Acta Math., 206 (2011), 127–204
DOI: 10.1007/s11511-011-0061-3

Random matrices: Universality of local eigenvalue statistics

by

Terence Tao

University of California, Los Angeles
Los Angeles, CA, U.S.A.

Van Vu

Rutgers
Piscataway, NJ, U.S.A.

Contents

T. Tao is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

V. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

1. Introduction

1.1. Wigner matrices and local statistics

The goal of this paper is to establish a universality property for the local eigenvalue statistics for random matrices. To simplify the presentation, we are going to focus on *Wigner Hermitian matrices*, which are perhaps the most prominent model in the field. We emphasize however that our main theorem (Theorem 15) is stated in a much more general setting, and can be applied to various other models of random matrices (such as random real symmetric matrices, for example).

Definition 1. (Wigner matrices) Let n be a large number. A *Wigner Hermitian matrix* (of size n) is defined as a random Hermitian $n\times n$ matrix M_n with upper triangular complex entries $\zeta_{ij}:=\xi_{ij}+\tau_{ij}\sqrt{-1}$, $1\leqslant i<j\leqslant n$, and diagonal real entries ξ_{ii}, $1\leqslant i\leqslant n$, where

- for $1\leqslant i<j\leqslant n$, ξ_{ij} and τ_{ij} are independent identically distrubuted (iid) copies of a real random variable ξ with mean zero and variance $\frac{1}{2}$;
- for $1\leqslant i\leqslant n$, ξ_{ii} are iid copies of a real random variable $\tilde{\xi}$ with mean zero and variance 1;
- ξ and $\tilde{\xi}$ have exponential decay, i.e., there exist constants C and C' such that $\mathbf{P}(|\xi|\geqslant t^C)\leqslant\exp(-t)$ and $\mathbf{P}(|\tilde{\xi}|\geqslant t^C)\leqslant\exp(-t)$ for all $t\geqslant C'$.

We refer to ξ and $\tilde{\xi}$ as the *atom distributions* of M_n, and to ξ_{ij} and τ_{ij} as the *atom variables*. We refer to the matrix $W_n:=M_n/\sqrt{n}$ as the *coarse-scale normalized Wigner Hermitian matrix*, and to $A_n:=M_n\sqrt{n}$ as the *fine-scale normalized Wigner Hermitian matrix*.

Example 2. An important special case of a Wigner Hermitian matrix is the *Gaussian unitary ensemble* (GUE), in which ξ and $\tilde{\xi}$ are Gaussian random variables with mean zero and variance $\frac{1}{2}$ and 1, respectively. The coarse-scale normalization W_n is convenient for placing all the eigenvalues in a bounded interval, while the fine-scale normalization A_n is convenient for keeping the spacing between adjacent eigenvalues to be roughly of unit size.

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

$$\lambda_1(A) \leqslant \ldots \leqslant \lambda_n(A),$$

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.

The study of the eigenvalues $\lambda_i(W_n)$ of (normalized) Wigner Hermitian matrices has been one of the major topics of study in random matrix theory. The properties of these eigenvalues are not only interesting in their own right, but also have been playing essential roles in many other areas of mathematics, such as mathematical physics, probability, combinatorics, and the theory of computing.

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

Definition 3. (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 $\mathbf{P}(E) \geqslant 1 - O(n^{-c})$ for some constant $c > 0$.
- E holds *with overwhelming probability* if $\mathbf{P}(E) \geqslant 1 - O_C(n^{-C})$ for *every* constant $C > 0$ (or equivalently, if $\mathbf{P}(E) \geqslant 1 - \exp(-\omega(\log n))$).
- E holds *almost surely* if $\mathbf{P}(E) = 1$.

Remark 4. Note from the union bound that the intersection of $O(n^{O(1)})$ many events with uniformly overwhelming probability, still has overwhelming probability. Unfortunately, the same is not true for events which are merely of high probability, which will cause some technical difficulties in our arguments.

A cornerstone of this theory is the *Wigner semi-circular law*. Denote by ϱ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$\varrho_{\mathrm{sc}}(x) := \begin{cases} \dfrac{\sqrt{4 - x^2}}{2\pi}, & \text{if } |x| \leqslant 2, \\ 0, & \text{if } |x| > 2. \end{cases} \tag{1}$$

([1]) See §1.7 for our conventions on asymptotic notation.

THEOREM 5. (Semi-circular law) *Let M_n be a Wigner Hermitian matrix. Then, for any real number x,*

$$\lim_{n\to\infty}\frac{1}{n}|\{1\leqslant i\leqslant n:\lambda_i(W_n)\leqslant x\}|=\int_{-2}^{x}\varrho_{\mathrm{sc}}(y)\,dy,$$

in the sense of probability (and also in the almost sure sense, if the M_n's are all minors of the same infinite Wigner Hermitian matrix), where we use $|I|$ to denote the cardinality of a finite set I.

Remark 6. Wigner [47] proved this theorem for special ensembles. The general version above is due to Pastur [36] (see [1] and [2] for detailed discussions). The semi-circular law in fact holds under substantially more general hypotheses than those given in Definition 1, but we will not discuss this matter further here. One consequence of Theorem 5 is that we expect most of the eigenvalues of W_n to lie in the interval $(-2+\varepsilon,2+\varepsilon)$ for $\varepsilon>0$ small; we shall thus informally refer to this region as the *bulk* of the spectrum.

Several stronger versions of Theorem 5 are known. For instance, it is known (see e.g. [3] and [4]) that asymptotically almost surely, one has

$$\lambda_j(W_n)=t\left(\frac{j}{n}\right)+O(n^{-\delta}) \tag{2}$$

for all $1\leqslant j\leqslant n$ and some absolute constant $\delta>0$, where $-2\leqslant t(a)\leqslant 2$ is defined by

$$a=:\int_{-2}^{t(a)}\varrho_{\mathrm{sc}}(x)\,dx. \tag{3}$$

In particular we have

$$\sup_{1\leqslant i\leqslant n}|\lambda_i(M_n)|=2\sqrt{n}\,(1+o(1)) \tag{4}$$

asymptotically almost surely (see [4] for further discussion).

Theorem 5 addressed the global behavior of the eigenvalues. The local properties are much harder and their studies require much more sophisticated tools. Most of the precise theorems have been obtained for the GUE, defined in Example 2. In the next few paragraphs, we mention some of the most famous results concerning this model.

1.2. Distribution of the spacings (gaps) of the eigenvalues of GUE

In this section M_n is understood to have the GUE distribution.

For a vector $x=(x_1,...,x_n)$, with $x_1<x_2...<x_n$, define the normalized gap distribution $S_n(s;x)$ by the formula

$$S_n(s;x):=\frac{1}{n}|\{1\leqslant i\leqslant n:x_{i+1}-x_i\leqslant s\}|.$$

For the GUE ensemble it is known [34] that

$$\lim_{n\to\infty} \mathbf{E}S_n(s,\lambda(A_n)) = \int_0^s p(\sigma)\,d\sigma, \tag{5}$$

where $A_n := M_n\sqrt{n}$ is the fine-scale normalization of M_n, and $p(\sigma)$ is the *Gaudin distribution*, given by the formula

$$p(s) := \frac{d^2}{ds^2}\det(I-K)_{L^2(0,s)},$$

where K is the integral operator on $L^2((0,s))$ with the *Dyson sine kernel*

$$K(x,y) := \frac{\sin\pi(x-y)}{\pi(x-y)}. \tag{6}$$

In fact a stronger result is known in the bulk of the spectrum. Let l_n be any sequence of numbers tending to infinity such that l_n/n tends to zero. Define

$$\widetilde{S}_n(s;x,u) := \frac{1}{l_n}\left|\left\{1\leqslant i\leqslant n : x_{i+1}-x_i \leqslant \frac{s}{\varrho_{\mathrm{sc}}(u)} \text{ and } |x_i-nu| \leqslant \frac{l_n}{\varrho_{\mathrm{sc}}(u)}\right\}\right|. \tag{7}$$

It is proven in [15] that for any fixed $-2<u<2$, we have

$$\lim_{n\to\infty} \mathbf{E}\widetilde{S}_n(s;\lambda(A_n),u) = \int_0^s p(\sigma)\,d\sigma. \tag{8}$$

The eigenvalue gap distribution has received much attention in the mathematics community, partially due to the fascinating (numerical) coincidence with the gap distribution of the zeros of the zeta functions. For more discussions, we refer to [13], [14], [29] and the references therein.

1.3. *k*-point correlation for GUE

Given a fine-scale normalized Wigner Hermitian matrix A_n, we can define the symmetrized distribution function $\varrho_n^{(n)}:\mathbb{R}^n\to\mathbb{R}^+$ to be the symmetric function on n variables such that the distribution of the eigenvalues $\lambda(A_n)$ is given by the restriction of $n!\varrho_n^{(n)}(x_1,...,x_n)\,dx_1\,...\,dx_n$ to the region $\{x_1\leqslant...\leqslant x_n\}$. For any $1\leqslant k\leqslant n$, the *k-point correlation function* $\varrho_n^{(k)}:\mathbb{R}^k\to\mathbb{R}^+$ is defined as the marginal integral of $\varrho_n^{(n)}$:

$$\varrho_n^{(k)}(x_1,...,x_k) := \frac{n!}{(n-k)!}\int_{\mathbb{R}^{n-k}} \varrho_n^{(n)}(x)\,dx_{k+1}\,...\,dx_n.$$

In the GUE case, one has an explicit formula for $\varrho_n^{(n)}$, obtained by Ginibre [23]:

$$\varrho_n^{(n)}(x) := \frac{n^{-n/2}}{Z_n^{(2)}}\prod_{1\leqslant i<j\leqslant n}|x_i-x_j|^2\exp\left(-\frac{x_1^2+...+x_n^2}{2n}\right), \tag{9}$$

where $Z_n^{(2)}>0$ is a normalizing constant, known as the *partition function*. From this formula, one can compute $\varrho_n^{(k)}$ explicitly. Indeed, it was established by Gaudin and Mehta [35] that

$$\varrho_n^{(k)}(x_1,...,x_k)=\det(K_n(x_i,x_j))_{1\leqslant i,j\leqslant k}, \tag{10}$$

where the kernel $K_n(x,y)$ is given by the formula

$$K_n(x,y):=\frac{1}{\sqrt{2n}}\exp\left(-\frac{x^2+y^2}{4n}\right)\sum_{j=0}^{n-1}h_j\left(\frac{x}{\sqrt{2n}}\right)h_j\left(\frac{y}{\sqrt{2n}}\right),$$

and $h_0,...,h_{n-1}$ are the first n Hermite polynomials, normalized to be orthonormal with respect to $\exp(-x^2)\,dx$. From this and the asymptotics of Hermite polynomials, it was shown by Dyson [16] that

$$\lim_{n\to\infty}\frac{1}{\varrho_{\mathrm{sc}}(u)^k}\varrho_n^{(k)}\left(nu+\frac{t_1}{\varrho_{\mathrm{sc}}(u)},...,nu+\frac{t_k}{\varrho_{\mathrm{sc}}(u)}\right)=\det(K(t_i,t_j))_{1\leqslant i,j\leqslant k}, \tag{11}$$

for any fixed $-2<u<2$ and real numbers $t_1,...,t_k$, where the Dyson sine kernel K was defined in (6).

1.4. The universality conjecture and previous results

It has been conjectured, since the 1960s, by Wigner, Dyson, Mehta and many others, that the local statistics (such as the above limiting distributions) are *universal*, in the sense that they hold not only for the GUE, but for any other Wigner random matrix also. This conjecture was motivated by similar phenomena in physics, such as the same laws of thermodynamics, which should emerge no matter what the details of atomic interaction.

The universality conjecture is one of the central questions in the theory of random matrices. In many cases, it is stated for a specific local statistics (such as the gap distribution or the k-point correlation, see [34, p. 9] for example). These problems have been discussed in numerous books and surveys (see [13], [14] and [34]).

Despite conjecture's long and distinguished history and the overwhelming supporting numerical evidence, rigorous results on this problem for general Wigner random matrices have only begun to emerge recently. At the edge of the spectrum, Soshnikov [41] proved the universality of the joint distribution of the largest k eigenvalues (for any fixed k), under the extra assumption that the atom distribution is symmetric.

THEOREM 7. ([41]) *Let k be a fixed integer and M_n be a Wigner Hermitian matrix, whose atom distribution is symmetric. Set $W_n := M_n\sqrt{n}$. Then the joint distribution of the k-dimensional random vector*

$$((\lambda_n(W_n)-2)n^{2/3}, ..., (\lambda_{n-k}(W_n)-2)n^{2/3})$$

has a weak limit as $n\to\infty$, which coincides with that in the GUE *case. The result also holds for the smallest eigenvalues $\lambda_1, ..., \lambda_k$.*

Note that this significantly strengthens (4) in the symmetric case. (For the non-symmetric case, see [39] and [40] for some recent results).

Returning to the bulk of the spectrum, Johansson [28] proved (11) and (8) for random Hermitian matrices whose entries are *Gauss divisible*. (See also the paper [5] by Ben Arous and Péché, where they discussed the removal of a technical condition in [28].) More precisely, Johansson considered the model $M_n = (1-t)^{1/2}M_n^1 + t^{1/2}M_n^2$, where $0<t\leqslant 1$ is fixed (i.e. independent of n), M_n^1 is a Wigner Hermitian matrix and M_n^2 is a GUE matrix independent of M_n^1. We will refer to such matrices as *Johansson matrices*.

THEOREM 8. ([28]) *Formulae* (11) (*in the weak sense*) *and* (8) (*and hence* (5)) *hold for Johansson matrices, as $n\to\infty$. By "weak sense", we mean that*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{\varrho_{\mathrm{sc}}(u)^k} \int_{\mathbb{R}^k} f(t_1, ..., t_k)\varrho_n^{(k)}\left(nu+\frac{t_1}{\varrho_{\mathrm{sc}}(u)}, ..., nu+\frac{t_k}{\varrho_{\mathrm{sc}}(u)}\right) dt_1 ... dt_k \\ = \int_{\mathbb{R}^k} f(t_1, ..., t_k) \det(K(t_i, t_j))_{1\leqslant i,j\leqslant k}\, dt_1 ... dt_k \end{aligned} \tag{12}$$

for any test function $f\in C_c(\mathbb{R}^k)$.

The property of being Gauss divisible can be viewed as a strong regularity assumption on the atom distribution. Very recently, Erdős, Schlein, Ramirez and Yau [17], [19] have relaxed this regularity assumption significantly. In particular in [17] an analogue of Theorem 8 (with $k=2$ for the correlation and l_n polynomial in n for the gap distribution) is proven assuming that the atom distribution is of the form

$$\nu\, dx = \exp(-V(x)) \exp(-x^2)\, dx,$$

where $V(x)\in C^6$, $\sum_{i=1}^6 |V^j(x)| \leqslant C(1+x^2)^k$ and $\nu(x)\leqslant C'\exp(-\delta x^2)$ for some fixed k, δ, C and C'. It was remarked in [17] that the last (exponential decay) assumption can be weakened somewhat.

Finally, let us mention that in a different direction, universality was established by Deift, Kriecherbauer, McLaughlin, Venakides and Zhou [15], Pastur and Shcherbina [37], and Bleher and Its [6] for a different model of random matrices, where the joint distribution of the eigenvalues is given explicitly by the formula

$$\varrho_n^{(n)}(x_1, ..., x_n) := c_n \prod_{1 \leqslant i < j \leqslant n} |x_i - x_j|^2 \exp(-V(x)), \tag{13}$$

where V is a general function and $c_n > 0$ is a normalization factor. The case $V = x^2$ corresponds to (9). For a general V, the entries of the matrix are correlated, and so this model differs from the Wigner model. (See [32] for some recent developments concerning these models, which are studied using the machinery of orthogonal polynomials.)

One of the main difficulties in establishing universality for general matrix ensembles lies in the fact that most of the results obtained in the GUE case (and the case in Johansson's theorem and those in [6], [15] and [37]) came from heavy use of the explicit joint distribution of the eigenvalues such as (9) and (13). The desired limiting distributions were proven using estimates on integrals with respect to these measures. Very powerful tools have been developed to handle this task (see [13] and [34] for example), but they cannot be applied for general Wigner matrices where an explicit measure is not available.

Nevertheless, some methods have been developed which do not require the explicit joint distribution. For instance, Soshnikov's result [41] was obtained using the (combinatorial) trace method rather than from an explicit formula from the distribution, although it is well understood that this method, while efficient for the studying of the edge, is of much less use in the study of the spacing distribution in the bulk of the spectrum. The recent argument in [19] also avoid explicit formulae, relying instead on an analysis of the *Dyson Brownian motion*, which describes the stochastic dynamics of the spectrum of Johansson matrices $M_n = (1-t)^{1/2} M_n^1 + t^{1/2} M_n^2$ in the t variable. (On the other hand, the argument in [17] uses explicit formulae for the joint distribution.) However, it appears that their method still requires a high degree of regularity on the atom distribution, whereas here we shall be interested in methods that do not require any regularity hypotheses at all (and in particular will be applicable to *discrete* atom distributions(2)).

1.5. Universality theorems

In this paper, we introduce a new method to study the local statistics. This method is based on the *Lindeberg strategy* [33] of replacing non-Gaussian random variables

(2) Subsequently to the release of this paper, we have realized that the two methods can in fact be combined to address the gap distribution problem and the k-point correlation problem even for discrete distributions without requiring moment conditions; see [18] for details.

with Gaussian ones. (For more modern discussions about Lindeberg's method, see [8] and [38].) Using this method, we are able to prove universality for general Wigner matrices under very mild assumptions. For instance, we have the following result.

THEOREM 9. (Universality of gap) *The limiting gap distribution* (5) *holds for Wigner Hermitian matrices whose atom distribution* ξ *has support on at least three points. The stronger version* (8) *holds for Wigner Hermitian matrices whose atom distribution* ξ *has support on at least three points and is such that the third moment* $\mathbf{E}\xi^3$ *vanishes.*

Remark 10. Our method also enables us to prove the universality of the variance and higher moments. Thus, the whole distribution of $S_n(s,\lambda)$ is universal, not only its expectation. See Remark 31.

THEOREM 11. (Universality of correlation) *The k-point correlation* (11) (*in the weak sense*) *holds for Wigner Hermitian matrices whose atom distribution* ξ *has support on at least three points and is such that the third moment* $\mathbf{E}\xi^3$ *vanishes.*

These theorems (and several others, see §1.6) are consequences of our more general main theorem below (Theorem 15). Roughly speaking, Theorem 15 states that the local statistics of the eigenvalues of a random matrix is determined by the first four moments of the atom distributions.

Theorem 15 applies in a very general setting. We will consider random Hermitian matrix M_n with entries ξ_{ij} obeying the following condition.

Definition 12. (Condition **C0**) A random Hermitian matrix $A_n=(\zeta_{ij})_{1\leqslant i,j\leqslant n}$ is said to obey *condition* **C0** if

• the variables ζ_{ij} are independent (but not necessarily identically distributed) for $1\leqslant i\leqslant j\leqslant n$, and have mean zero and variance 1;

• (uniform exponential decay) there exist constants $C,C'>0$ such that

$$\mathbf{P}(|\zeta_{ij}|\geqslant t^C)\leqslant \exp(-t) \tag{14}$$

for all $t\geqslant C'$ and $1\leqslant i,j\leqslant n$.

Clearly, all Wigner Hermitian matrices obey condition **C0**. However, the class of matrices obeying condition **C0** is much richer. For instance the Gaussian orthogonal ensemble (GOE), in which $\zeta_{ij}\equiv N(0,1)$ independently for all $i<j$ and $\zeta_{ii}\equiv N(0,2)$, is also essentially of this form,([3]) and so are all Wigner real symmetric matrices (the definition of which is given at the end of this section).

([3]) Note that for GOE a diagonal entry has variance 2 rather than 1. We thank Sean O'Rourke for pointing out this issue. On the other hand, Theorem 15 still holds if we change the variances of the diagonal entries, see Remark 16.

Definition 13. (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 $m, l \geqslant 0$ such that $m+l \leqslant k$.

Example 14. Given two random matrices $A_n = (\zeta_{ij})_{1\leqslant i,j\leqslant n}$ and $A'_n = (\zeta'_{ij})_{1\leqslant i,j\leqslant n}$ obeying condition **C0**, ζ_{ij} and ζ'_{ij} automatically match to order 1. If they are both Wigner Hermitian matrices, then they automatically match to order 2. If furthermore they are also symmetric (i.e. A_n has the same distribution as $-A_n$, and similarly for A'_n and $-A'_n$), then ζ_{ij} and ζ'_{ij} automatically match to order 3.

The following is our main result.

Theorem 15. (Four moment theorem) *There is a small positive constant c_0 such that for every $0<\varepsilon<1$ and $k\geqslant 1$ the following holds. Let*

$$M_n = (\zeta_{ij})_{1\leqslant i,j\leqslant n} \quad \text{and} \quad M'_n = (\zeta'_{ij})_{1\leqslant i,j\leqslant n}$$

be two random matrices satisfying **C0**. *Assume furthermore that ζ_{ij} and ζ'_{ij}, $1\leqslant i<j\leqslant n$, match to order* 4, *and that ζ_{ii} and ζ'_{ii}, $1\leqslant i\leqslant n$, match to order* 2. *Set $A_n := M_n\sqrt{n}$ and $A'_n := M'_n\sqrt{n}$, and let $G: \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$|\nabla^j G(x)| \leqslant n^{c_0} \tag{15}$$

for all $0\leqslant j\leqslant 5$ and $x\in\mathbb{R}^k$. Then, for any $\varepsilon n\leqslant i_1<...<i_k\leqslant(1-\varepsilon)n$ and for n sufficiently large depending on ε and k (and the constants C and C' in Definition 12), *we have*

$$|\mathbf{E}G(\lambda_{i_1}(A_n), ..., \lambda_{i_k}(A_n)) - \mathbf{E}G(\lambda_{i_1}(A'_n), ..., \lambda_{i_k}(A'_n))| \leqslant n^{-c_0}. \tag{16}$$

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* (16) *still holds provided that one strengthens* (15) *to*

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

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

The proof of this theorem begins in §3.3. As mentioned earlier, Theorem 15 asserts (roughly speaking) that the fine spacing statistics of a random Hermitian matrix in the bulk of the spectrum are only sensitive to the first four moments of the entries. It may be possible to reduce the number of matching moments in this theorem, but this seems to require a refinement of the method; see §3.2 for further discussion.

Remark 16. Theorem 15 still holds if we assume that the diagonal entries ζ_{ii} and ζ'_{ii} have the same mean and variance, for all $1 \leqslant i \leqslant n$, but these means and variances can be different at different i. The proof is essentially the same. In our analysis, we consider the random vector formed by non-diagonal entries of a row, and it is important that these entries have mean zero and the same variance, but the mean and variance of the diagonal entries never play a role. Details will appear elsewhere.

Remark 17. In a subsequent paper [46], we show that the condition

$$\varepsilon n \leqslant i_1 < \ldots < i_k \leqslant (1-\varepsilon)n$$

can be omitted. In other words, Theorem 15 also holds for eigenvalues at the edge of the spectrum.

Applying Theorem 15 to the special case when M'_n is GUE, we obtain the following.

COROLLARY 18. *Let M_n be a Wigner Hermitian matrix whose atom distribution ξ satisfies $\mathbf{E}\xi^3=0$ and $\mathbf{E}\xi^4=\frac{3}{4}$, and let M'_n be a random matrix sampled from* GUE. *Then, with G, A_n and A'_n as in the previous theorem, and n sufficiently large, one has*

$$|\mathbf{E}G(\lambda_{i_1}(A_n), \ldots, \lambda_{i_k}(A_n)) - \mathbf{E}G(\lambda_{i_1}(A'_n), \ldots, \lambda_{i_k}(A'_n))| \leqslant n^{-c_0}. \tag{17}$$

In the proof of Theorem 15, the following lower tail estimate on the consecutive spacings plays an important role. This theorem is of independent interest, and will also help in applications of Theorem 15.

THEOREM 19. (Lower tail estimates) *Let $0<\varepsilon<1$ be a constant, and let M_n be a random matrix obeying condition* **C0**. *Set $A_n := M_n\sqrt{n}$. Then, for every $c_0>0$, and for n sufficiently large depending on ε, c_0 and the constants C and C' in Definition* 12, *and for each $\varepsilon n \leqslant i \leqslant (1-\varepsilon)n$, one has $\lambda_{i+1}(A_n) - \lambda_i(A_n) \geqslant n^{-c_0}$ with high probability. In fact, one has*

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

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

The proof of this theorem begins in §3.5.

1.6. Applications

By using Theorems 15 and 19 in combination with existing results in the literature for GUE (or other special random matrix ensembles), one can establish universal asymptotic statistics for a wide range of random matrices. For instance, consider the ith eigenvalue $\lambda_i(M_n)$ of a Wigner Hermitian matrix. In the GUE case, Gustavsson [26], based on [11] and [42], proved that λ_i has Gaussian fluctuation:

Theorem 20. (Gaussian fluctuation for GUE [26]) *Let $i=i(n)$ be such that $i/n\to c$, as $n\to\infty$, for some $0<c<1$. Let M_n be drawn from the* GUE. *Set $A_n:=M_n\sqrt{n}$. Then*

$$\sqrt{\frac{4-t(i/n)^2}{2}}\,\frac{\lambda_i(A_n)-t(i/n)n}{\sqrt{\log n}}\to N(0,1)$$

in the sense of distributions, where $t(\cdot)$ is defined in (3). *More informally, we have*

$$\lambda_i(M_n)\approx t\Big(\frac{i}{n}\Big)\sqrt{n}+N\bigg(0,\frac{2\log n}{(4-t(i/n)^2)n}\bigg).$$

As an application of our main results, we have the following.

Corollary 21. (Universality of Gaussian fluctuation) *The conclusion of Theorem* 20 *also holds for any other Wigner Hermitian matrix M_n whose atom distribution ξ satisfies $\mathbf{E}\xi^3=0$ and $\mathbf{E}\xi^4=\frac{3}{4}$.*

Proof. Let M_n be a Wigner Hermitian matrix, and let M_n' be drawn from GUE. Let i, c and t be as in Theorem 20, and let c_0 be as in Theorem 19. In view of Theorem 20, it suffices to show that

$$\mathbf{P}(\lambda_i(A_n')\in I_-)-n^{-c_0}\leqslant\mathbf{P}(\lambda_i(A_n)\in I)\leqslant\mathbf{P}(\lambda_i(A_n')\in I_+)+n^{-c_0}\tag{18}$$

for all intervals $I=[a,b]$, and n sufficiently large depending on i and the constants C and C' in Definition 1, where

$$I_+:=[a-n^{-c_0/10},b+n^{-c_0/10}]\quad\text{and}\quad I_-:=[a+n^{-c_0/10},b-n^{-c_0/10}].$$

We will just prove the second inequality in (18), as the first is very similar. We define a smooth bump function $G:\mathbb{R}\to\mathbb{R}^+$ equal to 1 on I_- and vanishing outside I_+. Then we have

$$\mathbf{P}(\lambda_i(A_n)\in I)\leqslant\mathbf{E}G(\lambda_i(A_n))$$

and

$$\mathbf{E}G(\lambda_i(A_n'))\leqslant\mathbf{P}(\lambda_i(A_n')\in I).$$

On the other hand, one can choose G to obey (15). Thus, by Corollary 18, we have

$$|\mathbf{E}G(\lambda_i(A_n))-\mathbf{E}G(\lambda_i(A_n'))|\leqslant n^{-c_0},$$

and the second inequality in (18) follows by the triangle inequality. The first inequality is similarly proven using a smooth function which is 1 on I_- and vanishes outside I. $\square$

Remark 22. The same argument lets one establish the universality of the asymptotic joint distribution law for any k eigenvalues $\lambda_{i_1}(M_n), ..., \lambda_{i_k}(M_n)$ in the bulk of the spectrum of a Wigner Hermitian matrix for any fixed k (the GUE case is treated in [26]). In particular, we have the generalization

$$\begin{aligned}\mathbf{P}(\lambda_{i_j}(A_n') \in I_{j,-} \text{ for all } 1 \leqslant j \leqslant k) + O_k(n^{-c_0}) \\ \leqslant \mathbf{P}(\lambda_{i_j}(A_n) \in I_j \text{ for all } 1 \leqslant j \leqslant k) \\ \leqslant \mathbf{P}(\lambda_{i_j}(A_n') \in I_{j,+} \text{ for all } 1 \leqslant j \leqslant k) + O_k(n^{-c_0})\end{aligned} \tag{19}$$

for all $i_1, ..., i_k$ between εn and $(1-\varepsilon)n$ for some fixed $\varepsilon>0$, and all intervals $I_1, ..., I_k$, assuming n is sufficiently large depending on ε and k, and $I_{j,-} \subset I_j \subset I_{j,+}$ are defined as in the proof of Corollary 21. The details are left as an exercise to the interested reader.

Another quantity of interest is the least singular value

$$\sigma_n(M_n) := \inf_{1 \leqslant i \leqslant n} |\lambda_i(M_n)|$$

of a Wigner Hermitian matrix. In the GUE case, we have the following asymptotic distribution.

THEOREM 23. (Distribution of least singular value of GUE [1, Theorem 3.1.2], [27]) *For any fixed* $t>0$, *and* M_n *drawn from* GUE, *one has*

$$\mathbf{P}\left(\sigma_n(M_n) \leqslant \frac{t}{2\sqrt{n}}\right) \to \exp\left(\int_0^t \frac{f(x)}{x}\, dx\right)$$

as $n \to \infty$, *where* $f: \mathbb{R} \to \mathbb{R}$ *is the solution of the differential equation*

$$(tf'')^2 + 4(tf'-f)(tf'-f+(f')^2) = 0$$

with the asymptotics

$$f(t) = -\frac{t}{\pi} - \frac{t^2}{\pi^2} - \frac{t^3}{\pi^3} + O(t^4) \quad \textit{as } t \to 0.$$

Using our theorems, we can extend this result to more general ensembles:

COROLLARY 24. (Universality of the distribution of the least singular value) *The conclusions of Theorem* 23 *also hold for any other Wigner Hermitian matrix* M_n *whose atom distribution* ξ *satisfies* $\mathbf{E}\xi^3 = 0$ *and* $\mathbf{E}\xi^4 = \frac{3}{4}$.

Proof. Let M_n be a Wigner Hermitian matrix, and let M_n' be drawn from GUE. Let N_I be the number of eigenvalues of W_n' in an interval I. It is well known (see [1, Chapter 4]) that

$$N_I = \int_I \varrho_{\mathrm{sc}}(x)\,dx + O(\log n) \tag{20}$$

asymptotically almost surely (cf. (2) and Theorem 20). Applying this fact to the two intervals $I=[-\infty, \pm t/2\sqrt{n}\,]$, we conclude that

$$\mathbf{P}\biggl(\lambda_{n/2\pm(\log n)^2}(M_n') \in \biggl(-\frac{t}{2\sqrt{n}}, \frac{t}{2\sqrt{n}}\biggr)\biggr) = o(1)$$

for either choice of sign $\pm$. Using (18) (or (19)) (and modifying t slightly), we conclude that the same statement is true for M_n. In particular, we have

$$\mathbf{P}\biggl(\sigma_n(M_n) > \frac{t}{2\sqrt{n}}\biggr) = \sum_{i=n/2-(\log n)^2}^{n/2+(\log n)^2} \mathbf{P}\biggl(\lambda_i(M_n) < -\frac{t}{2\sqrt{n}} \text{ and } \lambda_{i+1}(M_n) > \frac{t}{2\sqrt{n}}\biggr) + o(1),$$

and similarly for M_n'. Using (19), we see that

$$\begin{aligned}\mathbf{P}\biggl(\lambda_i(M_n) < -\frac{t}{2\sqrt{n}} &\text{ and } \lambda_i(M_n) > -\frac{t}{2\sqrt{n}}\biggr) \\ &\leqslant \mathbf{P}\biggl(\lambda_i(M_n') < -\frac{t}{2\sqrt{n}} + n^{-c_0/10} \text{ and } \lambda_{i+1}(M_n') > \frac{t}{2\sqrt{n}} - n^{-c_0/10}\biggr) + O(n^{-c_0})\end{aligned}$$

and

$$\begin{aligned}\mathbf{P}\biggl(\lambda_i(M_n) < -\frac{t}{2\sqrt{n}} &\text{ and } \lambda_i(M_n) > -\frac{t}{2\sqrt{n}}\biggr) \\ &\geqslant \mathbf{P}\biggl(\lambda_i(M_n') < -\frac{t}{2\sqrt{n}} - n^{-c_0/10} \text{ and } \lambda_{i+1}(M_n') > \frac{t}{2\sqrt{n}} + n^{-c_0/10}\biggr) - O(n^{-c_0})\end{aligned}$$

for some $c>0$. Putting this together, we conclude that

$$\begin{aligned}\mathbf{P}\biggl(\sigma_n(M_n') > \frac{t}{2\sqrt{n}} - n^{-c_0/10}\biggr) + o(1) &\leqslant \mathbf{P}\biggl(\sigma_n(M_n) > \frac{t}{2\sqrt{n}}\biggr) \\ &\leqslant \mathbf{P}\biggl(\sigma_n(M_n') > \frac{t}{2\sqrt{n}} + n^{-c_0/10}\biggr) + o(1),\end{aligned}$$

and the claim follows. □

Remark 25. A similar universality result for the least singular value of non-Hermitian matrices was recently established by the authors in [45]. Our arguments in [45] also used the Lindeberg strategy, but were rather different in many other respects (in particular, they proceeded by analyzing random submatrices of the inverse matrix M_n^{-1}). One consequence of Corollary 24 is that M_n is asymptotically almost surely invertible. For discrete random matrices, this is already a non-trivial fact, first proven in [9]. If Theorem 23 can be extended to the Johansson matrices considered in [28], then the arguments below would allow one to remove the fourth moment hypothesis in Corollary 24 (assuming that ξ is supported on at least three points).

Remark 26. The above corollary still holds under the weaker assumption that the first three moments of ξ match those of the Gaussian variable; in other words, we can omit the last assumption that $\mathbf{E}\xi^4=\frac{3}{4}$. Details will appear elsewhere.

Remark 27. By combining this result with (4), one also obtains a universal distribution for the condition number $\sigma_1(M_n)/\sigma_n(M_n)$ of Wigner Hermitian matrices (note that the non-independent nature of $\sigma_1(M_n)$ and $\sigma_n(M_n)$ is not relevant, because (4) gives enough concentration of $\sigma_1(M_n)$ that it can effectively be replaced with $2\sqrt{n}$). We omit the details.

Now we are going to prove the first part of Theorem 9. Note that in contrast to previous applications, we are making no assumptions on the third and fourth moments of the atom distribution ξ. The extra observation here is that we do not always need to compare M_n with GUE. It is sufficient to compare it with any model where the desired statistics have been computed. In this case, we are going to compare M_n with a Johansson matrix. The definition of Johansson matrices provides more degrees of freedom via the parameters t and M_n^1, and we can use this to remove the condition of the third and fourth moments.

LEMMA 28. (Truncated moment matching problem) *Let ξ be a real random variable with mean zero, variance* 1, *third moment $\mathbf{E}\xi^3=\alpha_3$ and fourth moment $\mathbf{E}\xi^4=\alpha_4<\infty$. Then $\alpha_4-\alpha_3^2-1\geqslant 0$, with equality if and only if ξ is supported on exactly two points. Conversely, if $\alpha_4-\alpha_3^2-1\geqslant 0$, then there exists a real random variable with the specified moments.*

Proof. For any real numbers a and b, we have

$$0\leqslant \mathbf{E}(\xi^2+a\xi+b)^2=\alpha_4+2a\alpha_3+a^2+2b+b^2.$$

Setting $b:=-1$ and $a:=-\alpha_3$, we obtain the inequality $\alpha_4-\alpha_3^2-1\geqslant 0$. Equality only occurs when $\mathbf{E}(\xi^2-\alpha_3\xi-1)^2=0$, which by the quadratic formula implies that ξ is supported on at most two points.

Now we show that every pair (α_3,α_4) with $\alpha_4-\alpha_3^2-1\geqslant 0$ arises as the moments of a random variable with mean zero and variance 1. The set of all such moments is clearly convex, so it suffices to check the case when $\alpha_4-\alpha_3^2-1=0$. But if one considers the random variable ξ which equals $\tan\theta$ with probability $\cos^2\theta$ and $-\cot\theta$ with probability $\sin^2\theta$ for some $-\frac{1}{2}\pi<\theta<\frac{1}{2}\pi$, one easily computes that ξ has mean zero, variance 1, third moment $-2\cot 2\theta$ and fourth moment $4\csc 2\theta-3$, and the claim follows from the trigonometric identity $\csc^2 2\theta=\cot^2 2\theta+1$. □

Remark 29. The more general truncated moment problem (i.e. the truncated version of the classical Hamburger moment sequence problem, see [12] and [30]) was solved by Curto and Fialkow [12].

Corollary 30. (Matching lemma) *Let ξ be a real random variable with mean zero and variance* 1, *which is supported on at least three points. Then ξ matches to order* 4 *with $(1-t)^{1/2}\xi'+t^{1/2}\xi_G$ for some $0<t<1$ and some independent ξ' and ξ_G of mean zero and variance* 1, *where $\xi_G\equiv N(0,1)$ is Gaussian.*

Proof. The formal characteristic function

$$\mathbf{E}\exp(s\xi):=\sum_{j=0}^{\infty}\frac{s^j}{j!}\mathbf{E}\xi^j$$

has the expansion

$$1+\tfrac{1}{2}s^2+\tfrac{1}{6}\alpha_3 s^3+\tfrac{1}{24}\alpha_4 s^4+O(s^5).$$

By Lemma 28, we have that $\alpha_4-\alpha_3^2-1>0$. Observe that ξ will match to order 4 with $(1-t)^{1/2}\xi'+t^{1/2}\xi_G$ if and only if one has the identity

$$\begin{aligned}1&+\tfrac{1}{2}s^2+\tfrac{1}{6}\alpha_3 s^3+\tfrac{1}{24}\alpha_4 s^4\\&=\big(1+\tfrac{1}{2}(1-t)s^2+\tfrac{1}{6}(1-t)^{3/2}\alpha_3' s^3+\tfrac{1}{24}(1-t)^2\alpha_4' s^4\big)\big(1+\tfrac{1}{2}ts^2+\tfrac{1}{8}ts^4\big)+O(s^5),\end{aligned}$$

where α_3' and α_4' are the moments of ξ'. Formally dividing out by $1+\frac{1}{2}ts^2+\frac{1}{8}ts^4$, one can thus solve for α_3' and α_4' in terms of α_3 and α_4. Observe that, as $t\to 0$, α_3' and α_4' must converge to α_3 and α_4, respectively. Thus, for t sufficiently small, we will have $\alpha_4'-(\alpha_3')^2-1>0$. The claim now follows from Lemma 28. □

Proof of the first part of Theorem 9. Let M_n be as in this theorem and consider (5). By Corollary 30, we can find a Johansson matrix M_n' which matches M_n to order 4. By Theorem 8, (5) already holds for M_n'. Thus it will suffice to show that

$$\mathbf{E}\widetilde{S}_n(s;\lambda(A_n))=\mathbf{E}\widetilde{S}_n(s;\lambda(A_n'))+o(1).$$

By (7) and linearity of expectation, it suffices to show that

$$\mathbf{P}(\lambda_{i+1}(A_n)-\lambda_i(A_n)\leqslant s)=\mathbf{P}(\lambda_{i+1}(A_n')-\lambda_i(A_n')\leqslant s)+o(1)$$

uniformly for all $\varepsilon n \leqslant i \leqslant (1-\varepsilon)n$, for each fixed $\varepsilon>0$. But this follows by a modification of the argument used to prove (18) (or (19)), using a function $G(x,y)$ of two variables which is a smooth approximant to the indicator function of the half-space $\{(x,y): y-x\leqslant s\}$ (and using Theorem 19 to errors caused by shifting s); we omit the details. The second part of the theorem will be treated together with Theorem 11. ☐

Remark 31. By considering

$$\mathbf{P}(\lambda_{i+1}(A_n)-\lambda_i(A_n)\leqslant s \text{ and } \lambda_{j+1}(A_n)-\lambda_j(A_n)\leqslant s),$$

we can prove the universality of the variance of $S_n(s,\lambda)$. The same applies for higher moments.

The proof of Theorem 11 is a little more complicated. We first need a strengthening of (2) which may be of independent interest.

THEOREM 32. (Convergence to the semi-circular law) *Let M_n be a Wigner Hermitian matrix whose atom distribution ξ has vanishing third moment. Then, for any fixed $c>0$ and $\varepsilon>0$, and any $\varepsilon n\leqslant j\leqslant (1-\varepsilon)n$, one has*

$$\lambda_j(W_n) = t\left(\frac{j}{n}\right) + O(n^{-1+c})$$

asymptotically almost surely, where $t(\cdot)$ was defined in (3).

Proof. It suffices to show that

$$\lambda_j(W_n)\in\left[t\left(\frac{j}{n}\right)-n^{-1+c}, t\left(\frac{j}{n}\right)+n^{-1+c}\right]$$

asymptotically almost surely. Let M_n' be drawn from GUE, and thus the off-diagonal entries of M_n and M_n' match to order 3. From (20) we have

$$\lambda_j(W_n')\in\left[t\left(\frac{j}{n}\right)-\frac{n^{-1+c}}{2}, t\left(\frac{j}{n}\right)+\frac{n^{-1+c}}{2}\right]$$

asymptotically almost surely. The claim now follows from the last part of Theorem 15, letting $G=G(\lambda_j)$ be a smooth cutoff function equal to 1 on

$$\left[t\left(\frac{j}{n}\right)n-\frac{n^c}{2}, t\left(\frac{j}{n}\right)n+\frac{n^c}{2}\right]$$

and vanishing outside

$$\left[t\left(\frac{j}{n}\right)n-n^c, t\left(\frac{j}{n}\right)n+n^c\right].$$ ☐

Proof of Theorem 11. Fix k and u, and let M_n be as in Theorem 11. By Corollary 30, we can find a Johansson matrix M_n' whose entries match M_n to order 4. By Theorem 8 (and a slight rescaling), it suffices to show that the quantity

$$\int_{\mathbb{R}^k} f(t_1, ..., t_k)\varrho_n^{(k)}(nu+t_1, ..., nu+t_k)\, dt_1 \dots dt_k \tag{21}$$

only changes by $o(1)$ when the matrix M_n is replaced by M_n', for any fixed test function f. By an approximation argument, we can take f to be smooth.

We can rewrite the expression (21) as

$$\sum_{1\leqslant i_1,...,i_k\leqslant n} \mathbf{E} f(\lambda_{i_1}(A_n)-nu, ..., \lambda_{i_k}(A_n)-nu). \tag{22}$$

Applying Theorem 15, we already have

$$\mathbf{E} f(\lambda_{i_1}(A_n)-nu, ..., \lambda_{i_k}(A_n)-nu) = \mathbf{E} f(\lambda_{i_1}(A_n')-nu, ..., \lambda_{i_k}(A_n')-nu)+O(n^{-c_0})$$

for each individual $i_1, ..., i_k$ and some absolute constant $c_0>0$. At the same time, by Theorem 32, we see that asymptotically almost surely, the only $i_1, ..., i_k$ which contribute to (22) lie within $O(n^c)$ of $t^{-1}(u)n$, where $c>0$ can be made arbitrarily small. The claim then follows from the triangle inequality (choosing c small enough compared to c_0). □

Proof of the second part of Theorem 9. This proof is similar to the one above. We already know that $\mathbf{P}(\lambda_{i+1}-\lambda_i\leqslant s)$ is basically the same in the two models (M_n and M_n'). Theorem 32 now shows that after fixing a small neighborhood of u, the interval of indices i that involve fluctuates by at most n^c, where c can be made arbitrarily small. □

Remark 33. In fact, in the above applications, we only need Theorem 32 to hold for Johansson matrices. Thus, in order to remove the third moment assumption, it suffices to have this theorem for Johansson matrices (without the third moment assumption). We believe that this is within the power of the determinant process method, but do not pursue this direction here.

As another application, we can prove the following asymptotic for the determinant (and more generally, the characteristic polynomial) of a Wigner Hermitian matrix. The detailed proof is deferred to Appendix A.

Theorem 34. (Asymptotic for determinant) *Let M_n be a Wigner Hermitian matrix whose atom distribution ξ has vanishing third moment and is supported on at least three points. Then there is a constant $c>0$ such that*

$$\mathbf{P}\big(\big|\log|\det M_n|-\log\sqrt{n!}\,\big|\geqslant n^{1-c}\big)=o(1).$$

More generally, for any fixed complex number z, one has

$$\mathbf{P}\left(\left|\log|\det(M_n-zI\sqrt{n})|-\frac{n\log n}{2}-n\int_{-2}^{2}\log|y-z|\varrho_{\mathrm{sc}}(y)\,dy\right|\geqslant n^{1-c}\right)=o(1),$$

where the decay rate of $o(1)$ is allowed to depend on z.

Remark 35. A similar result was established for iid random matrices in [44] (see also [10] for a refinement), based on controlling the distance from a random vector to a subspace. That method relied heavily on the joint independence of all entries and does not seem to extend easily to the Hermitian case. We also remark that a universality result for correlations of the characteristic polynomial has recently been established in [24].

Let us now go beyond the model of Wigner Hermitian matrices. As already mentioned, our main theorem also applies for real symmetric matrices. In the next paragraphs, we formulate a few results that one can obtain in this direction.

Definition 36. (Wigner symmetric matrices) Let n be a large number. A *Wigner symmetric matrix* (of size n) is a random symmetric matrix $M_n=(\xi_{ij})_{1\leqslant i,j\leqslant n}$ where for $1\leqslant i<j\leqslant n$, ξ_{ij} are iid copies of a real random variable ξ with mean zero, variance 1, and exponential decay (as in Definition 1), while for $1\leqslant i\leqslant n$, ξ_{ii} are iid copies of a real random variable ξ' with mean zero, variance 2 and exponential decay. We set $W_n:=M_n/\sqrt{n}$ and $A_n:=M_n\sqrt{n}$ as before.

Example 37. The *Gaussian orthogonal ensemble* (GOE) is the Wigner symmetric matrix in which the off-diagonal atom distribution ξ is the Gaussian $N(0,1)$, and the diagonal atom distribution ξ' is $N(0,2)$.

As remarked earlier, while the Wigner symmetric matrices do not, strictly speaking, obey Condition **C0**, due to the diagonal variance being 2 instead of 1, it is not hard to verify that all the results in this paper continue to hold after changing the diagonal variance to 2. As a consequence, we can easily deduce the following analogue of Theorems 9 and 11.

THEOREM 38. (Universality for random symmetric matrices) *The limiting gap distribution and k-correlation function of Wigner symmetric real matrices with atom variable σ satisfying $\mathbf{E}\sigma^3=0$ and $\mathbf{E}\sigma^4=3$ are the same as those for* GOE. (*The explicit formulae for the limiting gap distribution and k-correlation function for* GOE *can be found in* [1] *and* [34]. *The limit of the k-correlation function is again in the weak sense.*)

The proof of Theorem 38 is similar to that of Theorems 9 and 11 and is omitted. The reason that we need to match the moments to order 4 here (compared to lower orders in Theorems 9 and 11) is that there is currently no analogue of Theorem 8 for the

GOE. Once such a result becomes available, the order automatically reduces to those in Theorems 9 and 11, respectively.

Finally let us mention that our results can be refined and extended in several directions. For instance, we can handle Hermitian matrices whose upper triangular entries are still independent, but having a non-trivial covariance matrix (the real and imaginary parts need not be independent). The diagonal entries can have mean different from zero (which, in the case when the off-diagonal entries are Gaussian, corresponds to Gaussian matrices with external field and has been studied in [7]) and we can obtain universality results in this case as well. We can also refine our argument to prove universality near the edge of the spectrum. These extensions and many others will be discussed in a subsequent paper.

1.7. 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 $X \leqslant CY$ for all sufficiently large n and for some constant C. Notation like $X \ll_k Y$, or $X = O_k(Y)$, means 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. The eigenvalues are always ordered increasingly.

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 $\|x\| := (x^* x)^{1/2}$. The *Frobenius norm* $\|A\|_F$ of a matrix is defined as

$$\|A\|_F = \operatorname{trace}(AA^*)^{1/2}.$$

Note that this bounds the operator norm

$$\|A\|_{\mathrm{op}} := \sup\{\|Ax\| : \|x\| = 1\}$$

of the same matrix. We will also use the following simple inequalities without further comment:

$$\|AB\|_F \leqslant \|A\|_F \|B\|_{\mathrm{op}} \quad \text{and} \quad \|B\|_{\mathrm{op}} \leqslant \|B\|_F,$$

and hence

$$\|AB\|_F \leqslant \|A\|_F \|B\|_F.$$

2. Preliminaries: Tools from linear algebra and probability

2.1. Tools from linear algebra

It is useful to keep in mind the (Courant–Fisher) minimax characterization of the eigenvalues

$$\lambda_i(A) = \min_V \max_{u \in V} u^* A u$$

of a Hermitian $n \times n$ matrix A, where V ranges over the i-dimensional subspaces of $\mathbb{C}^n$, and u ranges over unit vectors in V.

From this, one easily obtain *Weyl's inequality*:

$$\lambda_i(A) - \|B\|_{\mathrm{op}} \leqslant \lambda_i(A+B) \leqslant \lambda_i(A) + \|B\|_{\mathrm{op}}. \tag{23}$$

Another consequence of the minimax formula is the *Cauchy interlacing inequality*:

$$\lambda_i(A_{n-1}) \leqslant \lambda_i(A_n) \leqslant \lambda_{i+1}(A_{n-1}) \tag{24}$$

for all $1 \leqslant i < n$, where A_n is an $n \times n$ Hermitian matrix and A_{n-1} is the top $(n-1) \times (n-1)$ minor. In a similar spirit, one has

$$\lambda_i(A) \leqslant \lambda_i(A+B) \leqslant \lambda_{i+1}(A)$$

for all $1 \leqslant i < n$, whenever A and B are $n \times n$ Hermitian matrices with B being positive semi-definite and of rank 1. If B is instead negative semi-definite, one has

$$\lambda_i(A) \leqslant \lambda_{i+1}(A+B) \leqslant \lambda_{i+1}(A).$$

In either event, we conclude the following.

LEMMA 39. *Let A and B be Hermitian matrices of the same size, with B of rank* 1. *Then, for any interval I,*

$$|N_I(A+B) - N_I(A)| \leqslant 1,$$

where $N_I(M)$ is the number of eigenvalues of M in I.

One also has the following more precise version of the Cauchy interlacing inequality.

LEMMA 40. (Interlacing identity) *Let A_n be an $n \times n$ Hermitian matrix, let A_{n-1} be the top $(n-1) \times (n-1)$ minor, let a_{nn} be the bottom right component, and let $X \in \mathbb{C}^{n-1}$ be the rightmost column with the bottom entry a_{nn} removed. Suppose that X is not orthogonal to any of the unit eigenvectors $u_j(A_{n-1})$ of A_{n-1}. Then we have*

$$\sum_{j=1}^{n-1} \frac{|u_j(A_{n-1})^* X|^2}{\lambda_j(A_{n-1}) - \lambda_i(A_n)} = a_{nn} - \lambda_i(A_n) \tag{25}$$

for every $1 \leqslant i \leqslant n$.

Proof. By diagonalizing A_{n-1} (noting that this does not affect either side of (25)), we may assume that $A_{n-1}=\operatorname{diag}(\lambda_1(A_{n-1}),...,\lambda_{n-1}(A_{n-1}))$ and that $u_j(A_{n-1})=e_j$ for $j=1,...,n-1$. One then easily verifies that the characteristic polynomial $\det(A_n-\lambda I)$ of A_n is equal to

$$\left((a_{nn}-\lambda)-\sum_{j=1}^{n-1}\frac{|u_j(A_{n-1})^*X|^2}{\lambda_j(A_{n-1})-\lambda}\right)\prod_{j=1}^{n-1}(\lambda_j(A_{n-1})-\lambda),$$

when λ is distinct from $\lambda_1(A_{n-1}),...,\lambda_{n-1}(A_{n-1})$. Since $u_j(A_{n-1})^*X$ is non-zero by hypothesis, we see that this polynomial does not vanish at any of the $\lambda_j(A_{n-1})$. Substituting $\lambda_i(A_n)$ for λ, we obtain (25). □

The following lemma will be useful to control the coordinates of eigenvectors.

LEMMA 41. ([21]) *Let*

$$A_n=\begin{pmatrix} a & X^* \\ X & A_{n-1}\end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} x \\ v\end{pmatrix}$$

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

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

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 that $\lambda_i(A)=0$. The eigenvector equation then gives

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

and thus

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

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

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

As

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

the claim follows. □

The *Stieltjes transform* $s_n(z)$ of a Hermitian matrix W is defined for complex z by the formula

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

It has the following alternative representation (see e.g. [2, Chapter 11]).

LEMMA 42. *Let $W=(\zeta_{ij})_{1\leqslant i,j\leqslant n}$ be a Hermitian matrix and let z be a complex number which is not in the spectrum of W. Then we have*

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\zeta_{kk}-z-a_k^*(W_k-zI)^{-1}a_k},$$

where W_k is the $(n-1)\times(n-1)$ matrix with the k-th row and column removed, and $a_k\in\mathbb{C}^{n-1}$ is the k-th column of W with the k-th entry 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. □

2.2. Tools from probability

We will make frequent use of the following lemma, whose proof is presented in Appendix B. This lemma is a generalization of a result in [44].

LEMMA 43. (Distance between a random vector and a subspace) *Let*

$$X = (\xi_1, ..., \xi_n) \in \mathbb{C}^n$$

be a random vector whose entries are independent with mean zero, variance 1, and are bounded in magnitude by K almost surely for some $K\geqslant 10(\mathbf{E}|\xi|^4+1)$. Let H be a subspace of dimension d and π_H be the orthogonal projection onto H. Then

$$\mathbf{P}\big(\big|\|\pi_H(X)\|-\sqrt{d}\,\big|\geqslant t\big)\leqslant 10\exp\biggl(-\frac{t^2}{10K^2}\biggr).$$

In particular, one has

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

with overwhelming probability.

Another useful tool is the following theorem, which is a corollary of a more general theorem proven in Appendix D.

THEOREM 44. (Tail bounds for complex random walks) *Let $1 \leqslant N \leqslant n$ be integers, and let $A=(a_{ij})_{1\leqslant i\leqslant N, 1\leqslant j\leqslant n}$ be an $N\times n$ complex matrix, whose N rows are orthonormal in $\mathbb{C}^n$, obeying the incompressibility condition*

$$\sup_{\substack{1\leqslant i\leqslant N\\ 1\leqslant j\leqslant n}} |a_{ij}| \leqslant \sigma \tag{26}$$

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

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

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

- (Upper tail bound on S_i) *For $t \geqslant 1$, we have*

$$\mathbf{P}(|S_i| \geqslant t) \ll \exp(-ct^2) + C\sigma$$

for some absolute constant $c>0$.

- (Lower tail bound on $\vec{S}$) *For any $t \leqslant \sqrt{N}$, one has*

$$\mathbf{P}(|\vec{S}| \leqslant t) \ll O\left(\frac{t}{\sqrt{N}}\right)^{\lfloor N/4 \rfloor} + CN^4 t^{-3}\sigma.$$

3. Overview of the argument

We now give a high-level proof of our main results, Theorems 15 and 19, contingent on several technical propositions that we prove in later sections.

3.1. Preliminary truncation

In the hypotheses of Theorems 15 and 19, it is assumed that one has the uniform exponential decay property (14) on the coefficients ζ_{ij} in the random matrix M_n. From this and the union bound, we thus see that

$$\sup_{1\leqslant i,j\leqslant n} |\zeta_{ij}| \leqslant (\log n)^{C+1}$$

with overwhelming probability. Since events of probability less than, say, $O(n^{-100})$ are negligible for the conclusion of either Theorem 15 or 19, we may thus apply a standard truncation argument (see e.g. [2]) and *redefine* the atom variables ζ_{ij} on the events where their magnitude exceeds $(\log n)^{C+1}$, so that one in fact has

$$\sup_{1\leqslant i,j\leqslant n} |\zeta_{ij}| \leqslant (\log n)^{C+1} \tag{27}$$

almost surely. (This modification may affect the first, second, third and fourth moments of the real and imaginary parts of the ζ_{ij} by a very small factor (e.g. $O(n^{-10})$), but one can easily compensate for this by further adjustment of the ζ_{ij}, using the Weyl inequalities (23) if necessary; we omit the details.) Thus we will henceforth assume that (27) holds when proving both Theorems 15 and 19.

Remark 45. If one only assumed some finite number of moment conditions on ζ_{ij}, rather than the exponential condition (14), then one could only truncate the $|\zeta_{ij}|$ to be of size n^{1/C_0} for some constant C_0 rather than polylogarithmic in n. While several of our arguments extend to this setting, there is a key induction on n argument in §3.5 that seems to require $|\zeta_{ij}|$ to be of size $n^{o(1)}$ or better, which is the main reason why our results are restricted to random variables of exponential decay. However, this appears to be a largely technical restriction, and it seems very plausible that the results of this paper can be extended to atom distributions that are only assumed to have a finite number of moments bounded.

For technical reasons, it is also convenient to make the qualitative assumption that the ζ_{ij} have an (absolutely) continuous distribution in the complex plane, rather than a discrete one. This is so that pathological events such as eigenvalue collision will only occur with probability zero and can thus be ignored (though one of course still must deal with the event that two eigenvalues have an extremely small but non-zero separation). None of our bounds will depend on any quantitative measure of how continuous the ζ_{ij} are, so one can recover the discrete case from the continuous one by a standard limiting argument (approximating a discrete distribution by a smooth one while holding n fixed, and using the Weyl inequalities (23) to justify the limiting process); we omit the details.

3.2. Proof strategy for Theorem 15

For sake of exposition let us restrict attention to the case $k=1$. Thus we wish to show that the expectation $\mathbf{E}G(\lambda_i(A_n))$ of the random variable $G(\lambda_i(A_n))$ only changes by $O(n^{-c_0})$ if one replaces A_n with another random matrix A_n' with moments matching up to fourth order off the diagonal (and up to second order on the diagonal). To further simplify the exposition, let us suppose that the coefficients ζ_{pq} of A_n (or A_n') are real-valued rather than complex-valued.

At present, A_n' differs from A_n in all n^2 components. But suppose we make a much milder change to A_n, namely replacing a single entry $\zeta_{pq}\sqrt{n}$ of A_n with its counterpart $\zeta_{pq}'\sqrt{n}$ for some $1\leqslant p\leqslant q\leqslant n$. If $p\neq q$, one also needs to replace the companion entry $\zeta_{qp}\sqrt{n}=\bar{\zeta}_{pq}\sqrt{n}$ with $\zeta_{qp}'\sqrt{n}=\bar{\zeta}_{pq}'\sqrt{n}$, to maintain the Hermitian property. This creates another random matrix $\tilde{A}_n$ which differs from A_n in at most two entries. Note that $\tilde{A}_n$ continues to obey Condition **C0**, and has matching moments with either A_n or A_n' up to fourth order off the diagonal, and up to second order on the diagonal.

Suppose that one could show that $\mathbf{E}G(\lambda_i(A_n))$ differed from $\mathbf{E}G(\lambda_i(\tilde{A}_n))$ by at most n^{-2-c_0}, when $p\neq q$, and by at most n^{-1-c_0}, when $p=q$. Then, by applying this swapping procedure once for each pair $1\leqslant p\leqslant q\leqslant n$ and using the triangle inequality, one would obtain the desired bound $|\mathbf{E}G(\lambda_i(A_n))-\mathbf{E}G(\lambda_i(A_n'))|=O(n^{-c_0})$.

Now let us see why we would expect $\mathbf{E}G(\lambda_i(A_n))$ to differ from $\mathbf{E}G(\lambda_i(\tilde{A}_n))$ by such a small amount. For sake of concreteness let us restrict attention to the off-diagonal case $p\neq q$, where we have four matching moments; the diagonal case $p=q$ is similar but one only assumes two matching moments, which is ultimately responsible for the n^{-1-c_0} error rather than n^{-2-c_0}.

Let us freeze (or condition on) all the entries of A_n except for the (p,q) and (q,p) entries. For any complex number z, let $A(z)$ denote the matrix which equals A_n except at the (p,q) and (q,p) entries, where it equals z and $\bar{z}$, respectively. (Actually, with our hypotheses, we only need to consider real-valued z.) Thus it would suffice to show that

$$\mathbf{E}F(\zeta_{pq}\sqrt{n}\,)=\mathbf{E}F(\zeta_{pq}'\sqrt{n}\,)+O(n^{-2-c_0}) \tag{28}$$

for all (or at least most) choices of the frozen entries of A_n, where $F(z):=G(\lambda_i(A(z)))$. Note from (27) that we only care about values of z of size $O(n^{1/2+o(1)})$.

Suppose we could show the derivative estimates

$$\frac{d^l}{dz^l}F(z)=O(n^{-l+O(c_0)+o(1)}) \tag{29}$$

for $l=1,...,5$. (If z were complex-valued rather than real-valued, we would need to differentiate with respect to the real and imaginary parts of z separately, as F is not holomorphic, but let us ignore this technicality for now.) Then, by Taylor's theorem with remainder, we would have

$$F(z)=F(0)+F'(0)z+...+\frac{1}{4!}F^{(4)}(0)z^4+O(n^{-5+O(c_0)+o(1)}|z|^5),$$

and so in particular (using (27))

$$F(\zeta_{pq}\sqrt{n}\,)=F(0)+F'(0)\zeta_{pq}\sqrt{n}+...+\frac{1}{4!}F^{(4)}(0)\zeta_{pq}^4(\sqrt{n}\,)^4+O(n^{-5/2+O(c_0)+o(1)})$$

and similarly for $F(\zeta'_{pq}\sqrt{n}\,)$. Since $n^{-5/2+O(c_0)+o(1)}=O(n^{-2-c_0})$ for n large enough and c_0 small enough, we thus obtain the claim (28) due to the hypothesis that the first four moments of ζ_{pq} and ζ'_{pq} match. (Note how this argument barely fails if only three moments are assumed to match, though it is possible that some refinement of this argument might still succeed by exploiting further cancellations in the fourth-order term $F^{(4)}(0)\zeta_{pq}^4(\sqrt{n}\,)^4/4!$.)

Now we discuss why one would expect an estimate such as (29) to be plausible. For simplicity, we first focus attention on the easiest case $l=1$. Thus we now wish to show that $F'(z)=O(n^{-1+O(c_0)+o(1)})$. By (15) and the chain rule, it suffices to show that

$$\frac{d}{dz}\lambda_i(A(z))=O(n^{-1+O(c_0)+o(1)}).$$

A crude application of the Weyl bound (23) gives

$$\frac{d}{dz}\lambda_i(A(z))=O(1),$$

which is not good enough for what we want (although in the actual proof, we will take advantage of a variant of this crude bound to round z off to the nearest multiple of n^{-100}, which is useful for technical reasons relating to the union bound). But we can do better by recalling the *Hadamard first variation formula*:

$$\frac{d}{dz}\lambda_i(A(z))=u_i(A(z))^*A'(z)u_i(A(z)),$$

where we recall that $u_i(A(z))$ is the ith eigenvector of $A(z)$, normalized to be of unit magnitude. By construction, $A'(z)=e_pe_q^*+e_qe_p^*$, where $e_1,...,e_n$ are the basis vectors of $\mathbb{C}^n$. So to obtain the claim, one needs to show that the coefficients of $u_i(A(z))$ have size $O(n^{-1/2+o(1)})$. This type of delocalization result for eigenvalues has recently been established (with overwhelming probability) by Erdős, Schlein and Yau in [20], [21], [22] for Wigner Hermitian matrices, assuming some quantitative control on the continuous distribution of the ζ_{pq}. (A similar, but weaker, argument was used in [45] with respect to non-Hermitian random matrices; see [45, §4 and Appendix F].) With some extra care and a new tool (Lemma 43), we are able to extend their arguments to cover the current more general setting (see Proposition 62 and Corollary 63), with a slightly simpler proof. Also, z ranges over uncountably many possibilities, so one cannot apply the union bound to each instance of z separately; instead, one must perform the rounding trick mentioned earlier.

Now suppose we wish to establish the $l=2$ version of (29). Again applying the chain rule, we would now seek to establish the bound

$$\frac{d^2}{dz^2}\lambda_i(A(z)) = O(n^{-2+O(c_0)+o(1)}). \tag{30}$$

For this, we apply the Hadamard second variation formula

$$\frac{d^2}{dz^2}\lambda_i(A(z)) = -2u_i(A(z))^*A'(z)(A(z)-\lambda_i(A(z))I)^{-1}\pi_{u_i(A(z))^\perp}A'(z)u_i(A(z)),$$

where $\pi_{u_i(A(z))^\perp}$ is the orthogonal projection onto the orthogonal complement $u_i(A(z))^\perp$ of $u_i(A(z))$, and $(A(z)-\lambda_i(A(z))I)^{-1}$ is the inverse of $A(z)-\lambda_i(A(z))$ on that orthogonal complement. (This formula is valid as long as the eigenvalues $\lambda_j(A(z))$ are simple, which is almost surely the case due to the hypothesis of continuous distribution.) One can expand out the right-hand side in terms of the other (unit-normalized) eigenvectors $u_j(A(z))$, $j\neq i$, as

$$\frac{d^2}{dz^2}\lambda_i(A(z)) = -2\sum_{j\neq i}\frac{|u_j(A(z))^*A'(z)u_i(A(z))|^2}{\lambda_j(A(z))-\lambda_i(A(z))}.$$

By using Erdős–Schlein–Yau type estimates, one expects $|u_j(A(z))^*A'(z)u_i(A(z))|$ to be of size about $O(n^{-1+o(1)})$, while from Theorem 19 we expect $|\lambda_j(z)-\lambda_i(z)|$ to be bounded below by n^{-c_0} with high probability, and so the claim (30) is plausible (one still needs to sum over j, of course, but one expects $\lambda_j(z)-\lambda_i(z)$ to grow roughly linearly in j and so this should only contribute a logarithmic factor $O(\log n)=O(n^{o(1)})$ at worst). So we see for the first time how Theorem 19 is going to be an essential component in the proof of Theorem 15. Similar considerations also apply to the third, fourth and fifth derivatives of $\lambda_i(A(z))$, though as one might imagine the formulae become more complicated.

There is however a technical difficulty that arises, namely that the lower bound

$$|\lambda_j(A(z))-\lambda_i(A(z))| \geqslant n^{-c_0}$$

holds with *high* probability, but not with *overwhelming* probability (see Definition 3 for definitions). Indeed, given that eigenvalue collision is a codimension-2 event for real symmetric matrices and codimension-3 for Hermitian ones, one expects the failure probability to be about n^{-2c_0} in the real case and n^{-3c_0} in the complex case (this heuristic is also supported by the gap statistics for GOE and GUE). As one needs to take the union bound over many values of z (about n^{100} or so), this presents a significant problem.

However, this difficulty can be avoided by going back to the start of the argument and replacing the quantity $G(\lambda_i(z))$ with a "regularized" variant which vanishes whenever $\lambda_i(z)$ gets too close to another eigenvalue. To do this, it is convenient to introduce the quantity

$$Q_i(A(z)) := \sum_{j\neq i} \frac{1}{|\lambda_j(A(z))-\lambda_i(A(z))|^2} = \|(A(z)-\lambda_i(z)I)^{-1}\|_F^2;$$

this quantity is normally of size $O(1)$, but becomes large precisely when the gap between $\lambda_i(A(z))$ and other eigenvalues becomes small. The strategy is then to replace $G(\lambda_i(A(z)))$ by a truncated variant $G(\lambda_i(A(z)), Q_i(A(z)))$ which is supported on the region where Q_i is not too large (e.g. of size at most n^{c_0}), and apply the swapping strategy to the latter quantity instead. (For this, one needs control on derivatives of $Q_i(A(z))$ as well as on $\lambda_i(A(z))$, but it turns out that such bounds are available; this smoothness of Q_i is one reason why we work with Q_i in the first place, rather than more obvious alternatives such as $\inf_{j\neq i} |\lambda_j(A(z))-\lambda_i(A(z))|$.) Finally, to remove the truncation at the beginning and end of the iterated swapping process, one appeals to Theorem 19. Notice that this result is now only used twice, rather than $O(n^2)$ or $O(n^{100})$ times, and so the total error probability remains acceptably bounded.

One way to interpret this truncation trick is that while the "bad event" that Q_i is large has reasonably large probability (of order about n^{-c_0}), which makes the union bound ineffective, the Q_i does not change too violently when swapping one or more of the entries of the random matrix, and so one is essentially faced with the *same* bad event throughout the $O(n^2)$ different swaps (or throughout the $O(n^{100})$ or so different values of z). Thus the union bound is actually far from the truth in this case.

3.3. High-level proof of Theorem 15

We now begin the rigorous proof of Theorem 15, breaking it down into simpler propositions which will be proven in subsequent sections.

The heart of the argument consists of 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 46. (Replacement given a good configuration) *There exists a positive constant C_1 such that the following holds. Let $k\geqslant 1$ and $\varepsilon_1>0$, and assume that n is sufficiently large depending on these parameters. Let $1\leqslant i_1<...<i_k\leqslant n$. 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 *and* e_q *are unit vectors. We assume that for every* $1\leqslant j\leqslant k$ *and every* z, *with* $|z|\leqslant n^{1/2+\varepsilon_1}$, *whose real and imaginary parts are multiples of* n^{-C_1}, *the following properties are satisfied*:

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

$$|\lambda_i(A(z))-\lambda_{i_j}(A(z))|\geqslant n^{-\varepsilon_1}|i-i_j|. \tag{31}$$

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

$$\|P_{i_j}(A(z))e_p\|, \|P_{i_j}(A(z))e_q\|\leqslant n^{-1/2+\varepsilon_1}. \tag{32}$$

- *For every* $\alpha\geqslant 0$,

$$\|P_{i_j,\alpha}(A(z))e_p\|, \|P_{i_j,\alpha}(A(z))e_q\|\leqslant 2^{\alpha/2}n^{-1/2+\varepsilon_1}, \tag{33}$$

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

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

$$\mathbf{E}F(\zeta)=\mathbf{E}F(\zeta')+O(n^{-(r+1)/2+O(\varepsilon_1)}), \tag{34}$$

where

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

with

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

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

$$Q_{i_1}, ..., Q_{i_k}\leqslant n^{\varepsilon_1}$$

and obeys the derivative bounds

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

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

If G *obeys the improved derivative bounds*

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

for $0\leqslant j\leqslant 5$ *and some sufficiently large absolute constant* C, *then we can strengthen* $n^{-(r+1)/2+O(\varepsilon_1)}$ *in* (34) *to* $n^{-(r+1)/2-\varepsilon_1}$.

Remark 47. The need to restrict z to multiples of n^{-C_1}, as opposed to all complex z in the disk of radius $n^{1/2+\varepsilon_1}$, is so that we can verify the hypotheses in the next proposition using the union bound (as long as the events involved hold with overwhelming probability). For C_1 large enough, we will be able to use rounding methods to pass from the discrete setting of multiples of n^{-C_1} to the continuous setting of arbitrary complex numbers in the disk without difficulty.

We prove this proposition in §4. To use this result, we of course need to have the good configuration property holding often. This leads to the second key proposition.

PROPOSITION 48. (Good configurations occur very frequently) *Let* $\varepsilon, \varepsilon_1 > 0$ *and* $C, C_1, k \geqslant 1$. *Let* $\varepsilon n \leqslant i_1 < \ldots < i_k \leqslant (1-\varepsilon)n$, *let* $1 \leqslant p, q \leqslant n$, *let* $e_1, \ldots, e_n$ *be the standard basis of* $\mathbb{C}^n$ *and let* $A(0) = (\zeta_{ij})_{1 \leqslant i,j \leqslant n}$ *be a random Hermitian matrix with independent upper-triangular entries and* $|\zeta_{ij}| \leqslant n^{1/2}(\log n)^C$ *for all* $1 \leqslant i, j \leqslant n$, *with* $\zeta_{pq} = \zeta_{qp} = 0$, *but with* ζ_{ij} *having mean zero and variance* 1 *for all other* i *and* j, *and also being distributed continuously in the complex plane. Then* $A(0), e_p, e_q$ *obey the good configuration condition in Theorem* 46 *for* $i_1, \ldots, i_k$ *and with the indicated values of* ε_1 *and* C_1 *with overwhelming probability.*

We will prove this proposition in §5.

Given these two propositions (and Theorem 19) we can now prove Theorem 15. As discussed at the beginning of the section, we may assume that the ζ_{ij} are continuously distributed in the complex plane and obey the bound (27).

Let $0 < \varepsilon < 1$ and $k \geqslant 1$, and assume that c_0 is sufficiently small and C_1 sufficiently large. Let M_n, M'_n, ζ_{ij}, ζ'_{ij}, A_n, A'_n, G, $i_1, \ldots, i_k$ be as in Theorem 15.

We first need the following lemma.

LEMMA 49. *For each* $1 \leqslant j \leqslant k$, *one has* $Q_{i_j}(A_n) \leqslant n^{c_0}$ *with high probability.*

Proof. For brevity, we omit the variable A_n. Fix j, and suppose that $Q_{i_j} > n^{c_0}$. Then

$$\sum_{i \neq i_j} \frac{1}{|\lambda_i - \lambda_{i_j}|^2} > n^{c_0},$$

and so, by the pigeonhole principle, there exists an integer $0 \leqslant m \ll \log n$ such that

$$\sum_{2^m \leqslant |i - i_j| < 2^{m+1}} \frac{1}{|\lambda_i - \lambda_{i_j}|^2} \gg 2^{-m/2} n^{c_0},$$

which implies that

$$|\lambda_{i_j + 2^m} - \lambda_{i_j}| \ll 2^{3m/4} n^{-c_0/2} \quad \text{or} \quad |\lambda_{i_j - 2^m} - \lambda_{i_j}| \ll 2^{3m/4} n^{-c_0/2}.$$

It thus suffices to show that

$$\mathbf{P}(|\lambda_{i_j+2^m}-\lambda_{i_j}|\ll 2^{3m/4}n^{-c_0/2})\leqslant n^{-c_1}$$

uniformly in m (and similarly for $\lambda_{i_j-2^m}$), since the $\log n$ loss caused by the number of m's can easily be absorbed into the right-hand side.

Fix m. Suppose that $|\lambda_{i_j+2^m}-\lambda_{i_j}|\ll 2^{3m/4}n^{-c_0/2}$; then expressing the left-hand side as $\sum_{k=0}^{2^m-1}(\lambda_{i_j+k+1}-\lambda_{i_j+k})$ and using Markov's inequality, we see that

$$\lambda_{i_j+k+1}-\lambda_{i_j+k}\ll n^{-c_0/2}$$

for $\gg 2^m$ values of k, and thus

$$\mathbf{P}(|\lambda_{i_j+2^m}-\lambda_{i_j}|\ll 2^{3m/4}n^{-c_0/2})\ll \mathbf{E}\frac{1}{2^m}\sum_{k=0}^{2^m-1}\mathbf{I}(\lambda_{i_j+k+1}-\lambda_{i_j+k}\ll n^{-c_0/2}),$$

and hence, by linearity of expectation,

$$\mathbf{P}(|\lambda_{i_j+2^m}-\lambda_{i_j}|\ll 2^{3m/4}n^{-c_0/2})\ll \frac{1}{2^m}\sum_{k=0}^{2^m-1}\mathbf{P}(\lambda_{i_j+k+1}-\lambda_{i_j+k}\ll n^{-c_0/2}).$$

The claim now follows from Theorem 19. (There is a slight issue when $2^m\sim n$, so that the index i_j+k may leave the bulk; but then one works with, say, $\lambda_{i_j+2^{m-1}}-\lambda_{i_j}$ instead of $\lambda_{i_j+2^m}-\lambda_{i_j}$). □

Remark 50. One can also use Theorem 60 below to control all terms in the sum with $|i-i_j|\gg(\log n)^{C'}$ for some C', leading to a simpler proof of Lemma 49.

Of course, Lemma 49 also applies with A_n replaced by A_n'.

Let $\widetilde{G}:\mathbb{R}^k\times\mathbb{R}_+^k\to\mathbb{R}$ be the function

$$\widetilde{G}(\lambda_{i_1},...,\lambda_{i_k},Q_{i_1},...,Q_{i_k}):=G(\lambda_{i_1},...,\lambda_{i_k})\prod_{j=1}^{k}\eta(Q_{i_j}),$$

where $\eta(x)$ is a smooth cutoff function vanishing outside the region $x\leqslant n^{c_0}$ which equals 1 for $x\leqslant\frac{1}{2}n^{c_0}$. From (15) and the chain rule, we see that

$$|\nabla^j\widetilde{G}|\ll n^{c_0}$$

for $j=0,...,5$. Also, from Lemma 49, we have

$$|\mathbf{E}G(\lambda_{i_1}(A_n),...,\lambda_{i_k}(A_n))-\mathbf{E}\widetilde{G}(\lambda_{i_1}(A_n),...,\lambda_{i_k}(A_n),Q_{i_1}(A_n),...,Q_{i_k}(A_n))|\ll n^{-c}$$

for some $c>0$, and similarly with A_n replaced by A_n'. Thus (by choosing c_0 small enough) to prove (16) it will suffice to show that the quantity

$$\mathbf{E}\widetilde{G}(\lambda_{i_1}(A_n), ..., \lambda_{i_k}(A_n), Q_{i_1}(A_n), ..., Q_{i_k}(A_n)) \tag{35}$$

only changes by at most $\frac{1}{2}n^{-c_0}$ when one replaces A_n by A_n'.

As discussed in §3.2, it will suffice to show that the quantity (35) changes by at most $\frac{1}{4}n^{-2-c_0}$ when one swaps the ζ_{pq} entry with $1\leqslant p<q\leqslant n$ to ζ_{pq}' (and ζ_{qp} with ζ_{qp}'), and changes by at most $\frac{1}{4}n^{-1-c_0}$ when one swaps a diagonal entry ζ_{pp} with ζ_{pp}'. But these claims follow from Propositions 48 and 46. (The last part of Proposition 46 is used in the case when one only has three moments matching rather than four.)

The proof of Theorem 15 is now complete (contingent on Theorem 19 and Propositions 46 and 48).

3.4. Proof strategy for Theorem 19

We now informally discuss the proof of Theorem 19.

The machinery of Erdős, Schlein and Yau [20], [21], [22], which is useful in particular for controlling the Stieltjes transform of Wigner matrices, will allow us to obtain good lower bounds on the spectral gap $\lambda_i(A_n)-\lambda_{i-1}(A_n)$ in the bulk, as soon as $k\gg(\log n)^{C'}$ for a sufficiently large C'; see Theorem 60 for a precise statement. The difficulty here is that k is exactly 1. To overcome this difficulty, we will try to amplify the value of k by looking at the top left $(n-1)\times(n-1)$ minor A_{n-1} of A_n, and observing the following "backwards gap propagation" phenomenon:

If $\lambda_i(A_n)-\lambda_{i-k}(A_n)$ is very small, then $\lambda_i(A_{n-1})-\lambda_{i-k-1}(A_{n-1})$ will also be small with reasonably high probability.

If one accepts this phenomenon, then, by iterating it about $(\log n)^{C'}$ times, one can enlarge the spacing k to be of the size large enough so that an Erdős–Schlein–Yau type bound can be invoked to obtain a contradiction. (There will be a technical difficulty caused by the fact that the failure probability of this phenomenon, when suitably quantified, can be as large as $1/(\log n)^{O(1)}$, and thus apparently precluding the ability to get a polynomially strong bound on the failure rate, but we will address this issue later.)

Note that the converse of this statement follows from the Cauchy interlacing property (24). To explain why this phenomenon is plausible, observe from (24) that if $\lambda_i(A_n)-\lambda_{i-k}(A_n)$ is small, then $\lambda_i(A_n)-\lambda_{i-k-1}(A_{n-1})$ is also small. On the other hand, from Lemma 40 one has the identity

$$\sum_{j=1}^{n-1}\frac{|u_j(A_{n-1})^*X|^2}{\lambda_j(A_{n-1})-\lambda_i(A_n)}=\zeta_{nn}\sqrt{n}-\lambda_i(A_n), \tag{36}$$

where X is the rightmost column of A_n (with the bottom entry $\zeta_{nn}\sqrt{n}$ removed).

One expects $|u_j(A_{n-1})^*X|^2$ to have size about n on average (cf. Lemma 43). In particular, if $\lambda_i(A_n)-\lambda_{i-k}(A_{n-1})$ is small (e.g. of size $O(n^{-c})$), then the $j=i-1$ term is expected to give a large negative contribution (of size $\gg n^{1+c}$) to the left-hand side of the identity (36). At the same time, the right-hand side is much smaller, of size $O(n)$ or so on average; so we expect to have the large negative contribution mentioned earlier to be counterbalanced by a large positive contribution from some other index. The index which is most likely to supply such a large positive contribution is $j=i$, and so one expects $\lambda_i(A_{n-1})-\lambda_i(A_n)$ to be small (also of size $O(n^{-c})$, in fact). A similar argument also leads one to expect $\lambda_{i-k}(A_n)-\lambda_{i-k-1}(A_n)$ to be small, and the claimed phenomenon then follows from the triangle inequality.

In order to make the above strategy rigorous, there are a number of technical difficulties. The first is that the counterbalancing term mentioned above need not come from $j=i$, but could instead come from another value of j, or perhaps a "block" of several j put together, and so one may have to replace the gap $\lambda_i(A_{n-1})-\lambda_{i-k-1}(A_{n-1})$ by a more general type of gap. A second problem is that the gap $\lambda_i(A_{n-1})-\lambda_{i-k-1}(A_{n-1})$ is going to be somewhat larger than the gap $\lambda_i(A_n)-\lambda_{i-k}(A_n)$, and one is going to be iterating this gap growth about $(\log n)^{O(1)}$ times. In order to be able to contradict Theorem 60 at the end of the argument, the net gap growth should only be at most $O(n^c)$ for some small $c>0$. So one needs a reasonable control on the ratio between the gap for A_{n-1} and the gap for A_n; in particular, if one can keep the former gap to be at most $(1+1/k)^{O((\log n)^{0.9})}$ times the latter gap, then the net growth in the gap telescopes to $((\log n)^{O(1)})^{O((\log n)^{0.9})}$, which is indeed less than $O(n^c)$ and thus acceptable. To address these issues, we fix a base value n_0 of n, and for any $1\leqslant i-l<i\leqslant n\leqslant n_0$, we define the *regularized gap*

$$g_{i,l,n} := \inf_{1\leqslant i_-\leqslant i-l<i\leqslant i_+\leqslant n} \frac{\lambda_{i_+}(A_n)-\lambda_{i_-}(A_n)}{\min\{i_+-i_-, (\log n_0)^{C_1}\}^{(\log n_0)^{0.9}}}, \tag{37}$$

where $C_1>1$ is a large constant (depending on C) to be chosen later. (We need to cap i_+-i_- off at $(\log n_0)^{C_1}$ to prevent the large values of i_+-i_- from overwhelming the infimum, which is not what we want.)

We will shortly establish a rigorous result that asserts, roughly speaking, that if the gap $g_{i,l,n+1}$ is small, then the gap $g_{i,l+1,n}$ is also likely to be small, and thus giving a precise version of the phenomenon mentioned earlier.

There is one final obstacle, which has to do with the failure probability when $g_{i,l,n+1}$ is small but $g_{i,l+1,n}$ is large. If this event could be avoided with overwhelming probability (or even a high probability), then one would be done by the union bound (note that we only need to take the union over $O((\log n)^{O(1)})$ different events). While many of the events that could lead to failure can indeed be avoided with high probability, there is one type of event which does cause a serious problem, namely that the inner products $u_j(A_{n-1})^*X$ for $i_-\leqslant j\leqslant i_+$ could be unexpectedly small. Talagrand's inequality (Lemma 43) can be used to control this event effectively when i_+-i_- is large, but when i_+-i_- is small the probability of failure can be as high as $1/(\log n)^c$ for some $c>0$. However, one can observe that such high failure rates only occur when $g_{i,l+1,n}$ is only *slightly* larger than $g_{i,l,n+1}$. Indeed, one can show that the probability that $g_{i,l,n+1}$ is much higher than $g_{i,l+1,n}$, say of size $2^m g_{i,l,n+1}$ or more, is only

$$O\biggl(\frac{2^{-m/2}}{(\log n)^c}\biggr)$$

(for reasonable values of m), and in fact (due to Talagrand's inequality) the constant c can be increased to be much larger when l is large. This is still not quite enough for a union bound to give a total failure probability of $O(n^{-c})$, but one can exploit the martingale-type structure of the problem (or more precisely, the fact that the column X remains random, with independent entries, even after conditioning out all of the block A_{n-1}) to multiply the various bad failure probabilities together to end up with the final bound of $O(n^{-c})$.

3.5. High-level proof of Theorem 19

We now prove Theorem 19. Fix ε and c_0. We write i_0 and n_0 for i and n, respectively. Thus

$$\varepsilon n_0 \leqslant i_0 \leqslant (1-\varepsilon)n_0$$

and the task is to show that $|\lambda_{i_0}(A_{n_0})-\lambda_{i_0}(A_{n_0-1})|\geqslant n_0^{-c_0}$ with high probability. We may of course assume that n_0 is large compared to all other parameters. We may also assume the bound (27), and that the distribution of the A_n is continuous, so that events such as repeated eigenvalues occur with probability zero and can thus be neglected.

Let C_1 be a large constant to be chosen later. For any l and n with $1\leqslant i-l<i\leqslant n\leqslant n_0$, we define the normalized gap $g_{i,l,n}$ by (37). It will suffice to show that

$$g_{i_0,1,n_0}\leqslant n^{-c_0} \tag{38}$$

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

The first main tool for this is the following (deterministic) lemma, proven in §6.

Lemma 51. (Backwards propagation of gap) *Let* $\frac{1}{2}n_0 \leqslant n < n_0$ *and* $l \leqslant \frac{1}{10}\varepsilon n$ *be such that*

$$g_{i_0,l,n+1} \leqslant \delta \tag{39}$$

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

$$g_{i_0,l+1,n} \geqslant 2^m g_{i_0,l,n+1} \tag{40}$$

for some $m \geqslant 0$ *with*

$$2^m \leqslant \delta^{-1/2}. \tag{41}$$

Then, *one of the following statements hold*:

(i) (Macroscopic spectral concentration) *There exist* i_+ *and* i_-, $1 \leqslant i_- < i_+ \leqslant n+1$, *with* $i_+ - i_- \geqslant (\log n)^{C_1/2}$, *such that*

$$|\lambda_{i_+}(A_{n+1}) - \lambda_{i_-}(A_{n+1})| \leqslant \delta^{1/4} \exp((\log n)^{0.95})(i_+ - i_-).$$

(ii) (Small inner products) *There are* i_+ *and* i_-, $\frac{1}{2}\varepsilon n \leqslant i_- \leqslant i_0 - l < i_0 \leqslant i_+ \leqslant \left(1 - \frac{1}{2}\varepsilon\right)n$, *with* $i_+ - i_- \leqslant (\log n)^{C_1/2}$, *such that*

$$\sum_{i_- \leqslant j < i_+} |u_j(A_n)^* X_n|^2 \leqslant \frac{n(i_+ - i_-)}{2^{m/2}(\log n)^{0.01}}. \tag{42}$$

(iii) (Large coefficient) *We have*

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

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

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

(v) (Large inner product in bulk) *There exists* $\frac{1}{10}\varepsilon n \leqslant i \leqslant \left(1 - \frac{1}{10}\varepsilon\right)n$ *such that*

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

(vi) (Large row) *We have*

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

(vii) (Large inner product near i_0) *There exists* $\frac{1}{10}\varepsilon n \leqslant i \leqslant \left(1 - \frac{1}{10}\varepsilon\right)n$, *satisfying* $|i - i_0| \leqslant (\log n)^{C_1}$, *such that*

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

Remark 52. In the applications, δ will be a small negative power of n. The main bad event here is (ii) (and to a lesser extent (vii)); the other events will have a polynomially small probability of occurrence in practice (as a function of n) and so can be easily discarded. The events (ii) and (vii) are more difficult to discard, since their probability is not polynomially small in n, if m is small. On the other hand, these probabilities decay exponentially in m, and furthermore are independent in a martingale sense, and this will be enough for us to obtain a proper control. The exact numerical values of the exponents such as 0.9, 0.95, 0.8, etc. are not particularly important, though of course they need to lie between 0 and 1.

The second key proposition bounds the probability that each of the bad events (i)–(vii) occur, proven in §7.

PROPOSITION 53. (Bad events are rare) *Suppose that* $\frac{1}{2}n_0 \leqslant n < n_0$ *and* $l \leqslant \frac{1}{10}\varepsilon n$, *and set* $\delta := n_0^{-\varkappa}$ *for some sufficiently small fixed* $\varkappa > 0$. *Then, the following facts hold*:

(a) *The events* (i), (iii), (iv), (v) *and* (vi) *in Lemma* 51 *all fail with high probability.*

(b) *There is a constant* C' *such that all the coefficients of the eigenvectors* $u_j(A_n)$ *for* $\frac{1}{2}\varepsilon n \leqslant j \leqslant \left(1-\frac{1}{2}\varepsilon\right)n$ *are of magnitude at most* $n^{-1/2}(\log n)^{C'}$ *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* $2^{-\varkappa m} + n^{-\varkappa}$.

(c) *Furthermore, there is a constant* C_2 *(depending on* C', $\varkappa$ *and* C_1*) such that if* $l \geqslant C_2$ *and* A_n *is conditioned as in* (b), *then* (ii) *and* (vii) *in fact occur with a conditional probability of at most* $2^{-\varkappa m}(\log n)^{-2C_1} + n^{-\varkappa}$.

Let us assume these two propositions for now and conclude the proof of Theorem 19.

We may assume that c_0 is small. Set $\varkappa := \frac{1}{10}c_0$. For each $n_0 - (\log n_0)^{2C_1} \leqslant n \leqslant n_0$, let E_n be the event that one of the eigenvectors $u_j(A_n)$ for $\frac{1}{2}\varepsilon n \leqslant j \leqslant \left(1-\frac{1}{2}\varepsilon\right)n$ has a coefficient of magnitude more than $n^{-1/2}(\log n)^{C'}$, and let E_0 be the event that at least one of the exceptional events (i), (iii)–(vi), or E_n hold for some $n - (\log n_0)^{2C_1} \leqslant n \leqslant n_0$. Then, by Proposition 53 and the union bound, we have

$$\mathbf{P}(E_0) \leqslant n_0^{-\varkappa/2}. \tag{43}$$

It thus suffices to show that the event

$$g_{i_0,1,n_0} \leqslant n^{-10\varkappa} \quad \text{and} \quad E_0^c$$

is avoided with high probability.

To bound this, the first step is to increase the parameter l from 1 to C_2, in order to use Proposition 53 (c). Set $2^m := n_0^{\varkappa/C_2}$. From Proposition 53 (b), we see that the event that E_n fails, but (ii) or (vii) hold for $n=n_0-1$, $l=1$ and some i_- and i_+, occurs with probability $O(2^{-\varkappa m}+n_0^{-\varkappa})$. Applying Lemma 51 (noting that $n_0^{-10\varkappa}\leqslant\delta$), we conclude that

$$\mathbf{P}(g_{i_0,1,n_0}\leqslant n_0^{-10\varkappa} \text{ and } E_0^c)\leqslant\mathbf{P}(g_{i_0,2,n_0-1}\leqslant 2^m n_0^{-10\varkappa} \text{ and } E_0^c)+O(2^{-\varkappa m}+n_0^{-\varkappa}).$$

We can iterate this process C_2 times and conclude that

$$\begin{aligned}\mathbf{P}(g_{i_0,1,n_0}\leqslant n_0^{-10\varkappa} \text{ and } E_0^c)\leqslant\mathbf{P}(g_{i_0,C_2+1,n_0-C_2}\leqslant 2^{C_2 m} n_0^{-10\varkappa} \text{ and } E_0^c)&\\ +O(C_2 2^{-\varkappa m}+C_2 n_0^{-\varkappa});&\end{aligned}$$

substituting in the definition of m, we conclude that

$$\mathbf{P}(g_{i_0,1,n_0}\leqslant n_0^{-10\varkappa} \text{ and } E_0^c)\leqslant\mathbf{P}(g_{i_0,C_2+1,n_0-C_2}\leqslant n_0^{-9\varkappa} \text{ and } E_0^c)+n_0^{-\varkappa^2/2C_2}.$$

So it will suffice to show that

$$\mathbf{P}(g_{i_0,C_2+1,n_0-C_2}\leqslant n_0^{-9\varkappa} \text{ and } E_0^c)\leqslant n_0^{-\varkappa^2/2C_2}.$$

By Markov's inequality, it suffices to show that

$$\mathbf{E}Z_{n_0-C_2}^{-\varkappa/2}\mathbf{I}(E_0^c)\leqslant n_0^{\varkappa^2}, \tag{44}$$

where, for each $n_0-(\log n_0)^{C_1}\leqslant n\leqslant n_0-C_2$, Z_n is the random variable

$$Z_n:=\max\{\min\{g_{i_0,n_0-n+1,n},\delta\},n_0^{-9\varkappa}\}.$$

Indeed, we have

$$\begin{aligned}\mathbf{P}(g_{i_0,C_2+1,n_0-C_2}\leqslant n_0^{-9\varkappa} \text{ and } E_0^c)&\leqslant\mathbf{P}(Z_{n_0-C_2}=n_0^{-9\varkappa} \text{ and } E_0^c)\\ &\leqslant n_0^{-9\varkappa^2/2}\mathbf{E}Z_{n_0-C_2}^{-\varkappa/2}\mathbf{I}(E_0^c),\end{aligned}$$

whence the claim.

We now establish a recursive inequality for $\mathbf{E}Z_n^{-\varkappa/2}\mathbf{I}(E_0^c)$. Let n be such that $n_0-(\log n_0)^{C_1}\leqslant n<n_0-C_2$. Suppose we condition A_n so that E_n fails. Then, for any $m\geqslant 0$, we see from Proposition 53 (c) that (ii) or (vii) hold for $l=n_0-n$ and some i_- and i_+, with (conditional) probability at most $O(2^{-\varkappa m}(\log n)^{-2C_1}+n^{-\varkappa})$. Applying Lemma 51, we conclude that

$$\begin{aligned}\mathbf{P}(g_{i_0,n_0-n,n+1}\leqslant\delta \text{ and } g_{i_0,n_0-n+1,n}\geqslant 2^m g_{i_0,n_0-n,n+1} \text{ and } E_0^c\,|\,A_n)&\\ \ll 2^{-\varkappa m}(\log n)^{-2C_1}+n^{-\varkappa}.&\end{aligned}$$

Note that this inequality is also vacuously true if A_n is such that E_n holds, since the event E_0^c is then empty.

Observe that, if $Z_n > 2^m Z_{n+1}$ for some $m \geqslant 0$, then

$$g_{i_0,n_0-n,n+1} \leqslant \delta \quad \text{and} \quad g_{i_0,n_0-n+1,n} \geqslant 2^m g_{i_0,n_0-n,n+1}.$$

Thus

$$\mathbf{P}(Z_n > 2^m Z_{n+1} \text{ and } E_0^c \mid A_n) \ll 2^{-\varkappa m} (\log n)^{-2C_1} + n^{-\varkappa},$$

or equivalently

$$\mathbf{P}(Z_{n+1}^{-\varkappa/2} > 2^{m\varkappa/2} Z_n^{-\varkappa/2} \text{ and } E_0^c \mid A_n) \ll 2^{-\varkappa m} (\log n)^{-2C_1} + n^{-\varkappa}.$$

Since we are conditioning on A_n, Z_n is deterministic. Also, from the definition of Z_n, this event is vacuous for $2^m > n^{8\varkappa}$, and thus we can simplify the above bound as

$$\mathbf{P}(Z_{n+1}^{-\varkappa/2} > 2^{m\varkappa/2} Z_n^{-\varkappa/2} \text{ and } E_0^c \mid A_n) \leqslant 3 \cdot 2^{-\varkappa m} (\log n)^{-2C_1}.$$

Now we multiply this by $2^{m\varkappa/2}$ and sum over $m \geqslant 0$ to obtain

$$\mathbf{E}(Z_{n+1}^{-\varkappa/2} \mathbf{I}(E_0^c) \mid A_n) \leqslant Z_n^{-\varkappa/2} (1 + (\log n)^{-(2C_1-1)}).$$

Undoing the conditioning on A_n, we conclude that

$$\mathbf{E} Z_{n+1}^{-\varkappa/2} \mathbf{I}(E_0^c) \leqslant (1 + (\log n)^{-(2C_1-1)}) \mathbf{E} Z_n^{-\varkappa/2}.$$

Applying (43) (and the trivial bound $Z_n^{-\varkappa/2} \leqslant n^{9\varkappa^2/2}$), we have

$$\mathbf{E} Z_{n+1}^{-\varkappa/2} \mathbf{I}(E_0^c) \leqslant (1 + (\log n)^{-(2C_1-1)}) \mathbf{E} Z_n^{-\varkappa/2} \mathbf{I}(E_0^c) + n^{-\varkappa/4}.$$

Iterating this, we conclude that

$$\mathbf{E} Z_{n_0-C_2}^{-\varkappa/2} \mathbf{I}(E_0^c) \leqslant 2 \mathbf{E} Z_{n_0-\lfloor (\log n_0)^{C_1} \rfloor}^{-\varkappa/2} \mathbf{I}(E_0^c) + n_0^{-\varkappa/8}. \tag{45}$$

On the other hand, if E_0^c holds, then by (i) we have

$$\left| \lambda_{i_+}^{n_0 - \lfloor (\log n_0)^{C_1} \rfloor} - \lambda_{i_-}^{n_0 - \lfloor (\log n_0)^{C_1} \rfloor} \right| < n_0^{-\varkappa} \exp((\log n)^{0.95}) (i_+ - i_-)$$

whenever $1 \leqslant i_- \leqslant i - (\log n_0)^{C_1/2} < i \leqslant i_+ \leqslant n$. From this we have

$$g_{i_0, \lfloor (\log n_0)^{C_1} \rfloor + 1, n_0 - \lfloor (\log n_0)^{C_1} \rfloor} \leqslant n_0^{-\varkappa/2},$$

and hence

$$Z_{n_0 - \lfloor (\log n_0)^{C_1} \rfloor} = n_0^{-\varkappa}.$$

Inserting this into (45), we obtain (44) as required.

4. Good configurations have stable spectra

The purpose of this section is to prove Proposition 46. The first stage is to obtain some equations for the derivatives of eigenvalues of Hermitian matrices with respect to perturbations.

4.1. Spectral dynamics of Hermitian matrices

Suppose that $\lambda_i(A)$ is a *simple* eigenvalue, which means that $\lambda_i(A)\neq\lambda_j(A)$ for all $j\neq i$; note that almost all Hermitian matrices have simple eigenvalues. We then define $P_i(A)$ to be the orthogonal projection onto the 1-dimensional eigenspace corresponding to $\lambda_i(A)$. Thus, if $u_i(A)$ is a unit eigenvector for the eigenvalue $\lambda_i(A)$, then $P_i(A)=u_i(A)u_i(A)^*$. We also define the *resolvent* $R_i(A)$ to be the unique Hermitian matrix inverting $A-\lambda_i(A)I$ on the range of $I-P_i(A)$, and vanishing on the range of $P_i(A)$. If $u_1(A), ..., u_n(A)$ form an orthonormal eigenbasis associated with $\lambda_1(A), ..., \lambda_n(A)$, then we can write $R_i(A)$ explicitly as

$$R_i(A)=\sum_{j\neq i}\frac{1}{\lambda_j-\lambda_i}u_j(A)u_j(A)^*.$$

It is clear that

$$R_i(A-\lambda_i I)=I-P_i. \tag{46}$$

We also need the quantity

$$Q_i(A):=\|R_i(A)\|_F^2=\sum_{j\neq i}\frac{1}{|\lambda_j-\lambda_i|^2}.$$

By (23), each eigenvalue function $A\mapsto\lambda_i(A)$ for $1\leqslant i\leqslant n$ is continuous. However, we will need a quantitative control on the derivatives of this function. The first observation is that λ_i (as well as P_i, R_i and Q_i) depends smoothly on A whenever that eigenvalue is simple (even if other eigenvalues have multiplicity).

Lemma 54. *Let $1\leqslant i\leqslant n$ and let A_0 be a Hermitian matrix which has a simple eigenvalue at $\lambda_i(A_0)$. Then λ_i, P_i, R_i and Q_i are smooth for A in a neighborhood of A_0.*

Proof. By the Weyl inequality (23), $\lambda_i(A_0)$ stays away from the other $\lambda_j(A_0)$ by a bounded distance for all A_0 in a neighborhood of A_0. In particular, the characteristic polynomial $\det(A-\lambda I)$ has a simple zero at $\lambda_i(A)$ for all such A. Since this polynomial also depends smoothly on A, the smoothness of λ_i now follows. As $A-\lambda_i(A)I$ depends smoothly on A, has a single zero eigenvalue, and has all other eigenvalues bounded away from zero, we see that the 1-dimensional kernel $\ker(A-\lambda_i(A)I)$ also depends smoothly on A near A_0. Since P_i is the orthogonal projection onto this kernel, the smoothness of P_i now follows.

As $A-\lambda_i(A)I$ depends smoothly on A, and has eigenvalues bounded away from zero on the range of $1-P_i$ (which is also smoothly dependent on A), we see that R_i (and hence Q_i) also depends smoothly on A. □

Now we turn to more quantitative estimates on the smoothness of λ_i, P_i, R_i and Q_i for fixed $1\leqslant i\leqslant n$. For our applications to Proposition 46, we consider matrices $A=A(z)$ which are parameterized smoothly (though not holomorphically, of course) by some complex parameter z in a domain $\Omega\subset\mathbb{C}$. We assume that $\lambda_i(A(z))$ is simple for all $z\in\Omega$, which, by the above lemma, implies that $\lambda_i:=\lambda_i(A(z))$, $P_i:=P_i(A(z))$, $R_i:=R_i(A(z))$ and $Q_i:=Q_i(A(z))$ all depend smoothly on z in Ω.

It will be convenient to introduce some more notation, to deal with the technical fact that z is complex rather than real. For any smooth function $f(z)$ (which may be scalar, vector, or matrix-valued), we use

$$\nabla^m f(z) := \left(\frac{\partial^m f}{\partial \mathrm{Re}(z)^l \partial \mathrm{Im}(z)^{m-l}}\right)_{l=0}^m$$

to denote the mth gradient with respect to the real and imaginary parts of z. (Thus, $\nabla^m f$ is an $(m+1)$-tuple, each of whose components is of the same type as f; for instance, if f is matrix-valued, so are all the components of $\nabla^m f$.) If f is matrix-valued, we define $\|\nabla^m f\|_F$ to be the ℓ^2 norm of the Frobenius norms of the various components of $\nabla^m f$, and similarly for other norms.

We observe the Leibniz rule

$$\nabla^k(fg) = \sum_{m=0}^{k} (\nabla^m f)*(\nabla^{k-m} g) = f(\nabla^k g)+(\nabla^k f)g+\sum_{m=1}^{k-1}(\nabla^m f)*(\nabla^{k-m} g), \tag{47}$$

where the $(k+1)$-tuple $(\nabla^m f)*(\nabla^{k-m} g)$ is defined as

$$\left(\sum_{l'=\max\{0,l+m-k\}}^{\min\{l,m\}} \binom{l}{l'}\binom{k-l}{m-l'} \frac{\partial^{k-m} f}{\partial \mathrm{Re}(z)^{l'}\partial \mathrm{Im}(z)^{m-l'}} \frac{\partial^m g}{\partial \mathrm{Re}(z)^{l-l'}\partial \mathrm{Im}(z)^{k-m-l+l'}}\right)_{l=0}^k.$$

The exact coefficients here are not important, and one can view $(\nabla^m f)*(\nabla^{k-m} g)$ simply as a bilinear combination of $\nabla^m f$ and $\nabla^{k-m} g$. Note that (47) is valid for matrix-valued f and g as well as scalar f and g. For a tuple $(A_1, ..., A_l)$ of matrices, we define

$$\mathrm{trace}(A_1, ..., A_l) := (\mathrm{trace}(A_1), ..., \mathrm{trace}(A_l)).$$

We can now give the higher-order Hadamard variation formulae.

Proposition 55. (Recursive formula for derivatives of λ_i, P_i and R_i) *For any integer* $k\geqslant 1$, *we have*

$$\nabla^k\lambda_i=\sum_{m=1}^{k}\operatorname{trace}((\nabla^m A)*(\nabla^{k-m}P_i)P_i)-\sum_{m=1}^{k-1}(\nabla^m\lambda_i)*\operatorname{trace}((\nabla^{k-m}P_i)P_i) \tag{48}$$

and

$$\begin{aligned}\nabla^k P_i=&-R_i(\nabla^k A)P_i-P_i(\nabla^k A)R_i\\&-\sum_{m=1}^{k-1}[R_i((\nabla^m A)-(\nabla^m\lambda_i)I)*(\nabla^{k-m}P_i)P_i\\&\qquad+P_i(\nabla^{k-m}P_i)*((\nabla^m A)-(\nabla^m\lambda_i)I)R_i]\\&+\sum_{m=1}^{k-1}(\nabla^m P_i)*(\nabla^{k-m}P_i)(I-2P_i).\end{aligned} \tag{49}$$

Furthermore,

$$(\nabla^k R_i)P_i=-\sum_{m=0}^{k-1}(\nabla^m R_i)*(\nabla^{k-m}P_i) \tag{50}$$

and

$$(\nabla^k R_i)(I-P_i)=-(\nabla^k P_i)R_i-\sum_{m=0}^{k-1}(\nabla^m R_i)*((\nabla^{k-m}A)-(\nabla^{k-m}\lambda_i)I)R_i, \tag{51}$$

and thus

$$\begin{aligned}(\nabla^k R_i)=&-(\nabla^k P_i)R_i-\sum_{m=0}^{k-1}(\nabla^m R_i)*((\nabla^{k-m}A)-(\nabla^{k-m}\lambda_i)I)R_i\\&-\sum_{m=0}^{k-1}(\nabla^m R_i)*(\nabla^{k-m}P_i).\end{aligned} \tag{52}$$

Proof. Our starting point is the identities

$$\lambda_i P_i=AP_i \tag{53}$$

and

$$P_iP_i=P_i. \tag{54}$$

We differentiate these identities k times using the Leibniz rule (47) to obtain

$$(\nabla^k\lambda_i)P_i+\sum_{m=0}^{k-1}(\nabla^m\lambda_i)*(\nabla^{k-m}P_i)=\sum_{m=0}^{k}(\nabla^m A)*(\nabla^{k-m}P_i) \tag{55}$$

and

$$(\nabla^k P_i)P_i+P_i(\nabla^k P_i)+\sum_{m=1}^{k-1}(\nabla^m P_i)*(\nabla^{k-m}P_i)=(\nabla^k P_i). \quad (56)$$

Multiplying (55) by P_i and taking traces, one obtains (48) (the $m=0$ terms cancel because of (53), which implies that $\operatorname{trace}(A(\nabla^m P_i)P_i)=\lambda_i \operatorname{trace}((\nabla^m P_i)P_i))$.

We next compute $\nabla^k P_i$ using the decomposition

$$\nabla^k P_i=P_i(\nabla^k P_i)P_i+(I-P_i)(\nabla^k P_i)(I-P_i)+(I-P_i)(\nabla^k P_i)P_i+P_i(\nabla^k P_i)(I-P_i). \quad (57)$$

Multiplying both sides of (56) by P_i (on the right) and using the identity $P_iP_i=P_i$, we get a cancelation which implies that

$$P_i(\nabla^k P_i)P_i=-\sum_{m=1}^{k-1}(\nabla^m P_i)*(\nabla^{k-m}P_i)P_i. \quad (58)$$

Repeating the same trick with $I-P_i$ instead of P_i, we have

$$(I-P_i)(\nabla^k P_i)(I-P_i)=\sum_{m=1}^{k-1}(\nabla^m P_i)*(\nabla^{k-m}P_i)(I-P_i). \quad (59)$$

This gives two of the four components of $\nabla^k P_i$. To obtain the other components, multiply (55) on the left by $I-P_i$ and notice that the $(I-P_i)(\nabla^k \lambda_i)P_i$ term vanishes because of (54). Rearranging the terms, we obtain

$$(I-P_i)(A-\lambda_i)(\nabla^k P_i)=-\sum_{m=1}^{k-1}(I-P_i)((\nabla^m A)-(\nabla^m\lambda_i)I)*(\nabla^{k-m}P_i)-(I-P_i)(\nabla^k A)P_i.$$

Applying R_i on the left and P_i on the right and using (46), we get

$$(I-P_i)(\nabla^k P_i)P_i=-\sum_{m=1}^{k-1}R_i((\nabla^m A)-(\nabla^m\lambda_i)I)*(\nabla^{k-m}P_i)P_i-R_i(\nabla^k A)P_i.$$

By taking adjoints, we obtain

$$P_i(\nabla^k P_i)(I-P_i)=-\sum_{m=1}^{k-1}P_i(\nabla^{k-m}P_i)*((\nabla^m A)-(\nabla^m\lambda_i)I)R_i-P_i(\nabla^k A)R_i.$$

These, together with (57), (58) and (59), imply (49).

We now turn to R_i. Here, we use the identities (46),

$$R_i(A-\lambda_i I)=I-P_i \quad \text{and} \quad R_iP_i=0.$$

Differentiating the second identity k times gives (50). On the other hand, differentiating the first identity k times gives

$$(\nabla^k R_i)(A-\lambda_i I) = -(\nabla^k P_i) - \sum_{m=0}^{k-1} (\nabla^m R_i) * ((\nabla^{k-m} A) - (\nabla^{k-m} \lambda_i) I);$$

multiplying on the right by R_i, we obtain (51), and then (52) follows. □

We isolate the $k=1$ case of Proposition 55, obtaining the *Hadamard variation formulae*:

$$\nabla \lambda_i = \operatorname{trace}(\nabla A P_i) \tag{60}$$

and

$$\nabla P_i = -R_i(\nabla A) P_i - P_i(\nabla A) R_i. \tag{61}$$

4.2. Bounding the derivatives

We now use the recursive inequalities obtained in the previous section to bound the derivatives of λ_i and P_i, assuming some quantitative control on the spectral gap between λ_i and other eigenvalues, and on the matrix A and its derivatives. Let us begin with a crude bound.

LEMMA 56. (Crude bound) *Let $A=A(z)$ be an $n\times n$ matrix varying (real)-linearly in z (thus $\nabla^k A=0$ for $k\geqslant 2$), with*

$$\|\nabla A\|_{\mathrm{op}} \leqslant V$$

for some $V>0$. Let $1\leqslant i\leqslant n$. At some fixed value of z, suppose we have the spectral gap condition

$$|\lambda_j(A(z)) - \lambda_i(A(z))| \geqslant r \tag{62}$$

for all $j\neq i$ and some $r>0$ (in particular, $\lambda_i(A(z))$ is a simple eigenvalue). Then for all $k\geqslant 1$ we have (at this fixed choice of z)

$$|\nabla^k \lambda_i| \ll_k V^k r^{1-k}, \tag{63}$$

$$\|\nabla^k P_i\|_{\mathrm{op}} \ll_k V^k r^{-k}, \tag{64}$$

$$\|\nabla^k R_i\|_{\mathrm{op}} \ll_k V^k r^{-k-1}, \tag{65}$$

$$|\nabla^k Q_i| \ll_k n V^k r^{-k-2}. \tag{66}$$

Proof. Observe that the spectral gap condition (62) ensures that

$$\|R_i\|_{\mathrm{op}} \leqslant \frac{1}{r}. \tag{67}$$

We also observe the easy inequality

$$|\mathrm{trace}(BP_i)| = |\mathrm{trace}(P_iB)| = |\mathrm{trace}(P_iBP_i)| \leqslant \|B\|_{\mathrm{op}} \tag{68}$$

for any Hermitian matrix B, which follows as P_i is a rank-1 orthogonal projection.

To prove (63) and (64), we induct on k. The case $k=1$ follows from (60), (61), (67) and (68); and then, for $k>1$, the claim follows from the induction hypotheses and (48), (49), (67) and (68).

To prove (65), we also induct on k. The case $k=0$ follows from (67). For $k\geqslant 1$, the claim then follows from the induction hypothesis and (52).

To prove (66), we use the product rule to bound

$$|\nabla^k Q_i| \ll_k \sum_{m=0}^{k} |\mathrm{trace}((\nabla^m R_i)*(\nabla^{k-m} R_i))| \ll_k n \sum_{m=0}^{k} \|\nabla^m R_i\|_{\mathrm{op}} \|\nabla^{k-m} R_i\|_{\mathrm{op}},$$

and the claim follows from (65). □

This crude bound is insufficient for our applications, and we will need to supplement it with one that strengthens the spectral condition, and also assumes a "delocalization" property for the projections P_α and P_β relative to the perturbation $\dot{A}$.

LEMMA 57. (Better bound) *Let $A=A(z)$ be an $n\times n$ matrix varying real-linearly in z. Let $1\leqslant i\leqslant n$. At some fixed value of z, suppose that $\lambda_i=\lambda_i(A(z))$ is a simple eigenvalue, and that we have a partition*

$$I = P_i + \sum_{\alpha\in J} P_\alpha,$$

where J is a finite index set (not containing i) and P_α are orthogonal projections onto invariant subspaces for A (i.e. onto spans of eigenvectors not corresponding to λ_i). Suppose that, on the range of each P_α, the eigenvalues of $A-\lambda_i$ have magnitude at least r_α for some $r_\alpha>0$; equivalently, we have

$$\|R_i P_\alpha\|_{\mathrm{op}} \leqslant \frac{1}{r_\alpha}. \tag{69}$$

Suppose also that we have the delocalization bounds

$$\|P_\alpha \dot{A} P_\beta\|_F \leqslant v c_\alpha c_\beta \tag{70}$$

for all $\alpha,\beta\in J$, some $v>0$ and some $c_\alpha\geqslant 1$ satisfying the strong spectral gap condition

$$\sum_{\alpha\in J}\frac{c_\alpha^2}{r_\alpha}\leqslant L \tag{71}$$

for some $L>0$. Then at this fixed choice of z, and for all $\alpha,\beta\in J$, we have

$$|\nabla^k\lambda_i|\ll_k L^{k-1}v^k, \tag{72}$$

$$\|P_i(\nabla^k P_i)P_i\|_F\ll_k L^k v^k, \tag{73}$$

$$\|P_\alpha(\nabla^k P_i)P_i\|_F=\|P_i(\nabla^k P_i)P_\alpha\|_F\ll_k \frac{c_\alpha}{r_\alpha}L^{k-1}v^k, \tag{74}$$

$$\|P_\alpha(\nabla^k P_i)P_\beta\|_F\ll_k \frac{c_\alpha c_\beta}{r_\alpha r_\beta}L^{k-2}v^k \tag{75}$$

for all $k\geqslant 1$, and

$$\|P_i(\nabla^k R_i)P_i\|_F\ll_k L^{k+1}v^k, \tag{76}$$

$$\|P_\alpha(\nabla^k R_i)P_i\|_F=\|P_i(\nabla^k R_i)P_\alpha\|_F\ll_k \frac{c_\alpha}{r_\alpha}L^k v^k, \tag{77}$$

$$\|P_\alpha(\nabla^k R_i)P_\beta\|_F\ll_k \frac{c_\alpha c_\beta}{r_\alpha r_\beta}L^{k-1}v^k \tag{78}$$

for all $k\geqslant 0$.

We remark that we can unify the bounds (73)–(75) and (76)–(78) by allowing α and β to vary in $J\cup\{i\}$ rather than J, and adopting the conventions that $r_i:=1/L$ and $c_i:=1$.

Proof. Note that the projections P_i and the P_α are idempotent, and all annihilate each other and commute with A and R_i. We will use these facts throughout this proof without further comment.

To prove (72)–(75), we again induct on k. When $k=1$, the claim (72) follows from (60), (68) and (70), while (73)–(75) follow from (61) and (70). (For (75) we in fact obtain that the left-hand side is zero.)

Now suppose inductively that $k>1$, and that the claims have already been proven for all smaller values of k.

We first prove (72). From (48) and the linear nature of A, we have

$$|\nabla^k\lambda_i|\ll_k |\mathrm{trace}((\nabla A)*(\nabla^{k-1}P_i)P_i)|+\sum_{m=1}^{k-1}|\nabla^m\lambda_i|\,|\mathrm{trace}((\nabla^{k-m}P_i)P_i)|.$$

From (68) and the inductive hypothesis (73), we have

$$|\mathrm{trace}((\nabla^{k-m}P_i)P_i)|\ll_k L^{k-m}v^{k-m}$$

for any $1 \leqslant m \leqslant k-1$, and thus, by the inductive hypothesis (72), we see that

$$\sum_{m=1}^{k-1} |\nabla^m \lambda_i| \, |\mathrm{trace}((\nabla^{k-m} P_i) P_i)| \ll_k L^{k-1} v^k.$$

Next, by splitting $(\nabla A) * (\nabla^{k-1} P_i) P_i$ as

$$\sum_{\alpha \in J \cup \{i\}} (\nabla A) * P_\alpha (\nabla^{k-1} P_i) P_i$$

and using (68), we have

$$|\mathrm{trace}((\nabla A) * (\nabla^{k-1} P_i) P_i)| \ll_k \sum_{\alpha \in J \cup \{i\}} \|P_i (\nabla A) P_\alpha\|_F \, \|P_\alpha (\nabla^{k-1} P_i) P_i\|_F.$$

Using (70) and the inductive hypotheses (73) and (74), we thus have

$$|\mathrm{trace}((\nabla A) * (\nabla^{k-1} P_i) P_i)| \ll_k v L^{k-1} v^{k-1} + \sum_{\alpha \in J} v c_\alpha \frac{c_\alpha}{r_\alpha} L^{k-2} v^{k-1}.$$

Applying (71), we conclude (72) as desired.

We now prove (73). From (49) (or (58)) we have

$$\|P_i (\nabla^k P_i) P_i\|_F \ll_k \sum_{m=1}^{k-1} \|P_i (\nabla^m P_i) * (\nabla^{k-m} P_i) P_i\|_F.$$

We can split

$$\|P_i (\nabla^m P_i) * (\nabla^{k-m} P_i) P_i\|_F \ll_k \sum_{\alpha \in J \cup \{i\}} \|P_i (\nabla^m P_i) P_\alpha\|_F \, \|P_\alpha (\nabla^{k-m} P_i) P_i\|_F.$$

Applying the inductive hypotheses (73) and (74), we conclude that

$$\|P_i (\nabla^k P_i) P_i\|_F \ll_k \sum_{m=1}^{k-1} L^m v^m L^{k-m} v^{k-m} + \sum_{\alpha \in J} \frac{c_\alpha}{r_\alpha} L^{m-1} v^m \frac{c_\alpha}{r_\alpha} L^{k-m-1} v^{k-m};$$

bounding one of the $1/r_\alpha$ factors crudely by L and then summing using (71), we obtain (73) as desired.

Now we prove (75). From (49) (or (59)), we have

$$\begin{aligned} \|P_\alpha (\nabla^k P_i) P_\beta\|_F &\ll_k \sum_{m=1}^{k-1} \|P_\alpha (\nabla^m P_i) * (\nabla^{k-m} P_i) P_\beta\|_F \\ &\ll_k \sum_{m=1}^{k-1} \sum_{\gamma \in J \cup \{i\}} \|P_\alpha (\nabla^m P_i) P_\gamma\|_F \, \|P_\gamma (\nabla^{k-m} P_i) P_\beta\|_F. \end{aligned}$$

Using the inductive hypotheses (74) and (75), we obtain

$$\|P_\alpha(\nabla^k P_i)P_\beta\|_F \ll_k \sum_{m=1}^{k-1} \frac{1}{r_\alpha} L^{m-1} v^m \frac{1}{r_\beta} L^{k-m-1} v^{k-m} + \sum_{\gamma \in J} \frac{c_\alpha c_\gamma}{r_\alpha r_\gamma} L^{m-2} v^m \frac{c_\gamma c_\beta}{r_\gamma r_\beta} L^{k-m-2} v^{k-m}.$$

Bounding one of the $1/r_\gamma$ factors crudely by L and applying (71), we obtain (75) as desired.

Finally, we prove (74). Since $P_\alpha(\nabla^k P_i)P_i$ has the same Frobenius norm as its adjoint $P_i(\nabla^k P_i)P_\alpha$, it suffices to bound $\|P_\alpha(\nabla^k P_i)P_i\|_F$. From (49), we have

$$\|P_\alpha(\nabla^k P_i)P_i\|_F \ll_k \|R_i P_\alpha(\nabla A)*(\nabla^{k-1} P_i)P_i\|_F + \sum_{m=1}^{k-1} |\nabla^m \lambda_i| \, \|R_i P_\alpha(\nabla^{k-m} P_i)P_i\|_F,$$

and hence, by (69),

$$\|P_\alpha(\nabla^k P_i)P_i\|_F \ll_k \frac{1}{r_\alpha} \|P_\alpha(\nabla A)(\nabla^{k-1} P_i)P_i\|_F + \frac{1}{r_\alpha} \sum_{m=1}^{k-1} |\nabla^m \lambda_i| \, \|P_\alpha(\nabla^{k-m} P_i)P_i\|_F.$$

From the inductive hypotheses (72) and (74), and crudely bounding $1/r_\alpha$ by L, we have

$$\frac{1}{r_\alpha} \sum_{m=1}^{k-1} |\nabla^m \lambda_i| \, \|P_\alpha(\nabla^{k-m} P_i)P_i\|_F \ll_k \frac{c_\alpha}{r_\alpha} L^{k-1} v^k.$$

Moreover, by splitting

$$\|P_\alpha(\nabla A)*(\nabla^{k-1} P_i)P_i\|_F \leqslant \sum_{\beta \in J \cup \{i\}} \|P_\alpha(\nabla A)P_\beta\|_F \, \|P_\beta(\nabla^{k-1} P_i)P_i\|_F,$$

and using (70) and the inductive hypothesis (74), we have

$$\frac{1}{r_\alpha} \|P_\alpha(\nabla A)*(\nabla^{k-1} P_i)P_i\|_F \ll_k \frac{1}{r_\alpha} \left(v c_\alpha L^{k-1} v^{k-1} + \sum_{\beta \in J} v c_\alpha c_\beta \frac{c_\beta}{r_\beta} L^{k-2} v^{k-1} \right).$$

Applying (71), we obtain (74) as required.

Having proven (72)–(75) for all $k \geqslant 1$, we now prove (76)–(78) by induction. The claim is easily verified for $k=0$ (note that the left-hand sides of (76) and (77) in fact vanish, as does the left-hand side of (78), unless $\alpha=\beta$), so suppose that $k \geqslant 1$ and that (76)–(78) have been proven for smaller values of k.

We first prove (76). From (50) we have

$$\begin{aligned}\|P_i(\nabla^k R_i)P_i\|_F &\ll_k \sum_{m=0}^{k-1} \|P_i(\nabla^m R_i)*(\nabla^{k-m}P_i)P_i\|_F \\ &\ll_k \sum_{m=0}^{k-1}\sum_{\alpha\in J\cup\{i\}} \|P_i(\nabla^m R_i)P_\alpha\|_F\,\|P_\alpha(\nabla^{k-m}P_i)P_i\|_F.\end{aligned}$$

Applying (73), (74) and the induction hypotheses (76) and (77), we conclude that

$$\|P_i(\nabla^k R_i)P_i\|_F \ll_k \sum_{m=0}^{k-1} L^{m+1}v^m L^{k-m}v^{k-m} + \sum_{\alpha\in J}\frac{c_\alpha}{r_\alpha}L^m v^m \frac{c_\alpha}{r_\alpha}L^{k-m-1}v^{k-m};$$

crudely bounding one of the $1/r_\alpha$ factors by L and using (71), we obtain the claim.

Similarly, to prove (77), we apply (50) as before to obtain

$$\begin{aligned}\|P_\alpha(\nabla^k R_i)P_i\|_F &\ll_k \sum_{m=0}^{k-1} \|P_\alpha(\nabla^m R_i)*(\nabla^{k-m}P_i)P_i\|_F \\ &\ll_k \sum_{m=0}^{k-1}\sum_{\beta\in J\cup\{i\}} \|P_\alpha(\nabla^m R_i)P_\beta\|_F\,\|P_\beta(\nabla^{k-m}P_i)P_i\|_F;\end{aligned}$$

applying (73), (74) and the induction hypotheses (77) and (78), we conclude that

$$\|P_\alpha(\nabla^k R_i)P_i\|_F \ll_k \sum_{m=0}^{k-1} \frac{c_\alpha}{r_\alpha}L^m v^m L^{k-m}v^{k-m} + \sum_{\beta\in J}\frac{c_\alpha c_\beta}{r_\alpha r_\beta}L^{m-1}v^m \frac{c_\beta}{r_\beta}L^{k-m-1}v^{k-m}.$$

Again, bounding one of the $1/r_\beta$ factors by L and using (71), we obtain the claim.

Finally, we prove (78). From (51), we have

$$\begin{aligned}\|P_\alpha(\nabla^k R_i)P_\beta\|_F &\ll_k \|P_\alpha(\nabla^k P_i)R_iP_\beta\|_F + \|P_\alpha(\nabla^{k-1}R_i)*(\nabla A)R_iP_\beta\|_F \\ &\quad + \sum_{m=0}^{k-1}|\nabla^{k-m}\lambda_i|\,\|P_\alpha(\nabla^m R_i)R_iP_\beta\|_F.\end{aligned}$$

As $R_iP_\beta = P_\beta R_i P_\beta$ has a norm of at most $1/r_\beta$, we conclude that

$$\begin{aligned}\|P_\alpha(\nabla^k R_i)P_\beta\|_F &\ll_k \frac{1}{r_\beta}\|P_\alpha(\nabla^k P_i)P_\beta\|_F + \frac{1}{r_\beta}\|P_\alpha(\nabla^{k-1}R_i)*(\nabla A)P_\beta\|_F \\ &\quad + \frac{1}{r_\beta}\sum_{m=0}^{k-1}|\nabla^{k-m}\lambda_i|\,\|P_\alpha(\nabla^m R_i)P_\beta\|_F.\end{aligned}$$

From (72), the crude bound $1/r_\beta \leqslant L$ and the induction hypothesis (78), we have

$$\frac{1}{r_\beta}\sum_{m=0}^{k-1} |\nabla^{k-m}\lambda_i|\, \|P_\alpha(\nabla^m R_i)P_\beta\|_F \ll_k \frac{c_\alpha c_\beta}{r_\alpha r_\beta} L^{k-1} v^k.$$

From (75) and $1/r_\beta \leqslant L$, we similarly have

$$\frac{1}{r_\beta}\|P_\alpha(\nabla^k P_i)P_\beta\|_F \ll_k \frac{c_\alpha c_\beta}{r_\alpha r_\beta} L^{k-1} v^k.$$

Finally, splitting

$$\|P_\alpha(\nabla^{k-1}R_i)*(\nabla A)P_\beta\|_F \leqslant \sum_{\gamma\in J\cup\{i\}} \|P_\alpha(\nabla^{k-1}R_i)P_\gamma\|_F\, \|P_\gamma(\nabla A)P_\beta\|_F$$

and using the induction hypotheses (77) and (78), and (70), we obtain

$$\|P_\alpha(\nabla^{k-1}R_i)*(\nabla A)P_\beta\|_F \ll_k \frac{c_\alpha}{r_\alpha} L^{k-1} v^{k-1} v c_\beta + \sum_{\gamma\in J} \frac{c_\alpha c_\gamma}{r_\alpha r_\gamma} L^{k-2} v^{k-1} v c_\gamma c_\beta,$$

and thus, by (71),

$$\|P_\alpha(\nabla^{k-1}R_i)*(\nabla A)P_\beta\|_F \ll_k \frac{c_\alpha c_\beta}{r_\alpha} L^{k-1} v^k.$$

Putting all this together, we obtain (78) as claimed. □

We extract a special case of the above lemma, in which the perturbation only affects a single entry (and its transpose).

Corollary 58. (Better bound, special case) *Let $A=A(z)$ be an $n\times n$ matrix depending on a complex parameter z of the form*

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

for some vectors e_p and e_q. Fix an index $1 \leqslant i \leqslant n$. At some fixed value of z, suppose that $\lambda_i = \lambda_i(A(z))$ is a simple eigenvalue, and that we have a partition

$$I = P_i + \sum_{\alpha\in J} P_\alpha,$$

where J is a finite index set and P_α are orthogonal projections onto invariant subspaces for A (i.e. onto spans of eigenvectors not corresponding to λ_i). Suppose that on the range of each P_α, the eigenvalues of $A-\lambda_i$ have magnitude at least r_α for some $r_\alpha>0$. Suppose also that we have the incompressibility bounds

$$\|P_\alpha e_p\|, \|P_\alpha e_q\| \leqslant w d_\alpha^{1/2}$$

for all $\alpha \in J \cup \{i\}$ *and some* $w>0$ *and* $d_\alpha \geqslant 1$. *Then, at this value of* z, *and for all* $k\geqslant 1$, *we have*

$$|\nabla^k \lambda_i| \ll_k \bigg(\sum_{\alpha\in J} \frac{d_\alpha}{r_\alpha}\bigg)^{k-1} w^{2k} \tag{79}$$

and

$$|\nabla^k Q_i| \ll_k \bigg(\sum_{\alpha\in J} \frac{d_\alpha}{r_\alpha}\bigg)^{k+2} w^{2k} \tag{80}$$

for all $k\geqslant 0$.

Proof. A short computation shows that the hypotheses of Lemma 57 are obeyed with v replaced by $O(w^2)$, c_α set equal to $d_\alpha^{1/2}$, and L equal to $O\big(\sum_{\alpha\in J} d_\alpha/r_\alpha\big)$. From (72), we then conclude (79). As for (80), we see from the product rule that

$$|\nabla^k Q_i| \ll_k \sum_{m=0}^{k} |\mathrm{trace}((\nabla^m R_i)(\nabla^{k-m} R_i^*))|$$

which we can split further, using the Cauchy–Schwarz inequality, as

$$|\nabla^k Q_i| \ll_k \sum_{m=0}^{k} \sum_{\alpha,\beta\in J\cup\{i\}} \|P_\alpha(\nabla^m R_i)P_\beta\|_F \|P_\alpha(\nabla^{k-m} R_i)P_\beta\|_F.$$

Applying (76), (77) and (78), we conclude that

$$|\nabla^k Q_i| \ll_k L^{k+2} v^k + \sum_{\alpha\in J} \frac{c_\alpha^2}{r_\alpha^2} L^k v^k + \sum_{\alpha,\beta\in J} \frac{c_\alpha^2 c_\beta^2}{r_\alpha^2 r_\beta^2} L^{k-2} v^k.$$

Bounding one of the factors of $1/r_\alpha$ in the first sum by L, and $1/r_\alpha r_\beta$ in the second sum by L^2, and using (71) and the choices for v and L, we obtain the claim. $\square$

4.3. Conclusion of the argument

We can now prove Proposition 46. Fix $k\geqslant 1$, $r=2,3,4$ and $\varepsilon_1>0$, and suppose that C_1 is sufficiently large. We assume that $A(0)$, e_p, e_q, $i_1, ..., i_k$, G, F, ζ and ζ' are as in the proposition.

We may of course assume that $F(z_0)\neq 0$ for at least one z_0 with $|z_0|\leqslant n^{1/2+\varepsilon_1}$, since the claim is vacuous otherwise.

Suppose we can show that

$$\nabla^m F(z) = O(n^{-m+O(\varepsilon_1)}) \tag{81}$$

for all $|z|\leqslant n^{1/2+\varepsilon_1}$ and $0\leqslant m\leqslant 5$. Then, by Taylor expansion, one has

$$F(\zeta)=P(\zeta,\bar{\zeta})+O(n^{-(r+1)/2+O(\varepsilon_1)}),$$

where P is a polynomial of degree at most r whose coefficients are of size at most $n^{O(\varepsilon_1)}$. Taking expectations for both $F(\zeta)$ and $F(\zeta')$, we obtain the claim (34) when ζ and ζ' match to order r. A similar argument gives the improved version of (34) at the end of Proposition 46 if one can improve the right-hand side of (81) to $O(n^{-m-C'm\varepsilon_1})$ for some sufficiently large absolute constant C'.

It remains to show (81). By up to five applications of the chain rule, the above claims follow from the following lemma.

Lemma 59. *Suppose that $F(z_0)\neq 0$ for at least one z_0 with $|z_0|\leqslant n^{1/2+\varepsilon_1}$. Then*

$$|\nabla^k\lambda_{i_j}(z)|\ll_k n^{5\varepsilon_1 k}n^{-k}\quad \textit{and}\quad |\nabla^k Q_{i_j}(z)|\ll_k n^{5\varepsilon_1(k+2)}n^{-k}$$

for all z with $|z|\leqslant n^{1/2+\varepsilon_1}$ and all $1\leqslant j\leqslant k\leqslant 10$.

Proof. Fix j. Since $F(z_0)\neq 0$, we have

$$Q_{i_j}(A(z_0))\leqslant n^{\varepsilon_1}. \tag{82}$$

For a technical reason having to do with a subsequent iteration argument, we will replace (82) with the slightly weaker bound

$$Q_{i_j}(A(z_0))\leqslant 2n^{\varepsilon_1}. \tag{83}$$

By the definition of Q_i, we have, as a consequence, that

$$|\lambda_{i'}(A(z_0))-\lambda_{i_j}(A(z_0))|\gg n^{-\varepsilon_1/2} \tag{84}$$

for all $i'\neq i_j$.

By the Weyl inequalities (23), we thus have

$$|\lambda_{i'}(A(z))-\lambda_{i_j}(A(z))|\gg n^{-\varepsilon_1/2}$$

whenever $|z-z_0|\ll n^{-1-2\varepsilon_1}$. From Lemma 56, we conclude that

$$|\nabla^m\lambda_{i_j}(A(z))|\ll_m n^{\varepsilon_1(m+1)/2} \tag{85}$$

and

$$|\nabla^m Q_{i_j}(A(z))|\ll_m n^{\varepsilon_1(m+2)/2}n \tag{86}$$

for all $m\geqslant 1$, whenever $|z-z_0|\ll n^{-1-2\varepsilon_1}$. In particular, from (83) and the fundamental theorem of calculus, we have

$$Q_{i_j}(A(z))\ll n^{\varepsilon_1} \tag{87}$$

for all such z.

Note that, by setting C_1 sufficiently large, we can find z such that $|z-z_0|\ll n^{-1-2\varepsilon_1}$ and $|z|\leqslant n^{1/2+\varepsilon_1}$, and such that the real and imaginary parts of z are integer multiples of n^{-C_1}. Then (32) and (33) hold for this value of z. Applying Corollary 58, we conclude that

$$|\nabla^k\lambda_{i_j}(z)|\ll_k n^{2\varepsilon_1 k}n^{-k}\biggl(\sum_{\alpha=0}^{\log n}\frac{2^\alpha}{r_\alpha}\biggr)^{k-1}$$

and

$$|\nabla^k Q_{i_j}(z)|\ll_k n^{2\varepsilon_1 k}n^{-k}\biggl(\sum_{\alpha=0}^{\log n}\frac{2^\alpha}{r_\alpha}\biggr)^{k+2}$$

for all $k\geqslant 1$, where r_α is the minimal value of $|\lambda_i-\lambda_{i_j}|$ for $|i-i_j|\geqslant 2^\alpha$. Note that

$$\sum_{\alpha=0}^{\log n}\frac{2^\alpha}{r_\alpha}\ll\sum_{i\neq i_j}\frac{1}{|\lambda_i-\lambda_{i_j}|}.$$

At the same time, from (87) we have

$$\sum_{i\neq i_j}\frac{1}{|\lambda_i-\lambda_{i_j}|^2}\ll n^{\varepsilon_1}.$$

From the Cauchy–Schwarz inequality, this implies that

$$\sum_{\substack{i\neq i_j\\ |i-i_j|\leqslant n^{\varepsilon_1}}}\frac{1}{|\lambda_i-\lambda_{i_j}|}\ll n^{\varepsilon_1},$$

while from (31) we have (with room to spare)

$$\sum_{\substack{i\neq i_j\\ |i-i_j|\geqslant n^{\varepsilon_1}}}\frac{1}{|\lambda_i-\lambda_{i_j}|}\ll n^{3\varepsilon_1},$$

and thus

$$\sum_{\alpha=0}^{\log n}\frac{2^\alpha}{r_\alpha}\ll n^{3\varepsilon_1}.$$

Hence

$$|\nabla^k\lambda_{i_j}(z)|\ll_k n^{5\varepsilon_1 k}n^{-k}\quad\text{and}\quad|\nabla^k Q_{i_j}(z)|\ll_k n^{5\varepsilon_1(k+2)}n^{-k}$$

for all $k\geqslant 1$. Combining this with (85) and (86), we conclude that

$$|\nabla^k \lambda_{i_j}(z)| \ll_k n^{5\varepsilon_1 k} n^{-k}$$

and

$$|\nabla^k Q_{i_j}(z)| \ll_k n^{5\varepsilon_1 (k+2)} n^{-k} \tag{88}$$

for all z with $|z-z_0|\ll n^{-1+\varepsilon_1}$ and all $0\leqslant k\leqslant 10$.

This establishes the lemma in a ball $B(z_0, n^{-1-2\varepsilon_1})$ of radius $n^{-1-2\varepsilon_1}$ centered at z_0. To extend the result to the remainder of the region $\{z:|z|\leqslant n^{1/2+\varepsilon_1}\}$, we observe from (88) that Q_{i_j} varies by at most $O(n^{-1.9})$ (say) on this ball (instead of 1.9 we can write any constant less than 2, given that ε_1 is sufficiently small). Because of the gap between (82) and (83), we now see that (83) continues to hold for all other points z_1 in $B(z_0, n^{-1-2\varepsilon_1})$ with $|z_1|\leqslant n^{1/2+\varepsilon_1}$. Repeating the above arguments with z_0 replaced by z_1, and continuing this process, we can eventually cover the entire ball $\{z:|z|\leqslant n^{1/2+\varepsilon_1}\}$ by these estimates. The key point here is that at every point of the process (83) holds, since the length of the process is only $n^{3/2+3\varepsilon_1}$, while in each step the value of Q_{i_j} changes by at most $O(n^{-1.99})$. □

The proof of Proposition 46 is now complete.

5. Good configurations occur frequently

The purpose of this section is to prove Proposition 48. The arguments here are largely based on those in [20], [21] and [22].

5.1. Reduction to a concentration bound for the empirical spectral distribution

We will first reduce matters to the following concentration estimate for the empirical spectral distribution (ESD).

THEOREM 60. (Concentration for ESD) *For any $\varepsilon,\delta>0$, any random Hermitian matrix $M_n=(\zeta_{ij})_{1\leqslant i,j\leqslant n}$ whose upper-triangular entries are independent with mean zero and variance* 1, *such that $|\zeta_{ij}|\leqslant K$ almost surely for all i and j and some $1\leqslant K\leqslant n^{1/2-\varepsilon}$, and for any interval I in $[-2+\varepsilon, 2-\varepsilon]$ of width $|I|\geqslant K^2(\log n)^{20}/n$, the number of eigenvalues N_I of $W_n:=M_n/\sqrt{n}$ in I obeys the concentration estimate*

$$\left| N_I - n\int_I \varrho_{\mathrm{sc}}(x)\,dx \right| \leqslant \delta n|I|$$

with overwhelming probability.

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

Remark 61. Similar results were established in [20], [21] and [22] assuming stronger regularity hypotheses on the ζ_{ij}. The proof of this result follows their approach, but also uses Lemma 43 and a few other ideas which make the current more general setting possible. In our applications we will take $K=(\log n)^{O(1)}$, though Theorem 60 also has nontrivial content for larger values of K. The loss of $K^2(\log n)^{20}$ can certainly be improved, though for our applications any bound which is polylogarithmic for $K=(\log n)^{O(1)}$ will suffice.

Let us assume Theorem 60 for the moment. We can then conclude a useful bound on eigenvectors (which will also be applied to prove Theorem 19).

PROPOSITION 62. (Delocalization of eigenvectors) *Let* ε, M_n, W_n, ζ_{ij} *and* K *be as in Theorem* 60. *Then for any* $1\leqslant i\leqslant n$ *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* $O_\varepsilon(K^2(\log n)^{20}/n^{1/2})$.

Proof. By symmetry and the union bound, it suffices to establish this for the first coordinate of $u_i(W_n)$. By Lemma 41, it suffices to establish a lower bound

$$\sum_{j=1}^{n-1}\frac{|u_j(W_{n-1})^*X/\sqrt{n}\,|^2}{(\lambda_j(W_{n-1})-\lambda_i(W_n))^2}\gg_\varepsilon\frac{n}{K^2(\log n)^{20}}$$

with overwhelming probability, where W_{n-1} is the bottom right $(n-1)\times(n-1)$ minor of W_n and $X\in\mathbb{C}^{n-1}$ has entries ζ_{i1} for $i=2,...,n$. But by Theorem 60, we can (with overwhelming probability) find a set $J\subset\{1,...,n-1\}$ with $|J|\gg_\varepsilon K^2(\log n)^{20}$ such that $|\lambda_j(W_{n-1})-\lambda_i(W_n)|\ll_\varepsilon K^2(\log n)^{20}/n$ for all $n\in J$. Thus it will suffice to show that

$$\sum_{j\in J}|u_j(W_{n-1})^*X|^2\gg_\varepsilon|J|$$

with overwhelming probability. The left-hand side can be written as $\|\pi_H X\|^2$, where H is the span of all the eigenvectors associated with J. The claim now follows from Lemma 43. $\square$

We also have the following minor variant.

COROLLARY 63. *The conclusions of Theorem* 60 *and Proposition* 62 *continue to hold if one replaces a single diagonal entry* ζ_{pp} *of* M_n *by a deterministic real number* $x=O(K)$, *or if one replaces a single off-diagonal entry* ζ_{pq} *of* M_n *by a deterministic complex number* $z=O(K)$ (*and also replaces* ζ_{qp} *by* $\bar{z}$).

Proof. After the indicated replacement, the new matrix M_n' differs from the original matrix by a Hermitian matrix of rank at most 2. The modification of Theorem 60 then follows from Theorem 60 and Lemma 39. The modification of Proposition 62 then follows by repeating the proof. (One of the coefficients of X might now be deterministic rather than random, but it is easy to see that this does not significantly impact Lemma 43.) $\square$

Now we can prove Proposition 48. Let ε, ε_1, C, C_1, k, $i_1, ..., i_k$, p, q and $A(0)$ be as in that proposition. By the union bound, we may fix $1\leqslant j\leqslant k$, and also fix the $|z|\leqslant n^{1/2+\varepsilon_1}$ whose real and imaginary parts are multiples of n^{-C_1}. By the union bound again and Corollary 63 (with $K=(\log n)^C$), the eigenvalue separation condition (31) holds with overwhelming probability for every $1\leqslant i\leqslant n$ with $|i-j|\geqslant n^{\varepsilon_1}$, as does (32) (note that $\|P_{i_j}(A(z))e_p\|$ is the magnitude of the pth coordinate of a unit eigenvector $u_{i_j}(A(z))$ of $A(z)$). A similar argument using Pythagoras' theorem gives (33) with overwhelming probability, unless the eigenvalues $\lambda_i(A(z))$ contributing to (33) are not contained in the bulk region $[(-2+\varepsilon')n, (2-\varepsilon')n]$ for some $\varepsilon'>0$ independent of n. However, it is known (see [25]; one can also deduce this fact from Theorem 60) that $\lambda_i(A(z))$ will fall in this bulk region with overwhelming probability whenever $\frac{1}{2}\varepsilon n\leqslant i\leqslant\left(1-\frac{1}{2}\varepsilon\right)n$, if ε' is small enough depending on ε. Thus, with overwhelming probability, a contribution outside the bulk region can only occur if $2^\alpha\gg_\varepsilon n$, in which case the claim follows by estimating $\|P_{i_j,\alpha}(A(z))e_p\|$ crudely by $\|e_p\|=1$, and similarly for $\|P_{i_j,\alpha}(A(z))e_q\|$. This concludes the proof of Proposition 48 assuming Theorem 60.

5.2. Spectral concentration

It remains to prove Theorem 60.

Following [20], [21] and [22], we consider the *Stieltjes transform*

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

of W_n, together with its semi-circular counterpart

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

(which will be computed explicitly in (107)). We will primarily be interested in the imaginary part

$$\operatorname{Im}(s_n(x+\eta\sqrt{-1}\,)) = \frac{1}{n}\sum_{i=1}^n \frac{\eta}{\eta^2+(\lambda_i(W_n)-x)^2} > 0 \tag{89}$$

of the Stieltjes transform in the upper half-plane $\eta>0$.

It is well known that the convergence of the empirical spectral distribution of W_n to $\varrho_{\mathrm{sc}}(x)$ is closely tied to the convergence of s_n to s (see [2], for example). In particular, we have the following precise connection (cf. [21, Corollary 4.2]), whose proof is deferred to Appendix C.

LEMMA 64. (Control of Stieltjes transform implies control on ESD) *Let*

$$\frac{1}{10} \geqslant \eta \geqslant \frac{1}{n},$$

and $L, \varepsilon, \delta > 0$. *Suppose that one has the bound*

$$|s_n(z) - s(z)| \leqslant \delta, \tag{90}$$

with (uniformly) overwhelming probability for all z *with* $|\mathrm{Re}(z)| \leqslant L$ *and* $\mathrm{Im}(z) \geqslant \eta$. *Then for any interval* I *in* $[-L+\varepsilon, L-\varepsilon]$ *with*

$$|I| \geqslant \max\left\{2\eta, \frac{\eta}{\delta}\log\frac{1}{\delta}\right\},$$

one has

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

with overwhelming probability.

In view of this lemma, it suffices to show that for each complex number z with $\mathrm{Re}(z) \leqslant 2 - \frac{1}{2}\varepsilon$ and $\mathrm{Im}(z) \geqslant \eta := K^2(\log n)^{19}/n$, one has

$$|s_n(z) - s(z)| \leqslant o(1)$$

with (uniformly) overwhelming probability.

Fix z as above. From (107), $s(z)$ is the unique solution to the equation

$$s(z) + \frac{1}{s(z)+z} = 0, \tag{91}$$

with $\mathrm{Im}(s(z)) > 0$. The strategy is then to obtain a similar equation for $s_n(z)$ (note that one automatically has $\mathrm{Im}(s_n(z)) > 0$).

By Lemma 42, we may write

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\zeta_{kk}/\sqrt{n} - z - Y_k}, \tag{92}$$

where

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

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

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

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

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 (24), we have

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

and thus

$$\mathbf{E}(Y_k \mid W_{n,k}) = s_n(z)+O\left(\frac{1}{K^2(\log n)^{19}}\right). \tag{93}$$

We now claim that a similar estimate holds for Y_k itself.

Proposition 65. (Concentration of Y_k) *For each $1\leqslant k\leqslant n$, one has*

$$Y_k = s_n(z)+O\left(\frac{1}{\log n}\right)$$

with overwhelming probability.

Assume this proposition for the moment. By hypothesis, $\zeta_{kk}/\sqrt{n}\leqslant K/\sqrt{n}\leqslant n^{-\varepsilon}$ almost surely. Inserting these bounds into (92), we see that

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

with overwhelming probability (compare with (91)). This implies that with overwhelming probability either $s_n(z)=s(z)+o(1)$ or that $s_n(z)=-z+o(1)$. On the other hand, as $\operatorname{Im}(s_n(z))$ is necessarily positive, the second possibility can only occur when $\operatorname{Im}(z)=o(1)$. A continuity argument (as in [20]) then shows that the second possibility cannot occur at all (note that $s(z)$ stays a fixed distance away from $-z$ for z in a compact set) and the claim follows.

5.3. A preliminary concentration bound

It remains to prove Proposition 65. We begin with a preliminary bound (cf. [22, Theorem 5.1]).

PROPOSITION 66. *For all* $I \subset \mathbb{R}$ *with* $|I| \geqslant K^2(\log n)^2/n$, *one has*

$$N_I \ll n|I|$$

with overwhelming probability.

The proof, which follows the arguments from [22], but using Lemma 43 to simplify things somewhat, is presented in Appendix C.

Now we prove Proposition 65. Fix k, and write $z = x + \eta\sqrt{-1}$. From (93), it suffices to show that

$$Y_k - \mathbf{E}(Y_k \mid W_{n,k}) = O\left(\frac{1}{\log n}\right)$$

with overwhelming probability. Decomposing Y_k as in (114), it thus suffices to show that

$$\sum_{j=1}^{n-1} \frac{R_j}{\lambda_j(W_{n,k}) - (x + \eta\sqrt{-1}\,)} = O\left(\frac{1}{\log n}\right) \tag{94}$$

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

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

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

where H is the space spanned by the $u_j(W_{n,k})^*$ for $i_- \leqslant j \leqslant i_+$. From Lemma 43 and the union bound, we conclude that with overwhelming probability

$$\left|\sum_{j=i_-}^{i_+} R_j\right| \ll \frac{\sqrt{i_+ - i_-}\, K \log n + K^2(\log n)^2}{n}. \tag{95}$$

By the triangle inequality, this implies that

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

and hence, by a further application of the triangle inequality,

$$\sum_{j=i_-}^{i_+} |R_j| \ll \frac{(i_+ - i_-) + K^2(\log n)^2}{n} \tag{96}$$

with overwhelming probability.

Since $\eta \geqslant K^2(\log n)^{19}/n$, the bound (95) (together with Proposition 66) already lets one dispose of the contribution to (94) where $|\lambda_j(W_{n,k})-x| \leqslant K^2(\log n)^{10}/n$. For the remaining contributions, we subdivide into $O((\log n)^3)$ intervals $\{j : i_- \leqslant j \leqslant i_+\}$ such that in each interval

$$a \leqslant |\lambda_j(W_{n,k})-x| \leqslant \left(1+\frac{1}{(\log n)^2}\right)a$$

for some $a \geqslant K^2(\log n)^{10}/n$ (the value of a varies from interval to interval). For each such interval, the function

$$\frac{1}{\lambda_j(W_{n,k})-(x+\eta\sqrt{-1}\,)}$$

has magnitude $O(1/a)$ and fluctuates by at most $O(1/a(\log n)^2)$ as j ranges over the interval. From (95) and (96) we conclude that

$$\left|\sum_{j=i_-}^{i_+} \frac{R_j}{\lambda_j(W_{n,k})-(x+\eta\sqrt{-1}\,)}\right| \ll \frac{\sqrt{i_+-i_-}\,K\log n+K^2(\log n)^2}{an}+\frac{i_+-i_-}{an(\log n)^2}$$

with overwhelming probability. By Proposition 66, we have that $i_+-i_- \ll an$ with overwhelming probability. Thus,

$$\left|\sum_{j=i_-}^{i_+} \frac{R_j}{\lambda_j(W_{n,k})-(x+\eta\sqrt{-1}\,)}\right| \ll \frac{K\log n}{\sqrt{an}}+\frac{1}{(\log n)^4}$$

with overwhelming probability. Summing over the values of a (taking into account the lower bound for a) we obtain (94) as desired.

6. Propagation of narrow spectral gaps

We now prove Lemma 51. Fix i_0, l and n. Assume for contradiction that all of the conclusions fail. We will always assume that n_0 (and hence n) is sufficiently large.

By (37), we can find $1 \leqslant i_- \leqslant i_0-l < i_0 \leqslant i_+ \leqslant n+1$ such that

$$\lambda_{i_+}(A_{n+1})-\lambda_{i_-}(A_{n+1}) = g_{i_0,l,n+1}\min\{i_+-i_-,(\log n_0)^{C_1}\}^{(\log n_0)^{0.9}}.$$

If $i_+-i_- \geqslant (\log n)^{C_1/2}$, then conclusion (i) holds (for n large enough), so we may assume that

$$i_+-i_- < (\log n)^{C_1/2}. \tag{97}$$

We set

$$L := \lambda_{i_+}(A_{n+1})-\lambda_{i_-}(A_{n+1}) = g_{i_0,l,n+1}(i_+-i_-)^{(\log N)^{0.9}}. \tag{98}$$

In particular (by (39) and (97)) we have

$$L \leqslant \delta \exp((\log n)^{0.91}). \tag{99}$$

We now study the eigenvalue equation (25) for $i=i_-$, which we rearrange as

$$\sum_{j=i_-}^{n} \frac{|u_j(A_n)^*X_n|^2}{\lambda_j(A_n)-\lambda_{i_-}(A_{n+1})} = \sum_{j=1}^{i_-} \frac{|u_j(A_n)^*X_n|^2}{\lambda_{i_-}(A_{n+1})-\lambda_j(A_n)} + a_{n+1,n+1} - \lambda_{i_-}(A_{n+1}).$$

Observe that

$$\sum_{j=i_-}^{n} \frac{|u_j(A_n)^*X_n|^2}{\lambda_j(A_n)-\lambda_{i_-}(A_{n+1})} \geqslant \frac{1}{L} \sum_{i_- \leqslant j < i_+} |u_j(A_n)^*X_n|^2.$$

Since conclusion (ii) fails, we have

$$\sum_{j=i_-}^{n} \frac{|u_j(A_n)^*X_n|^2}{\lambda_j(A_n)-\lambda_{i_-}(A_{n+1})} \geqslant \frac{n(i_+-i_-)}{2^{m/2}L(\log n)^{0.01}}.$$

On the other hand, since conclusions (iii) and (iv) fail, we have

$$|a_{n+1,n+1}-\lambda_{i_-}(A_{n+1})| \leqslant \frac{n\exp(-(\log n)^{0.95})}{\delta^{1/2}} \leqslant \frac{n(i_+-i_-)}{2^{m/2+1}L(\log n)^{0.01}},$$

due to the bounds $i_+-i_- \geqslant 1$, (41) and (99). By the triangle inequality, we thus have

$$\sum_{1\leqslant j<i_-} \frac{|u_j(A_n)^*X_n|^2}{\lambda_{i_-}(A_{n+1})-\lambda_j(A_n)} \geqslant \frac{n(i_+-i_-)}{2^{m/2+1}L(\log n)^{0.01}}.$$

Note that all the summands on the left-hand side are non-negative. By a dyadic partition and the pigeonhole principle (using the convergence of the series $1/l^2$), we can thus find $k \geqslant 1$ such that

$$\sum_{\substack{1\leqslant j<i_- \\ 2^{k-1}\leqslant i_- - j < 2^k}} \frac{|u_j(A_n)^*X_n|^2}{\lambda_{i_-}(A_{n+1})-\lambda_j(A_n)} \gg \frac{n(i_+-i_-)}{2^{m/2}Lk^2(\log n)^{0.01}}. \tag{100}$$

In particular, $2^{k-1}<i_-$.

Let us first suppose that $2^{k-1} \geqslant (\log n)^{C_1/2}$. Then, by the failure of conclusion (i), we have

$$\lambda_{i_-}(A_{n+1})-\lambda_j(A_n) > \delta^{1/4}\exp((\log n)^{0.95})2^{k-1}$$

for all j in the summation in (100), and thus (by (41), (99) and the trivial bounds $i_+-i_- \geqslant 1$ and $k=O(\log n)$)

$$\sum_{\substack{1\leqslant j<i_- \\ 2^{k-1}\leqslant i_- - j < 2^k}} |u_j(A_n)^*X_n|^2 \gg \frac{n(i_+-i_-)}{2^{m/2}Lk^2\log^{0.01}} \delta^{1/4}\exp((\log n)^{0.95})2^{k-1} \gg \frac{n2^k}{\delta^{1/2}}. \tag{101}$$

On the other hand, from the failure of conclusion (v), we have

$$|u_j(A_n)^* X_n|^2 < \frac{n \exp(-(\log n)^{0.96})}{\delta^{1/2}} \tag{102}$$

for $\frac{1}{10}\varepsilon n \leqslant j \leqslant \left(1-\frac{1}{10}\varepsilon\right)n$. This already contradicts (101) when the range of summation in (101) is contained in the bulk region $\frac{1}{10}\varepsilon n \leqslant j \leqslant \left(1-\frac{1}{10}\varepsilon\right)n$. The only remaining case is when (101) approaches the edge, which only occurs when $2^k \gg \varepsilon n$. But in this case we note from Pythagoras' theorem and the failure of conclusion (vi) that

$$\sum_{j=1}^{n} |u_j(A_n)^* X_n|^2 < \frac{n^2 \exp(-(\log n)^{0.96})}{\delta^{1/2}},$$

leading again to a contradiction with (101). We may therefore assume that

$$2^{k-1} < (\log n)^{C_1/2},$$

and thus $k = O(C_1 \log\log n)$.

By the failure of conclusion (vii), we now have $|u_j(A_n)^* X_n|^2 \ll 2^{m/2} n (\log n)^{0.8}$ for all j in the summation in (100); we conclude that

$$\sum_{\substack{1 \leqslant j < i_- \\ 2^{k-1} \leqslant i_- - j < 2^k}} \frac{1}{\lambda_{i_-}(A_{n+1}) - \lambda_j(A_n)} \gg \frac{i_+ - i_-}{2^m L (\log n)^{0.82}}.$$

If we set $i_{--} := i_- - 2^{k-1}$, we conclude that $0 < i_- - i_{--} < (\log n)^{C_1/2}$ and

$$\lambda_{i_-}(A_{n+1}) - \lambda_{i_{--}}(A_n) \leqslant 2^m \frac{i_- - i_{--}}{i_+ - i_-} L (\log n)^{0.83}.$$

An analogous argument, starting with $i = i_+$ in (25) instead of $i = i_-$ and reflecting all the indices, allows us to find i_{++} with $0 \leqslant i_{++} - i_+ < (\log n)^{C_1/2}$ such that

$$\lambda_{i_{++}}(A_n) - \lambda_{i_+}(A_{n+1}) \leqslant 2^m \frac{i_- - i_{++}}{i_+ - i_-} L (\log n)^{0.83}.$$

Summing, we have

$$\lambda_{i_{++}}(A_n) - \lambda_{i_{--}}(A_n) \leqslant L(1 + 2^m \alpha (\log n)^{0.84}), \tag{103}$$

where

$$\alpha := \frac{i_{++} - i_{--}}{i_+ - i_-} - 1.$$

Note that $(1+\alpha)(i_+-i_-)=i_{++}-i_{--}\leqslant(\log N)^{C_1}$, and so, by (37) and (40),

$$\frac{\lambda_{i_{++}}(A_n)-\lambda_{i_{--}}(A_n)}{(1+\alpha)^{(\log N)^{0.9}}(i_+-i_-)^{(\log N)^{0.9}}}\geqslant 2^m g_{i_0,l,n+1}.$$

Combining this with (103) and (98), we conclude that

$$1+2^m\alpha(\log n)^{0.84}\geqslant 2^m(1+\alpha)^{(\log N)^{0.9}},$$

and hence

$$1+\alpha(\log n)^{0.84}\geqslant(1+\alpha)^{(\log N)^{0.9}}.$$

But this contradicts the elementary estimate $(1+\alpha)^x\geqslant 1+x\alpha$ for $\alpha>0$ and $x\geqslant 1$, and Lemma 51 follows.

7. Bad events are rare

We now prove Proposition 53. Let the notation and assumptions be as in that proposition.

We first prove (a). The truncation assumption (27) ensures that the events (iii), (v) and (vi) from Proposition 51 are empty for n large enough. The event (i) fails with overwhelming probability, due to Theorem 60. The event (iv) fails with overwhelming probability because of the well-known fact that the operator norm of A_n is $O(n)$ with overwhelming probability (see e.g. [1]; there are many proofs, for instance one can start by observing that $\|A_n\|_{\mathrm{op}}\leqslant 2\sup_x\|A_nx\|$, where x ranges over a $\frac{1}{2}$-net of the unit ball, and use the union bound followed by a standard concentration of measure result, such as the Chernoff inequality). This concludes the proof of (a).

Now we prove (b) and (c) jointly. By (27) and Proposition 62, we can find C' such that all the coefficients of the eigenvectors $u_j(A_n)$ for $\frac{1}{2}\varepsilon n\leqslant j\leqslant\left(1-\frac{1}{2}\varepsilon\right)n$ are of magnitude at most $n^{-1/2}(\log n)^{C'}$ with overwhelming probability.

Let us first consider (vii), in which we will be able to obtain the better upper bound of $2^{-\varkappa m}2^{-2C_1n}$ for the conditional probability of occurrence (and thus establishing (b) and (c) simultaneously for (vii)). If $2^m\geqslant(\log n)^{C_3}$ for some sufficiently large C_3, then the desired bound comes from (27) and Lemma 43. (In fact, the Chernoff bound would suffice as well, and the event fails with overwhelming probability.) Now suppose instead that $2^m\leqslant(\log n)^{O(1)}$. We wish to show that

$$\mathbf{P}(|S_i|\geqslant 2^{m/2}(\log n)^{0.8})\leqslant 2^{-\varkappa m}(\log n)^{-2C_1},\tag{104}$$

where $S_i\in\mathbb{C}$ is the random walk

$$S_i:=\zeta_{1,n+1}w_{i,1}+\ldots+\zeta_{n,n+1}w_{i,n}\tag{105}$$

and $w_{i,1}, ..., w_{i,n}$ are the coefficients of $u_i(A_n)$, which by hypothesis have magnitude $O(n^{-1/2}(\log n)^{C'})$ and square-sum to 1.

Observe that S_i has mean zero and variance 1. Applying Theorem 44 and (27), we conclude that

$$\mathbf{P}(|S_i|\geqslant t)\ll \exp(-ct^2)+n^{-1/2}(\log n)^{O(1)}$$

for any $t\geqslant 1$ and some absolute constant $c>0$, which easily yields (104) in the range $2^m\leqslant(\log n)^{O(1)}$.

The consideration of (ii) is similar. Write the left-hand side of (42) as $\|\pi_H(X_n)\|$, where H is the span of the $u_j(A_n)$ for $i_-\leqslant j<i_+$. Applying (27) and Lemma 43, we obtain the claim when $i_+-i_-\geqslant(\log n)^{C_3}$ for sufficiently large c_3 (in fact (ii) now fails with overwhelming probability), so we may assume instead that $i_+-i_-\leqslant(\log n)^{O(1)}$. In this case, the event (ii) can now be expressed as

$$|\vec{S}|\leqslant \frac{(i_+-i_-)^{1/2}}{2^{m/4}(\log n)^{0.005}}, \tag{106}$$

where $\vec{S}\in\mathbb{C}^{i_+-i_-}$ is the random vector with components S_j defined in (105).

From the orthonormality of the $u_i(A_n)$, we see that $\vec{S}$ has mean zero and has covariance matrix equal to the identity. Applying Theorem 44 again, we see that

$$\mathbf{P}(|\vec{S}|\leqslant t)\ll O\biggl(\frac{t}{(i_+-i_-)^{1/2}}\biggr)^{(i_+-i_-)/4}+n^{-1/2}t^{-3}(\log n)^{O(1)}.$$

Applying this with

$$t:=\frac{(i_+-i_-)^{1/2}}{2^{m/4}(\log n)^{0.005}},$$

and using the fact that $i_+-i_-\geqslant l\geqslant C_2$ by hypothesis, one concludes that (106) occurs with probability

$$\ll O(2^{m/4}(\log n)^{0.005})^{-C_2/4}+n^{-1/2}2^{3m/4}(\log n)^{O(1)},$$

which proves the claim as long as C_2 is large and $2^m\leqslant n^{1/100}$. But the case $2^m\geqslant n^{1/100}$ then follows by noting that the probability of the event (106) is non-increasing in m. The proof of Proposition 53 is now complete.

Appendix A. Concentration of determinant

In view of the standard identity

$$\int_{-2}^{2}\log|y|\varrho_{\mathrm{sc}}(y)\,dy=-\frac{1}{2}$$

(which can be verified for instance by applying contour integration to a branch cut of $(4-z^2)^{1/2}\log z$ around the slit $[-2,2]$; see also Remark 67 below) and Stirling's formula, it suffices to prove the latter claim.

Fix z; we allow implied constants to depend on z. We of course have

$$\log|\det(M_n-zI\sqrt{n}\,)|=\sum_{j=1}^n\log|\lambda_j(M_n)-z\sqrt{n}\,|=\frac{1}{2}n\log n+\sum_{j=1}^n\log|\lambda_j(W_n-z)|,$$

so it will suffice to show that

$$\left|\frac{1}{n}\sum_{j=1}^n\log|\lambda_j(W_n)-z|-\int_{-2}^2\log|y-z|\varrho_{\mathrm{sc}}(y)\,dy\right|\leqslant n^{-c}$$

asymptotically almost surely for some $c>0$. Making the change of variables $y=t(x)$, where t is defined in (3), it suffices to show that

$$\left|\frac{1}{n}\sum_{j=1}^n\log|\lambda_j(W_n)-z|-\int_0^1\log|t(x)-z|\,dx\right|\leqslant n^{-c}$$

asymptotically almost surely for some $c>0$.

From Theorem 11 (with $k=1$), we have

$$\inf_j|\lambda_j(W_n)-z|\geqslant n^{-2}$$

asymptotically almost surely (because the expected number of eigenvalues in the interval $[z-n^{-2},z+n^{-2}]$ is $o(1)$). From this (and (4)) we conclude that

$$\sup_j\log|\lambda_j(W_n)-z|=O(\log n)$$

asymptotically almost surely. Thus, the contribution of all j with $|t(j/n)-\mathrm{Re}(z)|\leqslant n^{-\varepsilon}$ will be negligible for any fixed $\varepsilon>0$, and it suffices to show that

$$\left|\frac{1}{n}\sum_{\substack{1\leqslant j\leqslant n\\|t(j/n)-\mathrm{Re}(z)|>n^{-\varepsilon}}}\log|\lambda_j(W_n)-z|-\int_0^1\log|t(x)-z|\,dx\right|\leqslant n^{-c}$$

asymptotically almost surely.

By (2), we see that with probability $1-o(1)$, one has $\lambda_j(W_n)=t(j/n)+O(n^{-\delta})$ for all $1\leqslant j\leqslant n$ and some absolute constant $\delta>0$, where $-2\leqslant t(a)\leqslant 2$ is defined by (3). By Taylor expansion, we thus have asymptotically almost surely that

$$\log|\lambda_j(W_n)-z|=\log\left|t\left(\frac{j}{n}\right)-z\right|+O(n^{-\delta/2})$$

for all $1\leqslant j\leqslant n$ with $|t(j/n)-\mathrm{Re}(z)|>n^{1-\varepsilon}$, if ε is chosen sufficiently small depending on δ. The claim then follows (for ε small enough) by approximating

$$\int_0^1 \log|t(x)-z|\,dx$$

by its Riemann integral away from the possible singularity at $\mathrm{Re}(z)$.

Remark 67. The logarithmic potential

$$\int_{-2}^2 \log|y-z|\varrho_{\mathrm{sc}}(y)\,dy$$

for the semi-circular distribution can be computed explicitly as

$$\int_{-2}^2 \log|y-z|\varrho_{\mathrm{sc}}(y)\,dy=\frac{1}{2}\mathrm{Re}\biggl(\frac{z-\sqrt{z^2-4}}{z+\sqrt{z^2-4}}\biggr)+\log\biggl|\frac{\sqrt{z^2-4}+z}{2}\biggr|,$$

where $\sqrt{z^2-4}$ is the branch of the square root of z^2-4 with cut at $[-2,2]$ which is asymptotic to z at infinity; this can be seen by integrating the formula

$$\int_{-2}^2 \frac{1}{y-z}\varrho_{\mathrm{sc}}(y)\,dy=\frac{-z+\sqrt{z^2-4}}{2} \tag{107}$$

for the Stieltjes formula for the semi-circular potential, which can be easily verified by the Cauchy integral formula.

Appendix B. The distance between a random vector and a subspace

The prupose of this appendix is to prove Lemma 43. We restate this lemma for the reader's convenience.

Lemma 68. *Let $X=(\xi_1,...,\xi_n)\in\mathbb{C}^n$ be a random vector whose entries are independent with mean zero, variance* 1, *and are bounded in magnitude by K almost surely for some $K\geqslant 10(\mathbf{E}|\xi|^4+1)$. Let H be a subspace of dimension d and π_H be the orthogonal projection onto H. Then*

$$\mathbf{P}\big(\big|\|\pi_H(X)\|-\sqrt{d}\,\big|\geqslant t\big)\leqslant 10\exp\biggl(-\frac{t^2}{10K^2}\biggr).$$

In particular, one has

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

with overwhelming probability.

It is easy to show that $\mathbf{E}\|\pi_H(X)\|^2=d$, so it is indeed natural to expect that with high probability $\pi_H(X)$ is around $\sqrt{d}$.

In a previous paper [44], the authors proved Lemma 68 for the special case when ξ_i are Bernoulli random variables (taking values ± 1 with probability half). This proof is a simple generalization of one in [44] (see also [45, Appendix E]). We use the following theorem, which is a consequence of Talagrand's inequality (see [45, Theorem E.2], [31], or [43]).

THEOREM 69. (Talagrand's inequality) *Let* $\mathbf{D}$ *be the unit disk* $\{z\in\mathbb{C}:|z|\leqslant 1\}$. *For every product probability* μ *on* $\mathbf{D}^n$, *every convex* 1*-Lipschitz function* $F:\mathbb{C}^n\to\mathbb{R}$ *and every* $r\geqslant 0$,

$$\mu(|F-M(F)|\geqslant r)\leqslant 4\exp\left(-\frac{r^2}{16}\right),$$

where $M(F)$ *denotes the median of* F.

Remark 70. In fact, the result still holds for the space $\mathbf{D}_1\times\ldots\times\mathbf{D}_n$, where the $\mathbf{D}_i$'s are complex regions with diameter 2.

An easy change of variables reveals the following generalization of this inequality: if μ is supported on a dilate $K\cdot\mathbf{D}^n$ of the unit disk for some $K>0$, rather than $\mathbf{D}^n$ itself, then for every $r>0$ we have

$$\mu(|F-M(F)|\geqslant r)\leqslant 4\exp\left(-\frac{r^2}{16K^2}\right). \tag{108}$$

In what follows, we assume that $K\geqslant g(n)$, where $g(n)$ is tending (arbitrarily slowly) to infinity with n. The map $X\mapsto|\pi_H(X)|$ is clearly convex and 1-Lipschitz. Applying (108), we conclude that

$$\mathbf{P}\big(\big||\pi_H(X)|-M(|\pi_H(X)|)\big|\geqslant t\big)\leqslant 4\exp\left(-\frac{t^2}{16K^2}\right) \tag{109}$$

for any $t>0$. To conclude the proof, it suffices to show that

$$\big|M(|\pi_H(X)|)-\sqrt{d}\,\big|\leqslant 2K. \tag{110}$$

Let $P=(p_{jk})_{1\leqslant j,k\leqslant n}$ be the $n\times n$ orthogonal projection matrix onto H. We have that $\operatorname{trace}(P^2)=\operatorname{trace}(P)=\sum_{i=1}^n p_{ii}=d$ and $|p_{ii}|\leqslant 1$. Furthermore,

$$|\pi_H(X)|^2=\sum_{1\leqslant i,j\leqslant n}p_{ij}\xi_i\bar{\xi}_j=\sum_{i=1}^n p_{ii}|\xi_i|^2+\sum_{1\leqslant i\neq j\leqslant n}p_{ij}\xi_i\bar{\xi}_j.$$

Consider the event $\mathcal{E}_+$ that $|\pi_H(X)|\geqslant\sqrt{d}+2K$. Since this implies that

$$|\pi_H(X)|^2\geqslant d+4K\sqrt{d}+4K^2,$$

we have

$$\mathbf{P}(\mathcal{E}_+)\leqslant\mathbf{P}\biggl(\sum_{i=1}^n p_{ii}|\xi_i|^2\geqslant d+2K\sqrt{d}\biggr)+\mathbf{P}\biggl(\biggl|\sum_{1\leqslant i\neq j\leqslant n}p_{ij}\xi_i\bar{\xi}_j\biggr|\geqslant 2K\sqrt{d}\biggr).$$

Set $S_1:=\sum_{i=1}^n p_{ii}(|\xi_i|^2-1)$. Then we have, by Chebyshev's inequality,

$$\mathbf{P}\biggl(\sum_{i=1}^n p_{ii}|\xi_i|^2\geqslant d+2K\sqrt{d}\biggr)\leqslant\mathbf{P}(|S_1|\geqslant 2K\sqrt{d}\,)\leqslant\frac{\mathbf{E}(|S_1|^2)}{4dK^2}.$$

On the other hand, by the assumption on K,

$$\mathbf{E}|S_1|^2=\sum_{i=1}^n p_{ii}^2\mathbf{E}(|\xi_i|^2-1)^2=\sum_{i=1}^n p_{ii}^2(\mathbf{E}|\xi_i|^4-1)\leqslant\sum_{i=1}^n p_{ii}^2K=dK.$$

Thus,

$$\mathbf{P}(|S_1|\geqslant 2K\sqrt{d}\,)\leqslant\frac{\mathbf{E}|S_1|^2}{4dK^2}\leqslant\frac{1}{K}\leqslant\frac{1}{10}.$$

Similarly, set $S_2:=|\sum_{i\neq j}p_{ij}\xi_i\bar{\xi}_j|$. Then we have $\mathbf{E}S_2^2=\sum_{i\neq j}|p_{ij}|^2\leqslant d$. So again, by Chebyshev's inequality,

$$\mathbf{P}(S_2\geqslant 2K\sqrt{d}\,)\leqslant\frac{d}{4dK^2}\leqslant\frac{1}{10}.$$

It follows that $\mathbf{P}(\mathcal{E}_+)\leqslant\frac{1}{5}$, and so $M(\|\pi_H(X)\|)\leqslant\sqrt{d}+2K$. To prove the lower bound, let $\mathcal{E}_-$ be the event that $\|\pi_H(X)\|\leqslant\sqrt{d}-2K$ and notice that

$$\mathbf{P}(\mathcal{E}_-)\leqslant\mathbf{P}(|\pi_H(X)|^2\leqslant d-2K\sqrt{d}\,)\leqslant\mathbf{P}(S_1\leqslant d-K\sqrt{d}\,)+\mathbf{P}(S_2\geqslant K\sqrt{d}\,).$$

Both terms on the right-hand side can be bounded by $\frac{1}{5}$ by the same argument as above. The proof is complete.

Appendix C. Controlling the spectral density by the Stieltjes transform

In this appendix we establish Lemma 64 and Proposition 66.

Proof of Lemma 64. From (90) we see that, with overwhelming probability, one has

$$|s_n(x+\eta\sqrt{-1}\,)| \ll 1$$

for all $-L \leqslant x \leqslant L$ which are multiples of n^{-100}. From (89), one concludes that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\eta}{\eta^2+|\lambda_i(W_n)-x|^2} \ll 1,$$

and we thus have the crude bound

$$N_I \ll \eta n$$

whenever $I \subset [-L, L]$ is an interval of length $|I| = \eta$. Summing in I, we thus obtain the bound

$$N_I \ll |I| n \tag{111}$$

with overwhelming probability whenever $I \subset [-L, L]$ has length $|I| \geqslant \eta$. (One could also invoke Proposition 66 for this step.)

Next, let $I \subset [-L+\varepsilon, L-\varepsilon]$ be such that $|I| \geqslant 2\eta$, and consider the function

$$F(y) := n^{-100} \sum_{\substack{x \in I \\ n^{-100} | x}} \frac{\eta}{\pi(\eta^2+|y-x|^2)},$$

where the sum ranges over all $x \in I$ that are multiples of n^{-100}. Observe that

$$\frac{1}{n}\sum_{i=1}^{n} F(\lambda_i(W_n)) = n^{-100}\frac{1}{\pi}\mathrm{Im}\Bigg(\sum_{\substack{x \in I \\ n^{-100} | x}} s_n(x+\eta\sqrt{-1}\,)\Bigg)$$

and

$$\int_{\mathbb{R}} F(y)\varrho_{\mathrm{sc}}(y)\,dy = n^{-100}\frac{1}{\pi}\mathrm{Im}\Bigg(\sum_{\substack{x \in I \\ n^{-100} | x}} s(x+\eta\sqrt{-1}\,)\Bigg).$$

With overwhelming probability, we have $s_n(x+\eta\sqrt{-1}\,) = s(x+\eta\sqrt{-1}\,)+O(\delta)$ for all x in the sum, by hypothesis, and hence

$$\frac{1}{n}\sum_{i=1}^{n} F(\lambda_i(W_n)) = \int_{\mathbb{R}} F(y)\varrho_{\mathrm{sc}}(y)\,dy + O(|I|\delta).$$

On the other hand, from Riemann integration, one sees that

$$F(y) = \int_I \frac{\eta}{\pi(\eta^2+|y-x|^2)}\,dx + O(n^{-10}).$$

One can then establish the pointwise bounds

$$F(y) \ll \frac{1}{1+(\operatorname{dist}(y,I)/\eta)}+n^{-10}$$

when $y \notin I$ and $\operatorname{dist}(y,I) \leqslant |I|$,

$$F(y) \ll \frac{\eta|I|}{\operatorname{dist}(y,I)^2}+n^{-10}$$

when $y \notin I$ and $\operatorname{dist}(y,I)>|I|$, and (since $\eta/\pi(\eta^2+|y-x|^2)$ has integral 1) in the remaining case

$$F(y) = 1+O\left(\frac{1}{1+\operatorname{dist}(y,I^c)/\eta}\right)+O(n^{-10}).$$

Using these bounds, one sees that

$$\int_{\mathbb{R}} F(y)\varrho_{\mathrm{sc}}(y)\,dy = \int_I \varrho_{\mathrm{sc}}(y)\,dy+O\left(\eta\log\frac{|I|}{\eta}\right),$$

and a similar argument using Riemann integration and (111) (as well as the trivial bound $N_J \leqslant n$ when J lies outside $[-L,L]$) gives

$$\frac{1}{n}\sum_{i=1}^n F(\lambda_i(W_n)) = \frac{1}{n}N_I+O_\varepsilon\left(\eta\log\frac{|I|}{\eta}\right).$$

Putting all this together, we conclude that

$$N_I = n\int_I \varrho_{\mathrm{sc}}(y)\,dy+O_\varepsilon(\delta n|I|)+O_\varepsilon\left(\eta n\log\frac{|I|}{\eta}\right).$$

The latter error term can be absorbed into the former, since

$$|I| \geqslant \frac{\eta}{\delta}\log\frac{1}{\delta},$$

and the claim follows. $\square$

Proof of Proposition 66. By the union bound it suffices to show this for

$$|I| = \eta := \frac{K^2(\log n)^2}{n}.$$

Let x be the center of I. Then, by (89), it suffices to show that the event that

$$N_I \geqslant Cn\eta \tag{112}$$

and

$$\operatorname{Im}(s_n(x+\eta\sqrt{-1}\,)) \geqslant C \tag{113}$$

for some large absolute constant C, fails with overwhelming probability.

Suppose that we have both (112) and (113). By (92) we have

$$\frac{1}{n}\sum_{k=1}^{n}\left|\operatorname{Im}\left(\frac{1}{\zeta_{kk}/\sqrt{n}-(x+\eta\sqrt{-1}\,)-Y_k}\right)\right| \geqslant C;$$

using the crude bound $|\operatorname{Im}(1/z)| \leqslant 1/|\operatorname{Im}(z)|$, we conclude that

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{|\eta+\operatorname{Im}(Y_k)|} \geqslant C.$$

At the same time, by writing $W_{n,k}$ in terms of an orthonormal basis $u_j(W_{n,k})$ of eigenfunctions, one sees that

$$Y_k = \sum_{j=1}^{n-1}\frac{|u_j(W_{n,k})^* a_k|^2}{\lambda_j(W_{n,k})-(x+\eta\sqrt{-1}\,)}, \tag{114}$$

and hence

$$\operatorname{Im}(Y_k) \geqslant \eta\sum_{j=1}^{n-1}\frac{|u_j(W_{n,k})^* a_k|^2}{\eta^2+(\lambda_j(W_{n,k})-x)^2}.$$

On the other hand, from (112) we can find $1 \leqslant i_- < i_+ \leqslant n$ with $i_+ - i_- \geqslant \eta n$ such that $\lambda_i(W_n) \in I$ for all $i_- \leqslant i \leqslant i_+$. By the Cauchy interlacing property (24), we thus have $\lambda_i(W_{n,k}) \in I$ for $i_- \leqslant i < i_+$. We conclude that

$$\operatorname{Im}(Y_k) \gg \frac{1}{\eta}\sum_{i_- \leqslant j < i_+}|u_j(W_{n,k})^* a_k|^2 = \frac{1}{\eta}\|P_{H_k} a_k\|^2,$$

where P_{H_k} is the orthogonal projection onto the $(i_+ - i_-)$-dimensional space H_k spanned by the eigenvectors $u_j(W_{n,k})$ for $i_- \leqslant j < i_+$. Putting all this together, we conclude that

$$\frac{1}{n}\sum_{k=1}^{n}\frac{\eta}{\eta^2+\|P_{H_k} a_k\|^2} \gg C.$$

On the other hand, from Lemma 68 we see that $\|P_{H_k} a_k\|^2 = O(\eta)$ with overwhelming probability. (One has to take the union bound over all possible choices of i_- and i_+, but there are only $O(n^2)$ such choices at most, so this is not a problem.) The claim then follows by taking C sufficiently large. $\square$

Appendix D. A multidimensional Berry–Esseen theorem

In this section, we prove Theorem 44. We will need the following multidimensional Berry–Esséen theorem, which is a generalisation of [45, Proposition D.2].

THEOREM 71. *Let* $N, n \geqslant 1$ *be integers, let* $v_1, ..., v_n \in \mathbb{C}^N$ *be vectors, let* $\zeta_1, ..., \zeta_n$ *be independent complex-valued variables with mean zero, variance* $\mathbf{E}|\zeta_j|^2 = 1$ *and the third moment bound*

$$\sup_{1 \leqslant i \leqslant n} \mathbf{E}|\zeta_i|^3 \leqslant C \tag{115}$$

for some constant $C \geqslant 1$. *Let* S *be the* $\mathbb{C}^N$*-valued random variable*

$$S := \sum_{i=1}^{n} v_i \zeta_i.$$

We identify $\mathbb{C}^N$ *with* $\mathbb{R}^{2N}$ *in the usual manner, and define the* covariance matrix M *of* S *to be the unique symmetric* $2N \times 2N$ *real matrix such that*

$$u^* M u := \mathbf{E}|\mathrm{Re}(u^* S)|^2 \tag{116}$$

for all $u \in \mathbb{C}^N \equiv \mathbb{R}^{2N}$.

Let G *be a Gaussian random variable on* $\mathbb{R}^{2N} \equiv \mathbb{C}^N$ *with mean zero and with the same covariance matrix* M *as* S, *and thus*

$$u^* M u = \mathbf{E}|G \cdot u|^2 = \mathbf{E}|\mathrm{Re}(u^* S)|$$

for all $u \in \mathbb{R}^{2n} \equiv \mathbb{C}^n$ (*where* $u \cdot v = \mathrm{Re}(v^* u)$ *denotes the dot product on* $\mathbb{R}^{2N}$). *More explicitly*, G *has the distribution function*

$$\frac{1}{(2\pi)^n (\det M)^{1/2}} \exp\left(-\frac{x^* M^{-1} x}{2}\right) dx_1 \ldots dx_{2N}$$

if M *is invertible, with an analogous limiting formula when* M *is singular. Then, for any* $\varepsilon > 0$ *and any measurable set* $\Omega \subset \mathbb{R}^{2N} \equiv \mathbb{C}^N$, *one has*

$$\mathbf{P}(S \in \Omega) \leqslant \mathbf{P}(G \in \Omega \cup \partial_\varepsilon \Omega) + O\left(CN^{3/2} \varepsilon^{-3} \sum_{j=1}^{n} |v_j|^3\right), \tag{117}$$

and similarly

$$\mathbf{P}(S \in \Omega) \geqslant \mathbf{P}(G \in \Omega \backslash \partial_\varepsilon \Omega) - O\left(CN^{3/2} \varepsilon^{-3} \sum_{j=1}^{n} |v_j|^3\right), \tag{118}$$

where

$$\partial_\varepsilon \Omega := \{x \in \mathbb{R}^{2N} : \mathrm{dist}_\infty(x, \partial\Omega) \leqslant \varepsilon\},$$

$\partial\Omega$ *is the topological boundary of* Ω *and* dist_∞ *is the distance with respect to the* ℓ^∞ *metric on* $\mathbb{R}^{2N}$.

Remark 72. The main novelty here, compared with that in [45, Proposition D.2], is that the random variable ζ_j is not assumed to be $\mathbb{C}$-normalized (which means that the real and imaginary parts of ζ_j have covariance matrix equal to half the identity). For instance, some of the ζ_j could be purely real, or supported on some other line through the origin, such as the imaginary axis.

Proof. We obtain the result by repeating the proof of [45, Proposition D.2] with some proper modification. For the readers convenience, we present all details.

It suffices to prove (117), as (118) follows by replacing Ω with its complement.

Let $\psi:\mathbb{R}\to\mathbb{R}^+$ be a bump function supported on the unit ball $\{x\in\mathbb{R}:|x|\leqslant 1\}$ of total mass $\int_{\mathbb{R}}\psi=1$, let $\Psi_{\varepsilon,N}:F^{\mathbb{R}}\to\mathbb{R}^+$ be the approximation of the identity

$$\Psi_{\varepsilon,\mathbb{R}}(x_1,...,x_N):=\prod_{i=1}^{\mathbb{R}}\frac{1}{\varepsilon}\psi\Big(\frac{x_i}{\varepsilon}\Big),$$

and let $f:\mathbb{R}^n\to\mathbb{R}^+$ be the convolution

$$f(x)=\int_{\mathbb{R}^N}\Psi_{\varepsilon,N}(y)1_\Omega(x-y)\,dy, \tag{119}$$

where 1_Ω is the indicator function of Ω. Observe that f equals 1 on $\Omega\backslash\partial_\varepsilon\Omega$, vanishes outside $\Omega\cup\partial_\varepsilon\Omega$, and is smoothly varying between 0 and 1 on $\partial_\varepsilon\Omega$. Thus it will suffice to show that

$$|\mathbf{E}f(S)-\mathbf{E}f(G)|\ll CN^{3/2}\varepsilon^{-3}\sum_{j=1}^n|v_j|^3.$$

We now use a Lindeberg replacement trick (cf. [33] and [38]). For each $1\leqslant i\leqslant n$, let g_i be a complex Gaussian with mean zero and with the same covariance matrix as ζ_i, and thus

$$\mathbf{E}\mathrm{Re}(g_i)^2=\mathbf{E}\mathrm{Re}(\zeta_i)^2,\quad \mathbf{E}\mathrm{Im}(g_i)^2=\mathbf{E}\mathrm{Im}(\zeta_i)^2$$

and

$$\mathbf{E}\mathrm{Re}(g_i)\mathrm{Im}(g_i)=\mathbf{E}\mathrm{Re}(\zeta_i)\mathrm{Im}(\zeta_i).$$

In particular g_i has mean zero and variance 1. We construct the $g_1,...,g_n$ to be jointly independent. Observe from (116) that the random variable

$$g_1v_1+...+g_nv_n\in\mathbb{C}^N$$

has mean zero and covariance matrix M, and thus has the same distribution as G. Thus, if we define the random variables

$$S_j:=\zeta_1v_1+...+\zeta_jv_j+g_{j+1}v_j+...+g_nv_n\in\mathbb{C}^N,$$

we have the telescoping triangle inequality

$$|\mathbf{E}f(S)-\mathbf{E}f(G)|\leqslant\sum_{j=1}^{n}|\mathbf{E}f(S_j)-\mathbf{E}f(S_{j-1})|. \tag{120}$$

For each $1\leqslant j\leqslant n$ we may write

$$S_j=S_j'+\zeta_j v_j \quad\text{and}\quad S_{j-1}=S_j'+g_j v_j,$$

where

$$S_j':=\zeta_1 v_1+\ldots+\zeta_{j-1}v_{j-1}+g_{j+1}v_j+\ldots+g_n v_n.$$

By Taylor's theorem with remainder, we thus have

$$f(S_j)=P_{S_j'}(\mathrm{Re}(\zeta_j),\mathrm{Im}(\zeta_j))+O\biggl(|\zeta_j|^3\sup_{x\in\mathbb{R}^n}\sum_{k=0}^{3}|(v_j\cdot\nabla)^k((v_j\sqrt{-1}\,)\cdot\nabla)^{3-k}f(x)|\biggr) \tag{121}$$

and

$$f(S_{j-1})=P_{S_j'}(\mathrm{Re}(g_j),\mathrm{Im}(g_j))+O\biggl(|g_j|^3\sup_{x\in\mathbb{R}^n}\sum_{k=0}^{3}|(v_j\cdot\nabla)^k((v_j\sqrt{-1}\,)\cdot\nabla)^{3-k}f(x)|\biggr), \tag{122}$$

where $P_{S_j'}$ is some quadratic polynomial depending on S_j', and v_j and $v_j\sqrt{-1}$ are viewed as vectors in $\mathbb{R}^{2N}$. A computation using (119) and the Leibniz rule reveals that all third partial derivatives of f have magnitude $O(\varepsilon^{-3})$, and so, by the Cauchy–Schwarz inequality, we have

$$\sum_{k=0}^{3}|(v_j\cdot\nabla)^k((v_j\sqrt{-1}\,)\cdot\nabla)^{3-k}f(x)|\ll|v_j|^3N^{3/2}\varepsilon^{-3}.$$

Observe that ζ_j and g_j are independent of S_j', and have the same mean and covariance matrix. Subtracting (121) from (122) and taking expectations using (115), we conclude that

$$|\mathbf{E}f(S_j)-\mathbf{E}f(S_{j-1})|\ll C|v_j|^3N^{3/2}\varepsilon^{-3},$$

and the claim follows from (120). □

Remark 73. The bounds here are not best possible, but are sufficient for our applications.

Now we are ready to prove Theorem 44.

Proof of Theorem 44. We first prove the upper tail bound on S_i. Here, the main tool is the $N=1$ case of Theorem 71. The variance of S_i is

$$\mathbf{E}|S_i|^2=\sum_{j=1}^n |a_{ij}|^2=1, \tag{123}$$

since the rows of A have unit size. Thus, the 2×2 covariance matrix of S_i is $O(1)$. Let G_i be a complex Gaussian with mean zero and the same covariance matrix as S_i. By Theorem 71, we have

$$\mathbf{P}(|S_i|\geqslant t)\leqslant \mathbf{P}(|G|\geqslant t-\varepsilon\sqrt{2}\,)+O\biggl(C\varepsilon^{-3}\sum_{i=1}^N\sum_{j=1}^n |a_{ij}|^3\biggr)$$

for any $\varepsilon>0$. Selecting $\varepsilon:=\frac{1}{10}$, and using the fact that G has variance 1, we conclude that

$$\mathbf{P}(|S_i|\geqslant t)\leqslant \exp(-ct^2)+O\biggl(C\sum_{i=1}^N\sum_{j=1}^n |a_{ij}|^3\biggr),$$

and the claim follows from (26) and (123).

Now we prove the lower tail bound on $\vec{S}$, using Theorem 71 in full generality. Observe that for any unit vector $u\in\mathbb{C}^N\equiv\mathbb{R}^{2N}$, one has

$$\mathbf{E}|u^*S|^2=\mathbf{E}\|u^*A\|^2=1,$$

by the orthonormality of the rows of A. Thus, by (116), the operator norm of the covariance matrix M of S has operator norm at most 1. On the other hand, we have

$$\operatorname{trace}(M)=\mathbf{E}|S|^2=\|A\|_F^2=N.$$

Thus, the $2N$ eigenvalues of M range between 0 and 1 and add up to N. This implies that at least $\frac{1}{2}N$ of them are at least $\frac{1}{4}$, and so one can find a $\lfloor\frac{1}{2}N\rfloor$-dimensional real subspace V of $\mathbb{R}^{2N}$ such that M is invariant on V and has all eigenvalues at least $\frac{1}{4}$ on V.

Now let G be a Gaussian in $\mathbb{R}^{2N}\equiv\mathbb{C}^N$ with mean zero and covariance matrix M. By Theorem 71, we have

$$\mathbf{P}(|\vec{S}|\leqslant t)\leqslant \mathbf{P}(|G|\leqslant t+\varepsilon\sqrt{2N}\,)+O\biggl(CN^{3/2}\varepsilon^{-3}\sum_{i=1}^N\sum_{j=1}^n |a_{ij}|^3\biggr)$$

for any $\varepsilon>0$. By (26) and (123) we have

$$\sum_{i=1}^N\sum_{j=1}^n |a_{ij}|^3\leqslant N\sigma.$$

Setting $\varepsilon:=t/\sqrt{2N}$, we conclude that

$$\mathbf{P}(|\vec{S}|\leqslant t)\leqslant\mathbf{P}(|G|\leqslant 2t)+O(CN^4t^{-3}\sigma).$$

Let G_V be the orthogonal projection of G onto V. Clearly

$$\mathbf{P}(|G|\leqslant 2t)\leqslant\mathbf{P}(|G_V|\leqslant 2t).$$

The Gaussian G_V has mean zero and covariance matrix at least $\frac{1}{4}I_V$ (i.e. all eigenvalues are at least $\frac{1}{4}$). By applying a linear transformation to reduce the covariance, we see that the quantity $\mathbf{P}(|G_V|\leqslant 2t)$ is maximized when the covariance matrix is *exactly* $\frac{1}{4}I_V$. Thus, in any orthonormal basis of G_V, the $\lfloor\frac{1}{2}N\rfloor$ components of G_V, $g_1,...,g_{\lfloor N/2\rfloor}$, are independent real Gaussians of variance $\frac{1}{4}$. If $|G_V|\leqslant 2t$, then $g_1^2+...+g_{\lfloor N/2\rfloor}^2\leqslant 4t^2$, and thus (by Markov's inequality) $g_i^2\leqslant 8t^2/N$ for at least $\lfloor\frac{1}{4}N\rfloor$ of the indices i. The number of choices of these indices is at most $2^{\lfloor N/2\rfloor}$, and the events $g_i^2\leqslant 2t^2/N$ are independent and occur with probability $O(t/\sqrt{N}\,)$, so we conclude from the union bound that

$$\mathbf{P}(|G_V|\leqslant 2t)\leqslant O(t/\sqrt{N}\,)^{\lfloor N/4\rfloor}$$

and the claim follows. □

Acknowledgments. We would like to thank P. Sarnak for bringing this beautiful problem to our attention. We would also like to thank G. Ben Arous, M. Krishnapur, K. Johansson, J. Lebowitz, S. Péché, B. Rider, A. Soshnikov and O. Zeitouni for useful conversations regarding the state-of-the-art of GUE/GOE and Johansson matrices. Part of the paper was written while the second author was visiting Microsoft Research in Boston and he would like to thank J. Chayes and C. Borgs for their hospitality. We thank Alain-Sol Sznitman and the anonymous referees for corrections.

References

[1] Anderson, G., Guionnet, A. & Zeitouni, O., *An Introduction to Random Matrices.* To appear by Cambridge University Press.

[2] Bai, Z. & Silverstein, J. W., *Spectral Analysis of Large Dimensional Random Matrices.* Springer Series in Statistics. Springer, New York, 2010.

[3] Bai, Z. D. & Yin, Y. Q., Convergence to the semicircle law. *Ann. Probab.*, 16 (1988), 863–875.

[4] — Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16 (1988), 1729–1741.

[5] Ben Arous, G. & Péché, S., Universality of local eigenvalue statistics for some sample covariance matrices. *Comm. Pure Appl. Math.*, 58 (2005), 1316–1357.

[6] Bleher, P. & Its, A., Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model. *Ann. of Math.*, 150 (1999), 185–266.

[7] Bleher, P. & Kuijlaars, A. B. J., Large n limit of Gaussian random matrices with external source. I. *Comm. Math. Phys.*, 252 (2004), 43–76.

[8] Chatterjee, S., A simple invariance principle. Preprint, 2005. `arXiv:math/0508213 [math.PR]`.

[9] Costello, K. P., Tao, T. & Vu, V., Random symmetric matrices are almost surely nonsingular. *Duke Math. J.*, 135 (2006), 395–413.

[10] Costello, K. P. & Vu, V., Concentration of random determinants and permanent estimators. *SIAM J. Discrete Math.*, 23 (2009), 1356–1371.

[11] Costin, O. & Lebowitz, J., Gaussian fluctuations in random matrices. *Phys. Rev. Lett.*, 75 (1995), 69–72.

[12] Curto, R. E. & Fialkow, L. A., Recursiveness, positivity, and truncated moment problems. *Houston J. Math.*, 17 (1991), 603–635.

[13] 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, 1999.

[14] — Universality for mathematical and physical systems, in *International Congress of Mathematicians*. Vol. I, pp. 125–152. Eur. Math. Soc., Zurich, 2007.

[15] Deift, P., Kriecherbauer, T., McLaughlin, K. T. R., Venakides, S. & Zhou, X., Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52 (1999), 1335–1425.

[16] Dyson, F. J., Correlations between eigenvalues of a random matrix. *Comm. Math. Phys.*, 19 (1970), 235–250.

[17] Erdős, L., Péché, S., Ramírez, J. A., Schlein, B. & Yau, H.-T., Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.*, 63 (2010), 895–925.

[18] Erdős, L., Ramírez, J. A., Schlein, B., Tao, T., Vu, V. & Yau, H.-T., Bulk universality for Wigner Hermitian matrices with subexponential decay. *Math. Res. Lett.*, 17 (2010), 667–674.

[19] Erdős, L., Ramírez, J. A., Schlein, B. & Yau, H.-T., Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electron. J. Probab.*, 15 (2010), 526–603.

[20] Erdős, L., Schlein, B. & Yau, H.-T., Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287 (2009), 641–655.

[21] — Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37 (2009), 815–852.

[22] — Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN*, 2010 (2010), 436–479.

[23] Ginibre, J., Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6 (1965), 440–449.

[24] Götze, F. & Kösters, H., On the second-order correlation function of the characteristic polynomial of a Hermitian Wigner matrix. *Comm. Math. Phys.*, 285 (2009), 1183–1205.

[25] Guionnet, A. & Zeitouni, O., Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5 (2000), 119–136.

[26] Gustavsson, J., Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist.*, 41 (2005), 151–178.

[27] Jimbo, M., Miwa, T., Môri, Y. & Sato, M., Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent. *Phys. D*, 1 (1980), 80–158.
[28] Johansson, K., Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.*, 215 (2001), 683–705.
[29] Katz, N. M. & Sarnak, P., *Random Matrices, Frobenius Eigenvalues, and Monodromy*. American Mathematical Society Colloquium Publications, 45. Amer. Math. Soc., Providence, RI, 1999.
[30] Landau, H. J. (ed.), *Moments in Mathematics*. Proceedings of Symposia in Applied Mathematics, 37. Amer. Math. Soc., Providence, RI, 1987.
[31] Ledoux, M., *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, 89. Amer. Math. Soc., Providence, RI, 2001.
[32] Levin, E. & Lubinsky, D. S., Universality limits in the bulk for varying measures. *Adv. Math.*, 219 (2008), 743–779.
[33] Lindeberg, J. W., Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, 15 (1922), 211–225.
[34] Mehta, M. L., *Random Matrices and the Statistical Theory of Energy Levels*. Academic Press, New York, 1967.
[35] Mehta, M. L. & Gaudin, M., On the density of eigenvalues of a random matrix. *Nuclear Phys.*, 18 (1960), 420–427.
[36] Pastur, L., On the spectrum of random matrices. *Teoret. Mat. Fiz.*, 10 (1972), 102–112 (Russian). English translation in *Theoret. and Math. Phys.*, 10 (1972), 67–74.
[37] Pastur, L. & Shcherbina, M., Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. *J. Stat. Phys.*, 86 (1997), 109–147.
[38] Paulauskas, V. & Račkauskas, A., *Approximation Theory in the Central Limit Theorem*. Mathematics and its Applications, 32. Kluwer, Dordrecht, 1989.
[39] Péché, S. & Soshnikov, A., Wigner random matrices with non-symmetrically distributed entries. *J. Stat. Phys.*, 129 (2007), 857–884.
[40] — On the lower bound of the spectral norm of symmetric random matrices with independent entries. *Electron. Commun. Probab.*, 13 (2008), 280–290.
[41] Soshnikov, A., Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.*, 207 (1999), 697–733.
[42] — Gaussian limit for determinantal random point fields. *Ann. Probab.*, 30 (2002), 171–187.
[43] Talagrand, M., A new look at independence. *Ann. Probab.*, 24 (1996), 1–34.
[44] Tao, T. & Vu, V., On random ± 1 matrices: singularity and determinant. *Random Structures Algorithms*, 28 (2006), 1–23.
[45] — Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20 (2010), 260–297.
[46] — Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.*, 298 (2010), 549–572.
[47] Wigner, E. P., On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, 67 (1958), 325–327.

Terence Tao
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095-1555
U.S.A.
tao@math.ucla.edu

Van Vu
Department of Mathematics
Rutgers
Piscataway, NJ 08854
U.S.A.
vanvu@math.rutgers.edu

Received June 8, 2009