Communications in
**Mathematical
Physics**

Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge

Terence Tao[1],[*], Van Vu[2],[**]

[1] Department of Mathematics, UCLA, Los Angeles CA 90095-1555, USA. E-mail: tao@math.ucla.edu
[2] Department of Mathematics, Rutgers, Piscataway, NJ 08854, USA. E-mail: vanvu@math.rutgers.edu

Abstract: This is a continuation of our earlier paper (Tao and Vu, http://arxiv.org/abs/0908.1982v4[math.PR], 2010) on the universality of the eigenvalues of Wigner random matrices. The main new results of this paper are an extension of the results in Tao and Vu (http://arxiv.org/abs/0908.1982v4[math.PR], 2010) from the bulk of the spectrum up to the edge. In particular, we prove a variant of the universality results of Soshnikov (Commun Math Phys 207(3):697–733, 1999) for the largest eigenvalues, assuming moment conditions rather than symmetry conditions. The main new technical observation is that there is a significant bias in the Cauchy interlacing law near the edge of the spectrum which allows one to continue ensuring the delocalization of eigenvectors.

1. Introduction

In our recent paper [27], a universality result (the Four Moment Theorem) was established for the eigenvalue spacings in the bulk of the spectrum of random Hermitian matrices. (See [6] for an extended discussion of the universality phenomenon, and [27] for further references on universality results in the context of Wigner Hermitian matrices.)

The main purpose of this paper is to extend this universality result to the edge of the spectrum as well.

1.1. Universality in the bulk. To recall the Four Moment Theorem, we need some notation.

Definition 1.1 *(Condition* **C0***). A random Hermitian matrix $M_n = (\zeta_{ij})_{1 \le i, j \le n}$ is said to obey* **condition C0** *if*

- *The ζ_{ij} are independent (but not necessarily identically distributed) for $1 \leq i \leq j \leq n$. For $1 \leq i < j \leq n$, they have mean zero and variance 1; for $i = j$, they have mean zero and variance c for some fixed $c > 0$ independent of n.*
- *(Uniform exponential decay) There exist constants $C, C' > 0$ such that*

$$\mathbf{P}(|\zeta_{ij}| \geq t^C) \leq \exp(-t) \tag{1}$$

for all $t \geq C'$ and $1 \leq i, j \leq n$.

Examples of random Hermitian matrices obeying Condition **C0** include the GUE and GOE ensembles, or the random symmetric Bernoulli ensemble in which each of the ζ_{ij} are equal to ± 1 with equal probability $1/2$. In GOE one has $c = 2$, but in the other two cases one has $c = 1$. The arguments in the previous paper [27] were largely phrased for the case $c = 1$, but it is not difficult to see that the arguments extend without difficulty to other values of c (the main point being that a modification of the variance of a single entry of a row vector does not significantly affect the Talagrand concentration inequality, [27, Lemma 43], or Lemma 2.1 below.).

Given an $n \times n$ Hermitian matrix A, we denote its n eigenvalues as

$$\lambda_1(A) \leq \cdots \leq \lambda_n(A),$$

and write $\lambda(A) := (\lambda_1(A), \ldots, \lambda_n(A))$. We also let $u_1(A), \ldots, u_n(A) \in \mathbb{C}^n$ be an orthonormal basis of eigenvectors of A with $Au_i(A) = \lambda_i(A)u_i(A)$; these eigenvectors $u_i(A)$ are only determined up to a complex phase even when the eigenvalues are simple, but this ambiguity will not cause a difficulty in our results as we will only be interested in the *magnitude* $|u_i(A)^* X|$ of various inner products $u_i(A)^* X$ of $u_i(A)$ with other vectors X.

It will be convenient to introduce the following notation for frequent events depending on n, in increasing order of likelihood:

Definition 1.2 *(Frequent events). Let E be an event depending on n.*

- *E holds **asymptotically almost surely** if [1] $\mathbf{P}(E) = 1 - o(1)$.*
- *E holds **with high probability** if $\mathbf{P}(E) \geq 1 - O(n^{-c})$ for some constant $c > 0$.*
- *E holds **with overwhelming probability** if $\mathbf{P}(E) \geq 1 - O_C(n^{-C})$ for **every** constant $C > 0$ (or equivalently, that $\mathbf{P}(E) \geq 1 - \exp(-\omega(\log n))$).*
- *E holds **almost surely** if $\mathbf{P}(E) = 1$.*

Definition 1.3 *(Moment matching). We say that two complex random variables ζ and ζ' **match to order** k if*

$$\mathbf{E}\mathrm{Re}(\zeta)^m \mathrm{Im}(\zeta)^l = \mathbf{E}\mathrm{Re}(\zeta')^m \mathrm{Im}(\zeta')^l$$

for all $m, l \geq 0$ such that $m + l \leq k$.

The first main result [27] can now be stated as follows:

Theorem 1.4 *(Four Moment Theorem) [27, Theorem 15]. There is a small positive constant c_0 such that for every $0 < \varepsilon < 1$ and $k \geq 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \leq i,j \leq n}$ and $M'_n = (\zeta'_{ij})_{1 \leq i,j \leq n}$ be two random matrices satisfying **C0**. Assume furthermore that for any $1 \leq i < j \leq n$, ζ_{ij} and ζ'_{ij} match to order 4 and for any $1 \leq i \leq n$,*

---

[1] See Sect. 1.4 for our conventions on asymptotic notation.

ζ_{ii} and ζ'_{ii} match to order 2. Set $A_n := \sqrt{n} M_n$ and $A'_n := \sqrt{n} M'_n$, and let $G : \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds

$$|\nabla^j G(x)| \le n^{c_0} \tag{2}$$

for all $0 \le j \le 5$ and $x \in \mathbb{R}^k$. Then for any $\varepsilon n \le i_1 < i_2 \cdots < i_k \le (1 - \varepsilon)n$, and for n sufficiently large depending on ε, k (and the constants C, C' in Definition 1.2) we have

$$|\mathbf{E}(G(\lambda_{i_1}(A_n), \ldots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \ldots, \lambda_{i_k}(A'_n)))| \le n^{-c_0}. \tag{3}$$

If ζ_{ij} and ζ'_{ij} only match to order 3 rather than 4, then there is a positive constant C independent of c_0 such that the conclusion (3) still holds provided that one strengthens (2) to

$$|\nabla^j G(x)| \le n^{-Cjc_0}$$

for all $0 \le j \le 5$ and $x \in \mathbb{R}^k$.

Informally, this theorem asserts that the distribution of any bounded number of eigenvalues in the bulk of the spectrum of a random Hermitian matrix obeying condition **C0** depends only on the first four moments of the coefficients.

There is also a useful companion result to Theorem 1.4, which is used both in the proof of that theorem, and in several of its applications:

Theorem 1.5 (*Lower tail estimates*) [27, Theorem 17]. *Let* $0 < \varepsilon < 1$ *be a constant, and let* M_n *be a random matrix obeying Condition* **C0**. *Set* $A_n := \sqrt{n} M_n$. *Then for every* $c_0 > 0$, *and for* n *sufficiently large depending on* ε, c_0 *and the constants* C, C' *in Definition 1.1, and for each* $\varepsilon n \le i \le (1 - \varepsilon)n$, *one has* $\lambda_{i+1}(A_n) - \lambda_i(A_n) \ge n^{-c_0}$ *with high probability. In fact, one has*

$$\mathbf{P}(\lambda_{i+1}(A_n) - \lambda_i(A_n) \le n^{-c_0}) \le n^{-c_1}$$

for some $c_1 > 0$ *depending on* c_0 *(and independent of* ε).

Theorem 1.4 (and to a lesser extent, Theorem 1.5) can be used to extend the range of applicability for various results on eigenvalue statistics in the bulk for Hermitian or symmetric matrices, for instance in extending results for special ensembles such as GUE or GOE (or ensembles obeying some regularity or divisibility conditions) to more general classes of matrices. See [27,13,10] for some examples of this type of extension. We also remark that a level repulsion estimate which has a similar spirit to Theorem 1.5 was established in [9, Theorem 3.5], although the latter result establishes repulsion of eigenvalues in a fixed small interval I, rather than at a fixed index i of the sequence of eigenvalues, and does not seem to be directly substitutable for Theorem 1.5 in the arguments of this paper.

The results of Theorem 1.4 and Theorem 1.5 only control eigenvalues $\lambda_i(A_n)$ in the *bulk* region $\varepsilon n \le i \le (1 - \varepsilon)n$ for some fixed $\varepsilon > 0$ (independent of n). The reason for this restriction was technical, and originated from the use of the following two related results (which are variants of previous results of Erdős, Schlein, and Yau[7–9]), whose proof relied on the assumption that one was in the bulk:

Theorem 1.6 *(Concentration for ESD)* [27, Theorem 56]. *For any $\varepsilon, \delta > 0$ and any random Hermitian matrix $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ whose upper-triangular entries are independent with mean zero and variance 1, and such that $|\zeta_{ij}| \leq K$ almost surely for all i, j and some $1 \leq K \leq n^{1/2-\varepsilon}$, and any interval I in $[-2+\varepsilon, 2-\varepsilon]$ of width $|I| \geq \frac{K^2 \log^{20} n}{n}$, the number of eigenvalues N_I of $W_n := \frac{1}{\sqrt{n}} M_n$ in I obeys the concentration estimate*

$$|N_I - n \int_I \rho_{sc}(x) \, dx| \ll \delta n |I|$$

with overwhelming probability, where ρ_{sc} is the semicircular distribution

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases} \tag{4}$$

In particular, $N_I = \Theta_\varepsilon(n|I|)$ with overwhelming probability.

Proposition 1.7 *(Delocalization of eigenvectors)* [27, Prop. 58]. *Let ε, M_n, W_n, ζ_{ij}, K be as in Theorem 1.6. Then for any $1 \leq i \leq n$ with $\lambda_i(W_n) \in [-2 + \varepsilon, 2 - \varepsilon]$, if $u_i(W_n)$ denotes a unit eigenvector corresponding to $\lambda_i(W_n)$, then with overwhelming probability each coordinate of $u_i(M_n)$ is $O_\varepsilon(\frac{K^2 \log^{20} n}{n^{1/2}})$.*

In the bulk region $[-2 + \varepsilon, 2 - \varepsilon]$, the semicircular function ρ_{sc} is bounded away from zero. Thus, Theorem 1.6 ensures that the eigenvalues of W_n in the bulk tend to have a mean spacing of $\Theta_\varepsilon(1/n)$ on the average. Applying the Cauchy interlacing law

$$\lambda_i(W_n) \leq \lambda_i(W_{n-1}) \leq \lambda_{i+1}(W_n), \tag{5}$$

where W_{n-1} is an $n - 1 \times n - 1$ minor of W_n, this implies that the bulk eigenvalues of W_{n-1} are within $\Theta_\varepsilon(1/n)$ of the corresponding eigenvalues of W_n on the average. Using linear algebra to express the coordinates of the eigenvector $u_i(M_n)$ in terms of W_n and a minor W_{n-1} (see Lemma 4.1 below), and using some concentration of measure results concerning the projection of a random vector to a subspace (see Lemma 2.1), we eventually obtain Proposition 1.7.

1.2. Universality up to the edge. The main results of this paper are that the above four theorems can be extended to the edge of the spectrum (thus effectively sending ε to zero). Let us now state these results more precisely. Firstly, we have the following extension of Theorem 1.6:

Theorem 1.8 *(Concentration for ESD up to edge). Consider a random Hermitian matrix $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ whose upper-triangular entries are independent with mean zero and variance 1, and such that $|\zeta_{ij}| \leq K$ almost surely for all i, j and some $K \geq 1$. Let $0 < \delta < 1/2$ be a quantity which can depend on n, and let I be an interval such that*

$$|I| \geq \frac{K^2 \log^4 n}{n \delta^{10}}.$$

We also make the mild assumption $K = o(n^{1/2} \delta^2)$. Then the number of eigenvalues N_I of $W_n := \frac{1}{\sqrt{n}} M_n$ in I obeys the concentration estimate

$$|N_I - n \int_I \rho_{sc}(x) \, dx| \ll \delta n |I|$$

with overwhelming probability.

Remark 1.9. The powers of K, δ and $\log n$ here are probably not best possible, but this will not be relevant for our purposes. In our applications K will be a power of $\log n$, and δ will be a negative power of $\log n$. (This allows the error term $O(\delta n|I|)$ in the above estimate for N_I to exceed the main term $n \int_I \rho_{sc}(x)\, dx$ when one is very near the edge, but this will not impact our arguments.)

We prove this theorem in Sect. 3, using the same (standard) Stieltjes transform method that was used to prove Theorem 1.6 in [27] (see also [9]), with a somewhat more careful analysis. We next use it to obtain the following extension of Proposition 1.7:

Proposition 1.10 (*Delocalization of eigenvectors up to the edge*). *Let M_n be a random matrix obeying Condition* **C0***. Then with overwhelming probability, every unit eigenvector $u_i(M_n)$ of M_n has coefficients at most $n^{-1/2} \log^{O(1)} n$, thus*

$$\sup_{1 \le i, j \le n} |u_i(M_n)^* e_j| \ll n^{-1/2} \log^{O(1)} n,$$

where $e_1, \ldots, e_n$ is the standard basis.

The deduction of Proposition 1.10 from Theorem 1.8 differs significantly from the analogous deduction of Proposition 1.7 in Theorem 1.6 in [27]. The main difference is that in the current case ρ_{sc} is no longer bounded away from zero, which causes the average eigenvalue spacing between $\lambda_i(W_n)$ and $\lambda_{i+1}(W_n)$ to be considerably larger than $1/n$. For instance, the gap between the second largest eigenvalue $\lambda_{n-1}(W_n)$ and the largest eigenvalue $\lambda_n(W_n)$ is typically of size $n^{-2/3}$.

The key new ingredient that helps us to deal with this problem is the following observation: the Cauchy interlacing law (5), when applied to the eigenvalues of the edge, is strongly *bias*. In particular, the gap between $\lambda_i(W_{n-1})$ and $\lambda_i(W_n)$ is significantly smaller than the gap between $\lambda_i(W_{n-1})$ and $\lambda_{i+1}(W_n)$. We can show that (with high probability), the first gap is of order $n^{-1+o(1)}$ while the second can be as large as $n^{-2/3}$ (and similarly for the gap between $\lambda_{i+1}(W_n)$ and $\lambda_{i-1}(W_{n-1})$ when $n/2 \le i \le n$). This new ingredient will be sufficient to recover Proposition 1.10; see Sect. 4, where the above proposition is proved.

Using Theorem 1.8 and Proposition 1.10, one can continue the arguments from [27] to establish the following extensions of Theorem 1.4 and Theorem 1.5:

Theorem 1.11 (*Four Moment Theorem up to the edge*). *There is a small positive constant c_0 such that for every $k \ge 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \le i, j \le n}$ and $M'_n = (\zeta'_{ij})_{1 \le i, j \le n}$ be two random matrices satisfying* **C0***. Assume furthermore that for any $1 \le i < j \le n$, ζ_{ij} and ζ'_{ij} match to order 4 and for any $1 \le i \le n$, ζ_{ii} and ζ'_{ii} match to order 2. Set $A_n := \sqrt{n} M_n$ and $A'_n := \sqrt{n} M'_n$, and let $G : \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds (2) for all $0 \le j \le 5$ and $x \in \mathbb{R}^k$. Then for any $1 \le i_1 < i_2 \cdots < i_k \le n$, and for n sufficiently large depending on k (and the constants C, C' in Definition 1.1) we have (3). If ζ_{ij} and ζ'_{ij} only match to order 3 rather than 4, then there is a positive constant C independent of c_0 such that the conclusion (3) still holds provided that one strengthens (2) to*

$$|\nabla^j G(x)| \le n^{-Cjc_0} \tag{6}$$

for all $0 \le j \le 5$ and $x \in \mathbb{R}^k$.

Theorem 1.12 *(Lower tail estimates up to the edge). Let M_n be a random matrix obeying Condition* **C0**. *Set $A_n := \sqrt{n} M_n$. Then for every $c_0 > 0$, and for n sufficiently large depending on c_0 and the constants C, C' in Definition 1.1, and for each $1 \leq i \leq n$, one has $\lambda_{i+1}(A_n) - \lambda_i(A_n) \geq n^{-c_0}$ with high probability, uniformly in i.*

The novelty here is that we have no assumption on the indices i_j and i. We present the proof of these theorems in Sects. 5, 6, following the arguments in [27] closely.

1.3. Applications. As Theorems 1.11, 1.12 extend Theorems 1.4, 1.5, all the applications of the latter theorems in [27] (concerning the bulk of the spectrum) can also be viewed as applications of these theorems. But because these results extend all the way to the edge, we can now obtain some results on the edge of the spectrum as well. For instance, we can prove

Theorem 1.13. *Let k be a fixed integer and M_n be a matrix obeying Condition* **C0**, *and suppose that the real and imaginary part of the atom variables have the same covariance matrix as the GUE ensemble (i.e. both components have variance $1/2$, and have covariance 0). Assume furthermore that all third moments of the atom variables vanish. Set $W_n := \frac{1}{\sqrt{n}} M_n$. Then the joint distribution of the k dimensional random vector*

$$\left((\lambda_n(W_n) - 2)n^{2/3}, \ldots, (\lambda_{n-k+1}(W_n) - 2)n^{2/3} \right) \tag{7}$$

has a weak limit as $n \to \infty$, which coincides with that in the GUE case (in particular, the limit for $k = 1$ is the GUE Tracy-Widom distribution [28], and for higher k is controlled by the Airy kernel [14]). The result also holds for the smallest eigenvalues $\lambda_1, \ldots, \lambda_k$, with the offset -2 replaced by $+2$.

If the atom variables have the same covariance matrix as the GOE ensemble (i.e. they are real with variance 1 off the diagonal, and 2 on the diagonal), instead of the GUE ensemble, then the same conclusion applies but with the GUE distribution replaced of course by the GOE distribution (see [29] for the $k = 1$ case).

This result was previously established by Soshnikov [25] (see also [23,24]) in the case when M_n is a Wigner Hermitian matrix (i.e. the off-diagonal entries are iid, and the matrix matches GUE to second order at least) with symmetric distribution (which implies, but is stronger than, matching to third order). For some additional partial results in the non-symmetric case see [20,21]. The exponential decay condition in Soshnikov's result has been lowered to a finite number of moments; see [22,18]. It is reasonable to conjecture that the exponential decay conditions in this current paper can similarly be lowered, but we will not pursue this issue here. It also seems plausible that the third moment matching conditions could be dropped, though this is barely beyond the reach of the current method[2].

Proof. We just prove the claim for the largest k eigenvalues and for GUE, as the claim for the smallest k and/or GOE is similar.

Set $A_n := \sqrt{n} M_n$. It suffices to show that for every smooth function $G : \mathbb{R}^k \to \mathbb{R}$, that the expectation

$$\mathbf{E} G((\lambda_n(A_n) - 2n)/n^{1/3}, \ldots, (\lambda_{n-k+1}(A_n) - 2n)/n^{1/3}) \tag{8}$$

---

[2] *Note added in proof.* The third moment condition has recently been dropped in [16], by combining the four moment theorem here with a new proof of universality for the distribution of the largest eigenvalue for gauss divisible matrices.

only changes by $o(1)$ when the matrix M_n is replaced with GUE. But this follows from the final conclusion of Theorem 1.11, thanks to the extra factor $n^{-1/3}$. □

Remark 1.14. Notice that there is some room to spare in this argument, as the $n^{-1/3}$ gain in (8) is far more than is needed for (6). Because of this, one can obtain similar universality results for suitably normalised eigenvalues $\lambda_i(A_n)$ with $i \leq n^{1-\varepsilon}$ or $i \geq n - n^{1-\varepsilon}$ for any $\varepsilon > 0$ (where the normalisation factor for $\lambda_i(A_n)$ is now $n^{2/3} \min(i, n-i)^{1/3}$, and the offset -2 is replaced by $-t$, where $\int_{-2}^{t} \rho_{sc}(x)\, dx = \frac{i}{n}$). We omit the details.

Remark 1.15. In analogy with [13], one should be able to drop the third moment condition in Theorem 1.13 if one can control the distribution of the largest (or smallest) eigenvalues from random matrices obtained from a suitable Ornstein-Uhlenbeck process, as in [12].

1.4. Notation. We consider n as an asymptotic parameter tending to infinity. We use $X \ll Y$, $Y \gg X$, $Y = \Omega(X)$, or $X = O(Y)$ to denote the bound $X \leq CY$ for all sufficiently large n and for some constant C. Notations such as $X \ll_k Y$, $X = O_k(Y)$ mean that the hidden constant C depend on another constant k. $X = o(Y)$ or $Y = \omega(X)$ means that $X/Y \to 0$ as $n \to \infty$; the rate of decay here will be allowed to depend on other parameters. We write $X = \Theta(Y)$ for $Y \ll X \ll Y$.

We view vectors $x \in \mathbb{C}^n$ as column vectors. The Euclidean norm of a vector $x \in \mathbb{C}^n$ is defined as $\|x\| := (x^*x)^{1/2}$.

Eigenvalues are always ordered in increasing order, thus for instance $\lambda_n(A_n)$ is the largest eigenvalue of a Hermitian matrix A_n, and $\lambda_1(A_n)$ is the smallest.

2. General Tools

In this section we record some general tools (proven in [27]) which we will use repeatedly in the sequel. We begin with a very useful concentration of measure result that describes the projection of a random vector to a subspace.

Lemma 2.1 (*Projection Lemma*). *Let $X = (\xi_1, \ldots, \xi_n) \in \mathbb{C}^n$ be a random vector whose entries are independent with mean zero, variance 1, and are bounded in magnitude by K almost surely for some K, where $K \geq 10(\mathbf{E}|\xi|^4 + 1)$. Let H be a subspace of dimension d and π_H the orthogonal projection onto H. Then*

$$\mathbf{P}(|\|\pi_H(X)\| - \sqrt{d}| \geq t) \leq 10 \exp(-\frac{t^2}{10K^2}).$$

In particular, one has

$$\|\pi_H(X)\| = \sqrt{d} + O(K \log n)$$

with overwhelming probability.

The same conclusion holds (with 10 replaced by another explicit constant) if one of the entries ξ_j of X is assumed to have variance c instead of 1, for some absolute constant $c > 0$.

Proof. See [27, Lem. 40]. (The main tool in the proof is Talagrand's concentration inequality.) It is clear from the triangle inequality that the modification of variance in a single entry does not significantly affect the conclusion except for constants. □

Next, we record a crude but useful upper bound on the number of eigenvalues in a short interval.

Lemma 2.2 *(Upper bound on ESD). Consider a random Hermitian matrix $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ whose upper-triangular entries are independent with mean zero and variance 1 (with variance c on the diagonal for some absolute constant $c > 0$), and such that $|\zeta_{ij}| \leq K$ almost surely for all i, j and some $K \geq 1$. Set $W_n := \frac{1}{\sqrt{n}} M_n$. Then for any interval $I \subset \mathbb{R}$ with $|I| \geq \frac{K^2 \log^2 n}{n}$,*

$$N_I \ll n|I|$$

with overwhelming probability, where N_I is the number of eigenvalues of W_n in I.

Proof. See [27, Prop. 62]. (The main tools in the proof are the Stieltjes transform method, Lemma 3.3 below, and Lemma 2.1.) Again, the generalisation to variances other than 1 on the diagonal do not cause significant changes to the argument. □

Finally, we recall a Berry-Esséen type theorem:

Theorem 2.3 *(Tail bounds for complex random walks). Let $1 \leq N \leq n$ be integers, and let $A = (a_{i,j})_{1 \leq i \leq N; 1 \leq j \leq n}$ be an $N \times n$ complex matrix whose N rows are orthonormal in $\mathbb{C}^n$, and obeying the incompressibility condition*

$$\sup_{1 \leq i \leq N; 1 \leq j \leq n} |a_{i,j}| \leq \sigma \tag{9}$$

for some $\sigma > 0$. Let $\zeta_1, \ldots, \zeta_n$ be independent complex random variables with mean zero, variance $\mathbf{E}|\zeta_j|^2$ equal to 1, and obeying $\mathbf{E}|\zeta_i|^3 \leq C$ for some $C \geq 1$. For each $1 \leq i \leq N$, let S_i be the complex random variable

$$S_i := \sum_{j=1}^{n} a_{i,j} \zeta_j$$

and let $\vec{S}$ be the $\mathbb{C}^N$-valued random variable with coefficients $S_1, \ldots, S_N$:

- *(Upper tail bound on S_i) For $t \geq 1$, we have $\mathbf{P}(|S_i| \geq t) \ll \exp(-ct^2) + C\sigma$ for some absolute constant $c > 0$.*
- *(Lower tail bound on $\vec{S}$) For any $t \leq \sqrt{N}$, one has $\mathbf{P}(|\vec{S}| \leq t) \ll O(t/\sqrt{N})^{\lfloor N/4 \rfloor} + CN^4 t^{-3} \sigma$.*

The same claim holds if one of the ζ_i is assumed to have variance c instead of 1 for some absolute constant $c > 0$.

Proof. See [27, Th. 41]. Again, the modification of the variance on a single entry can be easily seen to have no substantial effect on the conclusion. □

3. Asymptotics for the ESD

In this section we prove Theorem 1.8, using the Stieltjes transform method (see [2] for a general discussion of this method). We may assume throughout that n is large, since the claim is vacuous for n small.

It is known by the moment method (see e.g. [2] or [4]) that with overwhelming probability, all eigenvalues of W_n have magnitude at most $2 + o(1)$. Because of this, we may restrict attention to the case when I lies in interval $[-3, 3]$ (say).

We recall the *Stieltjes transform* $s_n(z)$ of a Hermitian matrix W_n, defined for complex z by the formula

$$s_n(z) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(W_n) - z}. \tag{10}$$

We also introduce the semicircular counterpart

$$s(z) := \int_{-2}^{2} \frac{1}{x - z} \rho_{sc}(x)\, dx,$$

which by a standard contour integral computation can be given explicitly as

$$s(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4}), \tag{11}$$

where we use the branch of the square root of $z^2 - 4$ with cut at $[-2, 2]$ which is asymptotic to z at infinity.

It is well known that one can control the empirical spectral distribution N_I via the Stieltjes transform. We will use the following formalization of this principle:

Lemma 3.1 (*Control of Stieltjes transform implies control on ESD*). *There is a positive constant C such that the following holds for any Hermitian matrix W_n. Let $1/10 \geq \eta \geq 1/n$ and $L, \varepsilon, \delta > 0$. Suppose that one has the bound*

$$|s_n(z) - s(z)| \leq \delta \tag{12}$$

with (uniformly) overwhelming probability for all z with $|\mathrm{Re}(z)| \leq L$ and $\mathrm{Im}(z) \geq \eta$. Then for any interval I in $[-L + \varepsilon, L - \varepsilon]$ with $|I| \geq \max(2\eta, \frac{\eta}{\delta} \log \frac{1}{\delta})$, one has

$$\left| N_I - n \int_I \rho_{sc}(x)\, dx \right| \ll_\varepsilon \delta n |I|$$

with overwhelming probability, where N_I is the number of eigenvalues of W_n in I.

Proof. See [27, Lem. 60]. □

As a consequence of this lemma (with $L = 4$ and $\varepsilon = 1$, say), we see that Theorem 1.8 follows from

Theorem 3.2 (*Concentration for the Stieltjes transform up to edge*). *Consider a random Hermitian matrix $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ whose upper-triangular entries are independent with mean zero and variance 1, with variance c on the diagonal for some absolute constant $c > 0$, and such that $|\zeta_{ij}| \leq K$ almost surely for all i, j and some $K \geq 1$. Set $W_n := \frac{1}{\sqrt{n}} M_n$. Let $0 < \delta < 1/2$ (which can depend on n), and suppose that*

$K = o(n^{1/2}\delta^2)$. *Then* (12) *holds with (uniformly) overwhelming probability for all* z
with $|\mathrm{Re}(z)| \leq 4$ *and*

$$\mathrm{Im}(z) \geq \frac{K^2 \log^{3.5} n}{\delta^8 n}.$$

The remainder of this section is devoted to proving Theorem 3.2. Fix z as in
Theorem 3.2, thus $|\mathrm{Re}(z)| \leq 4$ and $\mathrm{Im}(z) = \eta$, where

$$\eta n \geq \frac{K^2 \log^{3.5} n}{\delta^8}. \tag{13}$$

Our objective is to show (12) with (uniformly) overwhelming probability.

As in previous works (in particular [9,27]), the key is to exploit the fact that when
$\mathrm{Im} z > 0$, $s(z)$ is the unique solution to the equation

$$s(z) + \frac{1}{s(z) + z} = 0 \tag{14}$$

with $\mathrm{Im} s(z) > 0$; this is immediate from (11).

We now seek a similar relation for s_n. Note that $\mathrm{Im} s_n(z) > 0$ by (10). We use the
following standard matrix identity (cf. [27, Lem. 39], or [2, Chap. 11]):

Lemma 3.3. *We have*

$$s_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\frac{1}{\sqrt{n}}\zeta_{kk} - z - Y_k}, \tag{15}$$

where

$$Y_k := a_k^* (W_{n,k} - zI)^{-1} a_k,$$

$W_{n,k}$ *is the matrix* W_n *with the* k^{th} *row and column removed, and* a_k *is the* k^{th} *row of*
W_n *with the* k^{th} *element removed.*

Proof. By Schur's complement, $\frac{1}{\zeta_{kk} - z - a_k^*(W_k - zI)^{-1}a_k}$ is the kth diagonal entry of
$(W - zI)^{-1}$. Taking traces, one obtains the claim. □

Proposition 3.4 (*Concentration of* Y_k). *For each* $1 \leq k \leq n$, *one has* $Y_k = s_n(z) + o(\delta^2)$
with overwhelming probability.

Proof Fix k, and write $z = x + \sqrt{-1}\eta$.

The entries of a_k are independent of each other and of $W_{n,k}$, and have mean zero and
variance $\frac{1}{n}$. By linearity of expectation we thus have, on conditioning on $W_{n,k}$,

$$\mathbf{E}(Y_k | W_{n,k}) = \frac{1}{n} \mathrm{trace}(W_{n,k} - zI)^{-1} = \left(1 - \frac{1}{n}\right) s_{n,k}(z),$$

where

$$s_{n,k}(z) := \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{1}{\lambda_i(W_{n,k}) - z}$$

is the Stieltjes transform of $W_{n,k}$. From the Cauchy interlacing law (5) and (13), we have

$$s_n(z) - (1 - \frac{1}{n})s_{n,k}(z) = O\left(\frac{1}{n}\int_{\mathbb{R}}\frac{1}{|x - z|^2}\, dx\right) = O\left(\frac{1}{n\eta}\right) = o(\delta^2).$$

It follows that

$$\mathbf{E}(Y_k|W_{n,k}) = s_n(z) + o(\delta^2),$$

and so it will remain to show the concentration estimate

$$Y_k - \mathbf{E}(Y_k|W_{n,k}) = o(\delta^2)$$

with overwhelming probability.

Rewriting Y_k, it suffices to show that

$$\sum_{j=1}^{n-1}\frac{R_j}{\lambda_j(W_{n,k}) - (x + \sqrt{-1}\eta)} = o(\delta^2) \tag{16}$$

with overwhelming probability, where $R_j := |u_j(W_{n,k})^* a_k|^2 - 1/n$.

Let $1 \leq i_- < i_+ \leq n$, then

$$\sum_{i_- \leq j \leq i_+} R_j = \|P_H a_k\|^2 - \frac{\dim(H)}{n},$$

where H is the space spanned by the $u_j(W_{n,k})^*$ for $i_- \leq j \leq i_+$. From Lemma 2.1 and the union bound, we conclude that with overwhelming probability

$$\left|\sum_{i_- \leq j \leq i_+} R_j\right| \ll \frac{\sqrt{i_+ - i_-}K \log n + K^2 \log^2 n}{n}. \tag{17}$$

By the triangle inequality, this implies that

$$\sum_{i_- \leq j \leq i_+} \|P_H a_k\|^2 \ll \frac{i_+ - i_-}{n} + \frac{\sqrt{i_+ - i_-}K \log n + K^2 \log^2 n}{n},$$

and hence by a further application of the triangle inequality

$$\sum_{i_- \leq j \leq i_+} |R_j| \ll \frac{(i_+ - i_-) + K^2 \log^2 n}{n} \tag{18}$$

with overwhelming probability.

The plan is to use (17) and (18) to establish (16). Accordingly, we split the LHS of (16), into several subsums according to the distance $|\lambda_j - x|$. Lemma 2.2 provides a sharp estimate on the number of terms of each subsum which will allow us to obtain a good upper bound on the absolute value.

We turn to the details. From (13) we can choose two auxiliary parameters $0 < \delta', \alpha < 1$ such that

$$\delta' = o(\delta^2);$$
$$\alpha \log n = o(\delta^2);$$
$$\alpha \delta' \eta n \geq K^2 \log^2 n; \tag{19}$$
$$\frac{K \log n}{\sqrt{\alpha \delta' \eta n}} = o(\delta^2).$$

Indeed, one could set $\delta' := \delta^2 \log^{-0.01} n$ and $\alpha := \delta^2 \log^{-1.01} n$ and use (13).

Fix such parameters, and consider the contribution to (16) of the indices j for which

$$|\lambda_j(W_n) - x| \leq \delta' \eta.$$

By Lemma 2.2 and (19), the interval of j for which this occurs has cardinality $O(\delta' \eta n)$ (with overwhelming probability). On this interval, the quantity $\frac{1}{\lambda_j(W_{n,k}) - (x+\sqrt{-1}\eta)}$ has magnitude $O(\frac{1}{\eta})$. Applying (18) (and (19)), we see that the contribution of this case is thus

$$\ll \frac{1}{\eta} \frac{\delta' \eta n}{n} = o(\delta^2),$$

which is acceptable.

Next, we consider the contribution to (16) of those indices j for which

$$(1+\alpha)^l \delta' \eta < |\lambda_j(W_n) - x| \leq (1+\alpha)^{l+1} \delta' \eta$$

for some integer $0 \leq l \ll \log n/\alpha$, and then sum over l. By Lemma 2.2 and (19), the set of j for which this occurs is contained (with overwhelming probability) in at most two intervals of cardinality $O((1+\alpha)^l \alpha \delta' \eta n)$. On each of these intervals, the quantity $\frac{1}{\lambda_j(W_{n,k}) - (x+\sqrt{-1}\eta)}$ has magnitude $O\left(\frac{1}{(1+\alpha)^l \delta' \eta}\right)$ and fluctuates by $O\left(\frac{\alpha}{(1+\alpha)^l \delta' \eta}\right)$. Applying (17), (18) (and noting that $(1+\alpha)^l \alpha \delta' \eta n$ exceeds $K^2 \log^2 n$, by (19)) we see that the contribution of a single l to (16) is at most

$$\ll \frac{1}{(1+\alpha)^l \delta' \eta} \frac{\sqrt{\alpha(1+\alpha)^l \delta' \eta n} K \log n}{n} + \frac{\alpha}{(1+\alpha)^l \delta' \eta} \frac{\alpha(1+\alpha)^l \delta' \eta n}{n},$$

which simplifies to

$$\ll \alpha(1+\alpha)^{-l/2} \frac{K \log n}{\sqrt{\alpha \delta' \eta n}} + \alpha^2.$$

Summing over l we obtain a bound of

$$\ll \frac{K \log n}{\sqrt{\alpha \delta' \eta n}} + \alpha \log n,$$

which is acceptable by (19). $\quad\square$

We now conclude the proof of Theorem 1.8. By hypothesis,

$$\left| \frac{1}{\sqrt{n}} \zeta_{kk} \right| \leq K/\sqrt{n} = o(\delta^2)$$

almost surely. Inserting these bounds into (15), we see that with overwhelming probability

$$s_n(z) + \frac{1}{n} \sum_{k=1}^{n} \frac{1}{s_n(z) + z + o(\delta^2)} = 0.$$

By the triangle inequality (and square rooting the $o()$ decay), we can assume that either the error term $o(\delta^2)$ is $o(\delta^2 |s_n(z) + z|)$, or that $|s_n(z) + z|$ is $o(1)$. Suppose the former holds. Then by Taylor expansion

$$\frac{1}{s_n(z) + z + o(\delta^2)} = \frac{1}{s_n(z) + z} + o(\delta^2),$$

and thus

$$s_n(z) + \frac{1}{s_n(z) + z} = o(\delta^2).$$

If we assume $|z| \leq 10$ (say), we conclude that $|s_n(z)| \leq 100$. Multiplying out by $s_n(z) + z$ and rearranging, we obtain

$$\left(s_n(z) + \frac{z}{2} \right)^2 = \frac{z^2 - 4}{4} + o(\delta^2).$$

Thus

$$s_n(z) + \frac{z}{2} = \pm \sqrt{\frac{z^2 - 4}{4}} + o(\delta)$$

(treating the case when $z^2 - 4 = o(\delta^2)$ separately).

To summarise, we have shown (with overwhelming probability) in the region

$$|z| \leq 10; \quad |\mathrm{Re}(z)| \leq 4; \quad \mathrm{Im}(z) \geq \frac{K^2 \log^{3.5} n}{\delta^8 n}$$

that one either has $s_n(z) = s(z) + o(\delta)$, $s_n(z) = -z - s(z) + o(1) = s(z) - \sqrt{z^2 - 4} + o(1)$, or $|s_n(z) + z| = o(1)$. It is not hard to see that the first two cases are disconnected from the third (for n large enough) in this region, because $s(z) = \frac{-1}{s(z)+z}$ is bounded away from zero, as is $s(z) + z = \frac{-1}{s(z)}$. Furthermore, the first and second possibilities are also disconnected from each other except when $z^2 - 4 = o(\delta^2)$. Also, the second and third possibilities can only hold for $\mathrm{Im}(z) = o(1)$ since $s_n(z)$ and z both have positive real part. A continuity argument thus shows that the first possibility must hold throughout the region except when $z^2 - 4 = o(\delta^2)$, in which case either the first or second possibility can hold; but in that region, the first and second possibility are equivalent, and (12) follows. The proof of Theorem 1.8 is now complete.

4. Delocalization of Eigenvectors

Without loss of generalization, we can assume that the entries are continuously distributed. Having established Theorem 1.8, we now use this theorem to establish Proposition 1.10.

Let M_n obey Condition **C0**. Then by Markov's inequality, one has $|\zeta_{ij}| \ll \log^{O(1)} n$ with overwhelming probability (here and in the sequel we allow implied constants in the $O()$ notation to depend on the constants C, C' in (1)). By conditioning the ζ_{ij} to this event[3], we may thus assume that

$$|\zeta_{ij}| \leq K \tag{20}$$

almost surely for some $K = O(\log^{O(1)} n)$.

Fix $1 \leq i \leq n$; by symmetry we may take $i \geq n/2$. By the union bound and another application of symmetry, it suffices to show that

$$|u_i(M_n)^* e_1| \ll n^{-1/2} \log^{O(1)} n$$

with overwhelming probability.

To compute $u_i(M_n)^* e_1$ we use the following identity from [7] (see also [27, Lem. 38]):

Lemma 4.1 *Let*

$$A_n = \begin{pmatrix} a & X^* \\ X & A_{n-1} \end{pmatrix}$$

be a $n \times n$ Hermitian matrix for some $a \in \mathbb{R}$ and $X \in \mathbb{C}^{n-1}$, and let $\begin{pmatrix} x \\ v \end{pmatrix}$ be a unit eigenvector of A with eigenvalue $\lambda_i(A)$, where $x \in \mathbb{C}$ and $v \in \mathbb{C}^{n-1}$. Suppose that none of the eigenvalues of A_{n-1} are equal to $\lambda_i(A)$. Then

$$|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(A_{n-1}) - \lambda_i(A_n))^{-2} |u_j(A_{n-1})^* X|^2},$$

where $u_j(A_{n-1})$ is a unit eigenvector corresponding to the eigenvalue $\lambda_j(A_{n-1})$.

Proof By subtracting $\lambda_i(A)I$ from A we may assume $\lambda_i(A) = 0$. The eigenvector equation then gives

$$xX + A_{n-1}v' = 0,$$

thus

$$v' = -xA_{n-1}^{-1}X.$$

Since $\|v'\|^2 + |x|^2 = 1$, we conclude

$$|x|^2(1 + \|A_{n-1}^{-1}X\|^2) = 1.$$

Since $\|A_{n-1}^{-1}X\|^2 = \sum_{j=1}^{n-1} (\lambda_j(A_{n-1}))^{-2} |u_j(A_{n-1})^* X|^2$, the claim follows. □

---

[3] Strictly speaking, this distorts the mean and variance of ζ_{ij} by an exponentially small amount, but one can easily check that this does not significantly impact any of the arguments in this section.

Let M_{n-1} be the bottom right $n-1 \times n-1$ minor of M_n. As we are assuming that the coefficients of M_n are continuously distributed, we see almost surely that none of the eigenvalues of M_{n-1} are equal to $\lambda_i(M_n)$. We may thus apply Lemma 4.1 and conclude that

$$|u_i(M_n)^* e_1|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} \frac{|u_j(M_{n-1})^* X|^2}{(\lambda_j(M_{n-1}) - \lambda_i(M_n))^2}},$$

where X is the bottom left $n-1 \times 1$ vector of M_n (and thus has entries ζ_{j1} for $1 < j \le n$). It thus suffices to show that

$$\sum_{j=1}^{n-1} \frac{|u_j(M_{n-1})^* X|^2}{(\lambda_j(M_{n-1}) - \lambda_i(M_n))^2} \gg n \log^{-O(1)} n$$

with overwhelming probability.

It will be convenient to eliminate the exponent 2 in the denominator, as follows. From Lemma 2.1, one has $|u_j(M_{n-1})^* X| \ll \log^{O(1)} n$ with overwhelming probability for each j (and hence for all j, by the union bound). It thus suffices to show that

$$\sum_{j=1}^{n-1} \frac{|u_j(M_{n-1})^* X|^4}{(\lambda_j(M_{n-1}) - \lambda_i(M_n))^2} \gg n \log^{-O(1)} n$$

with overwhelming probability. By the Cauchy-Schwarz inequality, it thus suffices to show that

$$\sum_{j:i-T_- \le j \le i+T_+} \frac{|u_j(M_{n-1})^* X|^2}{|\lambda_j(M_{n-1}) - \lambda_i(M_n)|} \gg n^{1/2} \log^{-O(1)} n$$

with overwhelming probability for some $1 \le T_-, T_+ \ll \log^{O(1)} n$. It is convenient to work with the normalized matrix $W_n := \frac{1}{\sqrt{n}} M_n$, thus we need to show

$$\sum_{j:i-T_- \le j \le i+T_+} \frac{|u_j(W_{n-1})^* Y|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|} \gg \log^{-O(1)} n \qquad (21)$$

with overwhelming probability for some $1 \le T_-, T_+ \ll \log^{O(1)} n$, where $Y := \frac{1}{\sqrt{n}} X$ has entries $\frac{1}{\sqrt{n}} \zeta_{j1}$ for $1 < j \le n$.

There are two cases: the bulk case and the edge case; the former was already treated in [27], but the latter is new.

4.1. The bulk case. Suppose that $n/2 \le i < 0.999n$. Then from the semicircular law (or Theorem 1.8) we see that $\lambda_i(W_n) \in [-2 + \varepsilon, 2 + \varepsilon]$ with overwhelming probability for some absolute constant $\varepsilon > 0$. Let A be a large constant to be chosen later. A further application of Theorem 1.8 then shows that there is an interval I of length $\log^A n/n$ centered at $\lambda_i(W_n)$ which contains $\Theta(\log^A n)$ eigenvalues of W_n. If $\lambda_j(W_n), \lambda_{j+1}(W_n)$

lie in I, then by the Cauchy interlacing property (5), $|\lambda_j(W_{n-1}) - \lambda_i(W_n)| \ll \log^A n/n$. One can thus lower bound the left-hand side of (21) (for suitable values of T) by

$$\gg n \log^{-A} n \sum_{j:\lambda_j(W_n),\lambda_{j+1}(W_n)\in I} |u_j(W_{n-1})^* Y|^2.$$

One can rewrite this as $\log^{-A} n \|\pi_H X\|^2$, where H is the span of the $u_j(W_{n-1})$ for $\lambda_j(W_n), \lambda_{j+1}(W_n) \in I$. The claim then follows from Lemma 2.1 (for A large enough).

4.2. The edge case. We now turn to the more interesting edge case when $0.999n \le i \le n$. Using the semicircular law, we now see that

$$\lambda_i(W_n) \ge 1.9 \tag{22}$$

(say) with overwhelming probability.

Next, we can exploit the following identity:

Lemma 4.2 *(Interlacing identity)* [27, Lem. 37]. *If $u_j(W_{n-1})^* X$ is non-zero for every j, then*

$$\sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} = \frac{1}{\sqrt{n}} \zeta_{nn} - \lambda_i(W_n). \tag{23}$$

Proof By diagonalising W_{n-1} (noting that this does not affect either side of (23)), we may assume that $W_{n-1} = \mathrm{diag}(\lambda_1(W_{n-1}), \ldots, \lambda_{n-1}(W_{n-1}))$ and $u_j(W_{n-1}) = e_j$ for $j = 1, \ldots, n-1$. One then easily verifies that the characteristic polynomial $\det(W_n - \lambda I)$ of W_n is equal to

$$\prod_{j=1}^{n-1}(\lambda_j(W_{n-1}) - \lambda) \left[\left(\frac{1}{\sqrt{n}}\zeta_{nn} - \lambda \right) - \sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda} \right]$$

when λ is distinct from $\lambda_1(W_{n-1}), \ldots, \lambda_{n-1}(W_{n-1})$. Since $u_j(W_{n-1})^* X$ is non-zero by hypothesis, we see that this polynomial does not vanish at any of the $\lambda_j(W_{n-1})$. Substituting $\lambda_i(W_n)$ for λ, we obtain (23). □

Again, the continuity of the entries of M_n ensure that the hypothesis of Lemma 4.2 is obeyed almost surely. From (20), (22), (23) one has

$$\left| \sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} \right| \ge 1.9 - o(1)$$

with overwhelming probability, so to show (21), it will suffice by the triangle inequality to show that

$$\left| \sum_{j>i+T_+ \text{ or } j<i-T_-} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} \right| \le 1.8 + o(1) \tag{24}$$

(say) with overwhelming probability for some $T = \log^{O(1)} n$.

Let $A > 100$ be a large constant to be chosen later. By Theorem 1.8, we see (if A is large enough) that

$$N_I = n\alpha_I |I| + O(|I| n \log^{-A/20} n) \qquad (25)$$

with overwhelming probability for any interval I of length $|I| = \log^A n/n$, where $\alpha_I := \frac{1}{|I|} \int_I \rho_{sc}(x) \, dx$. For any such interval, we see from Lemma 2.1 (and Cauchy interlacing (5)) that with overwhelming probability

$$\sum_{j:\lambda_j(W_{n-1}) \in I} |u_j(W_{n-1})^* X|^2 = \frac{N_I}{n} + O\left(\frac{\log^{A/2+O(1)} n}{n}\right)$$

and thus by (25) (for A large enough)

$$\sum_{j:\lambda_j(W_{n-1}) \in I} |u_j(W_{n-1})^* X|^2 = \alpha_I |I| + O(|I| \log^{-A/20} n).$$

Set $d_I := \frac{\text{dist}(\lambda_i(W_n), I)}{|I|}$. If $d_I \geq \log n$ (say), then

$$\frac{1}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} = \frac{1}{d_I |I|} + O\left(\frac{1}{d_I^2 |I|}\right)$$

for all j in the above sum, thus

$$\sum_{j:\lambda_j(W_{n-1}) \in I} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} = \frac{\alpha_I}{d_I} + O\left(\frac{\log^{-A/20} n}{d_I}\right) + O\left(\frac{\alpha_I}{d_I^2}\right). \qquad (26)$$

We now partition the real line into intervals I of length $\log^A n/n$, and sum (26) over all I with $d_I \geq \log n$. Bounding α_I crudely by $O(1)$, we see that $\sum_I O\left(\frac{\alpha_I}{d_I^2}\right) = O\left(\frac{1}{\log n}\right) = o(1)$. Similarly, one has

$$\sum_I O\left(\frac{\log^{-A/20} n}{d_I}\right) = O(\log^{-A/20} n \log n) = o(1)$$

if A is large enough. Finally, Riemann integration of the principal value integral

$$p.v. \int_{-2}^{2} \frac{\rho_{sc}(x)}{x - \lambda_i(W_n)} \, dx := \lim_{\varepsilon \to 0} \int_{|x| \leq 2 : |x - \lambda_i(W_n)| > \varepsilon} \frac{\rho_{sc}(x)}{x - \lambda_i(W_n)} \, dx$$

shows that

$$\sum_I \frac{\alpha_I}{d_I} = p.v. \int_{-2}^{2} \frac{\rho_{sc}(x)}{x - \lambda_i(W_n)} \, dx + o(1).$$

The operator norm of W_n is $2 + o(1)$ with overwhelming probability (see e.g. [2,4]), so $|\lambda_i(W_n)| \leq 2 + o(1)$. Using the formula (11) for the Stieltjes transform, one obtains from residue calculus that

$$p.v. \int_{-2}^{2} \frac{\rho_{sc}(x)}{x - \lambda_i(W_n)} \, dx = -\lambda_i(W_n)/2$$

for $|\lambda_i(W_n)| \le 2$, with the right-hand side replaced by $-\lambda_i(W_n)/2 + \sqrt{\lambda_i(W_n)^2 - 4}/2$ for $|\lambda_i(W_n)| > 2$. In either event, we have

$$\left| p.v. \int_{-2}^{2} \frac{\rho_{sc}(x)}{x - \lambda_i(W_n)} \, dx \right| \le 1 + o(1).$$

Putting all this together, we see that

$$\left| \sum_{I:d_I \ge \log n} \sum_{j:\lambda_j(W_{n-1}) \in I} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} \right| \le 1 + o(1).$$

The intervals I with $d_I < \log n$ will contribute at most $\log^{A+O(1)} n$ eigenvalues, by (25) (and Cauchy interlacing (5)). The claim (24) now follows by setting T_- and T_+ appropriately. The proof of Proposition 1.10 is now complete.

Remark 4.3 From (21) and Lemma 2.1 one sees that

$$|\lambda_{i-1}(W_{n-1}) - \lambda_i(W_n)| \ll \log^{O(1)} n/n$$

with overwhelming probability for all $n/2 \le i \le n$, and similarly one has

$$|\lambda_i(W_{n-1}) - \lambda_i(W_n)| \ll \log^{O(1)} n/n$$

with overwhelming probability for all $1 \le i \le n/2$. On the other hand, according to the Tracy-Widom law, the gap between $\lambda_n(W_n)$ and $\lambda_{n-1}(W_n)$ (or between $\lambda_1(W_n)$ and $\lambda_2(W_n)$) can be expected to be as large as $n^{-2/3}$. Thus we see that there is a significant bias at the edge in the interlacing law (5), which can ultimately be traced to the imbalance of "forces" in the interlacing identity (23) at that edge.

5. Lower Bound on Eigenvalue Gap

We now give the proof of Theorem 1.12. Most of the proof will follow closely the proof of Theorem 1.5 in [27], so we shall focus on the changes needed to that argument. As such, this section will assume substantial familiarity with the material from [27], and will cite from it repeatedly (similarly for the next section).

For technical reasons relating to an induction argument, it turns out that one has to treat the extreme cases $i = 1, n$ separately:

Proposition 5.1 *(Extreme cases). Theorem 1.12 is true when $i = 1$ or $i = n$.*

Proof By symmetry it suffices to do this for $i = n$. By a limiting argument we may assume that the entries ζ_{ij} of M_n are continuously distributed. From Lemma 4.2 one has (almost surely) that

$$\sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^2}{\lambda_j(W_{n-1}) - \lambda_n(W_n)} = \frac{1}{\sqrt{n}} \zeta_{nn} - \lambda_n(W_n).$$

Recall that $\lambda_n(W_n) = 2 + o(1)$ with overwhelming probability; also, $\frac{1}{\sqrt{n}} \zeta_{nn} = o(1)$ with overwhelming probability. As all the terms in the left-hand side have the same sign, we conclude that

$$\frac{|u_{n-1}(W_{n-1})^* X|^2}{|\lambda_{n-1}(W_{n-1}) - \lambda_n(W_n)|} \ll 1.$$

From Theorem 2.3 and Proposition 1.10, we have $|u_{n-1}(W_{n-1})^* X| \geq n^{-c_0/10}$ (say) with high probability, and so

$$|\lambda_{n-1}(W_{n-1}) - \lambda_n(W_n)| \geq n^{-c_0}$$

with high probability. The claim now follows from the Cauchy interlacing property (5). $\square$

Remark 5.2 In fact, at the edge, one should be able to improve the lower bound on the eigenvalue gap substantially, from n^{-c_0} to $n^{1/3-c_0}$, in accordance to the Tracy-Widom law, but we will not need to do so here.

Now we handle the general case of Theorem 1.12. Fix M_n and c_0. We write n_0, i_0 for i, n, thus $1 \leq i_0 \leq n_0$ and our task is to show that

$$\lambda_{i_0+1}(A_n) - \lambda_{i_0}(A_{n_0}) \geq n^{-c_0}$$

with high probability. By Proposition 5.1 we may assume $1 < i_0 < n_0$. We may also assume n_0 to be large, as the claim is vacuous otherwise. As in previous sections, we may truncate so that all coefficients ζ_{ij} are of size $O(\log^{O(1)} n_0)$ (as before, the exponentially small corrections to the mean and variance of ζ_{ij} caused by this are easily controlled), and approximate so that the distribution is continuous rather than discrete.

For each $n_0/2 \leq n \leq n_0$, let A_n be the top left $n \times n$ minor of A_{n_0}. As in [27, Sect. 3.4], we introduce the *regularized gap*

$$g_{i,l,n} := \inf_{1 \leq i_- \leq i - l < i \leq i_+ \leq n} \frac{\lambda_{i_+}(A_n) - \lambda_{i_-}(A_n)}{\min(i_+ - i_-, \log^{C_1} n_0)^{\log^{0.9} n_0}}, \tag{27}$$

for all $n_0/2 \leq n \leq n_0$ and $1 \leq i - l < i \leq n$, where C_1 is a large constant to be chosen later. It will suffice to show that for each $1 < i_0 < n_0$,

$$g_{i_0,1,n_0} \leq n^{-c_0}$$

with high probability. By symmetry we may assume that $n_0/2 \leq i_0 < n_0$.

As before, we let $u_1(A_n), \ldots, u_n(A_n)$ be an orthonormal eigenbasis of A_n associated to the eigenvectors $\lambda_1(A_n), \ldots, \lambda_n(A_n)$. We also let $X_n \in \mathbb{C}^n$ be the rightmost column of A_{n+1} with the bottom coordinate $\sqrt{n}\zeta_{n+1,n+1}$ removed.

We will need two key lemmas. First, we have the following deterministic lemma (a minor variant of [27, Lem. 47]), showing that a narrow gap can be propagated backwards in n unless one of a small number of exceptional events happen:

Lemma 5.3 (*Backwards propagation of gap*). *Suppose that $i_0 \leq n + 1 \leq n_0$ and $l \leq n/10$ is such that*

$$g_{i_0,l,n+1} \leq \delta \tag{28}$$

for some $0 < \delta \leq 1$ (which can depend on n), and that

$$\lambda_{n+1}(A_{n+1}) - \lambda_n(A_{n+1}) \geq \delta \exp(\log^{0.91} n_0). \tag{29}$$

Then $i_0 \leq n$. Suppose further that

$$g_{i_0,l+1,n} \geq 2^m g_{i_0,l,n+1} \tag{30}$$

for some $m \geq 0$ with

$$2^m \leq \delta^{-1/2}. \tag{31}$$

Then one of the following statements hold:

  (i) *(Macroscopic spectral concentration) There exists $1 \leq i_- < i_+ \leq n+1$ with $i_+ - i_- \geq \log^{C_1/2} n$ such that $|\lambda_{i_+}(A_{n+1}) - \lambda_{i_-}(A_{n+1})| \leq \delta \exp(\log^{0.95} n)(i_+ - i_-)$.*
  (ii) *(Small inner products) There exists $1 \leq i_- \leq i_0 - l < i_0 \leq i_+ \leq n$ with $i_+ - i_- \leq \log^{C_1/2} n$ such that*

$$\sum_{i_- \leq j < i_+} |u_j(A_n)^* X_n|^2 \leq \frac{n(i_+ - i_-)}{2^{m/2} \log^{0.01} n}. \tag{32}$$

  (iii) *(Large coefficient) We have*

$$|\zeta_{n+1,n+1}| \geq n^{0.4}.$$

  (iv) *(Large eigenvalue) For some $1 \leq i \leq n+1$ one has*

$$|\lambda_i(A_{n+1})| \geq \frac{n \exp(-\log^{0.95} n)}{\delta^{1/2}}.$$

  (v) *(Large inner product) There exists $1 \leq i \leq n$ such that*

$$|u_i(A_n)^* X_n|^2 \geq \frac{n \exp(-\log^{0.96} n)}{\delta^{1/2}}.$$

  (vi) *(Large row) We have*

$$\|X_n\|^2 \geq \frac{n^2 \exp(-\log^{0.96} n)}{\delta^{1/2}}.$$

  (vii) *(Large inner product near i_0) There exists $1 \leq i \leq n$ with $|i - i_0| \leq \log^{C_1} n$ such that*

$$|u_i(A_n)^* X_n|^2 \geq 2^{m/2} n \log^{0.8} n.$$

Proof The proof of this proposition repeats the proof of [27, Lem. 47 in Sect. 6] almost exactly. Only the following changes have to be made:

- We have the upper bound $\lambda_{i_+}(A_{n+1}) - \lambda_{i_-}(A_{n+1}) \leq \delta(\log^{C_1} n)^{\log^{0.9} n_0}$, which together with (29) forces $i_+ \neq n+1$ and thus $i_0 \leq n$ as required.
- The variable j now lies in the range $1 \leq j \leq n$ rather than $\varepsilon n/10 \leq j \leq (1-\varepsilon/10)n$.
- i_{--} has to be defined as $\max(i_- - 2^{k-1}, 1)$ rather than just $i_- - 2^{k-1}$ (and similarly for i_{++}). □

Next, we need the following result that asserts that the events (i)-(vii) are rare:

Proposition 5.4 (*Bad events are rare*). *Suppose that $n_0/2 \leq n < n_0$ and $l \leq n/10$, and set $\delta := n_0^{-\kappa}$ for some sufficiently small fixed $\kappa > 0$. Let $m \geq 0$ be such that $2^m \leq \delta^{-1/2}$. Then:*

(a) *The events (i), (iii), (iv), (v), (vi) in Lemma 5.3 all fail with high probability.*
(b) *There is a constant C' such that all the coefficients of the eigenvectors $u_j(A_n)$ for $1 \leq j \leq n$ are of magnitude at most $n^{-1/2} \log^{C'} n$ with overwhelming probability. Conditioning A_n to be a matrix with this property, the events (ii) and (vii) occur with a conditional probability of at most $2^{-\kappa m} + n^{-\kappa}$.*
(c) *Furthermore, there is a constant C_2 (depending on C', κ, C_1) such that if $l \geq C_2$ and A_n is conditioned as in (b), then (ii) and (vii) in fact occur with a conditional probability of at most $2^{-\kappa m} \log^{-2C_1} n + n^{-\kappa}$.*

Proof The proof of this proposition repeats the proof of [27, Prop. 49 in Sect. 7] almost exactly. Only the following changes have to be made:

- All references to [27, Th. 56] (i.e. Theorem 1.6) need to be replaced with Theorem 1.8.
- The variable j now lies in the range $1 \leq j \leq n$ rather than $\varepsilon n/2 \leq j \leq (1 - \varepsilon/2)n$.
  □

Given Lemma 5.3 and Proposition 5.4, the proof of Theorem 1.12 exactly follows the proof of Theorem 1.5 in [27, Sect. 3.5], with the following minor changes:

- In the definition of the event E_n, the range $\varepsilon n/2 \leq j \leq (1 - \varepsilon/2)n$ needs to be expanded to $1 \leq j \leq n$.
- In the definition of the event E_0, the events that (29) fail for some $n_0 - \log^{2C_1} n_0 \leq n \leq n_0$ have to be included; but these events occur with polynomially small probability, thanks to Proposition 5.1 and the union bound.

This concludes the proof of Theorem 1.12.

6. The Four Moment Theorem

We now prove Theorem 1.11. As in [27, Sect. 3.3], the proof is based on two key propositions. The first proposition asserts that one can swap a single coefficient (or more precisely, two coefficients) of a (deterministic) matrix A as long as A obeys a certain "good configuration condition":

Proposition 6.1 (*Replacement given a good configuration*). *There exists a positive constant C_1 such that the following holds. Let $k \geq 1$ and $\varepsilon_1 > 0$, and assume n sufficiently large depending on these parameters. Let $1 \leq i_1 < \cdots < i_k \leq n$. For a complex parameter z, let $A(z)$ be a (deterministic) family of $n \times n$ Hermitian matrices of the form*

$$A(z) = A(0) + z e_p e_q^* + \bar{z} e_q e_p^*,$$

where e_p, e_q are unit vectors. We assume that for every $1 \leq j \leq k$ and every $|z| \leq n^{1/2+\varepsilon_1}$ whose real and imaginary parts are multiples of n^{-C_1}, we have

- *(Eigenvalue separation) For any $1 \leq i \leq n$ with $|i - i_j| \geq n^{\varepsilon_1}$, we have*

$$|\lambda_i(A(z)) - \lambda_{i_j}(A(z))| \geq n^{-\varepsilon_1}|i - i_j|. \tag{33}$$

- (Delocalization at i_j) If $P_{i_j}(A(z))$ is the orthogonal projection to the eigenspace associated to $\lambda_{i_j}(A(z))$, then

$$\|P_{i_j}(A(z))e_p\|, \|P_{i_j}(A(z))e_q\| \leq n^{-1/2+\varepsilon_1}. \tag{34}$$

- For every $\alpha \geq 0$,

$$\|P_{i_j,\alpha}(A(z))e_p\|, \|P_{i_j,\alpha}(A(z))e_q\| \leq 2^{\alpha/2} n^{-1/2+\varepsilon_1}, \tag{35}$$

whenever $P_{i_j,\alpha}$ is the orthogonal projection to the eigenspaces corresponding to eigenvalues $\lambda_i(A(z))$ with $2^\alpha \leq |i - i_j| < 2^{\alpha+1}$.

We say that $A(0), e_p, e_q$ are a good configuration for $i_1, \ldots, i_k$ if the above properties hold. Assuming this good configuration, then we have

$$\mathbf{E}(F(\zeta)) = \mathbf{E}F(\zeta') + O(n^{-(r+1)/2+O(\varepsilon_1)}), \tag{36}$$

whenever

$$F(z) := G(\lambda_{i_1}(A(z)), \ldots, \lambda_{i_k}(A(z)), Q_{i_1}(A(z)), \ldots, Q_{i_k}(A(z))),$$

and

$$G = G(\lambda_{i_1}, \ldots, \lambda_{i_k}, Q_{i_1}, \ldots, Q_{i_k})$$

is a smooth function from $\mathbb{R}^k \times \mathbb{R}_+^k \to \mathbb{R}$ that is supported on the region

$$Q_{i_1}, \ldots, Q_{i_k} \leq n^{\varepsilon_1}$$

and obeys the derivative bounds

$$|\nabla^j G| \leq n^{\varepsilon_1}$$

for all $0 \leq j \leq 5$, and ζ, ζ' are random variables with $|\zeta|, |\zeta'| \leq n^{1/2+\varepsilon_1}$ almost surely, which match to order r for some $r = 2, 3, 4$.

If G obeys the improved derivative bounds

$$|\nabla^j G| \leq n^{-Cj\varepsilon_1}$$

for $0 \leq j \leq 5$ and some sufficiently large absolute constant C, then we can strengthen $n^{-(r+1)/2+O(\varepsilon_1)}$ in (36) to $n^{-(r+1)/2-\varepsilon_1}$.

Proof See [27, Prop. 43]. □

The second proposition asserts that these good configurations occur very frequently:

Proposition 6.2 *(Good configurations occur very frequently). Let $\varepsilon_1 > 0$ and $C, C_1, k \geq 1$. Let $1 \leq i_1 < \cdots < i_k \leq n$, let $1 \leq p, q \leq n$, let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{C}^n$, and let $A(0) = (\zeta_{ij})_{1 \leq i,j \leq n}$ be a random Hermitian matrix with independent upper-triangular entries and $|\zeta_{ij}| \leq n^{1/2} \log^C n$ for all $1 \leq i, j \leq n$, with $\zeta_{pq} = \zeta_{qp} = 0$, but with ζ_{ij} having mean zero and variance 1 for all other ij, except on the diagonal where the variance is instead c for some absolute constant $c > 0$, and also being distributed continuously in the complex plane. Then $A(0), e_p, e_q$ obey the Good Configuration Condition in Theorem 6.1 for $i_1, \ldots, i_k$ and with the indicated value of ε_1, C_1 with overwhelming probability.*

Proof The proof of this proposition repeats the proof of [27, Prop. 44 in Sect. 5] almost exactly. Only the following changes have to be made:

- All references to [27, Th. 56] (i.e. Theorem 1.6) need to be replaced with Theorem 1.8.
- All references to [27, Prop. 58] (i.e. Proposition 1.7) need to be replaced with Proposition 1.10.
- The edge regions in which $\lambda_i(A(z))$ do not fall inside the bulk region $[(-2 + \varepsilon')n, (2 - \varepsilon')n]$ no longer need to be treated separately, thus simplifying the last paragraph of the proof somewhat. $\qquad\square$

Given these two propositions, the proof of Theorem 1.11 repeats the proof of [27, Th. 15 in Sect. 3.3] almost exactly. Only the following changes have to be made:

- All references to [27, Prop. 44] need to be replaced with Proposition 6.2.

The proof of Theorem 1.11 is now complete.

References

1. Anderson, G., Guionnet, A., Zeitouni, O.: An introduction to random matrices. *To be published by Cambridge Univ. Press*
2. Bai, Z.D., Silverstein, J.: Spectral analysis of large dimensional random matrices. Mathematics Monograph Series **2**, Beijing: Science Press, 2006
3. Bai, Z.D., Yin, Y.Q.: Convergence to the semicircle law. Ann. Probab. **16**, 863–875 (1988)
4. Bai, Z.D., Yin, Y.Q.: Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix. Ann. Probab. **16**, 1729–1741 (1988)
5. Deift, P.: Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; Providence, RI: Amer. Math. Soc., 1999
6. Deift, P.: Universality for mathematical and physical systems. In: *International Congress of Mathematicians* Vol. **I**, Zürich: Eur. Math. Soc., 2007, pp. 125–152
7. Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. Ann. Prob. **37**(3), 815–852 (2009)
8. Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. Commun. Math. Phys. **287**(2), 641–655 (2009)
9. Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. *Submitted*, available at http://arxiv.org/abs/0811.2591v3[math.ph], 2009
10. Erdős, L., Schlein, B., Yau, H.-T.: Universality of Random Matrices and Local Relaxation Flow. http://arxiv.org/abs/0907.5605v3[math-ph], 2009
11. Erdős, L., Ramirez, J., Schlein, B., Yau, H.-T.: Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. http://arxiv.org/abs/0905.2089v1[math-ph], 2009
12. Erdős, L., Ramirez, J., Schlein, B., Yau, H.-T.: Bulk universality for Wigner matrices. http://arxiv.org/abs/0905.4176v2[math-ph], 2009
13. Erdős, L., Ramirez, J., Schlein, B., Tao, T., Vu, V., Yau, H.-T.: Bulk universality for Wigner hermitian matrices with subexponential decay. http://arxiv.org/abs/0906.4400v1[math.PR], 2009
14. Forrester, P.: The spectral edge of random matrix ensembles. Nucl. Phys. B **402**, 709–728 (1993)
15. Johansson, K.: Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. Commun. Math. Phys. **215**(3), 683–705 (2001)
16. Johansson, K.: Universality for certain Hermitian Wigner matrices under weak moment conditions, preprint
17. Katz, N., Sarnak, P.: Random matrices, Frobenius eigenvalues, and monodromy. American Mathematical Society Colloquium Publications, **45**. Providence, RI: Amer. Math. Soc., 1999

18. Khorunzhiy, O.: High Moments of Large Wigner Random Matrices and Asymptotic Properties of the Spectral Norm. http://arxiv.org/abs/0907.3743v2[math.PR], 2009
19. Mehta, M.L.: Random Matrices and the Statistical Theory of Energy Levels. New York: Academic Press, 1967
20. Péché, S., Soshnikov, A.: On the lower bound of the spectral norm of symmetric random matrices with independent entries. Electron. Commun. Probab. **13**, 280–290 (2008)
21. Péché, S., Soshnikov, A.: Wigner random matrices with non-symmetrically distributed entries. J. Stat. Phys. **129**(5–6), 857–884 (2007)
22. Ruzmaikina, A.: Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries. Commun. Math. Phys. **261**(2), 277–296 (2006)
23. Sinai, Y., Soshnikov, A.: Central limit theorem for traces of large symmetric matrices with independent matrix elements. Bol. Soc. Brazil. Mat. **29**, 1–24 (1998)
24. Sinai, Y., Soshnikov, A.: A refinement of Wigners semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. Func. Anal. Appl. **32**, 114–131 (1998)
25. Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. Commun. Math. Phys. **207**(3), 697–733 (1999)
26. Soshnikov, A.: Gaussian limit for determinantal random point fields. Ann. Probab. **30**(1), 171–187 (2002)
27. Tao, T., Vu, V.: Random matrices: Universality of the local eigenvalue statistics. *Submitted*, available at http://arxiv.org/abs/0908.1982v4[math.PR], 2010
28. Tracy, C., Widom, H.: Level spacing distribution and Airy kernel. Commun. Math. Phys. **159**, 151–174 (1994)
29. Tracy, C., Widom, H.: On orthogonal and symplectic matrix ensembles. Commun. Math. Phys. **177**, 727–754 (1996)

Communicated by H.-T. Yau