}_c^\infty(\mathbb{R}; [0, 1])$ satisfying $\chi(y) = 1$ for $|y| \leq 1$, $\chi(y) = 0$ for $|y| \geq 2$, and $\|\chi'\|_\infty \leq C$.

Now using the Helffer-Sjöstrand formula from Proposition C.1 with $n = 1$, we get

$$\int f(\lambda)\, \hat{\mu}(\mathrm{d}\lambda) \;=\; \frac{1}{2\pi} \int \mathrm{d}x \int \mathrm{d}y\, (\partial_x + \mathrm{i}\partial_y)\big[\big(f(x) + \mathrm{i}yf'(x)\big)\chi(y)\big]\, \hat{s}(x + \mathrm{i}y)\,.$$

Since the left-hand side is real, we obtain

$$\int f(\lambda)\, \hat{\mu}(\mathrm{d}\lambda) \;=\; -\frac{1}{2\pi} \int \mathrm{d}x \int_{|y| \leq \eta} \mathrm{d}y\, f''(x)\chi(y) y\, \mathrm{Im}\, \hat{s}(x + \mathrm{i}y) \tag{8.2}$$

$$-\frac{1}{2\pi} \int \mathrm{d}x \int_{|y| > \eta} \mathrm{d}y\, f''(x)\chi(y) y\, \mathrm{Im}\, \hat{s}(x + \mathrm{i}y) \tag{8.3}$$

$$+\frac{\mathrm{i}}{2\pi} \int \mathrm{d}x \int \mathrm{d}y\, \big(f(x) + \mathrm{i}yf'(x)\big)\chi'(y)\hat{s}(x + \mathrm{i}y)\,. \tag{8.4}$$

Note the crucial cancellation of the terms proportional to $f'(x)\chi(y)$.

We shall estimate the three error terms (8.2)–(8.4) by using the estimate (2.14) from the local semicircle law, Theorem 2.6. In fact, (2.14) only holds at each $z \in \mathbf{S}$ individually, and in order to apply it to the integrals in (8.2)–(8.4) we need a *simultaneous* estimate for all $z \in \mathbf{S}$. This extension is the content of the following simple lemma, which follows from Theorem 2.6 using the argument in Remark 2.7 and the trivial identity $\hat{s}(x - \mathrm{i}y) = \overline{\hat{s}(x + \mathrm{i}y)}$.

Lemma 8.1. *For any fixed $\varepsilon > 0$, we have with high probability $|\hat{s}(x + \mathrm{i}y)| \leq N^\varepsilon/(N|y|)$ for $|x| \leq \varepsilon^{-1}$ and $|y| \in [\eta, \varepsilon^{-1}]$.*

Using Lemma 8.1, we may now estimate (8.2)–(8.4). First, using that χ' is supported in $[-2, 2] \setminus (-1, 1)$, we easily find

$$|(8.4)| \;\leq\; C\eta \tag{8.5}$$

with high probability.

Next, we estimate (8.2). Since y can be arbitrarily small, Lemma 8.1 does not apply. However, this can be easily remedied by the following monotonicity argument. For simplicity, we only deal with $y > 0$; the estimate for $y < 0$ is analogous. It is easy to see that the map $y \mapsto y\, \mathrm{Im}\, s(x + \mathrm{i}y)$ is nondecreasing for all x (this is a general property of the Stieltjes transform of any positive measure). Hence, for $y \in (0, \eta)$ we have

$$y\, \mathrm{Im}\, \hat{s}(x + \mathrm{i}y) \;\leq\; y\, \mathrm{Im}\, s(x + \mathrm{i}y) \;\leq\; \eta\, \mathrm{Im}\, s(x + \mathrm{i}\eta) \;\leq\; \eta(N^\varepsilon/(N\eta) + C) \;\leq\; C\eta$$

with high probability, where in the third step we used Lemma 8.1. We conclude that

$$|(8.2)| \;\leq\; \eta^{-1} \int_{|y| \leq \eta} \mathrm{d}y\, C\eta \;=\; C\eta \tag{8.6}$$

with high probability.

What remains is (8.3). Unlike in (8.2), the second derivative on f is not affordable (since it results in a factor η^{-1} after integration over x). The solution is to integrate by parts in x and then in y. Combined with $\hat{s}(x - \mathrm{i}y) = \overline{\hat{s}(x + \mathrm{i}y)}$, we therefore get

$$
|(8.3)| \leqslant \int \mathrm{d}x \int_\eta^\infty \mathrm{d}y \left| f'(x)\chi(y)\hat{s}(x + \mathrm{i}y) \right|
$$
$$
+ \int \mathrm{d}x \int_\eta^\infty \mathrm{d}y \left| f'(x)y\chi'(y)\hat{s}(x + \mathrm{i}y) \right| + \int \mathrm{d}x \left| f'(x)\eta\hat{s}(x + \mathrm{i}\eta) \right|. \quad (8.7)
$$

As above, using Lemma 8.1 we easily find that the second line of (8.7) is bounded by $C\eta$ with high probability. Moreover, the first term on the right-hand side of (8.7) is bounded by

$$
C \int_\eta^2 \mathrm{d}y \, \frac{N^\varepsilon}{Ny} \leqslant C\eta \log N \, .
$$

Recalling (8.5) and (8.6), we have therefore proved that

$$
\left| \int f(\lambda) \, \hat{\mu}(\mathrm{d}\lambda) \right| \leqslant CN^{-1+2\varepsilon} \quad (8.8)
$$

with high probability.

In order to conclude the proof of Theorem 2.8, we have to return from the smoothed indicator function f to the sharp indicator function of I. To that end, we note that if $I \subset [-3,3]$ we have the upper bound

$$
\mu(I) \leqslant \int f_{I,\eta}(\lambda) \, \mu(\mathrm{d}\lambda) \leqslant \int f_{I,\eta}(\lambda) \, \varrho(\mathrm{d}\lambda) + O(N^{-1+2\varepsilon}) \leqslant \varrho(I) + O(N^{-1+2\varepsilon})
$$

with high probability, where in the second step we used (8.8), and in the third that the density of ϱ is bounded (see (2.10)). Conversely, denoting by $I' := \{x \in \mathbb{R} : \mathrm{dist}(x, I^c) \geqslant \eta\}$, we get the lower bound

$$
\mu(I) \geqslant \int f_{I',\eta}(\lambda) \, \mu(\mathrm{d}\lambda) \geqslant \int f_{I',\eta}(\lambda) \, \varrho(\mathrm{d}\lambda) + O(N^{-1+2\varepsilon}) \geqslant \varrho(I) + O(N^{-1+2\varepsilon})
$$

with high probability. Since $\varepsilon > 0$ was arbitrary, we conclude that for any $I \subset [-3,3]$ we have $\hat{\mu}(I) = O_\prec(N^{-1})$.

Finally, in order to extend the result to arbitrary $I \subset \mathbb{R}$, we note that we have proved that $\hat{\mu}([-2,2]) = O_\prec(N^{-1})$. Since μ is a probability measure and $\varrho([-2,2]) = 1$, we therefore deduce that $\mu(\mathbb{R} \setminus [-2,2]) \prec N^{-1}$. The claim of Theorem 2.8 now easily follows for arbitrary $I \subset \mathbb{R}$ by splitting

$$
\mu(I) = \mu(I \cap [-2,2]) + \mu(I \cap (\mathbb{R} \setminus [-2,2])) = \mu(I \cap [-2,2]) + O_\prec(N^{-1}) = \varrho(I) + O_\prec(N^{-1}),
$$

where in the last step we used $\hat{\mu}(I) = O_\prec(N^{-1})$ for $I \subset [-2,2]$. (Note that this estimate is in fact simultaneous for all $I \subset \mathbb{R}$, i.e. for any fixed $\varepsilon > 0$ we have with high probability $\hat{\mu}(I) = O(N^{-1+\varepsilon})$ for all $I \subset \mathbb{R}$.) This concludes the proof of Theorem 2.8.

9. Eigenvalue rigidity: proof of Theorem 2.9

The first key input of the proof is the following estimate of the norm of H. (Note that the exponent of $N^{-2/3}$ is optimal.)

Proposition 9.1. *For a Wigner matrix H we have $\|H\| \leqslant 2 + O_\prec(N^{-2/3})$.*

The main tool in the proof of Proposition 9.1 is the following improved version of the local semicircle law (2.14) outside of the spectrum.

Lemma 9.2 (Improved local semicircle law outside of the spectrum). *For a Wigner matrix H we have*

$$s(z) = m(z) + O_\prec\left(\frac{1}{\sqrt{\kappa + \eta}}\left(\frac{\mathrm{Im}\, m}{N\eta} + \frac{1}{(N\eta)^2}\right)\right)$$

uniformly for $z \in \mathbf{S}$.

Proof. We use (6.6) with the a priori bound $\Theta \prec (N\eta)^{-1}$ from (2.14), which yields

$$\Theta \prec \frac{1}{\sqrt{\kappa + \eta}} \Phi_1^2,$$

where Φ_σ was defined in (6.3). The claim follows. $\qquad\square$

Note that the bound from Lemma 9.2 is better than (2.14) when E is sufficiently far outside of the spectrum, for large enough κ and small enough η, since in that case $\mathrm{Im}\, m$ is small by (3.3).

Proof of Proposition 9.1. For definiteness, we prove that the largest eigenvalue λ_1 of H satisfies $\lambda_1 \leqslant 2 + O_\prec(N^{-2/3})$; the smallest eigenvalue λ_N is handled similarly. From the Füredi-Komlós argument in Theorem G.1, we know that $\lambda_1 \leqslant 3$ with high probability. It therefore remains to show that, for any fixed $\varepsilon > 0$, there is no eigenvalue of H in

$$I := \left[2 + N^{-2/3+4\varepsilon}, 3\right] \tag{9.1}$$

with high probability. We do this using Lemma 9.2. The idea of the proof is to choose, for each $E \in I$, a scale $\eta(E)$ such that $\mathrm{Im}\, s(E + i\eta(E)) \leqslant \frac{N^{-\varepsilon}}{N\eta(E)}$ with high probability; such an estimate will be possible thanks to the improved bound from Lemma 9.2. (Note that the estimate (2.14) can never produce such a bound.) We shall then conclude the argument by noting that such an upper bound is strong enough to rule out the existence of an eigenvalue at E.

By an argument analogous to Remark 2.7, we find from Lemma 9.2 and (3.3) that, with high probability,

$$|s(z) - m(z)| \leqslant N^\varepsilon\left(\frac{\eta}{\kappa}\frac{1}{N\eta} + \frac{1}{\sqrt{\kappa}}\frac{1}{(N\eta)^2}\right) \tag{9.2}$$

for all $z \in \mathbf{S}$ satisfying $E \geqslant 2$. For each $E \in I$, we define

$$z(E) := E + i\eta(E), \qquad \eta(E) := N^{-1/2-\varepsilon}\kappa(E)^{1/4}. \tag{9.3}$$

Using (3.3), we find for $E \in I$ that

$$\mathrm{Im}\, m(z(E)) \leqslant \frac{\eta(E)}{\sqrt{\kappa(E)}} \leqslant \frac{N^{-\varepsilon}}{N\eta(E)}. \tag{9.4}$$

Moreover, from (9.2) we find that, with high probability,

$$\left|s(z(E)) - m(z(E))\right| \leqslant \frac{2N^{-\varepsilon}}{N\eta(E)} \tag{9.5}$$

for all $E \in I$. From (9.4) and (9.5) we conclude that, with high probability,

$$\operatorname{Im} s(z(E)) \leqslant \frac{3N^{-\varepsilon}}{N\eta(E)} \tag{9.6}$$

for all $E \in I$.

Now suppose that there is an eigenvalue, say λ_i, of H in I. Then we find

$$\operatorname{Im} s(z(\lambda_i)) = \frac{1}{N}\sum_j \frac{\eta(\lambda_i)}{(\lambda_j - \lambda_i)^2 + \eta(\lambda_i)^2} \geqslant \frac{1}{N\eta(\lambda_i)}. \tag{9.7}$$

Since (9.6) and (9.7) with $E = \lambda_i \in I$ are mutually exclusive, we conclude that, with high probability, there is no eigenvalue of H in I. Since $\varepsilon > 0$ was arbitrary, the claim follows. $\qquad\square$

Armed with Proposition 9.1, we may now complete the proof of Theorem 2.9. We only consider $i \leqslant N/2$; the indices $i > N/2$ are dealt with analogously. From Theorem 2.8 we get $\mu([-1,\infty)) \geqslant 1/2$ with high probability, which implies that $\lambda_i \geqslant -1$ for all $i \leqslant N/2$ with high probability; we shall use this fact tacitly in the following.

Fix $\varepsilon > 0$. We consider two cases. Suppose first that

$$\gamma_i, \lambda_i \in \left[2 - N^{-2/3+2\varepsilon}, \infty\right). \tag{9.8}$$

Then from (2.19) we find $i \leqslant CN^{3\varepsilon}$. Moreover, by Proposition 9.1, we have $|\lambda_i - 2| \leqslant N^{-2/3+2\varepsilon}$ with high probability. Since $\gamma_i \in [2 - N^{-2/3+2\varepsilon}, 2]$, we therefore deduce that

$$|\lambda_i - \gamma_i| \leqslant 2N^{-2/3+2\varepsilon} \leqslant CN^{3\varepsilon}N^{-2/3}i^{-1/3}. \tag{9.9}$$

Conversely, suppose that (9.8) does not hold. We define the function

$$f(E) := \varrho([E,\infty)).$$

Then from the definitions (2.4) and (2.18), we find

$$f(\gamma_i) + \frac{1}{2N} = \frac{i}{N} = \mu([\lambda_i,\infty)) = f(\lambda_i) + O\left(\frac{N^\varepsilon}{N}\right) \tag{9.10}$$

with high probability, where in the last step we used Theorem 2.8. By definition of f, we have

$$f(\gamma_i) \vee f(\lambda_i) \geqslant c(N^{-2/3+2\varepsilon})^{3/2} \geqslant N^\varepsilon \frac{N^\varepsilon}{N}.$$

From (9.10) we therefore get

$$f(\gamma_i) = f(\lambda_i)(1 + O(N^{-\varepsilon}))$$

with high probability. Since $f(\lambda) \asymp (2-\lambda)^{3/2}$, we deduce that $2 - \lambda_i \asymp 2 - \gamma_i$ with high probability. Moreover, since $f'(\lambda) \asymp (2-\lambda)^{1/2}$, we deduce that $f'(\lambda_i) \asymp f'(\gamma_i)$ with high probability, and hence

43

that $f'(\lambda) \asymp f'(\gamma_i)$ with high probability for any λ between λ_i and γ_i. Using the mean value theorem, (2.19), and (9.10), we therefore find

$$|\lambda_i - \gamma_i| \; \asymp \; \frac{|f(\lambda_i) - f(\gamma_i)|}{|f'(\gamma_i)|} \; \leqslant \; C \frac{N^\varepsilon}{N} \left(\frac{i}{N} \right)^{-1/3} \; = \; CN^\varepsilon N^{-2/3} i^{-1/3} \tag{9.11}$$

with high probability.

From the conclusions (9.9) and (9.11) for both cases, we conclude that $|\lambda_i - \gamma_i| \leqslant CN^{3\varepsilon} N^{-2/3} i^{-1/3}$ with high probability, for all $i \leqslant N/2$. Since $\varepsilon > 0$ was arbitrary, the proof of Theorem 2.9 is complete.

10. Extension of the spectral domain in the local law

The restriction of z to $\mathbf{S}$ in 2.6 is natural in the light of the requirement $\Psi \leqslant C$, but is in fact not necessary. Indeed, once Theorem 2.6 has been established, it is not too hard to extend it to a larger subset of $\mathbb{C}_+$, with sharper (in fact optimal) error bounds outside of the spectrum. (In fact, Theorem 2.6 may be extended to the whole upper-half plane $\mathbb{C}_+$ with optimal error bounds throughout. Since the region $\eta > \tau^{-1}$ is of limited practical interest, we shall not deal with it.)

10.1. Extension to all z down to the real axis. We begin by noting that the lower bound on η in (2.13) may be omitted.

Theorem 10.1. *Theorem 2.6 remains valid for $\mathbf{S}$ replaced with*

$$\left\{ E + \mathrm{i}\eta : |E| \leqslant \tau^{-1} , \, 0 < \eta \leqslant \tau^{-1} \right\}.$$

For an application of Theorem 10.1, see Section 11, where it is used to derive a simple universality result for the local eigenvalue statistics of Wigner matrices.

The rest of this subsection is devoted to the proof of Theorem 10.1. The key observation is the following simple deterministic monotonicity result. Define

$$\Gamma(z) \; := \; \max_{i,j} |G_{ij}(z)| . \tag{10.1}$$

Lemma 10.2. *For any $M > 1$ and $z \in \mathbb{C}_+$ we have $\Gamma(E + \mathrm{i}\eta/M) \leqslant M\Gamma(E + \mathrm{i}\eta)$.*

Proof. Fix $E \in \mathbb{R}$ and write $\Gamma(\eta) = \Gamma(E + \mathrm{i}\eta)$. For sufficiently small h, using the resolvent identity, the Cauchy-Schwarz inequality, and (3.6), we get

$$|\Gamma(\eta + h) - \Gamma(\eta)| \; \leqslant \; \max_{i,j} |G_{ij}(E + \mathrm{i}(\eta + h)) - G_{ij}(E + \mathrm{i}\eta)|$$

$$\leqslant \; |h| \max_{i,j} \sum_k |G_{ik}(E + \mathrm{i}(\eta + h)) G_{kj}(E + \mathrm{i}\eta)| \; \leqslant \; |h| \sqrt{\frac{\Gamma(\eta + h)\Gamma(\eta)}{(\eta + h)\eta}} .$$

Thus, Γ is locally Lipschitz continuous, and its almost everywhere defined derivative satisfies

$$\left| \frac{\mathrm{d}\Gamma}{\mathrm{d}\eta} \right| \; \leqslant \; \frac{\Gamma}{\eta} .$$

This implies $\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\Gamma(\eta)) \geqslant 0$ and therefore $\Gamma(\eta/M) \leqslant M\Gamma(\eta)$ as claimed. $\qquad\square$

In order to prove Theorem 10.1, we have to prove (2.14) and (2.15) for z satisfying $|E| \leqslant \tau^{-1}$ and $0 < \eta < N^{-1+\tau}$. Using that $|m(z)| \leqslant C$ by (3.2), we find

$$|G_{ij}(z) - m(z)\delta_{ij}| \leqslant \Gamma(z) + C \leqslant \frac{N^\tau}{N\eta}\Gamma(E + iN^{-1+\tau}) + C \prec \frac{N^\tau}{N\eta} \leqslant N^{2\tau}\Psi(z),$$

where in the second step we used Lemma 10.2, in the third step we used (2.15) with spectral parameter $E + iN^{-1+\tau} \in \mathbf{S}$, and in the last step we used that $\eta < N^{-1+\tau}$. Since $\tau > 0$ was arbitrary, we obtain

$$|G_{ij}(z) - m(z)\delta_{ij}| \prec \Psi(z),$$

which concludes the proof of (2.15) for $\eta < N^{-1+\tau}$. The proof of (2.14) is similar. This concludes the proof of Theorem 10.1.

10.2. Local law outside of the spectrum. Next, we extend Theorem 2.6 to all E outside of the spectrum, with optimal error bounds. Here, "outside of the spectrum" means that the distance from E to the limiting spectrum $[-2, 2]$ is more than $N^{-2/3}$, the scale on which the extreme eigenvalues of H fluctuate. Recall the definition of κ from (3.1).

Theorem 10.3. *Let H be a Wigner matrix. Fix $\tau > 0$ and define the domain*

$$\mathbf{S}^o \equiv \mathbf{S}_N^o(\tau) := \left\{ E + i\eta : |E| \geqslant 2 + N^{-2/3+\tau}, \eta > 0 \right\}.$$

Then we have

$$s(z) = m(z) + O_\prec\left(\frac{1}{N} \frac{1}{(\kappa + \eta) + (\kappa + \eta)^2} \right) \tag{10.2}$$

and

$$G_{ij}(z) = m(z)\delta_{ij} + O_\prec\left(\frac{1}{\sqrt{N}} \frac{1}{(\kappa + \eta)^{1/4} + (\kappa + \eta)^2} \right) \tag{10.3}$$

uniformly for $i, j = 1, \ldots, N$ and $z \in \mathbf{S}^o$.

Results of this type (typically in the stronger guise of *isotropic local laws*; see Section 12.1) are very useful for instance in the study of spiked random matrix models. See [10, 54, 55, 14] for more details.

The rest of this subsection is devoted to the proof of Theorem 10.3. We begin with (10.2), which is an easy consequence of the rigidity result from Theorem 2.9. Let $\varepsilon \in (0, \tau/2)$. From Theorem 2.9 we get, with high probability,

$$|\lambda_i - \gamma_i| \leqslant N^\varepsilon N^{-2/3}\big(i \wedge (N + 1 - i)\big)^{-1/3} \tag{10.4}$$

for all $i = 1, \ldots, N$. Next, it is convenient to define a slightly modified version $\tilde\gamma_i$ of the typical eigenvalue location γ_i from (2.18) through $N \int_{\tilde\gamma_i}^2 \varrho(\mathrm{d}x) = i$. With the convention $\tilde\gamma_0 := 2$, we may write

$$s(z) - m(z) = \sum_{i=1}^N \int_{\tilde\gamma_i}^{\tilde\gamma_{i-1}} \varrho(\mathrm{d}x)\left(\frac{1}{\lambda_i - z} - \frac{1}{x - z} \right).$$

Using (10.4), we find for $x \in [\tilde\gamma_i, \tilde\gamma_{i-1}]$ that

$$|\lambda_i - \gamma_i| + |x - \gamma_i| \leqslant 2N^\varepsilon N^{-2/3}\big(i \wedge (N + 1 - i)\big)^{-1/3}$$

45

with high probability. Since $|z - \gamma_i| \geqslant N^{-2/3+\tau}$ for all i and $\varepsilon > 0$ can be made arbitrarily small, we therefore get

$$|s(z) - m(z)| \prec \frac{1}{N} \sum_{i=1}^{N} N^{-2/3} \big(i \wedge (N+1-i)\big)^{-1/3} \frac{1}{|\gamma_i - z|^2}. \tag{10.5}$$

Now obtaining (10.2) is an elementary exercise in estimating the right-hand side of (10.5), using (2.19) (and its analogue for $i \geqslant N/2$); we omit the details.

What remains is the proof of (10.3). Unlike (10.2), the rigidity estimate from Theorem 2.9 is clearly not sufficient, since we need to control individual entries of G. We note first that, by polarization, it suffices to prove

$$\big|\langle \mathbf{v}, G(z)\mathbf{v} \rangle - m(z)\big| \prec \frac{1}{\sqrt{N}} \frac{1}{(\kappa + \eta)^{1/4} + (\kappa + \eta)^2} \tag{10.6}$$

for all $z \in \mathbf{S}^o$ and for all unit vectors $\mathbf{v}$ having at most two nonzero components[2]. In the rest of the proof, $\mathbf{v}$ denotes such a vector.

We split the proof into two cases: $\kappa + \eta \leqslant 1$ and $\kappa + \eta > 1$.

Case 1: $\kappa + \eta \leqslant 1$. Define $\eta_0 := N^{-1/2}\kappa^{1/4}$. By definition of the domain $\mathbf{S}^o$, we have $\eta_0 \leqslant \kappa/2$. Using (2.15) in Theorem 2.6 and (3.3), we find that (10.6) holds if $\eta \geqslant \eta_0$. For the following we therefore take

$$0 < \eta \leqslant \eta_0 \leqslant \kappa/2. \tag{10.7}$$

We proceed by comparison using the two spectral parameters

$$z := E + \mathrm{i}\eta, \qquad z_0 := E + \mathrm{i}\eta_0.$$

Since (10.6) holds at z_0 by Theorem 2.6, it is enough to prove the estimates

$$\big|m(z) - m(z_0)\big| \leqslant CN^{-1/2}\kappa^{-1/4} \tag{10.8}$$

and

$$\big|\langle \mathbf{v}, G(z)\mathbf{v} \rangle - \langle \mathbf{v}, G(z_0)\mathbf{v} \rangle\big| \prec N^{-1/2}\kappa^{-1/4}. \tag{10.9}$$

We start with (10.8). From the definition (2.6) and the square root decay of the density of ϱ near the spectral edges ± 2, it is not hard to derive the bound $m'(z) \leqslant C\kappa^{-1/2}$ for $z \in \mathbf{S}^o$ satisfying $\eta \leqslant \kappa \leqslant 2$. Therefore we get

$$\big|m(z) - m(z_0)\big| \leqslant C\kappa^{-1/2}\eta_0 = CN^{-1/2}\kappa^{-1/4},$$

which is (10.8).

What remains is to prove (10.9) under (10.7). By Theorem 2.9 and (10.7) we have $|E| \geqslant \|H\| + \eta_0$ with high probability. Thus we get

$$\mathrm{Im}\langle \mathbf{v}, G(z)\mathbf{v} \rangle = \sum_i \frac{|\langle \mathbf{v}, \mathbf{u}_i \rangle|^2 \eta}{(E - \lambda_i)^2 + \eta^2} \leqslant 2 \sum_i \frac{|\langle \mathbf{v}, \mathbf{u}_i \rangle|^2 \eta_0}{(E - \lambda_i)^2 + \eta_0^2} = 2\,\mathrm{Im}\langle \mathbf{v}, G(z_0)\mathbf{v} \rangle \prec N^{-1/2}\kappa^{-1/4} \tag{10.10}$$

---

[2]In fact, (10.6) is valid for arbitrary unit vectors $\mathbf{v}$; see Section 12.1.

by Theorem 2.6 and (3.3) at z_0.

Finally, we estimate the real part of the error in (10.9) using

$$
\begin{aligned}
\left|\mathrm{Re}\langle\mathbf{v},G(z)\mathbf{v}\rangle - \mathrm{Re}\langle\mathbf{v},G(z_0)\mathbf{v}\rangle\right| &= \sum_i \frac{(E-\lambda_i)(\eta_0^2-\eta^2)|\langle\mathbf{u}_i,\mathbf{v}\rangle|^2}{\left((E-\lambda_i)^2+\eta^2\right)\left((E-\lambda_i)^2+\eta_0^2\right)} \\
&\leqslant \frac{\eta_0}{E-\lambda_1}\sum_i \frac{\eta_0|\langle\mathbf{u}_i,\mathbf{v}\rangle|^2}{(E-\lambda_i)^2+\eta_0^2} \leqslant \mathrm{Im}\langle\mathbf{v},G(z_0)\mathbf{v}\rangle \quad (10.11)
\end{aligned}
$$

with high probability, where in the last step we used that $|E| \geqslant \|H\| + \eta_0$ with high probability. Combining (10.10) and (10.11) completes the proof of (10.9), and hence of (10.6) for the case $\kappa + \eta \leqslant 1$.

Case 2: $\kappa + \eta > 1$. Define the random signed measure

$$
\tilde{\mu} := \sum_{i=1}^N \langle\mathbf{v},\mathbf{u}_i\rangle\langle\mathbf{u}_i,\mathbf{v}\rangle\delta_{\lambda_i} - \varrho
$$

with Stieltjes transfrom

$$
\tilde{m}(z) := \int \frac{\tilde{\mu}(\mathrm{d}x)}{x-z} = \langle\mathbf{v},G(z)\mathbf{v}\rangle - m(z)\,.
$$

The basic idea of the proof is to apply the Helffer-Sjöstrand formula from Proposition C.1 to the function

$$
f_z(x) := \frac{1}{x-z} + \frac{1}{z}\,.
$$

To that end, we choose a smooth compactly supported cutoff function $\chi \in \mathcal{C}_c^\infty(\mathbb{C};[0,1])$ satisfying $\chi = 1$ in the $1/6$-neighbourhood of $[-2,2]$, $\chi = 0$ outside of the $1/3$-neighbourhood of $[-2,2]$, and $|\partial_{\bar{w}}\chi(w)| \leqslant C$. By Theorem 2.9 we have $\mathrm{supp}\,\tilde{\mu} \subset \{\chi = 1\}$ with high probability. By assumption of Case 2, $\mathrm{dist}(z,[-2,2]) \geqslant 1/2$. Since f_z is holomorphic in $\mathrm{supp}\,\chi$, the Helffer-Sjöstrand formula yields, for $x \in \mathrm{supp}\,\tilde{\mu}$,

$$
f_z(x) = \frac{1}{\pi}\int_{\mathbb{C}} \frac{\partial_{\bar{w}}(f_z(w)\chi(w))}{x-w}\,\mathrm{d}^2w \quad (10.12)
$$

with high probability. Noting that $\int \mathrm{d}\tilde{\mu} = 0$, we may therefore write

$$
\tilde{m}(z) = \int \tilde{\mu}(\mathrm{d}x)\,f_z(x) = \frac{1}{\pi}\int_{\mathbb{C}} f_z(w)\,\partial_{\bar{w}}\chi(w)\,\tilde{m}(w)\,\mathrm{d}^2w \quad (10.13)
$$

with high probability, where in second step we used (10.12) and the fact that f_z is holomorphic away from z. The integral is supported on the set $\mathrm{supp}\,\partial_{\bar{w}}\chi \subset \{w : \mathrm{dist}(w,[-2,2]) \in [1/6,1/3]\}$, on which we have the estimates $|f_z(w)| \leqslant C(\kappa(z)+\eta(z))^{-2}$ and $|\tilde{m}(w)| \prec N^{-1/2}$, as follows from Theorem 2.6 and Case 1. As in Remark 2.7, we may easily obtain the simultaneous bounds $|\tilde{m}(w)| \leqslant N^{\varepsilon-1/2}$ for all $w \in \mathrm{supp}\,\partial_{\bar{w}}$, with high probability, for any fixed $\varepsilon > 0$. Plugging these estimates into (10.13), we get

$$
|\tilde{m}(z)| \prec (\kappa+\eta)^{-2}N^{-1/2}\,,
$$

which is (10.6). This conclude the proof of Case 2, and hence also of Theorem 10.3.

47

11. Local law and comparison arguments

11.1. Overview of the Green function comparison method.
In this section we explain how the local law can be used to compare the local eigenvalue distribution of two random matrix ensembles whose entries are close. Here, the closeness is quantified using *moment matching.*

Definition 11.1. *Let $\mathbb{E}'$ and $\mathbb{E}''$ denote the expectations of two different Wigner ensembles. We say that $\mathbb{E}'$ and $\mathbb{E}''$ match to order n if we have, for all $i, j = 1, \ldots, N$ and $k, l = 0, 1, \ldots$ satisfying $k + l \leqslant n$,*

$$\mathbb{E}'(H_{ij})^k(\overline{H}_{ij})^l \;=\; \mathbb{E}''(H_{ij})^k(\overline{H}_{ij})^l. \tag{11.1}$$

Typically, we choose $\mathbb{E}'$ to be a Gaussian ensemble from Definition 2.3, for which the statistic that we are interested in can be explicitly computed, and $\mathbb{E}''$ to be a general Wigner ensemble. This allows us to obtain unversality results, expressing that if $\mathbb{E}'$ and $\mathbb{E}''$ match to a high enough order then they have the same asymptotic eigenvalue and eigenvector statistics.

Such comparison ideas go back to Lindeberg and his proof of the central limit theorem [57]. They were introduced into random matrix theory in [28] and used to derive a so-called *four-moment theorem* for the local eigenvalue statistics in [73, 74]. The use of Green functions greatly simplifies such comparison techniques in random matrix theory. This was first observed in [43], where the *Green function comparison method* was introduced. In this section we give a simple application of Green function comparison to the local eigenvalue statistics, given in Theorem 11.3 below.

Before explaining the Green function comparison method, it is instructive to recall Lindeberg's original proof of the central limit theorem, which forms the core idea of the Green function comparison method. (For its statement, we do not aim for optimal assumptions as the emphasis is on a clear and simple proof.)

Theorem 11.2 (CLT à la Lindeberg). *Let $n \in \mathbb{N}$, and suppose that $X_1, \ldots, X_n$ are independent real-valued random variables satisfying $\sum_{i=1}^n \mathbb{E}X_i^2 = n$ as well as $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^3 \leqslant K$ for all $1 \leqslant i \leqslant n$ with some constant $K > 0$. Then for any $f \in \mathcal{C}^3(\mathbb{R})$ we have*

$$\mathbb{E}f\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) - \mathbb{E}f(Y) \;=\; O\big(K\|f'''\|_\infty\, n^{-1/2}\big),$$

where Y is a standard normal random variable.

By a standard approximation argument of continuous bounded functions by smooth functions with bounded derivatives, Theorem 11.2 implies the convergence in distribution of $n^{-1/2}(X_1^{(n)} + \cdots + X_n^{(n)})$ taken from a triangular array $(X_i^{(n)})_{1 \leqslant i \leqslant n}$ satisfying the assumptions of Theorem 11.2 for each $n \in \mathbb{N}$.

Proof of Theorem 11.2. Let $(Y_i)_{1 \leqslant i \leqslant n}$ be a family of independent Gaussian random variables, independent of $(X_i)_{1 \leqslant i \leqslant n}$, satisfying $\mathbb{E}Y_i = 0$ and $\mathbb{E}Y_i^2 = \mathbb{E}X_i^2$ for $1 \leqslant i \leqslant n$. For $0 \leqslant \gamma \leqslant n$ define the new family $(Z_i^\gamma)_{1 \leqslant i \leqslant n}$ through

$$Z_i^\gamma \;:=\; \begin{cases} X_i & \text{if } i \leqslant \gamma \\ Y_i & \text{if } i > \gamma. \end{cases}$$

By telescoping we get

$$
\mathbb{E}f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right) - \mathbb{E}f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i\right) = \sum_{\gamma=1}^{n}\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i^{\gamma}\right) - f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i^{\gamma-1}\right)\right]. \qquad (11.2)
$$

Note that the two sums over i on the right-hand side only differ in the summand $i = \gamma$, where we have $Z_\gamma^\gamma = X_\gamma$ and $Z_\gamma^{\gamma-1} = Y_\gamma$. We estimate the difference using a Taylor expansion of order three around the random variable

$$
W_n^\gamma := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{1}(i \neq \gamma)Z_i^\gamma.
$$

Then the expectation on the right-hand side of (11.2) is equal to

$$
\mathbb{E}\left[f(W_n^\gamma + X_\gamma) - f(W_n^\gamma + Y_\gamma)\right] = \mathbb{E}\left[f'(W_n^\gamma)\frac{X_\gamma - Y_\gamma}{n^{1/2}} + \frac{1}{2}f''(W_n^\gamma)\frac{X_\gamma^2 - Y_\gamma^2}{n}\right] + O\left(K\|f'''\|_\infty\, n^{-3/2}\right),
$$

$$(11.3)$$

where we used the assumption $\mathbb{E}|X_i|^3 \leqslant K$ as well as $\mathbb{E}|Y_i|^3 \leqslant 3(\mathbb{E}Y_i^2)^{3/2} = 3(\mathbb{E}X_i^2)^{3/2} \leqslant 3\mathbb{E}|X_i|^3 \leqslant 3K$, by Hölder's inequality. Since X_γ and Y_γ are independent of W_n^γ, and using that the first two moments of X_γ and Y_γ match, we find that the right-hand side of (11.3) is $O\left(K\|f'''\|_\infty\, n^{-3/2}\right)$. Plugging this back into (11.2), we find that (11.2) is $O\left(K\|f'''\|_\infty\, n^{-1/2}\right)$. The claim then follows from the observation that $Y := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i$ is a standard normal random variable. $\qquad \square$

Note that the central limit theorem can be regarded as a universality result: the asymptotic distribution of $n^{-1/2}\sum_{i=1}^{N}X_i$ and $n^{-1/2}\sum_{i=1}^{N}Y_i$ coincide if variables X_i and Y_i have matching first two moments. Lindeberg's proof given above lays down the strategy for the Green function comparison method, which performs a similar comparison for the distribution of Green functions. Suppose that we have two random matrix ensembles $\mathbb{E}'$ and $\mathbb{E}''$ that match to order n (see Definition 11.1). As in Lindeberg's proof, we introduce the trivial independent coupling of these ensembles, where our probability space consists of the independent matrices H' and H'' whose laws have expectations $\mathbb{E}'$ and $\mathbb{E}''$ respectively.

In order to obtain a telescoping sum, we choose some (arbitrary) bijection

$$
\phi : \{(i,j) : 1 \leqslant i \leqslant j \leqslant N\} \to \{1, 2, \ldots, \gamma_N\},
$$

where $\gamma_N := \frac{N(N+1)}{2}$ is the number of independent matrix entries. We then define the matrices $H^0, H^1, \ldots, H^{\gamma_N}$ as the Wigner matrices satisfying

$$
H_{ij}^\gamma := \begin{cases} H_{ij}'' & \text{if } \phi(i,j) \leqslant \gamma \\ H_{ij}' & \text{if } \phi(i,j) > \gamma \end{cases}
$$

for $i \leqslant j$.

Now suppose that f is some statistic of Wigner matrices whose expectation we would like to understand. We telescope

$$
\mathbb{E}''f(H) - \mathbb{E}'f(H) = \mathbb{E}(f(H'') - f(H')) = \sum_{\gamma=1}^{\gamma_N}\mathbb{E}\left(f(H^\gamma) - f(H^{\gamma-1})\right). \qquad (11.4)
$$

49

As in Lindeberg's proof, we estimate each summand using Taylor's theorem. Note that there are $\gamma_N \asymp N^2$ terms, so that, if all derivatives of f were of order one, we would require *four moment matching* instead of *two moment matching* (as in Lindeberg's proof) for the sum to be $o(1)$. The importance of this four moment matching condition for random matrices was first realized in [28], and it was first applied to the question of universality of the local spectral statistics in [74].

It clear therefore that the key difficulty is to obtain good control on the derivatives of f. Especially for statistics that give information about the local eigenvalue distribution, f can be a very unstable function and controlling its derivatives is a highly nontrivial task. It is here that the local law enter the game, by providing such estimates.

A good choice of statistic f is some combination of Green functions, typically a well-behaved smooth function of some polynomial in the Green function entries. As outlined in Section 2.1 and detailed in Sections 11.2–11.3 below for the example of local eigenvalue statistics, such combinations cover all eigenvalue and eigenvector statistics of interest. Performing a Taylor expansion of each summand in (11.4), we find that we need to differentiate Green functions in the entries of H. By (2.3), this results in a sum over various polynomials in the Green function entries. The local law, for instance through (2.15), provides the required control of such polynomials. Note that in order to obtain information about the distribution of the eigenvalues on small scales, the imaginary part η of the Green function spectral parameter has to be chosen very small, and such bounds on the Green function are a highly nontrivial input. See Lemma 11.4 below for a precise statement and proof of Green function comparison.

The Green function comparison method has been successfully applied to many different problem in random matrix theory: local eigenvalue distribution in the bulk [43, 35], local eigenvalue distribution near the edge [45], distribution of the eigenvectors [53], distribution of outliers in deformed matrix models [54, 55] and even deriving large deviation estimates for the Green function [54, 24, 5]. In some of these applications, such as [54, 55], the Green function comparison method is used not only to obtain an upper bound on the difference between the two ensembles H' and H'' but to actually analyse this difference and hence understand the failure of universality. For instance, the distribution of outliers of deformed Wigner matrices in general depend on the third and fourth moments of the matrix entries, and the Green function comparison method can be used to compute the dependence of this distribution on the moments of H in full generality [54, 55].

In these notes we present the Green function comparison method as applied to the local eigenvalue statistics in the bulk spectrum.

11.2. Local eigenvalue statistics. A central problem in random matrix theory is to characterize the local eigenvalue statistics. It is best addressed using *correlation functions* of the eigenvalue process. To simplify the notation somewhat, we assume that H has an absolutely continuous law[3]. By the Weyl integration formula (see e.g. [3, Sec. 4.1], [72, Sec. 2.6.1], or [63, Sec. 4.1]), we find that the eigenvalues of H also have an absolutely continuous law.

For any $k = 1, 2, \ldots, N$ we define the *k-point intensity measure* of the eigenvalue process $\sum_{i=1}^{N} \delta_{\lambda_i}$ as the measure ρ_k on $\mathbb{R}^k$ defined by

$$\int f(x_1, \ldots, x_k)\, \rho_k(\mathrm{d}x_1, \ldots, \mathrm{d}x_k) := \mathbb{E} \sum_{i_1, \ldots, i_k=1}^{N}{}^{*} f(\lambda_{i_1}, \ldots, \lambda_{i_k})$$

---

[3]The case of general H may be easily recovered by considering $H^\varepsilon := H + \varepsilon V$ instead of H, where V is a GOE or GUE matrix and $\varepsilon > 0$. Then H^ε has an absolutely continuous law for all $\varepsilon > 0$, and the eigenvalue distribution of H coincides with that of H^ε after taking the limit $\varepsilon \downarrow 0$.

for bounded and measurable f, where the star next to the sum means that the sum is restricted to k-tuples $i_1, \ldots, i_k$ of distinct indices. Since the law of the eigenvalues of H is absolutely continuous, we find that $\rho_k(\mathrm{d}x_1, \ldots, \mathrm{d}x_k) = p_k(x_1, \ldots, x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$ has a density p_k, called the k-point correlation function of the eigenvalue process. Formally, p_k can be written as

$$p_k(x_1, \ldots, x_k) \;=\; \mathbb{E} \sum_{i_1, \ldots, i_k = 1}^{N}{}^{*} \delta(x_1 - \lambda_{i_1}) \cdots \delta(x_k - \lambda_{i_k}), \tag{11.5}$$

where δ is the Dirac delta function.

The interpretation of $\rho_k(\mathrm{d}x_1, \ldots, \mathrm{d}x_k)$ is the expected number of k-tuples of eigenvalues lying in the disjoint sets $\mathrm{d}x_1, \ldots, \mathrm{d}x_k$, so that $p_k(x_1, \ldots, x_k)$ is the expected density of k-tuples of eigenvalues near the distinct locations $x_1, \ldots, x_k$. Moreover, it is immediate that p_k is $\frac{N!}{(N-k)!}$ times the k-point marginal of the joint symmetrized probability density of the eigenvalues $\lambda_1, \ldots, \lambda_N$.

In order to see individual eigenvalues, we need to zoom into the spectrum to the scale of the typical eigenvalue spacing. We denote by

$$\varrho_E \;:=\; \frac{\varrho(\mathrm{d}E)}{\mathrm{d}E} \;=\; \frac{1}{2\pi}\sqrt{(4 - E^2)_+}$$

the density of the semicircle distribution at $E \in (-2, 2)$. By the semicircle law, the typical density of eigenvalues near the energy $E \in (-2, 2)$ is $N\varrho_E$, so that the typical eigenvalue spacing is of order $(N\varrho_E)^{-1}$. In order to perform the zooming in, we define the *local k-point intensity measure around the energy* $E \in (-2, 2)$, denoted by $\rho_{k,E}$, as the k-point intensity measure of rescaled eigenvalue process

$$\sum_{i=1}^{N} \delta_{N\varrho_E(\lambda_i - E)}. \tag{11.6}$$

Explicitly,

$$\rho_{k,E}(\mathrm{d}u_1, \ldots, \mathrm{d}u_k) \;=\; \rho_k\!\left(E + \frac{\mathrm{d}u_1}{N\varrho_E}, \ldots, E + \frac{\mathrm{d}u_k}{N\varrho_E}\right).$$

The measure $\rho_{k,E}$ has density

$$p_{k,E}(u_1, \ldots, u_k) \;:=\; \frac{1}{(N\varrho_E)^k}\, p_k\!\left(E + \frac{u_1}{N\varrho_E}, \ldots, E + \frac{u_k}{N\varrho_E}\right), \tag{11.7}$$

the *local k-point correlation function at energy E*. Note that $p_{k,E}(u_1, \ldots, u_k)$ has the interpretation of the expected density of k-tuples of eigenvalues near the distinct locations $u_1, \ldots, u_k$ *on the microscopic scale* where the typical separation of the eigenvalues is 1. We therefore expect $p_{k,E}$ to have a limit as $N \to \infty$, for fixed $E \in (-2, 2)$ and $k \in \mathbb{N}$. Indeed, in a series of pioneering works [48, 29, 60] Gaudin, Dyson, and Mehta proved that if H is the GUE then

$$p_{k,E}(u_1, \ldots, u_k) \;\to\; \det\!\big(K(u_i - u_j)\big)_{i,j=1}^{k} \tag{11.8}$$

weakly as $N \to \infty$, where

$$K(u) \;:=\; \frac{\sin(\pi u)}{\pi u}$$

is the *sine kernel*. Thus, the correlation functions in the limit have a determinantal structure. Generally, a point process whose correlation functions have a determinantal structure as on the

right-hand side of (11.8) is called *determinantal*. We have therefore seen that the rescaled eigenvalue process (11.6) converges as $N \to \infty$ to a determinantal point process with kernel K. A similar formula holds for the GOE with a more complicated kernel K.

Returning to the analogy of the central limit theorem, the limiting behaviour of $p_{k,E}$ for the Gaussian ensembles GUE and GOE can be determined thanks to an explicit computation using the "integrable" nature of the Gaussian distribution. The question of universality of the local eigenvalue statistics is whether (11.8) is true for other Wigner matrices as well.

11.3. Green function comparison for local eigenvalue statistics. We now explain how the Green function comparison method can be used to prove that the asymptotic behaviour of $p_{k,E}$ only depends on the first four moments of the Wigner ensembles.

Theorem 11.3. *Let $\mathbb{E}'$ and $\mathbb{E}''$ be Wigner ensembles with local k-point correlation functions (11.7) denoted by $p'_{k,E}$ and $p''_{k,E}$ respectively. Suppose that $\mathbb{E}'$ and $\mathbb{E}''$ match to order 4. Then for any fixed $k \in \mathbb{N}$, $E \in (-2, 2)$, and $f \in \mathcal{C}^1_c(\mathbb{R}^k)$ there exists a $\xi > 0$ such that*

$$\int (p'_{k,E} - p''_{k,E})(u_1, \ldots, u_k) \, f(u_1, \ldots, u_k) \, \mathrm{d}u_1 \cdots \mathrm{d}u_k \;=\; O(N^{-\xi}). \tag{11.9}$$

The rest of this subsection is devoted to the proof of Theorem 11.3 using Green function comparison. We have to express the local correlation function $p_{k,E}$ in terms of the Green function. To this end, we smear out $p_{k,E}$ using the approximate delta function θ_η from (2.8), which can be written using the Green function as in (2.7). Fix $\varepsilon > 0$ and choose a scale $\eta := N^{-1-\varepsilon}$ slightly smaller than the typical eigenvalue spacing $(N\varrho_E)^{-1}$. Then we define the corresponding microscopic scale $\tilde{\eta} := N\varrho_E \eta = \varrho_E N^{-\varepsilon}$ and the corresponding smeared-out observable

$$f_{\tilde{\eta}}(u_1, \ldots, u_k) \;:=\; \int \mathrm{d}v_1 \cdots \mathrm{d}v_k \, f(v_1, \ldots, v_k) \, \theta_{\tilde{\eta}}(u_1 - v_1) \cdots \theta_{\tilde{\eta}}(u_k - v_k). \tag{11.10}$$

The main work in the proof is to estimate (11.9) with f replaced by $f_{\tilde{\eta}}$. Having done this, going back to f will easily follow provided we have an upper bound on $p'_{k,E}$ and $p''_{k,E}$; such a bound will be obtained from the local semicircle law (see Lemma 11.5 below).

We transfer the smearing-out to the local k-point correlation function using

$$\int p_{k,E}(u_1, \ldots, u_k) \, f_{\tilde{\eta}}(u_1, \ldots, u_k) \, \mathrm{d}u_1 \cdots \mathrm{d}u_k \;=\; \int p_{k,E,\tilde{\eta}}(u_1, \ldots, u_k) \, f(u_1, \ldots, u_k) \, \mathrm{d}u_1 \cdots \mathrm{d}u_k, \tag{11.11}$$

where we defined the smeared-out local k-point correlation function

$$p_{k,E,\tilde{\eta}}(u_1, \ldots, u_k) \;:=\; \int p_{k,E}(v_1, \ldots, v_k) \theta_{\tilde{\eta}}(u_1 - v_1) \cdots \theta_{\tilde{\eta}}(u_1 - v_1) \, \mathrm{d}v_1 \cdots \mathrm{d}v_k. \tag{11.12}$$

Using the macroscopic variables

$$x_i \;:=\; E + \frac{u_i}{N\varrho_E} \tag{11.13}$$

associated with the microscopic variables u_i, we find

$$p_{k,E,\tilde{\eta}}(u_1, \ldots, u_k) \;=\; \frac{1}{(N\varrho_E)^k} \int p_k(y_1, \ldots, y_k) \, \theta_\eta(x_1 - y_1) \cdots \theta_\eta(x_k - y_k) \, \mathrm{d}y_1 \cdots \mathrm{d}y_k$$

$$=\; \frac{1}{(N\varrho_E)^k} \mathbb{E} \sum_{i_1, \ldots, i_k}^{*} \theta_\eta(x_1 - \lambda_{i_1}) \cdots \theta_\eta(x_k - \lambda_{i_k}).$$

52

From now on, to simplify notation, we only consider the case $k = 2$; other k are handled in the same way with more complicated notation. We find

$$
\begin{aligned}
p_{2,E,\tilde{\eta}}(u_1, u_2) &= \frac{1}{(N\varrho_E)^2}\mathbb{E}\sum_{i,j}\theta_\eta(x_1 - \lambda_i)\,\theta_\eta(x_2 - \lambda_j) - \frac{1}{(N\varrho_E)^2}\mathbb{E}\sum_{i}\theta_\eta(x_1 - \lambda_i)\,\theta_\eta(x_2 - \lambda_i) \\
&= \frac{1}{(N\varrho_E)^2}\mathbb{E}\operatorname{Tr}\theta_\eta(x_1 - H)\operatorname{Tr}\theta_\eta(x_2 - H) - \frac{1}{(N\varrho_E)^2}\mathbb{E}\operatorname{Tr}\big(\theta_\eta(x_1 - H)\theta_\eta(x_2 - H)\big) \\
&= \frac{1}{\pi^2(N\varrho_E)^2}\mathbb{E}\operatorname{Tr}\big(\operatorname{Im}G(z_1)\big)\operatorname{Tr}\big(\operatorname{Im}G(z_2)\big) - \frac{1}{\pi^2(N\varrho_E)^2}\mathbb{E}\operatorname{Tr}\big(\operatorname{Im}G(z_1)\operatorname{Im}G(z_2)\big),
\end{aligned}
$$

where $z_i := x_i + i\eta$ for $i = 1, 2$. Here in the last step we used (2.8). This formula is the desired representation of the smeared-out local two-point correlation function in terms of the Green function.

Recalling (11.11) and the definition (11.13), we find that Theorem 11.3 follows immediately from the following two lemmas.

Lemma 11.4 (Green function comparison). *Suppose that $\mathbb{E}'$ and $\mathbb{E}''$ match to order 4. Fix $E \in (-2, 2)$ and $\varepsilon \in (0, 1/20)$. Then there exists a $\xi > 0$ such that for $\eta := N^{-1-\varepsilon}$ we have*

$$
\frac{1}{N^2}(\mathbb{E}' - \mathbb{E}'')\operatorname{Tr}G(z)\operatorname{Tr}G(w) = O(N^{-\xi}) \tag{11.14}
$$

and

$$
\frac{1}{N^2}(\mathbb{E}' - \mathbb{E}'')\operatorname{Tr}\big(G(z)G(w)\big) = O(N^{-\xi}) \tag{11.15}
$$

for any $z, w \in \mathbb{C}$ satisfying $|\operatorname{Im}z| = |\operatorname{Im}w| = \eta$ and $|\operatorname{Re}z - E| + |\operatorname{Re}w - E| \leqslant C/N$.

Lemma 11.5 (Removal of smearing-out). *Fix $E \in (-2, 2)$ and $\varepsilon > 0$. Let $f \in \mathcal{C}_c^1(\mathbb{R}^2)$ and denote by $f_{\tilde{\eta}}$ its smeared-out version from (11.10) with $\tilde{\eta} = \varrho_E N^{-\varepsilon}$. Then there exists a $\xi > 0$ such that*

$$
\int p_{k,E}(u_1, u_2)(f - f_{\tilde{\eta}})(u_1, u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2 = O(N^{-\xi}),
$$

where $p_{k,E}$ is the local two-point correlation function at energy E of a Wigner ensemble.

Proof of Lemma 11.4. We write the left-hand sides of (11.14) and (11.15) as

$$
(\mathbb{E}' - \mathbb{E}'')\frac{1}{N^2}\sum_{i,j}G_{jj}(z)G_{ii}(w), \qquad (\mathbb{E}' - \mathbb{E}'')\frac{1}{N^2}\sum_{i,j}G_{ij}(z)G_{ji}(w) \tag{11.16}
$$

respectively. We only prove (11.14); the proof of (11.15) is analogous. We use the independent coupling (H', H'') from Section 11.1 and the telescopic sum from (11.4), with $f(H) := \frac{1}{N^2}\sum_{i,j}G_{ii}(z)G_{jj}(w)$. For each $\gamma = 1, \ldots, \gamma_N$ we introduce the Hermitian matrix $W^\gamma = (W_{ij}^\gamma)$ through

$$
W_{ij}^\gamma := \begin{cases} H_{ij}'' & \text{if } \phi(i,j) < \gamma \\ 0 & \text{if } \phi(i,j) = \gamma \\ H_{ij}' & \text{if } \phi(i,j) > \gamma \end{cases}
$$

for $i \leqslant j$. Thus, the matrices $H^\gamma - W^\gamma$ and $H^{\gamma-1} - W^\gamma$ have rank at most two, and they are independent of W^γ.

For $\gamma = 1, \ldots, \gamma_N$ we define the Green functions

$$R^\gamma(z) := (W^\gamma - z)^{-1}, \qquad S^\gamma(z) := (H^{\gamma-1} - z)^{-1}, \qquad T^\gamma(z) := (H^\gamma - z)^{-1}.$$

Since $\gamma_N = O(N^2)$, by (11.4) it suffices to prove that for each $\gamma = 1, \ldots, \gamma_N$ we have

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^\gamma(z)T_{jj}^\gamma(w) - \mathbb{E}\frac{1}{N^2}\sum_{i,j} S_{ii}^\gamma(z)S_{jj}^\gamma(w) = O(N^{-2-\xi}). \tag{11.17}$$

To prove (11.17), it suffices to prove

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^\gamma(z)T_{jj}^\gamma(w) = \mathcal{A}(z,w) + O(N^{-2-\xi}), \tag{11.18}$$

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} S_{ii}^\gamma(z)S_{jj}^\gamma(w) = \mathcal{A}(z,w) + O(N^{-2-\xi}), \tag{11.19}$$

for some $\mathcal{A}(z,w) \in \mathbb{C}$. We focus on (11.18). The idea is to perform a resolvent expansion of T^γ around R^γ. Denoting by $\Delta^\gamma := H^\gamma - W^\gamma$, we have

$$T^\gamma = \sum_{\ell=0}^{4} R^\gamma(-\Delta^\gamma R^\gamma)^\ell + T^\gamma(-\Delta^\gamma R^\gamma)^5. \tag{11.20}$$

From now on, we fix $\gamma = 1, \ldots, \gamma_N$ satisfying $\phi(a,b) = \gamma$ with $a \leqslant b$, and consistently omit γ from our notation. In particular,

$$\Delta_{ij} = \delta_{ia}\delta_{jb}H_{ij}'' + (1 - \delta_{ab})\delta_{ib}\delta_{ja}H_{ij}'', \tag{11.21}$$

so that Δ is a bounded-rank Hermitian matrix consisting of entries of H''; its rank is two if $a < b$ and one if $a = b$.

Plugging (11.20) into the left-hand side of (11.18), we get

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^\gamma(z)T_{jj}^\gamma(w) = \mathbb{E}\frac{1}{N^2}\sum_{i,j}\left(\sum_{\ell=0}^{4}\left(R(z)(-\Delta R(z))^\ell\right)_{ii} + \left(T(z)(-\Delta R(z))^5\right)_{ii}\right)$$
$$\times \left(\sum_{\ell=0}^{4}\left(R(w)(-\Delta R(w))^\ell\right)_{jj} + \left(T(w)(-\Delta R(w))^5\right)_{jj}\right). \tag{11.22}$$

We multiply out the parentheses to get a series of terms. We split it them into the *main terms*, which contain a total power of Δ at most four, and *remainder terms*. The main terms are large but their contribution depends only on the first four moments of Δ, i.e. of H'', and they will be put into $\mathcal{A}(z,w)$. The remainder terms are small thanks to the local semicircle laws for T and R and the estimate

$$\|\Delta_{ij}\|_\ell \leqslant C_\ell \mathbf{1}(\{i,j\} = \{a,b\})N^{-1/2}, \tag{11.23}$$

which follows from Definition 2.2 (iii). Explicitly, we define

$$\mathcal{A}(z,w) := \frac{1}{N^2}\sum_{i,j}\sum_{\ell_1,\ell_2 \geqslant 0} \mathbf{1}(\ell_1 + \ell_2 \leqslant 4)\,\mathbb{E}\left[\left(R(z)(-\Delta R(z))^{\ell_1}\right)_{ii}\left(R(w)(-\Delta R(w))^{\ell_2}\right)_{jj}\right]$$

Since Δ is independent of R, we find that $\mathcal{A}(z,w)$ depends on (H', H'') only through W and the first four moments of H''.

Repeating the same argument for (11.19), we find the *same* $\mathcal{A}(z,w)$ because the first four moments of H'' and H' coincide.

What remains, therefore, is to estimate

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^{\gamma}(z)T_{jj}^{\gamma}(w) - \mathcal{A}(z,w)\,, \qquad \mathbb{E}\frac{1}{N^2}\sum_{i,j} S_{ii}^{\gamma}(z)S_{jj}^{\gamma}(w) - \mathcal{A}(z,w)\,. \tag{11.24}$$

We again only concentrate on the first expression. It is given by a sum

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^{\gamma}(z)T_{jj}^{\gamma}(w) - \mathcal{A}(z,w) \;=\; \frac{1}{N^2}\sum_{i,j}\sum_{\alpha=1}^{C}\mathbb{E}X_{ij}(\alpha) \tag{11.25}$$

for some constant C, where $X_{ij}(\alpha)$ lies in the set $\mathcal{M}_{\ell,p}$, defined as the set monomials in the entries of $T(z), R(z), T(w), R(w)$, and Δ, such that the total degree of the entries of Δ is ℓ and total degree of the entries of $T(z), R(z), T(w), R(w)$ is p. Moreover, we immediately find that for every $X_{ij}(\alpha)\in\mathcal{M}_{\ell,p}$ we have $p\leqslant 2\ell$.

We estimate each such monomial by using the entrywise bounds (11.23) and

$$|R_{ij}(\zeta)| + |T_{ij}(\zeta)| \;\prec\; N^{\varepsilon} \tag{11.26}$$

for all i,j and $\zeta\in\{z,w\}$. The estimate (11.26) is the crucial estimate that makes the Green function comparison method work; it is proved using the local semicircle law. Before proving it, we explain how to use it to conclude the proof of Lemma 11.4.

Take a monomial $X\in\mathcal{M}_{\ell,p}$ with $\ell,p\leqslant C$. Using Hölder's inequality, (11.23), and (11.26), we find

$$|\mathbb{E}X| \;\leqslant\; \Big(\max_{i,j}\|\Delta_{ij}\|_{\ell+p}^{\ell}\Big)\max_{i,j}\sum_{\zeta\in\{z,w\}}\Big(\|R_{ij}(\zeta)\|_{\ell+p}^{p} + \|T_{ij}(\zeta)\|_{\ell+p}^{p}\Big)$$

$$\leqslant\; N^{\delta-\ell/2+p\varepsilon} \tag{11.27}$$

for any fixed $\delta > 0$ and large enough N. Here we estimated $\|R_{ij}(\zeta)\|_{p+\ell}\prec N^{\varepsilon}$ (and similarly for $\|T_{ij}(\zeta)\|_{p+\ell}$), by using the estimate $|R_{ij}(\zeta)|\leqslant N^{\varepsilon+\delta_1}$ on an event of probability at least $1 - N^{-D}$ for any large $D > 0$ and small $\delta_1 > 0$, for large enough N (as follows from (11.26)), and the trivial bound $|R_{ij}(\zeta)|\leqslant\eta^{-1}=N^{1+\varepsilon}$ for the complementary low-probability event.

Plugging the estimate (11.27) into (11.25) and recalling that $p\leqslant 2\ell$ and $\ell = 5,\ldots,10$ for all $X_{ij}(\alpha)$, we find

$$\mathbb{E}\frac{1}{N^2}\sum_{i,j} T_{ii}^{\gamma}(z)T_{jj}^{\gamma}(w) - \mathcal{A}(z,w) \;=\; O\big(N^{-5/2+10\varepsilon+\delta}\big)$$

for any $\delta > 0$ and large enough N. A similar bound holds for the second term of (11.24). We conclude that (11.18) and (11.19) hold with $\xi := 1/2 - 10\varepsilon - \delta$.

All that remains to complete the proof of Lemma 11.4 is the estimate (11.26). For definiteness, we suppose that $\zeta = z$ and $\operatorname{Im} z = \eta > 0$. (The case $\operatorname{Im} z = -\eta$ is obtained simply by complex conjugation.) Since $T(z)$ is the Green function of a Wigner matrix, we immediately get from Theorem 10.1 that $|T_{ij}(z)|\prec(N\eta)^{-1}\leqslant CN^{\varepsilon}$. For the analogous bound for $R(z)$ we cannot invoke

Theorem 10.1 because W is not a Wigner matrix. However, W differs from the Wigner matrix $H^\gamma \equiv H$ by Δ, so that we may again perform a resolvent expansion, except that we now expand around T instead of R:

$$R_{ij} = \sum_{\ell=0}^{2} \big(T(\Delta T)^\ell\big)_{ij} + \big(R(\Delta T)^3\big)_{ij}.$$

Using the above bound on the entries of T_{ij} and the trivial bound $|R_{ik}| \leqslant \eta^{-1} \leqslant CN^{1+\varepsilon}$, we therefore get

$$|R_{ij}| \prec \sum_{\ell=0}^{2} N^\varepsilon (N^{-1/2+\varepsilon})^\ell + N^{1+\varepsilon}(N^{-1/2+\varepsilon})^3 \prec N^\varepsilon,$$

as claimed. This concludes the proof of Lemma 11.4. □

Proof of Lemma 11.5. Using (11.7) and (11.5) we write

$$I := \int p_{k,E}(u_1, u_2)(f - f_{\tilde\eta})(u_1, u_2)\, \mathrm{d}u_1\, \mathrm{d}u_2 = \mathbb{E}\sum_{i \neq j}(f - f_{\tilde\eta})\big(N\varrho_E(\lambda_i - E), N\varrho_E(\lambda_j - E)\big). \tag{11.28}$$

We choose $\xi \in (0, \varepsilon/3)$ and set $\zeta := N^\xi$. We split

$$\begin{aligned}
(f - f_{\tilde\eta})&(u_1, u_2) \\
&= (f - f_{\tilde\eta})(u_1, u_2)\, \mathbf{1}(|u_1| \leqslant \zeta)\, \mathbf{1}(|u_2| \leqslant \zeta) \\
&\quad - f_{\tilde\eta}(u_1, u_2)\Big(\mathbf{1}(|u_1| \leqslant \zeta)\, \mathbf{1}(|u_2| > \zeta) + \mathbf{1}(|u_1| > \zeta)\, \mathbf{1}(|u_2| \leqslant \zeta) + \mathbf{1}(|u_1| > \zeta)\, \mathbf{1}(|u_2| > \zeta)\Big) \\
&= O\big(\tilde\eta \log \tilde\eta^{-1}\, \mathbf{1}(|u_1| \leqslant \zeta)\, \mathbf{1}(|u_2| \leqslant \zeta)\big) + O\left(\frac{\tilde\eta\, \mathbf{1}(|u_1| \leqslant \zeta)\, \mathbf{1}(|u_2| > \zeta)}{u_2^2 + \zeta^2}\right) \\
&\quad + O\left(\frac{\tilde\eta\, \mathbf{1}(|u_1| > \zeta)\, \mathbf{1}(|u_2| \leqslant \zeta)}{u_1^2 + \zeta^2}\right) + O\left(\frac{\tilde\eta^2\, \mathbf{1}(|u_1| > \zeta)\, \mathbf{1}(|u_2| > \zeta)}{(u_2^2 + \zeta^2)(u_1^2 + \zeta^2)}\right) \tag{11.29}
\end{aligned}$$

where in the second step we used that $\|f - f_{\tilde\eta}\|_\infty = O(\tilde\eta \log \tilde\eta^{-1})$ since f is bounded and Lipschitz, and that f is bounded with compact support.

Plugging (11.29) into (11.28) yields the decomposition $I = I_1 + I_2 + I_3 + I_4$ in self-explanatory notation. We estimate these terms using the eigenvalue rigidity estimates from Theorem 2.9, which imply that the event

$$\Xi := \big\{N\varrho_E|\lambda_i - E| \geqslant c\, |i - i_0| - N^{\xi/2} \text{ for all } i = 1, \ldots, N\big\} \tag{11.30}$$

has probability at least $1 - N^{-3}$, for some constant $c > 0$ and $i_0 \in \{1, \ldots, N\}$. We decompose the expectation on the right-hand side of (11.28) using $\mathbf{1}(\Xi) + \mathbf{1}(\Xi^c)$. The contribution of $\mathbf{1}(\Xi^c)$ is easily bounded by $O(N^{-1})$ because f is bounded.

In the following estimates it therefore suffices to estimate the expectation over the event Ξ. We find

$$I_1 \leqslant O(N^{-1}) + O(\zeta^2 \tilde\eta \log \tilde\eta^{-1}) = O(N^{-\xi}).$$

Moreover, we find

$$I_2 = O(N^{-1}) + O\left(\sum_i \frac{\zeta\tilde\eta}{(i - i_0)^2 + \zeta^2}\right) = O(N^{-\xi}).$$

The terms I_3 and I_4 are estimated analogously. This concludes the proof. □

We conclude this subsection by remarking that the moment matching assumptions from Definition 11.1 can be somewhat relaxed. Indeed, it is easy to check that Theorem 11.3 remains correct, with the same proof up to trivial adjustments, provided that, instead of the assumption that (11.1) hold for $k + l \leqslant 4$, we assume that

$$\mathbb{E}'(H_{ij})^k(\overline{H}_{ij})^l = \mathbb{E}''(H_{ij})^k(\overline{H}_{ij})^l + O\big(N^{(k+l)/2 - 2 + \delta_{ij} - \delta}\big)$$

for $k + l \leqslant 4$ for some constant $\delta > 0$. We leave the details of this extension to the interested reader.

11.4. Green function comparison for the spectral edge and eigenvectors. We conclude this section by sketching how the Green function comparison method can also be applied to the local eigenvalue statistics near the edge and to the distribution of the eigenvectors.

For the distribution of the extreme eigenvalues, we focus on the top edge 2. Based on Theorem 2.9, we expect the typical separation and scale of fluctuations of eigenvalues near the edge 2 to be of order $N^{-2/3}$. For any fixed $\ell \in \mathbb{N}$, we define the random vector

$$\mathbf{q}_\ell := N^{2/3}(\lambda_1 - 2, \lambda_2 - 2, \ldots, \lambda_\ell - 2).$$

For the case of the GUE, the asymptotic behaviour of $\mathbf{q}_\ell$ can be explicitly computed, and we have

$$\mathbf{q}_\ell \xrightarrow{\mathrm{d}} \mathbf{q}_{\ell,\infty} \tag{11.31}$$

as $N \to \infty$, where $\mathbf{q}_{\ell,\infty}$ is the *GUE-Tracy-Widom-Airy vector*. An analogous result holds for the GOE, with the *GOE-Tracy-Widom-Airy* random vector as the limit.

The Tracy-Widom-Airy distributions appear in several contexts in probability theory, usually at the boundary between the two phases of weakly versus strongly coupled components in a system. For brevity, here we only focus on the GUE-Tracy-Widom-Airy vector for the scalar case $\ell = 1$. More general descriptions can be found in [76, 77, 3, 66]. Let F_2 denote the cumulative distribution of the random variable $q_{1,\infty}$ from (11.31). Then F_2 is given by the Fredholm determinant

$$F_2(s) = \det(I - K_{\mathrm{Ai}})_{L^2([s,+\infty))},$$

where

$$K_{\mathrm{Ai}}(x, y) := \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x - y} = \int_0^{+\infty} \mathrm{Ai}(x + t)\,\mathrm{Ai}(y + t)\,\mathrm{d}t$$

is the *Airy kernel* and Ai denotes the *Airy function*. Alternatively, F_2 is also given by

$$F_2(s) = \exp\left(-\int_s^{+\infty} (x - s)q^2(x)\mathrm{d}x\right),$$

where q is the solution of the *Painlevé II* differential equation $q''(x) = xq(x) + 2q^3(x)$ such that $q(x)/\mathrm{Ai}(x)$ tends to 1 as $x \to +\infty$.

The question of *edge universality* is whether (11.31) holds for arbitrary Wigner matrices. It has seen much attention over the past 20 years, starting with the pioneering work [68, 69, 70, 64], where it was proved that (11.31) holds for any Wigner matrix whose entries have a symmetric distribution that matches the GUE to order two.

The Green function comparison method can be used to establish (11.31) for arbitrary Wigner matrices that match the GUE to order two. This was proved in [45], and a similar result holds

for real symmetric Wigner matrices and the GOE. The basic idea of the proof is similar to that of Theorem 11.3. Note, however, that instead of matching to order *four* we only need to assume matching to order *two*.

The matching to order four arose naturally in the proof of Theorem 11.3, as the number N^2 of terms in the telescopic sum (11.4) was compensated by the size $N^{-5/2}$ of each term. The factor $N^{-5/2}$ arose from the four-moment matching assumption and the bound of order 1 on the entries of the Green function. If we only assume matching to order two, a similar argument yields the bound $N^{-3/2}$ for each term of (11.4), which is clearly far from enough. The key observation leading to the solution to this problem is that near the edge the entries of the Green function are in fact small. Indeed, in order to resolve the individual eigenvalues near the edge we take $\eta = N^{-2/3-\varepsilon}$ instead of $N^{-1-\varepsilon}$ as in Lemma 11.4. For $|E - 2| \leqslant N^{-2/3+\varepsilon}$ and $z = E + i\eta$, we find from the local semicircle law and the elementary bound (3.3) that

$$|G_{ij}(z)| \prec \delta_{ij} + N^{-1/3+\varepsilon} .$$

Thus, off-diagonal Green function entries provide additional smallness near the spectral edges. Exploiting this mechanism requires a careful analysis of the various monomials generated in expansions of the type (11.22), classifying them in terms of the number of off-diagonal Green function entries. We refer to [45] for the full details.

Finally, the Green function comparison method can also be used to analyse the distribution of eigenvectors: in [53], it was used to prove a universality result for the distribution of individual eigenvectors, stating that the distribution of eigenvectors of Wigner matrices is universal near the spectral edges, as well as in the bulk under an additional four-moment matching condition. Subsequently, this latter result was also obtained in [75] using a different comparison method. Recently, in [22], the results of [53] were extended to the bulk eigenvectors by combining the comparison results of [53] with a novel eigenvector moment flow.

We now summarize the Green function comparison method for eigenvectors from [53]. In the case of the GUE, the distribution of eigenvectors is trivial: the law of the GUE is invariant under conjugation by unitary matrices, from which we deduce that the matrix of eigenvectors of the GUE is uniformly distributed on the group $SU(N)$. This implies that any eigenvector $\mathbf{u}_i = (u_i(k))_{k=1}^N$ is uniformly distributed on the unit sphere of $\mathbb{C}^N$, and the random variables

$$X_{ik} := \sqrt{N} u_i(k) \tag{11.32}$$

are asymptotically independent standard complex normals.

Using Green function comparison, one can prove, just like for the eigenvalues, a four-moment theorem in the bulk and a two-moment theorem near the edge. Indeed, it was proved in [53] that the distributions of a bounded number of eigenvector entries $(X_{i_1 k_1}, \ldots, X_{i_\ell k_\ell})$ for two different Wigner ensembles E' and E'' coincide if E' and E'' match to order four. Moreover, if E' and E'' match to order two, then the distributions of $(X_{i_1 k_1}, \ldots, X_{i_\ell k_\ell})$ coincide near the edge, i.e. provided that the indices $i_1, \ldots, i_\ell$ are bounded by N^ε for some small enough $\varepsilon > 0$. In fact, using more refined techniques, it was proved in [14, Section 7] that the distributions of $(X_{i_1 k_1}, \ldots, X_{i_\ell k_\ell})$ for two ensembles matching to order two coincide provided that $i_1, \ldots, i_\ell \leqslant N^{1-\varepsilon}$ for any fixed $\varepsilon > 0$.

The starting point of these proofs is to express the eigenvector components using the Green function. A helpful observation is that eigenvectors are only defined up to a global phase. To remove this ambiguity, it is therefore better to analyse the components of the rescaled *spectral*

projections $N\mathbf{u}_i\mathbf{u}_i^*$. They can be expressed, at least in an averaged form, in terms of the Green function using

$$\sum_j \frac{\eta/\pi}{(E - \lambda_j)^2 + \eta^2}\, N\mathbf{u}_j\mathbf{u}_j^* \;=\; \frac{N}{2\pi\mathrm{i}}\big(G(E + \mathrm{i}\eta) - G(E - \mathrm{i}\eta)\big). \tag{11.33}$$

In order to extract $N\mathbf{u}_i\mathbf{u}_i$ from the identity (11.33), we have to choose η to be much smaller than the typical eigenvalue separation near the eigenvalues λ_i, and integrate (11.33) in E over a random interval I centred around λ_i, whose length is much larger than η, such that all eigenvalues λ_j, $j \neq i$, are far from I on the scale η. Such an interval can be showed to exist with high probability [53]. Then the Green function comparison may be applied to integrals of entries of the right-hand side of (11.33). We refer to [53, 14] for the details.

12. Outlook: some further developments

In this concluding section we survey further developments of local laws beyond the simple case of Wigner matrices presented in these notes.

12.1. Isotropic laws. The local law from Theorem 2.6 states roughly that the matrices G and mI are close with high probability for $z \in \mathbf{S}$. As the dimension N of these matrices is very large, which notion of closeness to use is a nontrivial question. Theorem 2.6 establishes closeness in the sense of individual matrix entries. Other canonical (and stronger) notions are closeness in the *weak operator sense*, *strong operator sense*, and *norm sense*, which entail controlling

$$\langle \mathbf{w}, (G - mI)\mathbf{v} \rangle, \qquad |(G - mI)\mathbf{v}|, \qquad \|G - mI\|$$

respectively, for deterministic $\mathbf{v}, \mathbf{w} \in \mathbb{C}^N$. It is easy to see that already convergence in the strong operator sense must fail: $|(G - mI)\mathbf{v}|$ cannot be small for $z \in \mathbf{S}$ with high probability. To see this, write

$$|(G - mI)\mathbf{v}| \;=\; \sup_{|\mathbf{w}| \leqslant 1} |\langle \mathbf{w}, (G - mI)\mathbf{v} \rangle| \;\geqslant\; |\langle \mathbf{u}_i, \mathbf{v} \rangle| \left| \frac{1}{\lambda_i - z} - m \right|,$$

for any $i = 1, \ldots, N$, where in the last step we chose $\mathbf{w} = \mathbf{u}_i$ to be an eigenvector of H. Suppose that $z = E + \mathrm{i}\eta$ with E in the bulk spectrum and $\eta \geqslant N^{-1+\tau}$. Then from Theorem 2.9 we get that with high probability there exists $i \equiv i(E) = 1, \ldots, N$ such that $|\lambda_i - z| \leqslant 2\eta$. Thus,

$$|(G - mI)\mathbf{v}| \;\geqslant\; \frac{c|\langle \mathbf{u}_i, \mathbf{v} \rangle|}{\eta}.$$

Since $\mathbf{v}$ is deterministic, the numerator is at least of order $N^{-1/2}$ with probability $1 - o(1)$; for instance in the case of GUE this is trivial because $\mathbf{u}_i$ is uniformly distributed on the unit sphere. We conclude that $|(G - mI)\mathbf{v}|$ is with probability $1 - o(1)$ at least of order $N^{-1/2}\eta^{-1}$, which is much bigger than one for $\eta \ll N^{-1/2}$.

As it turns out, G and mI are close in the weak operator sense, and (2.15) generalizes to

$$\langle \mathbf{w}, G(z)\mathbf{v} \rangle - m(z)\langle \mathbf{w}, \mathbf{v} \rangle \;=\; O_{\prec}(\Psi(z)) \tag{12.1}$$

uniformly[4] for all deterministic unit vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^N$ and $z \in \mathbf{S}$. Such a result is called an *isotropic local law*, and was first obtained for Wigner matrices in [54] and subsequently generalized to other models in [13]. The isotropic local law is a nontrivial extension of Theorem 2.6, and both proofs of [54, 13] in fact rely crucially on the entrywise estimate from (2.15) as input.

Isotropic local laws have proved very useful in the analysis of deformed random matrix ensembles [54, 14] and the distribution of eigenvectors [22]. Here we only mention an immediate consequence: *isotropic delocalization* which extends the estimate from Theorem 2.10 to $|\langle \mathbf{v}, \mathbf{u}_i \rangle|^2 \prec N^{-1}$ for all $i = 1, \ldots, N$ and all deterministic unit vectors $\mathbf{v}$. The proof is the same as that of Theorem 2.10, using (12.1) instead of Theorem 2.6.

12.2. Beyond Wigner matrices. In these notes we only consider Wigner matrices from Definition 2.2. Although Wigner matrices occupy a central place in random matrix theory, there has recently been much interest in more general random matrix ensembles. One natural way to generalize Definition 2.2 is to admit matrix entries with general variances. Thus, we consider a matrix H as in Definition 2.2, except that we replace (ii) and (iii) with the following conditions.

(ii') For all i, j we have $EH_{ij} = 0$ and $\mathbb{E}|H_{ij}|^2 = S_{ij}$ for some matrix $S = (S_{ij})$.

(iii') If $S_{ij} > 0$ then $S_{ij}^{-1/2} H_{ij}$ is bounded in any L^p space, uniformly in N, i, j.

A common assumption on the matrix S is that it be (doubly) stochastic:

$$\sum_j S_{ij} = 1 \tag{12.2}$$

for all i. The assumption (12.2) guarantees that the limiting behaviour of H is still governed by the semicircle distribution [61, 15]. Two important classes of random matrices satisfying (12.2) are the following.

- *Generalized Wigner matrix.* We require (12.2) and that $NS_{ij} \asymp 1$ uniformly in N, i, j.

- *Band matrix.* Let f be a symmetric probability density on $\mathbb{R}$, called the *band profile function*, and $W \in [1, N]$ be the *band width*. Then we set

$$S_{ij} := \frac{1}{Z} f\left(\frac{[i-j]_N}{W}\right),$$

where $[i]_N := (i + N\mathbb{Z}) \cap [-N/2, N/2)$ is the periodic representative of $i \in \mathbb{Z}$ in $\mathbb{Z} \cap [-N/2, N/2)$, and $Z > 0$ is chosen so that (12.2) holds.

All of the results in these notes also hold for generalized Wigner matrices; see e.g. [45, 37] for more details. The proofs for generalized Wigner matrices are similar to the ones presented in these notes. The major difference is that in the proof of Theorem 2.6 one has to split the space $\mathbb{C}^N$ into the one-dimensional space spanned by $\mathbf{e} := N^{-1/2}(1, 1, \ldots, 1)^*$ and its orthogonal complement. In the direction of $\mathbf{e}$, the analysis is identical to that of Sections 5–6 thanks to the assumption (12.2). On the orthogonal complement of $\mathbf{e}$, the key extra difficulty is to obtain good bounds on the

---

[4]We emphasize that the estimate is uniform in $\mathbf{v}, \mathbf{w}$ in the sense of Definition 2.5, but of course not *simultaneous* in $\mathbf{v}, \mathbf{w}$. Indeed, a simultaneous statement would imply closeness in the norm sense, which is wrong as remarked above.

deterministic operator $(1 - m^2 S)^{-1}$. Near the spectral edge we have $m^2 \approx 1$ so that this operator is singular because, by (12.2), 1 is an eigenvalue of S with eigenvector **e**. However, thanks to the assumptions $N S_{ij} \asymp 1$ it is not hard to see that $(1 - m^2 S)^{-1}$ is in fact regular on the orthogonal complement of **e**. We refer to [37] for the full details.

Band matrices are a much more challenging, and interesting, model. They arise from solid state physics as an alternative to the Anderson model to describe the physics of a disordered quantum Hamiltonian. So far an optimal local law and eigenvector delocalization are open problems, although partial progress has been made in [31, 32, 37, 36, 5].

If the condition (12.2) does not hold, then the asymptotic eigenvalue distribution of H is no longer governed by the semicircle distribution, but by a much more complicated distribution that depends on S. This scenario was recently investigated in depth in [1, 2], where in particular a local law was established by a nontrivial extension of the methods presented in these notes, involving a stability analysis of a nonlinear quadratic vector equation.

Next, we say a few words about random matrices that are not of the above form but can nevertheless be studied using variants of the method presented in these notes. An important example is the *sample covariance matrix $H := XX^*$*, where X is an $M \times N$ matrix and $M, N \to \infty$ simultaneously. If the entries of X are independent with mean zero and variance N^{-1} then the asymptotic eigenvalue distribution of H is governed by the Marchenko-Pastur distribution [58, 4] instead of the semicircle distribution, and the local law was established in [42, 65, 26]. If the entries of X have a nontrivial covariance structure $\Sigma := \mathbb{E}H$, the asymptotic eigenvalue distribution becomes much more complicated and depends on Σ, and the basic strategy from these notes cannot be applied. A local law for such matrices was established in great generality in [56] using a new self-consistent comparison method.

We remark that a generalization of the methods of [42, 65] for sample covariance matrices may also be used to obtain local laws for non-Hermitian matrices with independent entries, using Girko's Hermitization trick [46]; see [23, 24, 80] for more details.

Local laws have also been derived for matrix models describing the free additive convolution [49, 50, 51, 62, 9, 6, 7] and, using different techniques from the ones presented in these notes, for general β-ensembles (or one-dimensional log-gases) [19, 20, 21]. We also mention that a local law was proved for Wigner matrices with heavy tailed entries in [17].

Finally, we note that local laws have been established for *sparse* random matrices, where most entries of H are with high probability zero. Such matrices typically arise as adjacency or Laplacian matrices of random graphs. A local law for the Erdős-Rényi graph was derived in [34]. Because the entries of the adjacency matrix of the Erdős-Rényi graph are independent up to the symmetry condition, the basic structure of the proof presented in these notes remains applicable. This is stark contrast to the case of *random regular graphs*, where the entries of the adjacency matrix exhibit strong nonlocal correlations and whole setup of the proof of Theorem 2.6 breaks down. For this case, a local law was recently derived in [8] using a new local resampling method.

A. Schur's complement formula and proof of Lemma 3.5

Lemma A.1 (Schur's complement formula). *Provided all of the following inverse matrices exist, the inverse of a block matrix is given by*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{A}^{-1} & -\mathcal{A}^{-1}BD^{-1} \\ -D^{-1}C\mathcal{A}^{-1} & D^{-1}C\mathcal{A}^{-1}BD^{-1} + D^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} A^{-1}B\mathcal{D}^{-1}CA^{-1} + A^{-1} & -A^{-1}B\mathcal{D}^{-1} \\ -\mathcal{D}^{-1}CA^{-1} & \mathcal{D}^{-1} \end{pmatrix},$$

where we defined

$$\mathcal{A} := A - BD^{-1}C, \qquad \mathcal{D} := D - CA^{-1}B.$$

Proof. This is routine verification. A good way of arriving at these formulas is to use Gaussian elimination to write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & BD^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ D^{-1}C & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 \\ CA^{-1} & 1 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} 1 & A^{-1}B \\ 0 & 1 \end{pmatrix},$$

from which the claim follows trivially. □

Proof of Lemma 3.5. Without loss of generality, we take $T = \emptyset$. (Otherwise consider the matrix $H^{(T)}$ instead of H.) It suffices to prove the claim for $G = M^{-1}$ for a general matrix M. As in Definition 3.1, we use the notation $M^{(k)} = (M_{ij})_{i,j\in\{1,\dots,N\}\setminus\{k\}}$.

First, from Schur's complement formula for the off-diagonal blocks we get for $i \neq j$

$$G_{ij} = -G_{ii} \sum_k^{(i)} M_{ik} G_{kj}^{(i)} = -G_{jj} \sum_k^{(j)} G_{ik}^{(j)} M_{kj}, \tag{A.1}$$

which immediately yields (3.5).

It remains to prove (3.4); there are several approaches. Perhaps the most natural one is based on the Woodbury matrix identity

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

which can itself be deduced for instance using the resolvent identity $Y^{-1} = X^{-1} + X^{-1}(X-Y)Y^{-1}$. Let $\mathbf{e}_k$ be the standard k-th column unit basis vector and define the matrices

$$M_{ij}^{[k]} := M_{ij}\mathbf{1}(i \neq k)\mathbf{1}(j \neq k), \qquad U := \begin{pmatrix} H\mathbf{e}_k & \mathbf{e}_k \end{pmatrix}, \qquad V := \begin{pmatrix} \mathbf{e}_k^* \\ \mathbf{e}_k^*H \end{pmatrix},$$

so that we have

$$M = M^{[k]} + UV - M_{kk}\mathbf{e}_k\mathbf{e}_k^*.$$

Then by the Woodbury matrix identity we have

$$\left(M^{[k]} - M_{kk}\mathbf{e}_k\mathbf{e}_k\right)^{-1} = G + GU(I - VGU)^{-1}VG.$$

A straightforward calculation yields

$$I - VGU = -\begin{pmatrix} 0 & G_{kk} \\ M_{kk} & 0 \end{pmatrix}, \qquad (I - VGU)^{-1} = -\frac{1}{M_{kk}G_{kk}}\begin{pmatrix} 0 & G_{kk} \\ M_{kk} & 0 \end{pmatrix},$$

from which we deduce, after a short calculation,

$$\left(M^{[k]} - M_{kk}\mathbf{e}_k\mathbf{e}_k\right)^{-1} = G - \frac{1}{M_{kk}}\mathbf{e}_k\mathbf{e}_k^* - \frac{1}{G_{kk}}G\mathbf{e}_k\mathbf{e}_k^*G.$$

We conclude

$$(M^{[k]})^{-1} = G - \frac{1}{G_{kk}}G\mathbf{e}_k\mathbf{e}_k^*G,$$

from which (3.4) follows.

For a second proof of (3.4), write, for $k \neq j$,

$$\sum_l^{(k)} M_{kl}\left(G_{lj} - \frac{G_{lk}G_{kj}}{G_{kk}}\right) = -M_{kk}G_{kj} - \frac{G_{kj}}{G_{kk}}(1 - M_{kk}G_{kk}) = -\frac{G_{kj}}{G_{kk}} = \sum_l^{(k)} M_{kl}G_{lj}^{(k)},$$

where in the last step we used (A.1). Taking the partial derivative $\frac{\partial}{\partial M_{ki}}$ of this identity (using the resolvent expansion to differentiate the entries of G) with $i \neq k$ yields

$$G_{ij} - \frac{G_{ik}G_{kj}}{G_{kk}} = G_{ij}^{(k)},$$

which is (3.4).

A third proof of (3.4) can be be obtained by explicit matrix inversion for $N = 3$, and then extended to arbitrary $N > 3$ using Schur's complement formula. $\qquad\square$

B. Basic properties of the semicircle distribution and proof of Lemma 5.5

For $k \geqslant 0$ define the *k-th Catalan number* $\mathcal{C}_k := \frac{1}{k+1}\binom{2k}{k}$.

Lemma B.1 (Moments of the semicircle distribution and Catalan numbers). *For each $k \geqslant 0$, we have*

$$\frac{1}{2\pi}\int_{-2}^2 x^k\sqrt{4 - x^2}\,\mathrm{d}x = \begin{cases} \mathcal{C}_{k/2} & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd}. \end{cases}$$

Proof. This is an easy computation using the change of variables $x = 2\cos\theta$ with $\theta \in [0, \pi]$. $\qquad\square$

Lemma B.2 (Stieltjes transform of the semicircle distribution). *The Stieltjes transform $m(z)$ of the semicircle distribution satisfies the relation (2.12), and is explicitly given by the formula (2.11).*

Proof. We prove (2.11), from which (2.12) follows immediately. Since both $m(z)$ and the right-hand side of (2.11) are analytic for $z \in \mathbb{C} \setminus [-2, 2]$, it suffices to prove (2.11) for z in some subset of $\mathbb{C}$ with a limit point. In fact, we establish (2.11) for $|z| > 2$. In this case, by Lemma B.1,

$$m(z) = -z^{-1}\int \frac{1}{1 - (x/z)}\varrho(\mathrm{d}x) = -z^{-1}\sum_{n \geqslant 0} \mathcal{C}_n z^{-2n},$$

where the power series is absolutely convergent. Now (2.11) follows directly from the series expansion of the right-hand side of (2.11). □

Proof of Lemma 5.5. We have
$$\tilde{m} \;=\; \frac{-z - \sqrt{z^2 - 4}}{2}\,,$$

with the conventions for the square root introduced after (2.11). From this we get
$$|m - \tilde{m}| \;=\; \left|\sqrt{z^2 - 4}\right| \;=\; \sqrt{|z - 2||z + 2|} \;\asymp\; \sqrt{\kappa + \eta}\,,$$

where in the last step we used that $|z| \leqslant C$ for $z \in \mathbf{S}$.

Next, if u is a solution of (5.11) then $u \in \{m_r, \tilde{m}_r\}$, where we defined
$$m_r \;:=\; \frac{-z + \sqrt{z^2 - 4(1 + r)}}{2}\,, \qquad \tilde{m}_r \;:=\; \frac{-z - \sqrt{z^2 - 4(1 + r)}}{2}\,.$$

(Note that the choice of the branch of the square root is immaterial.) Note that for any complex square root $\sqrt{\cdot}$ and $w, \zeta \in \mathbb{C}$ we have
$$\left|\sqrt{w + \zeta} - \sqrt{w}\right| \wedge \left|\sqrt{w + \zeta} + \sqrt{w}\right| \;\leqslant\; \frac{|\zeta|}{\sqrt{|w|}} \wedge \sqrt{|\zeta|}\,.$$

Hence (5.12) follows from $|\sqrt{z^2 - 4}| \asymp \sqrt{\kappa + \eta}$. □

C. The Helffer-Sjöstrand formula

Proposition C.1 (Helffer-Sjöstrand formula). *Let $n \in \mathbb{N}$ and $f \in \mathcal{C}^{n+1}(\mathbb{R})$. We define the almost analytic extension of f of degree n through*
$$\tilde{f}_n(x + \mathrm{i}y) \;:=\; \sum_{k=0}^{n} \frac{1}{k!}(\mathrm{i}y)^k f^{(k)}(x)\,.$$

Let $\chi \in \mathcal{C}_c^\infty(\mathbb{C}; [0, 1])$ be a smooth cutoff function. Then for any $\lambda \in \mathbb{R}$ satisfying $\chi(\lambda) = 1$ we have
$$f(\lambda) \;=\; \frac{1}{\pi} \int_{\mathbb{C}} \frac{\bar{\partial}(\tilde{f}_n(z)\chi(z))}{\lambda - z}\, \mathrm{d}^2 z\,,$$

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

Applied to the matrix H, the Helffer-Sjöstrand formula gives
$$f(H) \;=\; \frac{1}{\pi} \int_{\mathbb{C}} \bar{\partial}(\tilde{f}_n(z)\chi(z))\, G(z)\, \mathrm{d}^2 z\,, \tag{C.1}$$

provided χ is chosen so that $\chi = 1$ on $\mathrm{spec}(H)$. Compared to the functional calculus from (2.2), the Helffer-Sjöstrand representation (C.1) is more versatile and general. In particular, we only require that f be differentiable instead of analytic. In most applications, however, we shall require some additional smoothness of f (typically $f \in \mathcal{C}^2(\mathbb{R})$), since the integral on the right-hand side of (C.1) is more stable for small η if n is large. This generality is of great use in applications, since it allows one to choose f to be a (smoothed) indicator function, which is not possible with (2.2).

Proof of Proposition C.1. Abbreviate $g(z) := \tilde{f}_n(z)\chi(z)$. By assumption, we have $g(\lambda) = f(\lambda)$ for $\lambda \in \mathbb{R}$ satisfying $\chi(\lambda) = 1$. Choose $r > 0$ large enough that $g = 0$ outside the ball $B_r(0)$. For $\varepsilon > 0$ define the region $D_\varepsilon := B_r(0) \setminus B_\varepsilon(\lambda)$. Then we have

$$0 = \oint_{\partial B_r(0)} \frac{g(z)}{\lambda - z}\, dz = \int_{\partial D_\varepsilon} \frac{g(z)}{\lambda - z}\, dz + \oint_{\partial B_\varepsilon(\lambda)} \frac{g(z)}{\lambda - z}. \tag{C.2}$$

By Green's formula, the first term of (C.2) is equal to

$$\int_{\partial D_\varepsilon} \frac{g(z)}{\lambda - z}\, dz = \int_{D_\varepsilon} d\left(\frac{g(z)}{\lambda - z}\right) \wedge dz = \int_{D_\varepsilon} \bar{\partial}\left(\frac{g(z)}{\lambda - z}\right) d\bar{z} \wedge dz = \int_{D_\varepsilon} \frac{\bar{\partial}g(z)}{\lambda - z}\, 2i\, d^2 z\,,$$

where in the last step we used that $(\lambda - z)^{-1}$ is holomorphic in D_ε. Moreover, by Cauchy's theorem the second term of (C.2) is equal to $-2\pi i g(\lambda) + o(1)$ as $\varepsilon \downarrow 0$, since g is continuous. We conclude that

$$g(\lambda) = \frac{1}{\pi} \int_{D_\varepsilon} \frac{\bar{\partial}g(z)}{\lambda - z}\, d^2 z + o(1)$$

as $\varepsilon \downarrow 0$. Letting $\varepsilon \downarrow 0$ completes the proof. $\qquad\square$

D. Multilinear large deviation estimates: proof of Lemma 3.6

We first recall the following version of the Marcinkiewicz-Zygmund inequality.

Lemma D.1. *Let $X_1, \ldots, X_N$ be a family of independent random variables each satisfying (3.7) and suppose that the family (b_i) is deterministic. Then*

$$\left\|\sum_i b_i X_i\right\|_p \leqslant (Cp)^{1/2} \mu_p \left(\sum_i |b_i|^2\right)^{1/2} \tag{D.1}$$

Proof. The proof is a simple application of Jensen's inequality. Writing $B^2 := \sum_j |b_i|^2$, we get, by the classical Marcinkiewicz-Zygmund inequality [71] in the first line, that

$$\left\|\sum_i b_i X_i\right\|_p^p \leqslant (Cp)^{p/2} \left\|\left(\sum_i |b_i|^2 |X_i|^2\right)^{1/2}\right\|_p^p$$

$$= (Cp)^{p/2} B^p\, \mathbb{E}\left[\left(\sum_i \frac{|b_i|^2}{B^2}|X_i|^2\right)^{p/2}\right]$$

$$\leqslant (Cp)^{p/2} B^p\, \mathbb{E}\left[\sum_i \frac{|b_i|^2}{B^2}|X_i|^p\right]$$

$$\leqslant (Cp)^{p/2} B^p \mu_p^p\,. \qquad\square$$

Next, we prove the following intermediate result.

Lemma D.2. *Let $X_1, \ldots, X_N, Y_1, \ldots, Y_N$ be independent random variables each satisfying (3.7), and suppose that the family (a_{ij}) is deterministic. Then for all $p \geqslant 2$ we have*

$$\left\|\sum_{i,j} a_{ij} X_i Y_j\right\|_p \leqslant Cp\, \mu_p^2 \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}.$$

65

Proof. Write

$$\sum_{i,j} a_{ij} X_i Y_j = \sum_j b_j Y_j , \qquad b_j := \sum_i a_{ij} X_i .$$

Note that (b_j) and (Y_j) are independent families. By conditioning on the family (b_j), we therefore get from Lemma D.1 and the triangle inequality that

$$\left\| \sum_j b_j Y_j \right\|_p \leqslant (Cp)^{1/2} \mu_p \left\| \sum_j |b_j|^2 \right\|_{p/2}^{1/2} \leqslant (Cp)^{1/2} \mu_p \left(\sum_j \|b_j\|_p^2 \right)^{1/2} .$$

Using Lemma D.1 again, we have

$$\|b_j\|_p \leqslant (Cp)^{1/2} \mu_p \left(\sum_i |a_{ij}|^2 \right)^{1/2} .$$

This concludes the proof. $\qquad\square$

Lemma D.3. *Let $X_1, \ldots, X_N$ be independent random variables each satisfying (3.7), and suppose that the family (a_{ij}) is deterministic. Then we have*

$$\left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_p \leqslant Cp \mu_p^2 \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} .$$

Proof. The proof relies on the identity (valid for $i \neq j$)

$$1 = \frac{1}{Z_N} \sum_{I \sqcup J = [\![N]\!]} \mathbf{1}(i \in I) \mathbf{1}(j \in J) , \tag{D.2}$$

where the sum ranges over all partitions of $[\![N]\!] = \{1, \ldots, N\}$ into two sets I and J, and $Z_N := 2^{N-2}$ is independent of i and j. Moreover, we have

$$\sum_{I \sqcup J = [\![N]\!]} 1 = 2^N - 2 , \tag{D.3}$$

where the sum ranges over nonempty subsets I and J. Now we may estimate

$$\left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_p \leqslant \frac{1}{Z_N} \sum_{I \sqcup J = [\![N]\!]} \left\| \sum_{i \in I} \sum_{j \in J} a_{ij} X_i X_j \right\|_p \leqslant \frac{1}{Z_N} \sum_{I \sqcup J = [\![N]\!]} Cp \mu_p^2 \left(\sum_{i \neq j} |a_{ij}|^2 \right)^{1/2} ,$$

where we used that, for any partition $I \sqcup J = [\![N]\!]$, the families $(X_i)_{i \in I}$ and $(X_j)_{j \in J}$ are independent, and hence the Lemma D.2 is applicable. The claim now follows from (D.3). $\qquad\square$

Note that the proof of Lemma D.3 may be easily extended to multilinear expressions of the form $\sum_{i_1, \ldots, i_k}^* a_{i_1 \ldots i_k} X_{i_1} \cdots X_{i_k}$. We shall not pursue such extensions here.

We may now complete the proof of Lemma 3.6.

66

Proof of Lemma 3.6. The proof is a simple application of Markov's inequality. Part (i) follows from Lemma D.1, part (ii) from Lemma D.3, and part (iii) from Lemma D.2. We give the details for part (iii).

For $\varepsilon > 0$ and $D > 0$ we have

$$\mathbb{P}\left[\left|\sum_{i \neq j} a_{ij} X_i X_j\right| \geqslant N^{\varepsilon} \Psi\right] \leqslant \mathbb{P}\left[\left|\sum_{i \neq j} a_{ij} X_i X_j\right| \geqslant N^{\varepsilon} \Psi, \left(\sum_{i \neq j} |a_{ij}|^2\right)^{1/2} \leqslant N^{\varepsilon/2} \Psi\right]$$

$$+ \mathbb{P}\left[\left(\sum_{i \neq j} |a_{ij}|^2\right)^{1/2} \geqslant N^{\varepsilon/2} \Psi\right]$$

$$\leqslant \mathbb{P}\left[\left|\sum_{i \neq j} a_{ij} X_i X_j\right| \geqslant N^{\varepsilon/2}\left(\sum_{i \neq j} |a_{ij}|^2\right)^{1/2}\right] + N^{-D-1}$$

$$\leqslant \left(\frac{C p \mu_p^2}{N^{\varepsilon/2}}\right)^p + N^{-D-1}$$

for arbitrary D. In the second step we used the definition of $\left(\sum_{i \neq j} |a_{ij}|^2\right)^{1/2} \prec \Psi$ with parameters $\varepsilon/2$ and $D+1$. In the last step we used Lemma D.3 by conditioning on (a_{ij}). Given ε and D, there is a large enough p such that the first term on the last line is bounded by N^{-D-1}. Since ε and D were arbitrary, the proof is complete.

The claimed uniformity in u in the case that a_{ij} and X_i depend on an index u also follows from the above estimate. $\qquad\square$

E. Proof of Lemma 2.1

Proof of Lemma 2.1. A preliminary remark is that by (2.7) (and its natural extension to the set $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ of nonnegative measures on $\mathbb{R}$ with total mass at most 1), any element of $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ is determined by its Stieltjes transform.

Note that for any fixed $z \in \mathbb{C}_+$, the function which maps a probability measure on $\mathbb{R}$ to its Stieltjes transform evaluated at z is continuous in the topology of convergence in distribution. As a consequence, if $\mu \xrightarrow{\mathrm{d}} \varrho$ in probability, then $s(z) \to m(z)$ in probability.

To prove the converse, let us first treat the deterministic case. Suppose $\mu \equiv \mu_N$ to be deterministic. Then by Helly's theorem, from any of its subsequences, one can extract a subsequence having a limit θ in $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ for the vague topology (i.e. the topology generated by integration of continuous functions with null limits at infinity). By hypothesis, θ must have Stieltjes transform m, hence be equal to ϱ. This proves that μ converges to ϱ for the vague topology, hence for the convergence in distribution because μ and ϱ are both probability measures.

Let us now prove the converse in the general case where $\mu \equiv \mu_N$ is random. Let (z_n) be a sequence of pairwise distinct elements of $\mathbb{C}_+$ having a limit in $\mathbb{C}_+$. By what precedes, the distance d_S on the set $\mathcal{M}_1(\mathbb{R})$ of probability measures on $\mathbb{R}$ defined by

$$\mathrm{d}_S(\nu, \nu') := \sum_n 2^{-n} S_{\nu - \nu'}(z_n) \qquad \text{with } S_{\nu-\nu'}(z) := \int \frac{\mathrm{d}(\nu - \nu')(t)}{t - z}$$

defines the topology of convergence in distribution. By hypothesis, $\mathrm{d}_S(\mu, \varrho)$ tends to zero in probability, which closes the proof of the theorem.

Let us also give a more direct and more conceptual proof of the converse. For the vague topology defined above, the set $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ is a compact Polish space, hence so is the set $\mathcal{M}_1(\mathcal{M}_{\leqslant 1}(\mathbb{R}))$ of probability measures on $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ endowed with the topology of the convergence in distribution. Denoting by $P \equiv P_N \in \mathcal{M}_1(\mathcal{M}_{\leqslant 1}(\mathbb{R}))$ the distribution of μ, the statement we have to prove is that $P \to \delta_\varrho$. Let $P_\infty \in \mathcal{M}_1(\mathcal{M}_{\leqslant 1}(\mathbb{R}))$ be an accumulation point of P. By hypothesis, a P_∞-distributed random element of $\mathcal{M}_{\leqslant 1}(\mathbb{R})$ has Stieltjes transform equal to m with probability one, hence is equal to ϱ with probability one. As the topology induced by the vague topology on $\mathcal{M}_1(\mathbb{R})$ is the one of the convergence in distribution, this concludes the proof. $\qquad\square$

F. Global semicircle law and proof of Theorem 2.4

Theorem 2.4 follows from the following more general result.

Theorem F.1. *Suppose that the Hermitian matrix H satisfies the following conditions.*

(i) *The upper-triangular entries $(H_{ij} : 1 \leqslant i \leqslant j \leqslant N)$ are independent.*

(ii) *For all i, j we have $\mathbb{E}H_{ij} = 0$ and for all $i \neq j$ we have $\mathbb{E}|H_{ij}|^2 = N^{-1}$.*

(iii) *The sequence $\delta_N := N \sup_{i,j} \mathbb{E}|H_{ij}|^3$ tends to zero as $N \to \infty$.*

Then for any fixed $z \in \mathbb{C}_+$ and $\varepsilon > 0$ there exist constants $C, c > 0$ such that

$$\mathbb{P}(|s(z) - m(z)| \geqslant \varepsilon) \;\leqslant\; C\mathrm{e}^{-cN}.$$

To prove this theorem, we shall use several preliminary lemmas. The first one is a decorrelation result, generalizing the well-known Stein lemma (obtained by a straightforward integration by parts) to non-Gaussian variables.

Lemma F.2. *Let $f \in \mathcal{C}^2(\mathbb{R}^n; \mathbb{R})$ be bounded with bounded first and second derivatives. Let X be a centred random vector with values in $\mathbb{R}^n$ with finite second moment. Then there is $C \equiv C(n)$ such that*

$$\left|\mathbb{E}f(X)X - \mathbb{E}\langle \nabla f(X), Y\rangle Y\right| \;\leqslant\; C\|\nabla^2 f\|_\infty \mathbb{E}|X|^3,$$

where Y is an independent copy of X and $\nabla^2 f$ is the Hessian of f.

Proof. Note first that by the Taylor-Lagrange formula, for any x,

$$|f(x) - (f(0) + \langle \nabla f(0), x\rangle)| \;\leqslant\; \|\nabla^2 f\|_\infty |x|^2,$$

so it suffices to prove that

$$\left|\mathbb{E}(f(0) + \langle \nabla f(0), Y\rangle)Y - \mathbb{E}\langle \nabla f(X), Y\rangle Y\right| \;\leqslant\; C\|\nabla^2 f\|_\infty \mathbb{E}|X|^3.$$

Since X is centred, we can omit the term $f(0)Y$, and it suffices to prove that

$$|\nabla f(0) - \nabla f(X)| \;\leqslant\; C\|\nabla^2 f\|_\infty |X|,$$

which follows from the mean-value theorem. $\qquad\square$

This lemma implies easily that for H a centred complex random variable ($n = 2$) with finite third moment and g a $\mathcal{C}^2$ function on $\mathbb{C}$ which is bounded and has bounded first and second derivatives, we have

$$\left| \mathbb{E}g(H)H \; - \; \mathbb{E}\left(\frac{\partial g(H)}{\partial \operatorname{Re}(H)} \operatorname{Re}(H')H' + \frac{\partial g(H)}{\partial \operatorname{Im}(H)} \operatorname{Im}(H')H' \right) \right| \; \leqslant \; C\|\nabla^2 g\|_\infty \mathbb{E}|H|^3 \qquad \text{(F.1)}$$

for H' distributed as H, independent of H.

The following well known lemma can be found in [59] or [18, Th. 6.2]. It states that functions of many independent random variables X_i have sub-Gaussian tails, with variance bounded by a quantity which is additive in the variables X_i.

Lemma F.3 (McDiarmid). *Let $X_1, \ldots, X_n$ be independent random variables taking values in some spaces denoted by $E_1, \ldots, E_n$, let*

$$f : E_1 \times \cdots \times E_n \to \mathbb{R}$$

be a measurable function and set $Y = f(X_1, \ldots, X_n)$. Define, for each $k = 1, \ldots, n$,

$$c_k := \sup |f(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_n) - f(x_1, \ldots, x_{k-1}, z, x_{k+1}, \ldots, x_n)|, \qquad \text{(F.2)}$$

where the supremum is taken over $x_i \in E_i$ for all $i \neq k$ and $y, z \in E_k$. Then for each $r \geqslant 0$, we have

$$\mathbb{P}(|Y - \mathbb{E}Y| \geqslant r) \; \leqslant \; 2 \exp\left(-\frac{2r^2}{c_1^2 + \cdots + c_n^2} \right). \qquad \text{(F.3)}$$

The following elementary consequence of McDiarmid's lemma is pointed out in [16] (see also [67, 63]). It implies for example that the fluctuations of $s(z)$ around its mean have order at most $N^{-1/2}$, as if the eigenvalues had been some i.i.d. L^2 random variables. It is well known (see e.g. [4]) that in fact, for Wigner matrices, as a consequence of the eigenvalue repulsion phenomenon, these fluctuations have order N^{-1}. However, for heavy-tailed or Erdős-Rényi random matrices, the fluctuations can have any order between N^{-1} and $N^{-1/2}$ (see [11, 12]).

Lemma F.4. *Let H be a random $N \times N$ Hermitian matrix whose semi-columns $(H_{i,1})_{1\leqslant i\leqslant N}$, $(H_{i,2})_{2\leqslant i\leqslant N}$, $\ldots$, $(H_{i,N})_{N\leqslant i\leqslant N}$ are independent random vectors. Then for any $z \in \mathbb{C}_+$ and any $r \geqslant 0$,*

$$\mathbb{P}(|s(z) - \mathbb{E}s(z)| \geqslant r) \; \leqslant \; 2\mathrm{e}^{-\frac{N\eta^2 r^2}{8}}. \qquad \text{(F.4)}$$

Proof. It suffices to notice that, by (2.3), a variation of one of the semi-columns of H affects $G(z)$ by an additive perturbation by a matrix with rank at most 2 and operator norm at most $2/\eta$. $\quad\square$

Proof of Theorem F.1. Note that we have $G(z) = z^{-1}(-1 + HG(z))$, so that

$$\begin{aligned} \mathbb{E}s(z) \; &= \; -\frac{1}{z} + \frac{1}{Nz}\mathbb{E}\operatorname{Tr} HG \\ &= \; -\frac{1}{z} + \frac{1}{Nz}\sum_k \mathbb{E}H_{kk}G_{kk} + \frac{1}{Nz}\sum_{k\neq l}\mathbb{E}H_{kl}G_{lk}\,. \end{aligned}$$

Set $E_{kl\pm lk}$ to be the matrix with all entries equal to zero, except the (k,l)-th and (l,k)-th ones, which are respectively equal to 1 and ± 1. Set also E_{kk} to be the matrix with all entries equal to zero, except the (k,k)-th one, which is equal to 1. Finally, define

$$\alpha_{kl} := N\mathbb{E}\operatorname{Re}(H_{kl})H_{kl} \qquad \text{and} \qquad \beta_{kl} := N\mathbb{E}\operatorname{Im}(H_{kl})H_{kl}\,.$$

Using (2.3) to differentiate, it follows from (F.1) that

$$\mathbb{E}s(z) = -\frac{1}{z} - \frac{1}{N^2 z}\sum_k N\mathbb{E}H_{kk}^2\mathbb{E}(GE_{kk}G)_{kk} - \frac{1}{N^2 z}\sum_{k\neq l}\alpha_{kl}\mathbb{E}(GE_{kl+lk}G)_{lk}$$
$$- \frac{\mathrm{i}}{N^2 z}\sum_{k\neq l}\beta_{kl}\mathbb{E}(GE_{kl-lk}G)_{lk} + O(\mathcal{E}(z,N))$$

where we use the shorthand $\mathcal{E}(z,N) := \max\{\eta^{-2}, \eta^{-3}\}(N^{-1/2} + \delta_N)$. Hence

$$\mathbb{E}s(z) = -\frac{1}{z} - \frac{1}{N^2 z}\sum_{k,l}\alpha_{kl}\mathbb{E}[G_{lk}^2 + G_{kk}G_{ll}] - \frac{\mathrm{i}}{N^2 z}\sum_{k,l}\beta_{kl}\mathbb{E}[G_{lk}^2 - G_{kk}G_{ll}] + O(\mathcal{E}(z,N))\,.$$

Next, note that

$$\left|\sum_{k,l}G_{lk}^2\right| = \left|\operatorname{Tr}GG^T\right| \leqslant N\eta^{-2}\,.$$

Using the fact that $\alpha_{kl} - \mathrm{i}\beta_{kl} = 1$, it follows that

$$\mathbb{E}s(z) = -\frac{1}{z} - \frac{1}{N^2 z}\sum_{k,l}\mathbb{E}G_{kk}G_{ll} + O(\mathcal{E}(z,N)) = -\frac{1}{z} - \frac{1}{z}\mathbb{E}s(z)^2 + O(\mathcal{E}(z,N))\,.$$

But by (F.4),

$$\left|\mathbb{E}s(z)^2 - (\mathbb{E}s(z))^2\right| \leqslant \frac{16}{\eta^2 N}\,,$$

so

$$\mathbb{E}s(z) = -\frac{1}{z} - \frac{1}{z}(\mathbb{E}s(z))^2 + O(\mathcal{E}(z,N))\,.$$

We then conclude the proof using Lemma 5.5 and (F.4). $\qquad\square$

We conclude this appendix with a remark about Theorem F.1 applied to sparse random matrices. Sparse random matrices arise typically as adjacency matrices of random graphs. For instance, let $G(N,p)$ denote the *Erdős-Rényi graph*, defined as the (simple) random graph on N vertices where each edge $\{i,j\}$ is open with probability p and closed with probability $1-p$, independently of the other edges. The graph is encoded by its adjacency matrix $A = A^* = \{0,1\}^{N\times N}$, whereby $A_{ij} = A_{ji}$ is equal to one if and only if the edge $\{i,j\}$ is open in $G(N,p)$. The diagonal entries of A are zero, and the strict upper-triangular entries $(A_{ij} : 1 \leqslant i < j \leqslant N)$ are i.i.d. Bernoulli random variables with law $(1-p)\delta_0 + p\delta_1$. We define the normalized adjacency matrix $B := (Np(1-p))^{-1/2}A$, and denote by $s_B(z)$ the Stieltjes transform of its empirical eigenvalue distribution (see (2.5)). Then Theorem F.1 implies that for any fixed $z \in \mathbb{C}_+$ we have $s_B(z) \to m(z)$ in probability provided that $Np(1-p) \to \infty$ as $N \to \infty$.

For the proof, we define $H := B - \mathbb{E}B$. Then it is an easy exercise to check that H satisfies the assumptions (i)–(iii) of Theorem F.1 provided that $Np(1 - p) \to \infty$. Thus, Theorem F.1 implies that $s_H(z) \to m(z)$ in probability. To return to B, we note that

$$B = H - \frac{p}{\sqrt{Np(1-p)}}I + \sqrt{\frac{Np}{1-p}}\mathbf{e}\mathbf{e}^*,$$

where we defined the unit vector $\mathbf{e} := N^{-1/2}(1, 1, \ldots, 1)^*$. Thus, B differs from H by a small shift proportional to the identity plus a matrix of rank one, and the convergence $s_B(z) \to m(z)$ follows from the convergence $s_H(z) \to m(z)$ combined with the following general deterministic result.

Lemma F.5 (Small perturbations do not affect the global regime). *Let H and $\tilde{H}$ be Hermitian $N \times N$ matrices, and denote by $s(z)$ and $\tilde{s}(z)$ the Stieltjes transforms of their empirical eigenvalue distributions respectively (see (2.5)). Then we have*

$$|s(z) - \tilde{s}(z)| \leqslant \frac{\mathrm{rank}(H - \tilde{H})}{N} \min\left\{\frac{2}{\eta}, \frac{\|H - \tilde{H}\|}{\eta^2}\right\}.$$

Proof. We drop the arguments z and denote by G and $\tilde{G}$ the Green functions of H and $\tilde{H}$ respectively. Then the resolvent identity yields

$$s - \tilde{s} = \frac{1}{N}\mathrm{Tr}(G - \tilde{G}) = \frac{1}{N}\mathrm{Tr}(G(\tilde{H} - H)\tilde{G}).$$

Moreover, by the same resolvent identity we have $\mathrm{rank}(G - \tilde{G}) \leqslant \mathrm{rank}(H - \tilde{H})$. The claim now follows from the trivial bounds $\|G(z)\| \leqslant \eta^{-1}$ and $\|\tilde{G}(z)\| \leqslant \eta^{-1}$, combined with the general estimate $|\mathrm{Tr}\,X| \leqslant \mathrm{rank}(X)\|X\|$ for any matrix X, which can be proved for instance by singular value decomposition of X. This concludes the proof. $\qquad\square$

We note that also a local law has been established for the Erdős-Rényi graph; see Section 12.2 for references.

G. Extreme eigenvalues: the Füredi-Komlós argument

We state here a theorem, essentially due to Füredi and Komlós in [47] (see also [3, 4, 78, 72]), that we use in the proof of Proposition 9.1 to ensure that $\|H\| \leqslant C$ with high probability. As extreme eigenvalues are not the main subject of this text, in the proof given here a key combinatorial result, Vu's Lemma from G.4, is admitted without proof.

Theorem G.1. *Let H be a Wigner matrix. Then for any fixed $\varepsilon > 0$, we have*

$$\|H\| \leqslant 2 + \varepsilon$$

with high probability.

The rest of this appendix is devoted to the proof of Theorem G.1. We remark that if, instead of Definition 2.2 (iii), we make the stronger assumption that $\sqrt{N}H_{ij}$ are uniformly sub-Gaussian (see (H.1) below), then a simpler concentration argument, given in Appendix H, yields the bound

$\|H\| \leqslant C$ with high probability, which is also sufficient to complete the proof of Proposition 9.1 and may therefore be used to replace the arguments of this section.

Before proving Theorem G.1, let us state the following lemma, whose proof is postponed after the proof of the theorem.

Lemma G.2. *Let $\widetilde{H}$ be a Hermitian matrix $\widetilde{H} = \widetilde{H}^*$ whose entries $\widetilde{H}_{ij}$ satisfy the following conditions.*

   (i) The upper-triangular entries $(\widetilde{H}_{ij} : 1 \leqslant i \leqslant j \leqslant N)$ are independent.

   (ii) For all i,j we have $\mathbb{E}\widetilde{H}_{ij} = 0$, and for $i \neq j$, we have $\mathbb{E}|\widetilde{H}_{ij}|^2 \leqslant N^{-1}$.

   (iii) We have $(\log N)^2 \max\limits_{i,j} \|\widetilde{H}_{ij}\|_\infty \longrightarrow 0$.

Then for any fixed $\varepsilon > 0$, we have
$$\|\widetilde{H}\| \leqslant 2 + \varepsilon$$
with high probability.

Remark G.3. It follows obviously that if $\sigma = \sigma_N$ is a deterministic sequence tending to 1, then the conclusion of Lemma G.2 also holds for the matrix $\sigma\widetilde{H}$. We then deduce, by standard perturbation bounds (e.g. [3, Cor. A.6]), that for M a deterministic Hermitian matrix such that $\mathrm{Tr}(M^2) = o(1)$, the conclusion of the theorem also holds for the matrix $\sigma\widetilde{H} + M$.

Proof of Theorem G.1. Let us introduce several auxiliary matrices. We choose $\kappa \in (0, 1/2)$ and define

- $X = (X_{ij}) = \sqrt{N}H$,

- $Y = (Y_{ij})$ with $Y_{ij} := X_{ij}\mathbf{1}(|X_{ij}| \leqslant N^\kappa)$,

- $M = (M_{ij})$ with $M_{ij} := \mathbb{E}Y_{ij}$,

- $\widetilde{H} := \frac{1}{\sqrt{N}\sigma}(Y - M)$ with $\sigma := \max_{ij} \sigma_{ij}$ and σ_{ij} is the standard deviation of Y_{ij}.

Note first that by the Chebyshev inequality, for each $p \in \mathbb{N}$, if C_p is a uniform bound on $\|\sqrt{N}H_{ij}\|_p^p$, then for all $x > 0$,
$$\mathbb{P}(|X_{ij}| \geqslant x) \leqslant \frac{C_p}{x^p}. \tag{G.1}$$

The first consequence of (G.1) is that for any $L > 0$, there is $C = C(L)$ such that
$$\mathbb{P}(X \neq Y) \leqslant CN^{-L}. \tag{G.2}$$

Next, as
$$\begin{aligned}
|M_{ij}| &= |\mathbb{E}X_{ij} - \mathbb{E}X_{ij}\mathbf{1}(|X_{ij}| > N^\kappa)| \\
&= |\mathbb{E}X_{ij}\mathbf{1}(|X_{ij}| > N^\kappa)| \\
&\leqslant \mathbb{E}|X_{ij}|\mathbf{1}(|X_{ij}| > N^\kappa) \\
&= N^\kappa\mathbb{P}(|X_{ij}| > N^\kappa) + \int_{N^\kappa}^{+\infty} \mathbb{P}(|X_{ij}| > x)\mathrm{d}x,
\end{aligned}$$

we know that for any $L > 0$, there is $C = C(L)$ such that

$$|M_{ij}| \leqslant CN^{-L}. \tag{G.3}$$

In the same way, as $\mathbb{E}Y_{ij}^2 = 1 - \mathbb{E}|X_{ij}|^2 \mathbf{1}(|X_{ij}| > N^\kappa)$, we have

$$1 - \mathbb{E}Y_{ij}^2 = N^{2\kappa}\mathbb{P}(|X_{ij}| > N^\kappa) + \int_{N^\kappa}^{+\infty} 2x\mathbb{P}(|X_{ij}| > x)\mathrm{d}x\,,$$

so that for any $L > 0$, there is $C = C(L)$ such that

$$\max_{i,j} |\sigma_{ij} - 1| \leqslant CN^{-L}. \tag{G.4}$$

The first consequence of (G.3) and (G.4) is that for any $\eta \in (0, 1/2 - \kappa)$, for N large enough, for all i, j, we have

$$\|\widetilde{H}_{ij}\|_\infty \leqslant N^{-\eta}. \tag{G.5}$$

As the entries of $\widetilde{H}$ obviously satisfy the other hypotheses of Lemma G.2, we deduce that the conclusion of this lemma holds for the extreme eigenvalues of $\widetilde{H}$. By Remark G.3 and the estimates (G.4) and (G.3), we deduce that the conclusion of Lemma G.2 also holds for the extreme eigenvalues of $\frac{1}{\sqrt{N}}Y$. At last, (G.2) and the union bound allow to conclude. $\qquad\square$

Proof of Lemma G.2. We will prove the lemma thanks to the following equation, true for any $\varepsilon > 0$ and k even (the choice of k will be specified later):

$$\mathbb{P}(\|\widetilde{H}\| > 2 + \varepsilon) \leqslant \mathbb{P}(\operatorname{Tr}\widetilde{H}^k \geqslant (2 + \varepsilon)^k) \leqslant (2 + \varepsilon)^{-k}\mathbb{E}\operatorname{Tr}\widetilde{H}^k. \tag{G.6}$$

We have, for any k even,

$$\mathbb{E}\operatorname{Tr}\widetilde{H}^k \;=\; \sum_{\mathbf{i}=(i_1,\dots,i_k)} \mathbb{E}\widetilde{H}_{i_1 i_2} \cdots \widetilde{H}_{i_k i_1} \;=\; \sum_{p=1}^{k} \sum_{\substack{\mathbf{i}=(i_1,\dots,i_k) \\ |\{i_1,\dots,i_k\}|=p}} \mathbb{E}\widetilde{H}_{i_1 i_2} \cdots \widetilde{H}_{i_k i_1}\,.$$

For each $\mathbf{i} = (i_1, \dots, i_k)$, let $G_\mathbf{i}$ be the simple, non oriented graph with vertex set $V_\mathbf{i} := \{i_1, \dots, i_k\}$ and edge set $E_\mathbf{i} := \{\{i_t, i_{t+1}\}\,;\, t = 1, \dots, k\}$ for $i_{k+1} := i_1$. For $p := |V_\mathbf{i}|$, denote also by $W_\mathbf{i}$ the p-tuple deduced from $\mathbf{i}$ by removing any entry which has already appeared when running $\mathbf{i}$ from i_1 to i_k (for example, if $\mathbf{i} = (1, 2, 5, 1, 3, 2, 4)$, then $W_\mathbf{i} = (1, 2, 5, 3, 4)$). For $\mathbb{E}\widetilde{H}_{i_1 i_2} \cdots \widetilde{H}_{i_k i_1}$ to be non zero, each edge of $G_\mathbf{i}$ has to be visited at least twice by the path $\gamma_\mathbf{i} := (i_1, \dots, i_k, i_1)$. It implies that $|E_\mathbf{i}| \leqslant k/2$, so that, as $G_\mathbf{i}$ is connected, $|V_\mathbf{i}| \leqslant k/2 + 1$.

For any $p = 1, \dots, k/2 + 1$ and any $W \in \{1, \dots, N\}^p$ with pairwise distinct entries, let $I_k(W)$ denote the set k-tuples $\mathbf{i} \in \{1, \dots, N\}^k$ such that

■ $W_\mathbf{i} = W$,

■ each edge of $G_\mathbf{i}$ has to be visited at least twice by the path $\gamma_\mathbf{i}$.

Then we have

$$\mathbb{E}\operatorname{Tr}\widetilde{H}^k \;=\; \sum_{p=1}^{k/2+1} \sum_{\substack{W \in \{1, \dots, N\}^p \\ \text{with distinct entries}}} \sum_{\mathbf{i} \in I_k(W)} \mathbb{E}\widetilde{H}_{i_1 i_2} \cdots \widetilde{H}_{i_k i_1}\,.$$

Note that Hypothesis (ii) implies that for

$$K \equiv K_N := \sup_{i,j} \|\widetilde{H}_{ij}\|_\infty, \tag{G.7}$$

for any $m \geqslant 2$,

$$\mathbb{E}|\widetilde{H}_{ij}|^m \leqslant N^{-1}K^{m-2}. \tag{G.8}$$

Hence that for any $p = 1, \ldots, k/2 + 1$ and any $\mathbf{i}$ such that $|V_\mathbf{i}| = p$, by (G.8), we have

$$\mathbb{E}\widetilde{H}_{i_1 i_2} \cdots \widetilde{H}_{i_k i_1} \leqslant N^{-(p-1)}K^{k-2(p-1)}.$$

Indeed, there are at least $p - 1$ independent factors in the product (because each $\ell \in \{2, \ldots, k\}$ such that $i_\ell \notin \{i_1, \ldots, i_{\ell-1}\}$ gives rise to a new independent factor). It follows that

$$\mathbb{E}\operatorname{Tr}\widetilde{H}^k \leqslant \sum_{p=1}^{k/2+1} N^{-(p-1)}K^{k-2(p-1)} \sum_{\substack{W \in \{1, \ldots, N\}^p \\ \text{with distinct entries}}} |I_k(W)|$$

Then, the key result we shall rely on is the following one, due to Vu in [78, p. 735].

Lemma G.4. *For any $p = 1, \ldots, k/2 + 1$ and any $W \in \{1, \ldots, N\}^p$ with distinct entries, we have*

$$|I_k(W)| \leqslant \binom{k}{2p-2} 2^{k+s+1} p^s (s+2)^s, \tag{G.9}$$

where $s := k - 2(p-1)$.

It follows that, using the notation $s = k - 2(p-1)$, $p = \frac{k-s}{2} + 1$,

$$\begin{aligned}
\mathbb{E}\operatorname{Tr}\widetilde{H}^k &\leqslant \sum_{p=1}^{k/2+1} N^{-(p-1)}K^{k-2(p-1)}N^p \binom{k}{2p-2} 2^{k+s+1} p^s (s+2)^s \\
&= N \sum_{s \in \{0,2,4,\ldots,k\}} K^s \binom{k}{s} 2^{k+s+1} \left(\frac{k-s}{2} + 1\right)^s (s+2)^s
\end{aligned}$$

Note that for any $k \geqslant 16$ (with the notation $x := s/k$),

$$2k^{-2}\left(\frac{k-s}{2} + 1\right)(s+2) = (1 - x + 2k^{-1})(x + 4k^{-1}) = (1-x)x + 2k^{-1}(2-x) + 8k^{-2} \leqslant 1.$$

Hence for $k \geqslant 16$,

$$\mathbb{E}\operatorname{Tr}\widetilde{H}^k \leqslant 2^{k+1}N \sum_{s=0}^{k} \binom{k}{s} \left(Kk^2\right)^s = 2^{k+1}N \left(1 + Kk^2\right)^k \tag{G.10}$$

Let us now choose $k \equiv k_N$ even such that for K as in (G.7), we have

$$(\log N)/k \longrightarrow 0 \qquad \text{and} \qquad k^2 K \longrightarrow 0$$

(such a k exists by Hypothesis (iii)). Then by (G.10), for any $\varepsilon, D > 0$, we have

$$N^D (2 + \varepsilon)^{-k}\mathbb{E}\operatorname{Tr}\widetilde{H}^k \longrightarrow 0,$$

which, by (G.6), concludes the proof of Lemma G.2. $\qquad\square$

H. Extreme eigenvalues: sub-Gaussian entries

In this appendix we provide a simple alternative proof that the extreme eigenvalues of H are bounded, following [39]. This argument requires sub-Gaussian entries and gives a weaker bound than Theorem G.1, establishing that $\|H\| \leqslant C$ for some constant C with high probability. On the other hand, it is very simple and gives strong bounds on the error probabilities. Note that any bound of the form $\|H\| \leqslant C$ is sufficient for the proof of Proposition 9.1, so that the argument given in this appendix may be used to replace entirely that of Appendix G if one strengthens the decay assumption in (iii) of Definition 2.2 to sub-Gaussian decay as in (H.1).

Theorem H.1. *Let H be a Hermitian $N \times N$ matrix with independent centred upper-triangular entries such that for some positive constant τ we have*

$$\mathbb{E}e^{\tau N |H_{ij}|^2} \leqslant \tau^{-1} \tag{H.1}$$

for all i, j. Then there are positive constants c, C depending only on τ such that for all $t > 0$ we have

$$\mathbb{P}(\|H\| \geqslant t) \leqslant e^{-c(t^2 - C)N}.$$

Proof. It is elementary (see e.g. [27, Lem. 1.4.2]) that for any N and any $\varepsilon > 0$, one can find a (deterministic) ε-net B of the unit sphere of $\mathbb{C}^N$ with cardinality $|B| \leqslant ((2 + \varepsilon)/\varepsilon)^{2N}$. This means that for any $\mathbf{v}$ in the unit sphere there exists $\mathbf{w} \in B$ such that $|\mathbf{v} - \mathbf{w}| \leqslant \varepsilon$. Taking $\varepsilon < 1$, we deduce that

$$\|H\|^2 \leqslant \frac{\max_{\mathbf{v} \in B} |H\mathbf{v}|^2}{1 - \varepsilon}.$$

Fixing $\varepsilon \in (0, 1)$ and using a union bound, we therefore have to prove that for some positive constants c, C and for any deterministic unit vector $\mathbf{v}$ we have

$$\mathbb{P}(|H\mathbf{v}|^2 \geqslant u) \leqslant e^{-c(u - C)N}$$

for all $u > 0$. By Markov's inequality, it suffices to prove that for some positive constants c, C and for any deterministic unit vector $\mathbf{v}$ we have

$$\mathbb{E}e^{cN|H\mathbf{v}|^2} \leqslant e^{cNC}.$$

For a unit vector $\mathbf{v} = (v_1, \ldots, v_N)$, using the independence of the upper-triangular entries of H and the Cauchy-Schwarz inequality we obtain, for any $c > 0$,

$$
\begin{aligned}
\mathbb{E}e^{cN|H\mathbf{v}|^2} &\leqslant \mathbb{E}e^{2cN\sum_i |\sum_{j<i} H_{ij}v_j|^2 + 2cN\sum_i |\sum_{j\geqslant i} H_{ij}v_j|^2} \\
&\leqslant \left(\mathbb{E}e^{4cN\sum_i |\sum_{j<i} H_{ij}v_j|^2} \mathbb{E}e^{4cN\sum_i |\sum_{j\geqslant i} H_{ij}v_j|^2} \right)^{1/2} \\
&= \prod_{i=1}^{N} \left(\mathbb{E}e^{4cN|\sum_{j<i} H_{ij}v_j|^2} \mathbb{E}e^{4cN|\sum_{j\geqslant i} H_{ij}v_i|^2} \right)^{1/2},
\end{aligned}
$$

so that the claim follows from Lemma H.2 below. $\qquad \square$

Lemma H.2. *Let $\mathbf{Y} \in \mathbb{C}^N$ be a random vector with independent centred entries Y_i satisfying*

$$\mathbb{E}e^{\delta|Y_i|^2} \;\leqslant\; \delta^{-1}$$

for all i, where δ is a positive constant. Then there is a $\tau > 0$ depending only on δ such that for any unit vector $\mathbf{v} \in \mathbb{C}^N$ we have

$$\mathbb{E}e^{\tau|\langle \mathbf{Y},\mathbf{v}\rangle|^2} \;\leqslant\; \tau^{-1}\,.$$

Proof. By decomposing both $\mathbf{Y}$ and $\mathbf{v}$ into real and imaginary parts and using Hölder's inequality, we can assume without loss of generality that both $\mathbf{Y}$ and $\mathbf{v}$ have real entries. Let then g be a standard real Gaussian variable, independent of the other variables, let $\mathbb{E}_g$ denote the expectation with respect to g and let $\mathbb{E}$ denote the expectation with respect to all other variables than g. We have, for any $\lambda > 0$,

$$\mathbb{E}e^{\lambda|\langle \mathbf{Y},\mathbf{v}\rangle|^2} = \mathbb{E}\mathbb{E}_g e^{\sqrt{2\lambda}\langle \mathbf{Y},\mathbf{v}\rangle g} \;=\; \mathbb{E}_g\mathbb{E}e^{\sqrt{2\lambda}\langle \mathbf{Y},\mathbf{v}\rangle g} \;=\; \mathbb{E}_g\prod_i \mathbb{E}e^{\sqrt{2\lambda}Y_i v_i g},$$

so that, by Lemma H.3 below,

$$\mathbb{E}e^{\lambda|\langle \mathbf{Y},\mathbf{v}\rangle|^2} \;\leqslant\; \mathbb{E}_g\prod_i e^{6\lambda g^2 v_i^2/\delta^3} \;\leqslant\; \mathbb{E}_g e^{6\lambda g^2/\delta^3}\,,$$

which is finite as soon as $12\lambda < \delta^3$. $\qquad\square$

Lemma H.3. *For any real centred random variable X, $r \in \mathbb{R}$, and $\delta > 0$, we have*

$$\mathbb{E}e^{rX} \;\leqslant\; 1 + 3(e^{r^2/\delta^2} - 1)\mathbb{E}e^{\delta X^2} \;\leqslant\; \exp\!\big(3r^2\delta^{-2}\mathbb{E}e^{\delta X^2}\big)\,.$$

Proof. The second inequality follows from the fact that for any $y \geqslant 1$ we have

$$1 + 3(e^{r^2/\delta^2} - 1)y \;=\; 1 + \sum_{n\geqslant 1}\frac{3y(r^2/\delta^2)^n}{n!} \;\leqslant\; 1 + \sum_{n\geqslant 1}\frac{(3yr^2/\delta^2)^n}{n!} \;=\; \exp\!\big(3r^2\delta^{-2}y\big)\,.$$

To prove the first inequality, it suffices to consider the case $r = 1$ (otherwise consider $\tilde{X} := rX$ and $\tilde{\delta} := \delta/r$). Using that $e^x \leqslant 1 + x + 3\frac{e^x + e^{-x} - 2}{2}$ for all $x \in \mathbb{R}$, we get

$$\mathbb{E}e^{X} \;\leqslant\; 1 + 3\sum_{n\geqslant 1}\frac{\mathbb{E}X^{2n}}{(2n)!} \;\leqslant\; 1 + 3\sum_{n\geqslant 1}\frac{n!\delta^{-2n}\mathbb{E}e^{\delta X^2}}{(2n)!} \;\leqslant\; 1 + 3(e^{1/\delta^2} - 1)\mathbb{E}e^{\delta X^2}\,,$$

as claimed. $\qquad\square$

Acknowledgements

References

[1] O. Ajanki, L. Erdős, T. Krüger, *Quadratic vector equations on complex upper half-plane.* Preprint arXiv:1506.05095.

[2] O. Ajanki, L. Erdős, T. Krüger, *Universality for general Wigner-type matrices.* Preprint arXiv:1506.05098.

[3] G. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices.* Cambridge Studies in Advanced Mathematics 118, Cambridge, 2009.

[4] Z.D. Bai, J.W. Silverstein, *Spectral analysis of large dimensional random matrices.* Second Edition, Springer, New York, 2009.

[5] Z. Bao, L. Erdős, *Delocalization for a class of random block band matrices.* Preprint arXiv:1503.07510.

[6] Z. Bao, L. Erdős, K. Schnelli, *Local Stability of the Free Additive Convolution.* Preprint arXiv:1508.05905.

[7] Z. Bao, L. Erdős, K. Schnelli, *Local law of addition of random matrices on optimal scale.* Preprint arXiv:1509.07080.

[8] R. Bauerschmidt, A. Knowles, H.-T. Yau, *Local semicircle law for random regular graphs.* Preprint arXiv:1503.08702.

[9] F. Benaych-Georges, *Local Single Ring Theorem.* Preprint arXiv:1501.07840.

[10] F. Benaych-Georges, A. Guionnet, M. Maïda, *Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices.* Electron. J. Probab. Vol. 16 (2011), no. 60, 1621–1662.

[11] F. Benaych-Georges, A. Guionnet, C. Male, *Central limit theorems for linear statistics of heavy tailed random matrices.* Comm. Math. Phys. 329 (2014), no. 2, 641–686.

[12] F. Benaych-Georges, A. Maltsev, *Fluctuations of linear statistics of half-heavy-tailed random matrices.* Preprint arXiv:1410.5624.

[13] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices.* Electron. J. Probab. 19 (2014), no. 33, 1–53.

[14] A. Bloemendal, A. Knowles, H.-T. Yau, J. Yin, *On the principal components of sample covariance matrices.* To appear in Probab. Theory Related Fields.

[15] L. Bogachev, S.A. Molchanov, L.A. Pastur, *On the density of states of random band matrices.* (Russian) Mat. Zametki 50 (1991), no. 6, 31–42; translation in Math. Notes 50 (1992), no. 5–6, 1232–1242.

[16] C. Bordenave, P. Caputo, D. Chafaï, *Spectrum of non-Hermitian heavy tailed random matrices.* Comm. Math. Phys. 307 (2011), no. 2, 513–560.

[17] C. Bordenave, A. Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices.* Probab. Theory Related Fields 157 (2013), no. 3–4, 885–953.

[18] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities*. Oxford University Press, Oxford, 2013.

[19] P. Bourgade, L. Erdős, H.-T. Yau, *Bulk universality of general β-ensembles with non-convex potential*. J. of Math. Phys. 53, special issue in honor of E. Lieb's 80th birthday, (2012), no. 9.

[20] P. Bourgade, L. Erdős, H.-T. Yau, *Universality of general β-ensembles*. Duke Math. J. 163 (2014), no. 6, 1127–1190.

[21] P. Bourgade, L. Erdős, H.-T. Yau, *Edge Universality of β-ensembles*. Comm. Math. Phys. 332 (2014), no. 1, 261–353.

[22] P. Bourgade, H.-T. Yau, *The eigenvector moment flow and local quantum unique ergodicity*. Preprint arXiv:1312.1301.

[23] P. Bourgade, H.-T. Yau, J. Yin, *Local circular law for random matrices*, Probab. Theory Related Fields 159 (2014), no. 3, 545–595.

[24] P. Bourgade, H.-T. Yau, J. Yin, *The local circular law II: the edge case*. Probab. Theory Related Fields 159 (2014), no. 3–4, 619–660.

[25] C. Cacciapuoti , A. Maltsev, B. Schlein, *Bounds for the Stieltjes transform and the density of states of Wigner matrices*. Probab. Theory Related Fields 163 (2014), no. 1, 1–29.

[26] C. Cacciapuoti , A. Maltsev, B. Schlein, *Local Marchenko-Pastur law at the hard edge of sample covariance matrices*. J. Math. Phys. 54 (2013).

[27] D. Chafaï, O. Guédon, G. Lecué, A. Pajor, *Interactions between compressed sensing random matrices and high dimensional geometry*. Panoramas et Synthèses 37. Société Mathématique de France, Paris, 2012.

[28] S. Chatterjee, *A generalization of the Lindeberg principle*. Ann. Probab. 34 (2006), no. 6, 2061–2076.

[29] F.J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*. J. Mathematical Phys. 3 (1962) 119–1198.

[30] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*. Russian Mathematical Surveys 66 (2011), no. 3, 507–626.

[31] L. Erdős, A. Knowles, *Quantum diffusion and eigenfunction delocalization in a random band matrix model*. Comm. Math. Phys. 303 (2011), no. 2, 509–554.

[32] L. Erdős, A. Knowles, *Quantum diffusion and delocalization for band matrices with general distribution*. Ann. Henri Poincaré 12 (2011), no. 7, 1227–1319.

[33] L. Erdős, A. Knowles, H.-T. Yau, *Averaging Fluctuations in Resolvents of Random Band Matrices*. Ann. Henri Poincaré 14 (2013), no. 8, 1837–1926.

[34] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *Spectral Statistics of Erdős-Rényi Graphs I: Local Semicircle Law*. Ann. Probab. 41 (2013), no. 3B, 2279–2375.

[35] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues.* Comm. Math. Phys. 314 (2012), no. 3, 587–640.

[36] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *Delocalization and Diffusion Profile for Random Band Matrices.* Comm. Math. Phys. 323 (2013), no. 1, 367–416.

[37] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *The local semicircle law for a general class of random matrices.* Electron. J. Probab 18 (2013), 1–58.

[38] L. Erdős, J. Ramírez, B. Schlein, H.-T. Yau, *Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation.* Electron. J. Probab. 15 (2010), no. 18, 526–603.

[39] L. Erdős, B. Schlein, H.T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices.* Ann. Probab. 37 (2009), no. 3, 815–852.

[40] L. Erdős, B. Schlein, H.T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices.* Comm. Math. Phys. 287 (2009), 641–655.

[41] L. Erdős, B. Schlein, H.T. Yau, *Wegner estimate and level repulsion for Wigner random matrices.* Int. Math. Res. Not. 2010 (2009), 436–479.

[42] L. Erdős, B. Schlein, H.T. Yau, J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices.* Ann. Inst. Henri Poincaré Probab. Stat. 48 (2012), no. 1, 1–46.

[43] L. Erdős, H.-T. Yau, J. Yin, *Bulk universality for generalized Wigner matrices.* Probab. Theor. Rel. Fields 154 (2012), no. 1–2, 341–407.

[44] L. Erdős, H.-T. Yau, J. Yin, *Universality for generalized Wigner matrices with Bernoulli distribution.* J. of Combinatorics 1 (2011), no. 2, 15–85

[45] L. Erdős, H.-T. Yau, J. Yin, *Rigidity of Eigenvalues of Generalized Wigner Matrices.* Adv. Math. 229 (2012), no. 3, 1435–1515.

[46] V.L. Girko, *The circular law.* Teor. Veroyatnost. i Primenen. (Russian) 29 (1984), 669–679.

[47] Z. Füredi, J. Komlós, *The eigenvalues of random symmetric matrices.* Combinatorica 1 (1981), no. 3, 233–241.

[48] M. Gaudin, *Sur la loi limite de l'espacement des valeurs propres d'une matrice aléatoire.* Nuclear Physics 25 (1961), 447–458.

[49] V. Kargin, *A concentration inequality and a local law for the sum of two random matrices.* Probab. Theory Related Fields 154 (2012), no. 3–4, 677–702.

[50] V. Kargin, *An inequality for the distance between densities of free convolutions.* Ann. Probab. 41 (2013), no. 5, 3241–3260.

[51] V. Kargin, *Subordination of the resolvent for a sum of random matrices.* Ann. Probab. 43 (2015), no. 4, 2119–2150.

[52] A. Khorunzhy, *On smoothed density of states for Wigner random matrices*. Rand. Op. Stoch. Eq. 4 (1997), no. 2, 147–162.

[53] A. Knowles, J. Yin, *Eigenvector Distribution of Wigner Matrices*. Probab. Theory Related Fields 155 (2013), 543–582.

[54] A. Knowles, J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*. Comm. Pure Appl. Math. 66 (2013), no. 11, 1663–1750.

[55] A. Knowles, J. Yin, *The outliers of a deformed Wigner matrix*. Ann. Probab. 42 (2014), no. 5, 1980–2031.

[56] A. Knowles, J. Yin, *Anisotropic local laws for random matrices*. Preprint arXiv:1410.3516.

[57] J.W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*. Math. Z. 15 (1922), 211–225.

[58] V.A. Marchenko, L.A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*. Mat. Sb. 72 (1967), no. 114, 507–536.

[59] C. McDiarmid, *On the method of bounded differences*. Surveys in combinatorics, London Math. Soc. Lecture Note Ser. 141, Cambridge Univ. Press, Cambridge (1989), pp. 148–188.

[60] M.L. Mehta, *Random Matrices*. Second Edition, Academic Press, Boston, 1991.

[61] S.A. Molchanov, L.A. Pastur, A.M. Khorunzhii, *Distribution of the eigenvalues of random band matrices in the limit of their infinite order*. (Russian. English, Russian summary) Teoret. Mat. Fiz. 90 (1992), no. 2, 163–178; translation in Theoret. and Math. Phys. 90 (1992), no. 2, 108–118.

[62] S. O'Rourke, V. Vu, *Universality of local eigenvalue statistics in random matrices with external source*. Random Matrices Theory Appl. 3 (2014), no. 2, 37 pp.

[63] L.A. Pastur, M. Shcherbina, *Eigenvalue distribution of large random matrices*. Mathematical Surveys and Monographs, 171. American Mathematical Society, Providence, RI, 2011.

[64] S. Péché, A. Soshnikov, *Wigner random matrices with non-symmetrically distributed entries*. J. Stat. Phys. 129 (2007), no. 5–6, 857–884.

[65] N.S. Pillai, J. Yin, *Universality of covariance matrices*. Ann. Appl. Probab. 24 (2014), no. 3, 935–1001.

[66] J.A. Ramírez, B. Rider, B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*. J. Amer. Math. Soc. 24 (2011), no. 4, 919–944.

[67] M. Shcherbina, B. Tirozzi, *Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs*. J. Math. Phys. 51 (2010), no. 2, 20 pp.

[68] Y.G. Sinai, A. Soshnikov, *A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices*. (Russian) Funktsional. Anal. i Prilozhen. 32 (1998), no. 2, 56–79, 96; translation in Funct. Anal. Appl. 32 (1998), no. 2, 114–131.

[69] A. Soshnikov, *Universality at the edge of the spectrum in Wigner random matrices*. Comm. Math. Phys. 207 (1999), no. 3, 697–733.

[70] A. Soshnikov, *A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices*. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. J. Statist. Phys. 108 (2002), no. 5–6, 1033–1056.

[71] D. Stroock, *Probability theory, and analytic view*. Second edition. Cambridge University Press, Cambridge, 2011.

[72] T. Tao, *Topics in random matrix theory*. Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2012.

[73] T. Tao, V. Vu, *Random matrices: universality of local eigenvalue statistics up to the edge*. Comm. Math. Phys. 298 (2010), no. 2, 549–572.

[74] T. Tao, V. Vu, *Random matrices: universality of local eigenvalue statistics*. Acta Math. 206 (2011), no. 1, 127–204.

[75] T. Tao, V. Vu, *Random matrices: universal properties of eigenvectors*. Rand. Mat. Theor. Appl. 1 (2012), 1150001.

[76] C. Tracy, H. Widom, *Level spacing distribution and Airy kernel*. Commun. Math. Phys. 159 (1994), no.1, 151–174.

[77] C. Tracy, H. Widom, *On orthogonal and symplectic matrix ensembles*. Commun. Math. Phys. 177 (1996), no. 3, 727–754.

[78] V. Vu, *Spectral norm of random matrices*. Combinatorica 27 (2007), no. 6, 721–736.

[79] E.P. Wigner, *On the distribution of the roots of certain symmetric matrices*. Ann. of Math. 67 (1958), 325–327.

[80] J. Yin, *The local circular law III: general case*. Probab. Theory Related Fields 160 (2014), no. 3, 679–732.

Florent Benaych-Georges, MAP 5 (CNRS UMR 8145) - Université Paris Descartes, 45 rue des Saints-Pères 75270 Paris cedex 6, France. Email: florent.benaych-georges@parisdescartes.fr.

Antti Knowles, University of Geneva, Section of Mathematics, 2-4 Rue du Lièvre, 1211 Genève 4, Switzerland. Email: antti.knowles@unige.ch.