Wegner estimate and level repulsion for Wigner random matrices

László Erdős[1]*, Benjamin Schlein[2†] and Horng-Tzer Yau[3‡]

Institute of Mathematics, University of Munich,
Theresienstr. 39, D-80333 Munich, Germany[1]

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
Wilberforce Rd, Cambridge CB3 0WB, UK[2]

Department of Mathematics, Harvard University
Cambridge MA 02138, USA[3]

Nov 11, 2008

Abstract

We consider $N \times N$ Hermitian random matrices with independent identically distributed entries (Wigner matrices). The matrices are normalized so that the average spacing between consecutive eigenvalues is of order $1/N$. Under suitable assumptions on the distribution of the single matrix element, we first prove that, away from the spectral edges, the empirical density of eigenvalues concentrates around the Wigner semicircle law on energy scales $\eta \gg N^{-1}$. This result establishes the semicircle law on the optimal scale and it removes a logarithmic factor from our previous result [6]. We then show a Wegner estimate, i.e. that the averaged density of states is bounded. Finally, we prove that the eigenvalues of a Wigner matrix repel each other, in agreement with the universality conjecture.

AMS Subject Classification: 15A52, 82B44

Running title: Wegner estimate and level repulsion

Key words: Semicircle law, Wigner random matrix, level repulsion, Wegner estimate, density of states, localization, extended states.

*Partially supported by SFB-TR 12 Grant of the German Research Council

†Partially supported by Sofja-Kovalevskaya Award of the Humboldt Foundation.

‡Partially supported by NSF grants DMS-0602038, 0757425, 0804279

1 Introduction

Let $H = (h_{ij})$ be an $N \times N$ hermitian matrix with eigenvalues $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_N$. These matrices form a *hermitian Wigner ensemble* if the matrix elements,

$$h_{ij} = \bar{h}_{ji} = N^{-1/2} z_{ij} \in \mathbb{C}, \quad (1 \leq i < j \leq N), \quad \text{and} \quad h_{ii} = N^{-1/2} x_{ii} \in \mathbb{R}, \quad (1 \leq i \leq N) \tag{1.1}$$

are independent random variables with mean zero. We assume that z_{ij} $(i < j)$ all have a common distribution ν with variance $\int_{\mathbb{C}} |z|^2 \mathrm{d}\nu(z) = 1$ and with a strictly positive density function $h : \mathbb{R}^2 \to \mathbb{R}_+$, i.e.

$$\mathrm{d}\nu(z) = (\text{const.}) h(x, y) \mathrm{d}x \mathrm{d}y \quad \text{where} \quad x = \operatorname{Re} z, \quad y = \operatorname{Im} z.$$

We will often denote $g := -\log h$. Throughout the paper we also assume that

$$\text{either} \quad h(x, y) = h^*(x) h^*(y), \qquad \text{or} \quad h(x, y) = h^*(x^2 + y^2) \tag{1.2}$$

with some positive function $h^* : \mathbb{R} \to \mathbb{R}_+$, i.e. either the real and imaginary parts of the random variables z_{ij}, $i < j$, are independent and identically distributed, or the distribution depends only on the absolute value $|z_{ij}|$. The diagonal elements, x_{ii}, also have a common distribution, $\mathrm{d}\widetilde{\nu}(x) = (\text{const.}) e^{-\widetilde{g}(x)} \mathrm{d}x$ with $\widetilde{g} : \mathbb{R} \to \mathbb{R}$. Let $\mathbb{P}$ and $\mathbb{E}$ denote the probability and the expectation value, respectively, w.r.t the joint distribution of all matrix elements. The normalization (1.1) of the matrix elements and fixing the variance of $\mathrm{d}\nu$ to be one ensure that the spectrum of H is $[-2, 2] + o(1)$ with probability one in the limit as $N \to \infty$.

For the special case $g(x, y) = x^2 + y^2$, $\widetilde{g}(x) = x^2/2$, the hermitian Wigner ensemble is called the *Gaussian Unitary Ensemble (GUE)*. Due to the unitary invariance of the GUE matrices, the joint eigenvalue distribution can be explicitly expressed in terms of a Vandermonde determinant and all correlation functions are computable (see [12] for an overview). This approach can be applied for more general ensembles with unitary invariance, i.e. for ensembles where the distribution is invariant under the transformation $H \to U^{-1} H U$ for any unitary matrix U (for a general overview via the Riemann-Hilbert approach see [4]). In particular, the density of the eigenvalues converges to the Wigner semicircle law as $N \to \infty$ and the truncated two-point correlation function, appropriately rescaled, is given by the famous Wigner-Dyson sine-kernel in the bulk spectrum, see [4, 13] and references therein (near the spectral edges a different universal statistics holds). Higher order correlations can be expressed as determinants involving the sine-kernel. The order statistics of eigenvalues can also be computed. The most important one is the nearest-neighbor level statistics, or gap distribution, i.e. the distribution of the difference between two consecutive eigenvalues, $\mu_{\alpha+1} - \mu_\alpha$, in the bulk. With an appropriate rescaling, the density function $f(x)$ of the eigenvalue gap is universal. It is characterized by $f(x) \sim x^2$ near 0 that corresponds to a strong level repulsion. The large distance behavior, $f(x) \sim \exp(-x^2)$, expresses a strong supression of large eigenvalue gaps.

These properties of the eigenvalue statistics are conjectured to hold for much more general matrix ensembles beyond the invariant ensembles, in particular for general Wigner matrices. Numerical evidences very strongly support these conjectures, nevertheless only a few rigorous results are known for ensembles without unitary invariance (notable exceptions are the universality of the Tracy-Widom distribution for the extremal eigenvalues [14] and the Wigner-Dyson sine-kernel for Wigner matrices with Gaussian convoluted distributions [9]). In this paper we prove the strong level repulsion and a subexponential estimate for the large distance behavior of the gap distribution.

For Wigner matrices, the Wigner semicircle law has been long established on scale of order 1, i.e. the empirical counting measure of the eigenvalues (also called empirical density of states measure in physics),

$\varrho_N(E) = \frac{1}{N}\sum_{j=1}^N \delta(E-\mu_j)$, converges weakly to $\varrho_{sc}(E)\mathrm{d}E$ (see (2.4)) in probability as $N\to\infty$ (see [15] for the original result). The weak convergence does not allow one to identify the local density of eigenvalues on energy scales $\eta \ll 1$. Note that the number of eigenvalues in any interval of length η within $[-2,2]$ is typically of order $N\eta$, so the self-averaging property is expected to hold for the smoothed density of states as long as the smoothing is on scale $\eta \gg 1/N$. In Section 6, using a necessary a-priori bound from Section 5, we prove that the semicircle law holds on the smallest possible scales, i.e. for any interval of length $\eta \gg 1/N$ (Theorem 3.1). This removes the logarithmic factor in our previous work [6] and establishes the optimal result. As a corollary, we obtain an optimal result on the delocalization of the eigenvectors (Corollary 3.2). The proof is a bootstrap argument in η; it relies on (non-optimal) bounds on the supremum norm of the eigenvectors, which in turn, can be obtained by first establishing the semicircle law on a larger scale $\eta \geq (\log N)^4/N$. Although the semicircle law on a larger scale and bounds on the eigenvectors were already established in [6], the error bound was not sufficiently strong. Therefore, in Section 4, we first improve the results of [6].

In Section 7 we give an upper bound on the tail distribution of the distance between consecutive eigenvalues (Theorem 3.3). The bound is only subexponential in contrast to the expected Gaussian decay. In Section 8 we prove the Wegner estimate for Wigner matrices, i.e. that the averaged density of states, $\mathbb{E}\,\varrho(E)$, is bounded (Theorem 3.4). Note that the Wegner estimate is an information on arbitrarily short scales, i.e. it is uniform in η. On scales $\eta \lesssim 1/N$, however, the smoothed empirical density of states truly fluctuates since individual eigenvalues near E dominate, but the averaged density of states remains bounded.

Finally, in Section 9 we establish an upper bound $f(x) \leq Cx^2$ for the density function of the eigenvalue spacing in the regime where x is small (Theorem 3.5). Apart from the constant, this upper bound coincides with the prediction obtained from the universality conjecture on the level spacing distribution and it proves that the level repulsion in Wigner matrices is as strong as in the GUE ensemble. We also give an optimal estimate on higher order level repulsion. We show that the probability that there are k eigenvalues in a small spectral interval I, with $|I| = \varepsilon/N$ ($\varepsilon \ll 1$), is bounded from above by $C\varepsilon^{k^2}$ in accordance with the prediction from GUE that is based upon the explicit formula for the joint density function $\sim \prod_{j<\ell}(\mu_\ell - \mu_j)^2$ of k eigenvalues.

We work with hermitian Wigner matrices, but our method applies to symmetric Wigner matrices as well. In that case, the level repulsion is weaker, $f(x) \leq Cx$, in accordance with the explicit gap distribution function for Gaussian Orthogonal Ensemble (GOE).

We need to assume further conditions on the distributions of the matrix elements in addition to (1.1), (1.2):

C1) There exists a $\delta_0 > 0$ such that

$$D := \int_{\mathbb{C}} \exp\big[\delta_0 |z|^2\big]\mathrm{d}\nu(z) < \infty, \qquad \widetilde{D} := \int_{\mathbb{R}} \exp\big[\delta_0 x^2\big]\mathrm{d}\widetilde{\nu}(x) < \infty\,. \tag{1.3}$$

To establish the Wegner estimate and the level repulsion, we need some smoothness property of the density function h. We assume that

C2) The Fourier transform of the functions h and $h(\Delta g)$, with $g = -\log h$, satisfies the decay estimate

$$|\widehat{h}(t,s)| \leq \frac{1}{\big[1+\omega_a(t^2+s^2)\big]^a}, \qquad |\widehat{h\Delta g}(t,s)| \leq \frac{1}{\big[1+\widetilde{\omega}_a(t^2+s^2)\big]^a} \tag{1.4}$$

with some exponent $a \geq 1$ and constants $\omega_a, \widetilde{\omega}_a > 0$. (Note that $a\omega_a \leq \frac{1}{4}$ by the condition that the variance is 1.)

In our previous papers [5, 6] we assumed that dν satisfies the logarithmic Sobolev inequality for the proof of the analogue of Lemma 4.2 (Lemma 2.1 of [6]). M. Ledoux has kindly pointed out to us that by applying a theorem of Hanson and Wright [8], this lemma also holds under the moment condition **C1)** only. We remark that the original paper [8] assumed that dν was symmetric; this conditon was later removed by Wright [16].

Another assumption we made in [5, 6] states that either the Hessian of $g = -\log h$ is bounded from above or the distribution is compactly supported. This was needed because we used Lemma 2.3 of [5], whose original proof required the condition on Hess g. An alternative proof of this lemma was given by Bourgain (the proof reproduced in the Appendix of [6]) under the additional condition that the support of dν is compact. In this paper, we extend the results of [8, 16] and apply them to prove a weaker but for our purposes still sufficient version of Lemma 2.3 in [5]. This approach requires no additional condition apart from **C1)**. Condition **C2**) will play a role only in Theorem 3.4 and Theorem 3.5.

In our previous papers [5, 6] we assumed that the real and imaginary parts of z_{ij} are independent. It is straightforward to check that all results of [5, 6] hold for the case of radially symmetric distributions (second condition in (1.2)) as well.

Convention. We assume condition **C1)** throughout the paper and every constant may depend on the constants $\delta_0, D, \widetilde{D}$ from (1.3) without further notice.

Acknowledgement. The authors are grateful to M. Ledoux for his remark that Lemma 2.1 of [6] follows from a result of Hanson and Wright [8].

2 Notation and the basic formula

For any spectral parameter $z = E + i\eta \in \mathbb{C}$, $\eta > 0$, we denote the Green function by $G_z = (H - z)^{-1}$. Let $F(E) = F_N(E)$ be the empirical distribution function of the eigenvalues

$$F(E) := F_N(E) = \frac{1}{N}\Big| \{\alpha \ : \ \mu_\alpha \le E\} \Big| \tag{2.1}$$

(in physics it is called the integrated density of states). Its derivative is the empirical density of states measure

$$\varrho(E) := F'(E) = \frac{1}{N}\sum_{\alpha=1}^N \delta(E - \mu_\alpha).$$

Its statistical average, $\mathbb{E}\,\varrho(E)$, is called the *averaged density of states.* We define the Stieltjes transform of F as

$$m = m(z) = \frac{1}{N}\mathrm{Tr}\, G_z = \int_{\mathbb{R}} \frac{\mathrm{d}F(E)}{E - z}\,, \tag{2.2}$$

and we let

$$\varrho_\eta(E) = \frac{\mathrm{Im}\, m(z)}{\pi} = \frac{1}{N\pi}\mathrm{Im}\ \mathrm{Tr}\ G_z = \frac{1}{N\pi}\sum_{\alpha=1}^N \frac{\eta}{(\mu_\alpha - E)^2 + \eta^2} \tag{2.3}$$

be the normalized density of states of H around energy E and regularized on scale η. We note that $\varrho(E) = \lim_{\eta\to 0+0} \varrho_\eta(E)$. The random variable m and the random measures ϱ and ϱ_η also depend on N, when necessary, we will indicate this fact by writing m_N, ϱ_N and $\varrho_{\eta,N}$.

For any $z = E + i\eta$ we let

$$m_{sc} = m_{sc}(z) = \int_{\mathbb{R}} \frac{\varrho_{sc}(x)\mathrm{d}x}{x - z}$$

be the Stieltjes transform of the Wigner semicircle distribution function whose density is given by

$$\varrho_{sc}(E) = \frac{1}{2\pi}\sqrt{4 - E^2}\mathbf{1}(|E| \le 2) \; . \tag{2.4}$$

Let $B^{(k)}$ denote the $(N-1)\times(N-1)$ minor of H after removing the k-th row and k-th column. Note that $B^{(k)}$ is an $(N-1)\times(N-1)$ Hermitian Wigner matrix with a normalization factor off by $(1-\frac{1}{N})^{1/2}$. Let $\lambda_1^{(k)} \le \lambda_2^{(k)} \le \ldots \le \lambda_{N-1}^{(k)}$ denote its eigenvalues and $\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{N-1}^{(k)}$ the corresponding normalized eigenvectors.

Let $\mathbf{a}^{(k)} = (h_{k,1}, h_{k,2}, \ldots h_{k,k-1}, h_{k,k+1}, \ldots h_{k,N})^* \in \mathbb{C}^{N-1}$, i.e. the k-th column after removing the diagonal element $h_{k,k} = h_{kk}$. Computing the (k,k) diagonal element of the resolvent G_z, we have

$$G_z(k,k) = \frac{1}{h_{kk} - z - \mathbf{a}^{(k)}\cdot(B^{(k)} - z)^{-1}\mathbf{a}^{(k)}} = \Big[h_{kk} - z - \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\xi_\alpha^{(k)}}{\lambda_\alpha^{(k)} - z}\Big]^{-1} \tag{2.5}$$

where we defined

$$\xi_\alpha^{(k)} := \big|\sqrt{N}\mathbf{a}^{(k)}\cdot\mathbf{u}_\alpha^{(k)}\big|^2$$

and note that $\mathbb{E}\,\xi_\alpha^{(k)} = 1$. Thus

$$m(z) = \frac{1}{N}\sum_{k=1}^{N}\left[h_{kk} - z - \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\xi_\alpha^{(k)}}{\lambda_\alpha^{(k)} - z}\right]^{-1} \; . \tag{2.6}$$

Similarly to the definition of $m(z)$ in (2.2), we also define the Stieltjes transform of the density of states of $B^{(k)}$

$$m^{(k)} = m^{(k)}(z) = \frac{1}{N-1}\,\mathrm{Tr}\,\frac{1}{B^{(k)} - z} = \int_{\mathbb{R}}\frac{\mathrm{d}F^{(k)}(x)}{x - z}$$

with the empirical counting function

$$F^{(k)}(x) = \frac{1}{N-1}\big|\,\big\{\alpha\;:\;\lambda_\alpha^{(k)} \le x\big\}\big|.$$

The spectral parameter z is fixed in most of the proofs and we will often omit it from the argument of the Stieltjes transforms. Let $\mathbb{E}_k$ denote the expectation value w.r.t the random vector $\mathbf{a}^{(k)}$. The distribution of $B^{(k)}$, $\mathbf{a}^{(k)}$ and $\xi_\alpha^{(k)}$ does not depend on k, so we will often omit this superscript when it is unnecessary.

For any spectral interval $I \subset \mathbb{R}$, we denote

$$\mathcal{N}_I := \#\{\alpha\;:\;\mu_\alpha \in I\}$$

$$\mathcal{N}_I^{(k)} ;= \#\{\alpha\;:\;\lambda_\alpha^{(k)} \in I\}$$

the number of eigenvalues in I of H and $B^{(k)}$, respectively. When we are interested only in the distribution of $\mathcal{N}_I^{(k)}$, we drop the superscript k, but to avoid confusion with $\mathcal{N}_I$, we denote by $\mathcal{N}_I^\lambda$ a random variable with the common distribution of $\mathcal{N}_I^{(k)}$.

With these notations, the following basic upper bound on $\mathcal{N}_I$ follows immediately:

Proposition 2.1 *Let $I = [E - \eta/2, E + \eta/2]$ be an interval of length η about the spectral point E and let $z = E + i\eta$. Then we have the following estimate on the number of eigenvalues in I:*

$$\mathcal{N}_I \le C\eta \, Im \sum_{k=1}^{N} \left[h_{kk} - z - \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(k)}}{\lambda_\alpha^{(k)} - z} \right]^{-1} . \tag{2.7}$$

Proof. We have

$$\mathcal{N}_I = N \int_I \mathrm{d}F(x) \le \frac{5}{4} N\eta \int_{E-\eta/2}^{E+\eta/2} \frac{\eta \mathrm{d}F(x)}{(x-E)^2 + \eta^2} \le \frac{5}{4} N\eta \, \mathrm{Im}\, m(z)$$

that completes the proof of the first inequality using (2.6). □

3 Main results

The first main result establishes the semicircle law on the optimal scale $\eta \ge O(1/N)$; the proof will be given in Section 6.

Theorem 3.1 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)**. *Let $\kappa > 0$ and fix an energy $E \in [-2+\kappa, 2-\kappa]$. Set $c_0 = \pi \varrho_{sc}(E) > 0$ and let $K = 300/c_0$. Let $z = E + i\eta$ denote the spectral parameter with imaginary part satisfying $\eta \ge K/N$. Then there are constants C and c, depending on κ, such that*

$$\mathbb{P}(|m(z) - m_{sc}(z)| \ge \delta) \le C\, e^{-c\delta\sqrt{N\eta}} \tag{3.1}$$

for any sufficiently small δ (depending on κ) and any sufficiently large $N \ge N_0(\delta)$.

Furthermore, if $\mathcal{N}_{\eta^*}(E) = |\{\alpha \, : \, |\mu_\alpha - E| \le \eta^*/2\}|$ denotes the number of eigenvalues in the interval $[E - \eta^*/2, E + \eta^*/2]$, then for any $\delta > 0$ there is a constant K_δ such that

$$\mathbb{P}\Big\{ \Big| \frac{\mathcal{N}_{\eta^*}(E)}{N\eta^*} - \varrho_{sc}(E) \Big| \ge \delta \Big\} \le C\, e^{-c\delta\sqrt{N\eta^*}} \tag{3.2}$$

holds for all η^* satisfying $K_\delta \le N\eta^*$ and for all sufficiently large $N \ge N_0$ (depending on δ and κ).

As a corollary to this theorem, we can formulate a result on the eigenvectors:

Corollary 3.2 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)**, *then the following hold:*

(i) For any $\kappa > 0$ and $K > 0$ there exist constants $M_0 = M_0(K, \kappa)$, $N_0 = N_0(K, \kappa)$ and $c = c(K, \kappa)$ such that for any interval $I \subset [-2+\kappa, 2-\kappa]$ of length $|I| \le K/N$ we have

$$\mathbb{P}\left\{ \exists \, \mathbf{v} \text{ with } H\mathbf{v} = \mu\mathbf{v}, \, \|\mathbf{v}\| = 1, \, \mu \in I \text{ and } |v_1| \ge \frac{M}{N^{1/2}} \right\} \le e^{-c\sqrt{M}} \tag{3.3}$$

for all $M \ge M_0$ and $N \ge N_0$.

(ii) For any $\kappa > 0$ and $2 \leq p < \infty$ there exist $M_0 = M_0(\kappa, p)$, $N_0 = N_0(\kappa, p)$ and $c = c(\kappa, p) > 0$ such that for any interval $I \subset [-2+\kappa, 2-\kappa]$ of length $|I| = 1/N$ we have

$$\mathbb{P}\Big\{\exists\ \mathbf{v} \text{ with } H\mathbf{v} = \mu\mathbf{v},\ \|\mathbf{v}\| = 1,\ \mu \in I \text{ and } \|\mathbf{v}\|_p \geq MN^{\frac{1}{p}-\frac{1}{2}}\Big\} \leq e^{-c\sqrt{M}} \tag{3.4}$$

for all $M \geq M_0$, all $N \geq N_0$.

(iii) For any $\kappa > 0$ there exist $M_0 = M_0(\kappa)$, $N_0 = N_0(\kappa)$ and $c = c(\kappa)$ such that

$$\mathbb{P}\Big\{\exists\ \mathbf{v} \text{ with } H\mathbf{v} = \mu\mathbf{v},\ \|\mathbf{v}\| = 1,\ \mu \in I \text{ and } \|\mathbf{v}\|_\infty \geq \frac{M}{N^{1/2}}\Big\} \leq e^{-c\sqrt{M}} \tag{3.5}$$

for all $M \geq M_0(\log N)^2$, all $N \geq N_0$.

The second main result is an upper bound on the tail distribution of the eigenvalue gap; the proof is given in Section 7.

Theorem 3.3 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)***. Fix an energy $E \in [-2+\kappa, 2-\kappa]$. Denote by λ_α the largest eigenvalue below E and assume that $\alpha \leq N-1$. Then there are positive constants C and c depending on κ such that*

$$\mathbb{P}\Big(\lambda_{\alpha+1} - E \geq \frac{K}{N},\ \alpha \leq N-1\Big) \leq C\, e^{-c\sqrt{K}} \tag{3.6}$$

for any $N \geq 1$ and any $K \geq 0$.

The third main result is the Wegner estimate for the averaged density of states:

Theorem 3.4 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)** *and condition* **C2)** *with an exponent $a = 5$ in (1.4). Let $\kappa > 0$, choose an energy $|E| < 2-\kappa$ and consider an energy interval $I = [E - \eta/2, E + \eta/2]$ and set $\varepsilon = N\eta$, with $\varepsilon > 0$. Let $\mathcal{N}_I$ be the number of eigenvalues in I and assume $N \geq 10$. Then*

$$\mathbb{P}(\mathcal{N}_I \geq 1) \leq \mathbb{E}\,\mathcal{N}_I^2 \leq C\,\varepsilon \tag{3.7}$$

uniformly in N and E. In particular,

$$\sup_{I \subset [-2+\kappa, 2-\kappa]}\ \sup_{N \geq 10}\ \mathbb{E}\Big[\frac{\mathcal{N}_I}{N|I|}\Big] \leq C\,, \tag{3.8}$$

and therefore the averaged density of states, $\mathbb{E}\,\varrho_N(E)$, is an absolutely continuous measure with a uniformly bounded density, i.e.

$$\sup_{|E| \leq 2-\kappa}\ \sup_{N \geq 10}\ \mathbb{E}\,\varrho_N(E) \leq C \tag{3.9}$$

(with a slight abuse of notations, $\mathbb{E}\,\varrho_N(E)$ denotes the measure and its density as well). The constant C in (3.7), (3.8) and (3.9) depends only on κ and on the constants characterizing the distribution $\mathrm{d}\nu$ via the conditions **C1)**–**C2)***. The estimate (3.7) holds for $N \leq 10$ as well if, instead of* **C1)** *and* **C2)***, we assume that the density function, (const.) $\exp(-\widetilde{g})$, of the diagonal matrix elements satisfies $\int_{\mathbb{R}} |\widetilde{g}'(x)| \exp(-\widetilde{g}(x))\mathrm{d}x < \infty$.*

Finally, the following theorem establishes an upper bound on the level repulsion.

Theorem 3.5 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)**. *Let $\kappa > 0$, fix an energy E with $|E| < 2-\kappa$, set $\eta = \varepsilon/N$ and let $\mathcal{N}_\eta$ be the number of eigenvalues in $I_\eta = [E-\eta/2, E+\eta/2]$. Fix $k \in \mathbb{N}$, and assume that condition* **C2**) *holds with $a = k^2+5$. Then, there exists a constant $C > 0$, depending on k and κ, such that*

$$\mathbb{P}(\mathcal{N}_\eta \geq k) \leq C\, \varepsilon^{k^2} \tag{3.10}$$

for all $\varepsilon > 0$ and uniformly for all $N \geq N_0(k)$.

The common starting point of all proofs is Proposition 2.1. Using the estimate $\mathrm{Im}(a+bi)^{-1} \leq (a^2+b^2)^{-1/2}$ on the right hand side of (2.7), we have

$$\mathcal{N}_I \leq C\eta \sum_{k=1}^{N} \frac{1}{(a_k^2+b_k^2)^{1/2}} \tag{3.11}$$

with

$$a_k := \eta + \frac{1}{N}\sum_{\alpha=1}^{N-1} \frac{\eta \xi_\alpha^{(k)}}{(\lambda_\alpha^{(k)}-E)^2+\eta^2}, \qquad b_k := h_{kk} - E - \frac{1}{N}\sum_{\alpha=1}^{N-1} \frac{(\lambda_\alpha^{(k)}-E)\xi_\alpha^{(k)}}{(\lambda_\alpha^{(k)}-E)^2+\eta^2}\,,$$

where a_k and b_k are the imaginary and real part, respectively, of the reciprocal of the summands in (2.7). Theorems 3.1 and 3.3 rely only on the imaginary part, i.e. b_k in (3.11) will be neglected. In the proofs of Theorems 3.4 and 3.5, however, we make an essential use of b_k as well. Since typically $1/N \lesssim |\lambda_\alpha^{(k)} - E|$, we note that a_k^2 is much smaller than b_k^2 if $\eta \ll 1/N$ and this is the relevant regime for the Wegner estimate and for the level repulsion. Assuming a certain smoothness condition on the distribution $\mathrm{d}\nu$ (condition **C2)**), the distribution of the variables $\xi_\alpha^{(k)}$ will also be smooth. Although $\xi_\alpha^{(k)}$ are not independent for different α's, they are sufficiently decorrelated so that the distribution of b_k inherits some smoothness which will make the expectation value $(a_k^2+b_k^2)^{-p/2}$ finite for certain $p > 0$. This will give a bound on the p-th moment on $\mathcal{N}_I$ which will imply (3.7) and (3.10).

4 Semicircle law and delocalization on intermediate scales

In this section we review the proof of the convergence to the semicircle law on intermediate energy scales of the order $\eta \geq (\log N)^4/N$. This convergence has already been established in our previous work [6] but with a speed of convergence uniform in η, for $\eta \geq (\log N)^8/N$. Our new estimate shows that the speed of convergence becomes faster as η increases (and we also reduce the power of the logarithm from 8 to 4). Moreover, we show that the results hold under the condition **C1)** only. Thus we obtain a stronger version of our earlier results under weaker assumptions.

The following result is an analogue of Theorem 3.1 for intermediate scales. It states that the density of states regularized on any scale $\eta \geq N^{-1}(\log N)^4$ converges to the Wigner semicircle law in probability uniformly for all energies away from the spectral edges. Note, however, that the estimate for larger scales is sufficiently strong so that uniformity in the spectral parameter z can be obtained which is not expected for short scales.

Theorem 4.1 *Let H be an $N \times N$ hermitian Wigner matrix satisfying the condition* **C1)***. Let the energy scale η be chosen such that $(\log N)^4/N \leq \eta \leq 1$. Then for any $\kappa > 0$ there exists a constant $c = c(\kappa)$ such that the Stieltjes transform $m_N(z)$ (see (2.2)) of the empirical eigenvalue distribution of the $N \times N$ Wigner matrix satisfies*

$$\mathbb{P}\Big\{ \sup_{Re\, z \in [-2+\kappa, 2-\kappa]} |m_N(z) - m_{sc}(z)| \geq \delta \Big\} \leq e^{-c\delta\sqrt{N\eta}} \tag{4.1}$$

for any $\delta > 0$ if $N \geq N_0(\kappa, \delta)$ is large enough.

Furthermore, on the scale η^* with $(\log N)^4/N \ll \eta^* \ll 1$ we have the convergence of the counting function as well. More precisely, let $\mathcal{N}_{\eta^*}(E) = |\{\alpha \ : \ |\mu_\alpha - E| \leq \eta^*/2\}|$ denote the number of eigenvalues in the interval $[E - \eta^*/2, E + \eta^*/2]$. Then, for any $\delta > 0$ there is a constant K_δ such that

$$\mathbb{P}\Big\{ \sup_{|E| \leq 2-\kappa} \Big| \frac{\mathcal{N}_{\eta^*}(E)}{N\eta^*} - \varrho_{sc}(E) \Big| \geq \delta \Big\} \leq e^{-c\delta\sqrt{N\eta^*}} \tag{4.2}$$

holds for all η^* satisfying $K_\delta (\log N)^4/N \leq \eta^* \leq 1/K_\delta$ and for all sufficiently large $N \geq N_0$ (depending on δ and κ).

Proof. This theorem is proven exactly as Theorem 1.1 in [6] after replacing the key Lemma 2.1 of [6] by the following Lemma 4.2. M. Ledoux has informed us that Lemma 2.1 of [6] follows from a result of Hanson and Wright [8]. We will reproduce his argument in the proof of Proposition 4.3. This requires only Proposition 4.5 below, which is a mild extension of the Hanson-Wright theorem to the complex case.

Lemma 4.2 *Let $E \in [-2+\kappa, 2-\kappa]$. Suppose that $\mathbf{v}_\alpha$ and λ_α are eigenvectors and eigenvalues of an $N \times N$ random hermitian matrix B with a law satisfying the assumption of Theorem 4.1. Let*

$$X = \frac{1}{N} \sum_\alpha \frac{\xi_\alpha - 1}{\lambda_\alpha - z}$$

with $z = E + i\eta$, $\xi_\alpha = |\mathbf{b} \cdot \mathbf{v}_\alpha|^2$, where the components of $\mathbf{b}$ are i.i.d. random variables, independent of B and satisfying the condition **C1)***. Then there exists a positive constant c (depending on κ) so that for every $\delta > 0$, we have*

$$\mathbb{P}[|X| \geq \delta] \leq 5\, e^{-c \min\{\delta\sqrt{N\eta},\, \delta^2 N\eta\}} \tag{4.3}$$

if $N\eta \geq (\log N)^2$ and N is sufficiently large.

For simplicity, we formulated the lemma for $N \times N$ matrices, but it will be applied for the $(N-1) \times (N-1)$ minors of H.

Proof of Lemma 4.2. Define the intervals $I_n = [E - 2^{n-1}\eta, E + 2^{n-1}\eta]$ and let M and K_0 be sufficiently large fixed numbers. We have $[-K_0, K_0] \subset I_{n_0}$ with $n_0 = C\log(K_0/\eta) \leq C\log(NK_0)$. Denote by Ω the event

$$\Omega := \Omega(M, K_0) = \Big\{ \max_n \frac{\mathcal{N}_{I_n}}{N|I_n|} \geq M \Big\} \cup \{\max_\alpha |\lambda_\alpha| \geq K_0\}\ , \tag{4.4}$$

where $\mathcal{N}_{I_n} = |\{\alpha \ : \ \lambda_\alpha \in I_n\}|$ is the number of eigenvalues in the interval I_n. Therefore, if $\mathbb{P}_\mathbf{b}$ denotes the probability w.r.t. the variable $\mathbf{b}$, we find

$$\mathbb{P}[|X| \geq \delta] \leq \mathbb{E}\Big[\mathbf{1}_{\Omega^c} \cdot \mathbb{P}_\mathbf{b}\big[|X| \geq \delta\big]\Big] + \mathbb{P}(\Omega)\ .$$

We will prove below the following two propositions which complete the proof of Lemma 4.2. □

Proposition 4.3 *Assume condition* **C1)**. *Let* $\Omega = \Omega(M, K_0)$ *be given by* (4.4) *and let* $\eta \geq 1/N$. *Then for sufficiently large and fixed* M, K_0 *there is a positive* $c = c(M, K_0)$ *such that for any* $\delta > 0$

$$\mathbb{E}\Big[\mathbf{1}_{\Omega^c} \cdot \mathbb{P}_{\mathbf{b}}\big[|X| \geq \delta\big]\Big] \leq 4\, e^{-c \min\{\delta\sqrt{N\eta},\, \delta^2 N\eta\}} .$$

Proposition 4.4 *Assume condition* **C1)**. *Let* η *be chosen such that* $(\log N)^2/N \leq \eta \leq 1$. *Then for sufficiently large and fixed* M *and* K_0 *there is a positive constant* c *such that*

$$\mathbb{P}\big[\Omega(M, K_0)\big] \leq e^{-c\sqrt{MN\eta}} . \tag{4.5}$$

for all N *sufficiently large.*

Both results are based on a theorem of Hanson and Wright [8], extended to non-symmetric variables by Wright [16]. The result was formulated for real valued random variables. We do not know if their theorems hold for general complex random variables, but they hold true in two special cases, namely when either the real and imaginary parts of b_j are i.i.d. or if the distribution of b_j is rotationally symmetric (see (1.2)). We formulate this easy extension of their result and we give the proof in the Appendix.

Proposition 4.5 *Let* b_j, $j = 1, 2, \ldots N$ *be a sequence of complex i.i.d. random variables with distribution* $\mathrm{d}\nu$ *satisfying the Gaussian decay* (1.3) *for some* $\delta_0 > 0$. *Suppose that condition* (1.2) *holds, i.e. either both the real and imaginary parts are i.i.d. or the distribution* $\mathrm{d}\nu$ *is rotationally symmetric. Let* a_{jk}, $j, k = 1, 2, \ldots N$ *be arbitrary complex numbers and let* $\mathcal{A}$ *be the* $N \times N$ *matrix with entries* $\mathcal{A}_{jk} := |a_{jk}|$. *Define*

$$X = \sum_{j,k=1}^{N} a_{jk}\big[b_j \overline{b}_k - \mathbb{E} b_j \overline{b}_k\big] .$$

Then there exists a constant $c > 0$, *depending only on* δ_0, D *from* (1.3), *such that for any* $\delta > 0$

$$\mathbb{P}(|X| \geq \delta) \leq 4 \exp\big(-c \min\{\delta/A,\ \delta^2/A^2\}\big) ,$$

where $A := (Tr\, \mathcal{A}\mathcal{A}^t)^{1/2} = \big[\sum_{j,k} |a_{jk}|^2\big]^{1/2}$.

Proof of Proposition 4.3. Write X in the form

$$X = \sum_{j,k=1}^{N} a_{jk}\big[b_j \overline{b_k} - \mathbb{E} b_j \overline{b_k}\big] ,$$

where

$$a_{jk} = \frac{1}{N} \sum_{\alpha} \frac{\overline{u_\alpha(j)} u_\alpha(k)}{\lambda_\alpha - z} .$$

We have

$$A^2 := \sum_{j,k=1}^{N} |a_{jk}|^2 = \frac{1}{N^2} \sum_{\alpha} \frac{1}{|\lambda_\alpha - z|^2} .$$

On the set Ω^c we have

$$A^2 = \frac{1}{N^2} \sum_{n=0}^{n_0} \sum_{\lambda_\alpha \in I_n \setminus I_{n-1}} \frac{1}{|\lambda_\alpha - z|^2} \leq \frac{1}{N^2} \sum_{n=0}^{n_0} \frac{\mathcal{N}_{I_n}}{(2^n \eta)^2} \leq \frac{2M}{N\eta} \tag{4.6}$$

where we estimated the number of eigenvalues in $I_n \setminus I_{n-1}$ by $\mathcal{N}_{I_n}$ and we set $I_{-1} := \emptyset$. Using Proposition 4.5 we obtain that

$$\mathbb{E}\Big[\mathbf{1}_{\Omega^c} \cdot \mathbb{P}_{\mathbf{b}}[|X| \geq \delta]\Big] \leq 4 \exp\big(-c \min\{\delta\sqrt{N\eta},\, \delta^2 N\eta\}\big)$$

where the constant c depends on M and on δ_0, D from (1.3). This completes the proof of Proposition 4.3. $\square$

Remark. The same result can be proven by assuming that the distribution dν satisfies the logarithmic Sobolev inequality, see Lemma 2.1 of [6]; the bound $\exp(-c\delta(\log N)^2)$ obtained there can be easily improved to $C\exp(-c\delta\sqrt{N\eta})$ since the exceptional set Ω is defined differently.

Proof of Proposition 4.4. Under condition **C1)**, we showed in Lemma 7.4 of [5] that

$$\mathbb{P}\{\max_\alpha |\lambda_\alpha| \geq K_0\} \leq e^{-cK_0^2 N} \tag{4.7}$$

for sufficiently large K_0. To estimate the large deviation of $\mathcal{N}_{I_n}$, we use the following weaker version of Theorem 2.1 of [5]:

Theorem 4.6 *Assume condition* **C1)**. *Let $I \subset \mathbb{R}$ be an interval with length $|I| \geq (\log N)/N$. Then there is a positive constant $c > 0$ such that for any K large enough*

$$\mathbb{P}\big\{\mathcal{N}_I \geq KN|I|\big\} \leq e^{-c\sqrt{KN|I|}}\,. \tag{4.8}$$

Combining (4.7) and (4.8) and recalling $N\eta \geq (\log N)^2$, we have

$$\mathbb{P}(\Omega) \leq C\log(NK_0)e^{-c\sqrt{MN\eta}} + e^{-cK_0^2 N} \leq e^{-\widetilde{c}\sqrt{MN\eta}}$$

completing the proof of Proposition 4.4. $\square$

Proof of Theorem 4.6. The proof is the same as the proof of Theorem 2.1 in [5] but in the estimate (2.20) at the end of the proof we use the following lemma instead of Corollary 2.4 to Lemma 2.3 [5]:

Lemma 4.7 *Assume condition* **C1)**. *Let the components of the vector $\mathbf{b} \in \mathbb{C}^{N-1}$ be complex i.i.d. variables with a common distribution* dν *and let $\xi_\alpha = |\mathbf{b}\cdot\mathbf{v}_\alpha|^2$, where $\{\mathbf{v}_\alpha\}_{\alpha\in\mathcal{I}}$ is an orthonormal set in $\mathbb{C}^{N-1}$. Then for $\delta \leq 1/2$ there is a constant $c > 0$ such that*

$$\mathbb{P}\Big\{\sum_{\alpha\in\mathcal{I}} \xi_\alpha \leq \delta m\Big\} \leq e^{-c\sqrt{m}} \tag{4.9}$$

holds for any $\mathcal{I}$, where $m = |\mathcal{I}|$ is the cardinality of the index set $\mathcal{I}$.

We remark that a stronger bound of the form e^{-cm} was proven in Lemma 2.3 [5] under the condition that Hess g is bounded and in the special case when $g(x,y)$ was in the form $g(x)+g(y)$. An alternative proof under the condition that the support of dν is compact is due to J. Bourgain and it is reproduced in the Appendix of [6]. Using the stronger e^{-cm} bound in (4.9), the bound in (4.8) can be improved to $e^{-cKN|I|}$. Here we present a proof that gives the weaker bound but it uses no additional assumption apart from **C1)** and (1.2).

Proof of Lemma 4.7. Let

$$X := \sum_{i,j=1}^{N} a_{ij}\big[b_i\overline{b}_j - \mathbb{E}\, b_i\overline{b}_j\big], \qquad \text{with} \qquad a_{ij} := \sum_{\alpha\in\mathcal{I}} \overline{v}_\alpha(i)v_\alpha(j)\,.$$

Notice that $\sum_{\alpha\in\mathcal{I}}\xi_\alpha = X + |\mathcal{I}| = X+m$ since $\mathbb{E}\,\xi_\alpha = 1$. By $\delta\le 1/2$ we therefore obtain

$$\mathbb{P}\big\{\sum_{\alpha\in\mathcal{I}}\xi_\alpha \le \delta m\big\} \le \mathbb{P}\big\{|X|\ge \frac{m}{2}\big\}\,.$$

Since

$$A^2 := \sum_{i,j=1}^{N}|a_{ij}|^2 = \sum_{\alpha,\beta\in\mathcal{I}}\sum_{i,j=1}^{N}\overline{v}_\alpha(i)v_\alpha(j)v_\beta(i)\overline{v}_\beta(j) = m\,,$$

by Proposition 4.5, we obtain

$$\mathbb{P}\big\{\sum_{\alpha\in\mathcal{I}}\xi_\alpha \le \delta m\big\} \le \mathbb{P}\big\{|X|\ge \frac{m}{2}\big\} \;\le 4\exp\Big(-c\min\big\{\frac{m}{2A},\frac{m^2}{4A^2}\big\}\Big) \le e^{-c\sqrt{m}}.$$

for some $c>0$. $\square$

Using Theorem 4.1, we can prove delocalization of the eigenvectors of H. In Theorem 1.2 of [6] we proved that $\|\mathbf{v}\|_\infty \le (\log N)^{9/2}/N^{1/2}$ holds for all eigenvectors with probability bigger than $1-e^{-c(\log N)^2}$. The following theorem is a generalization of this result using the stronger estimates from Theorem 4.1.

Theorem 4.8 *Let H be an $N\times N$ hermitian Wigner matrix satisfying the condition* **C1)**. *For any $\kappa>0$ there exists $c>0$, depending on κ such that*

$$\mathbb{P}\left\{\exists\ \mathbf{v}\ \textit{with}\ H\mathbf{v}=\mu\mathbf{v},\ \|\mathbf{v}\|=1,\ \mu\in[-2+\kappa,2-\kappa]\ \textit{and}\ \|\mathbf{v}\|_\infty \ge \frac{M}{N^{1/2}}\right\} \le e^{-c\sqrt{M}}$$

for all $M\ge(\log N)^4$ and all $N\ge N_0(\kappa)$ large enough.

Proof. Let $\eta^* = M/N$ and partition the interval $[-2+\kappa,2-\kappa]$ into $n_1 = O(1/\eta^*)\le O(N)$ intervals $I_1,I_2,\ldots,I_{n_1}$ of length η^*. As before, let $\mathcal{N}_I = |\{\beta\ :\ \mu_\beta\in I\}|$ denote the number of eigenvalues in I. Let

$$c_1 := \varrho_{sc}(2-\kappa) = \min\big\{\varrho_{sc}(E)\ :\ E\in[-2+\kappa,2-\kappa]\big\} > 0\,.$$

By using (4.2) in Theorem 4.1 and the fact that $N\eta^*\ge(\log N)^4$, we have

$$\mathbb{P}\left\{\max_n \mathcal{N}_{I_n} \le \frac{c_1}{2}\,N\eta^*\right\} \le CNe^{-c\sqrt{N\eta^*}} \le e^{-\widetilde{c}\sqrt{N\eta^*}}\,. \tag{4.10}$$

Suppose that $\mu\in I_n$, and that $H\mathbf{v}=\mu\mathbf{v}$. Consider the decomposition

$$H = \begin{pmatrix} h & \mathbf{a}^* \\ \mathbf{a} & B\end{pmatrix} \tag{4.11}$$

where $\mathbf{a} = (h_{1,2}, \ldots h_{1,N})^*$ and B is the $(N-1) \times (N-1)$ matrix obtained by removing the first row and first column from H. Let λ_α and $\mathbf{u}_\alpha$ (for $\alpha = 1, 2, \ldots, N-1$) denote the eigenvalues and the normalized eigenvectors of B. Similarly to [6], from the eigenvalue equation $H\mathbf{v} = \mu\mathbf{v}$ and from (4.11) we find for the first component of $\mathbf{v} = (v_1, v_2, \ldots, v_N)$ that

$$|v_1|^2 = \frac{1}{1 + \mathbf{a} \cdot (\mu - B)^{-2}\mathbf{a}} = \frac{1}{1 + \frac{1}{N}\sum_\alpha \frac{\xi_\alpha}{(\mu - \lambda_\alpha)^2}} \leq \frac{4N[\eta^*]^2}{\sum_{\lambda_\alpha \in I_n} \xi_\alpha}, \tag{4.12}$$

where in the second equality we set $\xi_\alpha = |\sqrt{N}\mathbf{a} \cdot \mathbf{u}_\alpha|^2$ and used the spectral representation of B. We recall that the eigenvalues of H, $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_N$, and the eigenvalues of B are interlaced: $\mu_1 \leq \lambda_1 \leq \mu_2 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1} \leq \mu_N$ and the inequalities are strict with probability one (see Lemma 2.5 of [5]). This means that there exist at least $\mathcal{N}_{I_n} - 1$ eigenvalues of B in I_n. Therefore, using that the components of any eigenvector are identically distributed, we have

$$\begin{aligned}
&\mathbb{P}\Big(\exists\ \mathbf{v} \text{ with } H\mathbf{v} = \mu\mathbf{v},\ \|\mathbf{v}\| = 1,\ \mu \in [-2+\kappa, 2-\kappa] \text{ and } \|\mathbf{v}\|_\infty \geq \frac{M}{N^{1/2}}\Big) \\
&\quad \leq N n_1 \sup_n \mathbb{P}\Big(\exists\ \mathbf{v} \text{ with } H\mathbf{v} = \mu\mathbf{v},\ \|\mathbf{v}\| = 1,\ \mu \in I_n \text{ and } |v_1|^2 \geq \frac{M^2}{N}\Big) \\
&\quad \leq C\, N^2 \sup_n \mathbb{P}\left(\sum_{\lambda_\alpha \in I_n} \xi_\alpha \leq 4\right) \\
&\quad \leq C\, N^2 \sup_n \mathbb{P}\left(\sum_{\lambda_\alpha \in I_n} \xi_\alpha \leq 4 \text{ and } \mathcal{N}_{I_n} \geq \frac{c_1}{2}\, N\eta^*\right) + C\, N^2 \sup_n \mathbb{P}\left(\mathcal{N}_{I_n} \leq \frac{c_1}{2}\, N\eta^*\right) \\
&\quad \leq C\, N^2 e^{-\widetilde{c}\sqrt{N\eta^*}} + C\, N^2 e^{-\widetilde{c}\sqrt{N\eta^*}} \\
&\quad \leq e^{-c\sqrt{M}},
\end{aligned} \tag{4.13}$$

for a sufficiently small $\widetilde{c} > 0$ (we also used that $N\eta^* = M \geq (\log N)^4$ in the last step). Here we used Lemma 4.7 to estimate the first probability in the fourth line of (4.13) and (4.10) to estimate the second one. □

5 Upper bound for the density on short scales

Now we start our analysis on short scales $\eta \geq 1/N$. As before, we always assume condition **C1)** in addition to (1.2). We first show a large deviation upper bound on the number of eigenvalues on short scales about a fixed energy E away from the spectral edges. This complements the estimate in Theorem 4.6 that was valid for larger scales.

Theorem 5.1 *Let $\kappa > 0$ and fix an energy $E \in [-2+\kappa, 2-\kappa]$. Let $\eta > 0$ with $1 \leq N\eta \leq CN^{1/10}$. Let*

$$\mathcal{N} := \#\{\alpha\ :\ \mu_\alpha \in I_\eta := [E - \eta/2, E + \eta/2]\}\ .$$

Then for any $2 \leq M \leq \frac{CN^{1/10}}{N\eta}$, we have

$$\mathbb{P}\Big(\frac{\mathcal{N}}{N\eta} \geq M\Big) \leq \Big(\frac{C}{M}\Big)^{\frac{1}{8}MN\eta} \tag{5.1}$$

and for any $1 \le p \le CN^{1/20}$
$$\mathbb{E}\Big[\frac{\mathcal{N}}{N\eta}\Big]^p \le C^p\Big(1+\frac{p}{N\eta}\Big)^p . \tag{5.2}$$
All constants depend on κ.

Proof of Theorem 5.1. It is sufficent to prove (5.1), since (5.2) easily follows from it and from (4.8):
$$\begin{aligned}\mathbb{E}\Big[\frac{\mathcal{N}}{N\eta}\Big]^p &\le 2^p + p\int_2^\infty M^{p-1}\mathbb{P}\Big(\frac{\mathcal{N}}{N\eta}\ge M\Big)\mathrm{d}M \\ &\le 2^p + p\int_2^\Lambda M^{p-1-\frac{1}{16}MN\eta}\,\mathrm{d}M + \int_\Lambda^\infty M^{p-1-c\sqrt{MN\eta}}\,\mathrm{d}M \\ &\le C^p\Big(1+\frac{p}{N\eta}\Big)^p\end{aligned} \tag{5.3}$$
with $\Lambda = CN^{1/10}/N\eta$ and with a sufficiently large constant C. To prove (5.1), we use (2.7) to obtain
$$\frac{\mathcal{N}_I}{N\eta} \le C\sum_{k=1}^N \frac{1}{\eta + \frac{1}{N}\sum_\alpha \frac{\eta\xi_\alpha^{(k)}}{(\lambda_\alpha^{(k)}-E)^2+\eta^2}} \le \frac{C}{N}\sum_{k=1}^N \frac{1}{\eta+\frac{Z^{(k)}}{N\eta}} , \tag{5.4}$$
where we defined
$$Z^{(k)} := Z^{(k)}(\eta) = \sum_{\alpha:\lambda_\alpha^{(k)}\in I_\eta} \xi_\alpha^{(k)} . \tag{5.5}$$
To estimate the large deviation of $Z^{(k)}$, we will later prove the following lemma:

Lemma 5.2 *Let* ν,β *be nonnegative numbers such that* $\nu + 7\beta < 1$*. Then for any* $\delta \ge N^{-\nu}$ *and* $m \le N^\beta$*, we have*
$$\mathbb{P}\Big\{\frac{1}{m}\sum_{\alpha=1}^m \xi_\alpha \le \delta\Big\} \le (C\delta)^m \tag{5.6}$$
with a constant C *depending on* ν *and* β.

Note that the estimate in this lemma is more precise than (4.9), but the stronger estimate is valid only if m is not too large. In the proof we will use information about the eigenfunctions obtained in Theorem 4.8.

Let
$$\mathcal{N}^{(k)} := \mathcal{N}_{I_\eta}^{(k)} = \#\{\alpha : \lambda_\alpha^{(k)} \in I_\eta\}$$
denote the number of eigenvalues of the minor $B^{(k)}$ in the interval I_η (see Section 2 for the definitions). By the interlacing property of the eigenvalues, $\mathcal{N} \ge MN\eta$ implies $\mathcal{N}^{(k)} \ge MN\eta - 1 \ge \frac{1}{2}MN\eta \ge N\eta$ for any k

(since $M \geq 2$ thus $MN\eta \geq 2$). Therefore, from (5.4) we have for any $q \geq 1$ that

$$
\begin{aligned}
\mathbb{P}\Big(\frac{\mathcal{N}}{N\eta} \geq M\Big) &\leq \mathbb{P}\Big(\frac{C}{N}\sum_k \frac{\mathbf{1}(\mathcal{N}^{(k)} \geq \frac{1}{2}MN\eta)}{\frac{Z^{(k)}}{N\eta} + \eta} \geq M\Big) \\
&\leq \Big(\frac{C}{M}\Big)^q \mathbb{E}\bigg[\frac{\mathbf{1}(\mathcal{N}^{(1)} \geq \frac{1}{2}MN\eta)}{\frac{Z^{(1)}}{N\eta} + \eta}\bigg]^q \\
&\leq \Big(\frac{C}{M}\Big)^q \int_0^\infty \mathbb{P}\bigg[\mathcal{N}^{(1)} \geq \frac{1}{2}MN\eta,\ \frac{Z^{(1)}}{N\eta} + \eta \leq t^{-1/q}\bigg]\mathrm{d}t \\
&\leq \Big(\frac{C}{M}\Big)^q + \Big(\frac{C}{M}\Big)^q \int_1^{(1/\eta)^q} \mathbb{P}\bigg(\sum_{\alpha=1}^{MN\eta/2} \xi_\alpha^{(1)} \leq N\eta t^{-1/q}\bigg)\mathrm{d}t \\
&\leq \Big(\frac{C}{M}\Big)^q + \Big(\frac{C}{M}\Big)^q \int_1^{(1/\eta)^q} \big[C\max\{t^{-1/q}, N^{-\nu}\}\big]^{\frac{1}{2}MN\eta}\mathrm{d}t \\
&\leq \Big(\frac{C}{M}\Big)^{\frac{1}{8}MN\eta}
\end{aligned}
\tag{5.7}
$$

if we use Lemma 5.2 with e.g. $\nu = 1/4$ (noticing that $\frac{1}{2}MN\eta \leq N^\beta$ with $\beta = 1/10$) and we choose $q = \frac{1}{8}MN\eta$ in the last line (we use that $[\eta^{-1}N^{-4\nu}]^q \leq 1$). $\square$

Proof of Lemma 5.2. We will present the proof under the first condition in (1.2); the proof under the second condition is analogous. With the notation $\mathbf{b} = \sqrt{N}\mathbf{a}$, the components of $\mathbf{b}$ can thus be written as $b_j = x_j + iy_j$ where x_j, y_j are i.i.d. random variables with expectation zero and variance 1/2. Similarly we decompose the eigenvectors into real and imaginary parts, i.e. we write $\mathbf{u}_\alpha = \mathbf{v}_\alpha + i\mathbf{w}_\alpha$ and we have

$$
\xi_\alpha = |\mathbf{b}\cdot\mathbf{u}_\alpha|^2 = \Big(\sum_{j=1}^N (x_j v_\alpha(j) + y_j w_\alpha(j))\Big)^2 + \Big(\sum_{j=1}^N (x_j w_\alpha(j) - y_j v_\alpha(j))\Big)^2 .
$$

The probability and expectation w.r.t. $\mathbf{b}$ are denoted by $\mathbb{P}_{\mathbf{b}}$ and $\mathbb{E}_{\mathbf{b}}$. We define the event

$$
\Omega := \Big\{\|\mathbf{u}_\alpha\|_\infty \leq CN^{2\beta-1/2}(\log N)^4\ :\ \alpha = 1, 2, \ldots, m\Big\} ,
$$

where $\mathbf{u}_\alpha$ are the eigenvectors of $B = B^{(1)}$. Note that Ω is independent of the vector $\mathbf{b} = \sqrt{N}\mathbf{a}^{(1)}$, thus

$$
\mathbb{P}\{\sum_{\alpha=1}^m \xi_\alpha \leq m\delta\} \leq \mathbb{P}(\Omega^c) + \mathbb{E}\Big[\mathbf{1}(\Omega)\mathbb{P}_{\mathbf{b}}\big(\sum_{\alpha=1}^m \xi_\alpha \leq m\delta\big)\Big] .
$$

By Theorem 4.8,

$$
\mathbb{P}(\Omega^c) \leq e^{-cN^\beta(\log N)^2} \leq (C\delta)^m .
$$

On the event Ω, the probability $\mathbb{P}_{\mathbf{b}}\big(\sum_{\alpha=1}^m \xi_\alpha \le m\delta\big)$ will be estimated as follows, where we introduced $t := \delta^{-1} \le N^\nu$:

$$
\begin{aligned}
\mathbb{P}_{\mathbf{b}}\Big\{\sum_{\alpha=1}^m \xi_\alpha \le m\delta\Big\} &\le e^m \mathbb{E}_{\mathbf{b}} e^{-t\sum_{\alpha=1}^m \xi_\alpha} \\
&= e^m \mathbb{E}_{\mathbf{b}} \prod_{\alpha=1}^m e^{-t|\mathbf{b}\cdot\mathbf{u}_\alpha|^2} \\
&= e^m \mathbb{E}_{\mathbf{b}} \prod_{\alpha=1}^m \int_{\mathbb{R}^2} \frac{\mathrm{d}\tau_\alpha \mathrm{d}s_\alpha}{\pi} e^{-i\sqrt{t}\big[\tau_\alpha\sum_j (x_j v_\alpha(j)+y_j w_\alpha(j)) + s_\alpha \sum_j (x_j w_\alpha(j) - y_j v_\alpha(j))\big] - \tau_\alpha^2/4 - s_\alpha^2/4} \\
&= e^m \Big(\prod_{\alpha=1}^m \int_{\mathbb{R}^2} e^{-\frac14(\tau_\alpha^2+s_\alpha^2)} \frac{\mathrm{d}\tau_\alpha \mathrm{d}s_\alpha}{\pi}\Big) \\
&\qquad \times \prod_{j=1}^N \mathbb{E}_{x_j}\mathbb{E}_{y_j} e^{-i\sqrt{t}\sum_\alpha \Big[x_j(\tau_\alpha v_\alpha(j)+s_\alpha w_\alpha(j)) - y_j(s_\alpha v_\alpha(j)-\tau_\alpha w_\alpha(j))\Big]} \\
&\le e^m \Big(\prod_{\alpha=1}^m \int_{\mathbb{R}^2} e^{-\frac14(\tau_\alpha^2+s_\alpha^2)} \mathbf{1}\big(|\tau_\alpha|+|s_\alpha| \le N^{\beta/2}\log N\big) \frac{\mathrm{d}\tau_\alpha \mathrm{d}s_\alpha}{\pi}\Big) \\
&\qquad \times \prod_{j=1}^N \left(1 - \frac{t}{8}\Big[\big(\sum_\alpha(\tau_\alpha v_\alpha(j)+s_\alpha w_\alpha(j))\big)^2 + \big(\sum_\alpha(s_\alpha v_\alpha(j)-\tau_\alpha w_\alpha(j))\big)^2\Big]\right) \\
&\qquad\quad + m(Ce)^m e^{-cN^\beta(\log N)^2}\ .
\end{aligned}
\tag{5.8}
$$

The last term comes from the Gaussian tail of the restriction $|\tau_\alpha|, |s_\alpha| \le N^{\beta/2}\log N$ for all α. In estimate (5.8) we have used that $\Big|\mathbb{E}\big[e^{iY} - 1 - iY + \frac12 Y^2\big]\Big| \le \mathbb{E}|Y^3|$, thus for any real random variable Y with $\mathbb{E}\, Y = 0$ and $|Y| \le \frac14$ we have

$$
|\mathbb{E}\, e^{iY}| \le 1 - \frac12 \mathbb{E}\, Y^2 + \mathbb{E}|Y^3| \le 1 - \frac14 \mathbb{E}\, Y^2\ .
$$

We applied this to

$$
Y = Y_j = -\sqrt{t}\sum_{\alpha=1}^m \Big[x_j(\tau_\alpha v_\alpha(j)+s_\alpha w_\alpha(j)) + y_j(s_\alpha v_\alpha(j)-\tau_\alpha w_\alpha(j))\Big]
$$

with

$$
\mathbb{E}\, Y_j^2 = \frac{t}{2}\Big[\big(\sum_\alpha(\tau_\alpha v_\alpha(j)+s_\alpha w_\alpha(j))\big)^2 + \big(\sum_\alpha(s_\alpha v_\alpha(j)-\tau_\alpha w_\alpha(j))\big)^2\Big]
$$

and we also used

$$
|Y_j| \le \ Cm\sqrt{t}\, N^{5\beta/2-1/2}(\log N)^5 \le \ CN^{(\nu+7\beta-1)/2}(\log N)^5 \le \frac14 \tag{5.9}
$$

on the event Ω and in the regime where $|\tau_\alpha|, |s_\alpha| \le N^{\beta/2}\log N$ for all α.

Reexponentiating $1 - \frac{1}{4}\mathbb{E}\, Y_j^2 \le \exp(-\frac{1}{8}\mathbb{E} Y_j^2)$ and using that by the orthogonality of $\mathbf{u}_\alpha$ we have,

$$\sum_{j=1}^N \Big[\big(\sum_{\alpha=1}^m (\tau_\alpha v_\alpha(j) + s_\alpha w_\alpha(j)) \big)^2 + \big(\sum_{\alpha=1}^m (s_\alpha v_\alpha(j) - \tau_\alpha w_\alpha(j)) \big)^2 \Big] = \sum_{\alpha=1}^m (\tau_\alpha^2 + s_\alpha^2)$$

and we obtain (with $t = \delta^{-1}$)

$$\begin{aligned} \mathbb{P}_{\mathbf{b}} \Big\{ \sum_{\alpha=1}^m \xi_\alpha \le m\delta \Big\} &\le e^m \Big(\prod_{\alpha=1}^m \int_{\mathbb{R}^2} e^{-\frac{1}{4}(\tau_\alpha^2 + s_\alpha^2)} \frac{\mathrm{d}\tau_\alpha \mathrm{d}s_\alpha}{2\pi} \Big) e^{-\frac{t}{16}\sum_\alpha (\tau_\alpha^2+s_\alpha^2)} + Ce^{-cN^\beta (\log N)^2} \\ &\le \Big(\frac{C}{1+t} \Big)^m + Ce^{-cN^\beta(\log N)^2} \\ &\le (C\delta)^m . \qquad \square \end{aligned} \tag{5.10}$$

6 Proof of the semicircle law on short scales

In this section we prove the semicircle law on the shortest possible scale $\eta \ge O(1/N)$.

Proof of Theorem 3.1. We will prove only (3.1), the proof of (3.2) can be obtained from (3.1) exactly as in Corollary 4.2 of [5]. We can assume that $\eta \le (\log N)^4/N$, since the regime $\eta \ge (\log N)^4/N$ has been covered in Theorem 4.1. The constants in this proof depend on κ (in addition to δ_0, D from (1.3)) and we will not follow their precise dependence. For $k = 1, 2, \ldots, N$ define the random variables

$$X_k(z) = X_k := \mathbf{a}^{(k)} \cdot \frac{1}{B^{(k)} - z} \mathbf{a}^{(k)} - \mathbb{E}_k\, \mathbf{a}^{(k)} \cdot \frac{1}{B^{(k)} - z} \mathbf{a}^{(k)} = \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(k)} - 1}{\lambda_\alpha^{(k)} - z} , \tag{6.1}$$

where we used that $\mathbb{E}_k \xi_\alpha^{(k)} = \|\mathbf{u}_\alpha^{(k)}\|^2 = 1$ and we recall that $\mathbb{E}_k$ denotes the expectation w.r.t. the random vector $\mathbf{a}^{(k)}$ (see Section 2 for notation). We note that

$$\mathbb{E}_k\, \mathbf{a}^{(k)} \cdot \frac{1}{B^{(k)} - z} \mathbf{a}^{(k)} = \frac{1}{N} \sum_\alpha \frac{1}{\lambda_\alpha^{(k)} - z} = \Big(1 - \frac{1}{N} \Big) m^{(k)} .$$

It follows from (2.2) and (2.5) that

$$m = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_{kk} - z - \big(1 - \frac{1}{N}\big) m^{(k)} - X_k} . \tag{6.2}$$

We use that

$$\Big| m - \Big(1 - \frac{1}{N} \Big) m^{(k)} \Big| = \Big| \int \frac{\mathrm{d}F(x)}{x - z} - \Big(1 - \frac{1}{N} \Big) \int \frac{\mathrm{d}F^{(k)}(x)}{x - z} \Big| = \frac{1}{N} \Big| \int \frac{NF(x) - (N-1)F^{(k)}(x)}{(x-z)^2} \mathrm{d}x \Big| .$$

and we recall that the eigenvalues of H and $B^{(k)}$ are interlaced,

$$\mu_1 \le \lambda_1^{(k)} \le \mu_2 \le \lambda_2^{(k)} \le \ldots \le \lambda_{N-1}^{(k)} \le \mu_N , \tag{6.3}$$

(see e.g. Lemma 2.5 of [5]), therefore we have $\max_x |NF(x)-(N-1)F^{(k)}(x)| \le 1$. Thus

$$\Big| m - \Big(1-\frac{1}{N}\Big) m^{(k)}\Big| \le \frac{1}{N}\int \frac{\mathrm{d}x}{|x-z|^2} = \frac{\pi}{N\eta}\,. \tag{6.4}$$

Let $M \ge 2$ be sufficiently large and fixed. Fix $E \in [-2+\kappa, 2-\kappa]$ away from the spectral edge. Assume for the moment only that $1/N \le \eta \le 1$. Define $I_n = [E - 2^{n-1}\eta, E + 2^{n-1}\eta]$, and let K_0 be a sufficiently large fixed number. For some constant $C = C(K_0)$ we have $[-K_0, K_0] \subset \bigcup_{n=0}^{C\log N} I_n$. Denote by Ω the event

$$\Omega := \Big\{ \max_{n \le C\log N} \frac{\mathcal{N}_{I_n}}{N|I_n|} \ge M \Big\} \cup \{\max_\alpha |\lambda_\alpha| \ge K_0\}\,. \tag{6.5}$$

Let n_0 be the largest non-negative integer such that $2^{n_0}N\eta \le (\log N)^4$, recall that we assumed $N\eta \le (\log N)^4$. Similarly to the proof of Proposition 4.4, by using (5.1) for short scales and (4.8) for larger scales, we get

$$\begin{aligned} \mathbb{P}(\Omega) &\le e^{-cN} + \sum_{n=0}^{n_0} \Big(\frac{C}{M}\Big)^{2^{n-3}MN\eta} + \sum_{n=n_0+1}^{C\log N} e^{-c\sqrt{2^n MN\eta}} \\ &\le e^{-cN} + \Big(\frac{C}{M}\Big)^{cMN\eta} + e^{-c\sqrt{MN\eta}} \le 3\, e^{-c\sqrt{N\eta}} \end{aligned} \tag{6.6}$$

with some $c > 0$ (first term coming from the probability of $\max_\alpha |\lambda_\alpha| \ge K_0$).

From now on, we additionally assume that $K/N \le \eta \le (\log N)^4/N$. For $n \le n_0$ define $z_n = E + i\eta_n$ with $\eta_n = 2^n\eta$, i.e. $z = z_0$ and $2^n\eta \le (\log N)^4/N$ for all $n \le n_0$. We have from (6.2)

$$\begin{aligned} m(z_n) &= \frac{1}{N}\sum_{k=1}^N \frac{1}{-m(z_n) - z_n + \delta_k} \\ &= \frac{1}{-m(z_n) - z_n} - \frac{1}{N}\sum_{k=1}^N \frac{1}{-m(z_n)-z_n}\ \frac{\delta_k}{h_{kk} - z_n - \frac{1}{N}\sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(k)}}{\lambda_\alpha^{(k)} - z_n}}\,, \end{aligned} \tag{6.7}$$

where

$$\delta_k = \delta_k(z_n) := h_{kk} + m(z_n) - \Big(1 - \frac{1}{N}\Big) m^{(k)}(z_n) - X_k(z_n).$$

Recall $c_0 = \pi\varrho_{sc}(E)$, thus $\mathrm{Im}\ m_{sc}(z) = c_0 + O(\eta)$. Define the event

$$\Xi_n := \{\mathrm{Im}\ m(z_n) \ge c_0/10\}\,.$$

On the event Ξ_n, by using (6.4) and that $N\eta \ge K = 300/c_0$, we have $\mathrm{Im}\ m^{(k)}(z_n) \ge c_0/20$ for any k.

Thus, on the event $\Xi_n \cap \Omega^c$ and for any positive integer r, we have

$$\begin{aligned}
\frac{c_0}{20} &\le \frac{1}{N}\sum_\alpha \frac{\eta_n}{(\lambda_\alpha^{(k)}-E)^2+\eta_n^2} \\
&\le \frac{\mathcal{N}^{(k)}_{I_{n+r}}}{N\eta_n} + \frac{1}{N}\sum_{\ell=n+r+1}^{C\log N}\ \sum_{\alpha\,:\,\lambda_\alpha^{(k)}\in I_\ell\setminus I_{\ell-1}} \frac{\eta_n}{(\lambda_\alpha^{(k)}-E)^2+\eta_n^2} \\
&\le \frac{\mathcal{N}^{(k)}_{I_{n+r}}}{N\eta_n} + \frac{1}{N}\sum_{\ell=n+r+1}^{C\log N} \frac{2^n\eta\mathcal{N}^{(k)}_{I_\ell}}{(2^{\ell-2}\eta)^2} \qquad (6.8)\\
&\le \frac{\mathcal{N}^{(k)}_{I_{n+r}}}{N\eta_n} + 16\sum_{\ell=n+r+1}^{C\log N} \frac{\mathcal{N}_{I_\ell}+1}{N|I_\ell|}\frac{1}{2^{\ell-n}} \\
&\le \frac{\mathcal{N}^{(k)}_{I_{n+r}}}{N\eta_n} + 2^{5-r}M\ ,
\end{aligned}$$

where we used that from the interlacing property we have $\mathcal{N}^{(k)}_I \le \mathcal{N}_I + 1$, for any interval I.

Thus, on $\Xi_n \cap \Omega^c$, with the choice $r = [\log_2(1280M/c_0)] + 1$, we have the lower bound

$$\mathcal{N}^{(k)}_{I_{n+r}} \ge \gamma_n \quad \text{with} \quad \gamma_n := \frac{c_0}{40}N\eta_n$$

for any $n \le n_0$ and for any $k = 1, 2, \dots N$. Hence from (6.7) and recalling the definition (5.5) we get, for any $p \ge 1$, that

$$\begin{aligned}
\mathbb{E}\left|m(z_n) + \frac{1}{m(z_n)+z_n}\right|^p \mathbf{1}\big(\Xi_n\cap\Omega^c\big) &\le \mathbb{E}\left[\frac{10}{c_0}\ \frac{1}{N}\sum_{k=1}^N \frac{|\delta_k|\cdot\mathbf{1}(\mathcal{N}^{(k)}_{I_{n+r}}\ge\gamma_n)}{\eta_n + \frac{1}{N}\sum_\alpha \frac{\eta_n\xi_\alpha^{(k)}}{(\lambda_\alpha^{(k)}-E)^2+\eta_n^2}}\right]^p \\
&\le \mathbb{E}\left[\frac{10}{c_0}\ \frac{|\delta_1|\cdot\mathbf{1}(\mathcal{N}^{(1)}_{I_{n+r}}\ge\gamma_n)}{\eta_n + \frac{1}{2^{2r}N\eta_n}Z^{(1)}(\eta_{n+r})}\right]^p \qquad (6.9)\\
&\le 2^{2pr}C_1^p\big[\mathbb{E}\,|\delta_1|^{2p}\big]^{1/2}\left[\mathbb{E}\left|\frac{1}{\eta_n+\gamma_n^{-1}\sum_{\alpha=1}^{\gamma_n}\xi_\alpha^{(1)}}\right|^{2p}\right]^{1/2}
\end{aligned}$$

(with $C_1 = (const)c_0^{-2}$). The second term can be estimated similarly to (5.7). For any $1 \le p \le c_0N\eta/300$ we have that

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{\eta_n+\gamma_n^{-1}\sum_{\alpha=1}^{\gamma_n}\xi_\alpha^{(1)}}\right|^{2p} &\le \int_0^{(1/\eta_n)^{2p}} \mathbb{P}\Big(\sum_{\alpha=1}^{\gamma_n/2}\xi_\alpha^{(1)} \le \gamma_n t^{-1/2p}\Big)\mathrm{d}t \\
&\le 1 + \int_1^{(1/\eta)^{2p}} \big[C\max\{t^{-1/2p}, N^{-\nu}\}\big]^{\gamma_n/2}\mathrm{d}t \qquad (6.10)\\
&\le C_\nu
\end{aligned}$$

where we chose $\nu > 3/4$ and used that $4p/\nu \le \gamma_n \le C(\log N)^4$.

For the first term on the r.h.s. of (6.9), we use $\mathbb{E}|h_{kk}|^{2p} \le C^p N^{-p}$ and (6.4) to get

$$\mathbb{E}\,|\delta_1|^{2p} \le C^p N^{-p} + \Big(\frac{C}{N\eta_n}\Big)^{2p} + C^p \mathbb{E}|X_1(z_n)|^{2p}. \tag{6.11}$$

To estimate $\mathbb{E}|X_1(z_n)|^{2p}$, we will need the following extension of Lemma 4.2 to $\eta \ge O(1/N)$.

Lemma 6.1 *Let $E \in [-2+\kappa, 2-\kappa]$. Suppose that $\mathbf{v}_\alpha$ and λ_α are eigenvectors and eigenvalues of an $N \times N$ random matrix with a law satisfying the assumption of Theorem 4.1. Let*

$$X = \frac{1}{N}\sum_\alpha \frac{\xi_\alpha - 1}{\lambda_\alpha - z}$$

with $z = E + i\eta$, $\xi_\alpha = |\mathbf{b}\cdot\mathbf{v}_\alpha|^2$, where the components of $\mathbf{b}$ are i.i.d. random variables satisfying the condition **C1)***. Then there exist two positive constants K, C and c (depending on κ) so that for every $0 < \delta \le 1$, we have*

$$\mathbb{P}[|X| \ge \delta] \le C\, e^{-c\,\min\{\delta\sqrt{N\eta},\, \delta^2 N\eta\}} \tag{6.12}$$

if $K \le N\eta \le (\log N)^4$.

Proof of Lemma 6.1. We follow the proof of Lemma 4.2 but with the redefined set Ω (see (6.5) instead of (4.4)). Using the improved bounds from Theorem 5.1 we have already proved in (6.6) that $P(\Omega) \le 3\, e^{-c\sqrt{N\eta}}$. To estimate $\mathbb{E}\big[1_{\Omega^c}\cdot \mathbb{P}_{\mathbf{b}}(|X| \ge \delta)\big]$, we follow the proof of Proposition 4.3. The only difference is that in (4.6) the summation runs from $n = 0$ to $n = C\log N$, but the estimate on the right hand side of (4.6) is still valid. This completes the proof of Lemma 6.1. □

Given the bound (6.12), we have

$$\mathbb{E}|X_1(z_n)|^{2p} \le \frac{(Cp^2)^p}{(N\eta_n)^p}$$

and, from (6.11), we get

$$\mathbb{E}\,|\delta_1|^{2p} \le \frac{(Cp^2)^p}{(N\eta_n)^p}\,.$$

Thus

$$\mathbb{E}\Big|m(z_n) + \frac{1}{m(z_n)+z_n}\Big|^p \mathbf{1}\big(\Xi_n \cap \Omega^c\big) \le \frac{(Cp)^p}{(N\eta_n)^{p/2}}\,. \tag{6.13}$$

For any δ, set the event

$$\Lambda_n(\delta) = \Lambda_n := \Big\{\Big|m(z_n) + \frac{1}{m(z_n)+z_n}\Big| \ge \delta\Big\}$$

then from (6.13)

$$\mathbb{P}(\Lambda_n \cap \Omega^c) \le \mathbb{P}(\Xi_n^c \cap \Omega^c) + \frac{(Cp)^p}{(N\eta_n\delta^2)^{p/2}}\,.$$

We recall the stability of the equation $m + (m+z)^{-1} = 0$, i.e. that there exists a constant $C_\kappa \ge 1$, depending only on κ, so that for any δ

$$\Big|m + \frac{1}{m+z}\Big| \le \delta \quad\Longrightarrow\quad |m - m_{sc}(z)| \le C_\kappa \delta\,. \tag{6.14}$$

Choosing $\delta \le c_0/10C_\kappa$ and recalling that $\mathrm{Im}\, m_{sc}(z) = c_0 + o(1)$ as $N \to \infty$, we also see that

$$\Big| m + \frac{1}{m+z} \Big| \le \delta \quad \Longrightarrow \quad \mathrm{Im}\, m \ge c_0/2 \tag{6.15}$$

for sufficiently large N. We also know (e.g. from [5])

$$\mathrm{Im}\, m(z_n) \ge \frac{1}{2} \mathrm{Im}\, m(z_{n+1}) \, .$$

Thus, on the event Ξ_n^c we have $\mathrm{Im}\, m(z_{n+1}) \le c_0/5$, which by (6.15) implies that

$$\Xi_n^c \subset \Lambda_{n+1}$$

assuming that $\delta \le 4c_0/5C_\kappa$.

Thus, we get

$$\mathbb{P}(\Lambda_n \cap \Omega^c) \le \mathbb{P}(\Lambda_{n+1} \cap \Omega^c) + \frac{(Cp)^p}{(2^n \eta N \delta^2)^{p/2}} \, .$$

After summing up this inequality for $0 \le n \le n_0$, and using the result from [6] on the scale $\eta_{n_0} \sim (\log N)^4/N$, we get

$$\mathbb{P}(\Lambda_n \cap \Omega^c) \le \frac{(Cp)^p}{(\eta N \delta^2)^{p/2}} + e^{-c(\log N)^2}$$

for sufficiently large $N \ge N_0(\delta)$. Thus, combining this with (6.6), for sufficiently small δ, we have

$$\mathbb{P}\Big(\Big| m(z_0) + \frac{1}{m(z_0) + z_0} \Big| \ge \delta \Big) \le \frac{(Cp)^p}{(\eta N \delta^2)^{p/2}} + C\, e^{-c\sqrt{N\eta}} + e^{-c(\log N)^2}$$

Choosing $p = \min\{1, \ c\delta\sqrt{N\eta}\}$ with some small constant c and using the stability bound (6.14), we obtain Theorem 3.1 for the remaining case of $\eta \le (\log N)^4/N$. $\square$.

Proof of Corollary 3.2. Part (i) follows from (4.12) and (4.13) by noticing that no N^2 entropy factor in (4.13) is needed. In estimating $\mathbb{P}(\mathcal{N}_{I_n} \le \frac{1}{2} c_1 N \eta^*)$ in (4.13) we infer to the semicircle law (3.2) which now holds on the $O(1/N)$ scale. Part (ii) follows from part (i) and from

$$\mathbb{P}(\|\mathbf{v}\|_p^p \ge M^p N^{1-\frac{p}{2}}) = \mathbb{P}\Big(\frac{1}{N} \sum_{j=1}^N |v_j|^p \ge \frac{M^p}{N^{p/2}} \Big) \le \Big(\frac{N^{p/2}}{M^p} \Big)^q \mathbb{E} |v_1|^{pq} \le e^{-c\sqrt{M}} \, .$$

with the choice of $q = c\sqrt{M}$ where $c = c(\kappa, p) > 0$ is sufficiently small. Here we used that from part (i) we have that for any $m \ge 1$

$$\mathbb{E}\, (N^{1/2}|v_1|)^m \le M_0^m + m \int_{M_0}^\infty t^{m-1} e^{-c\sqrt{t}} \mathrm{d}t \le (Cm)^{2m}$$

where $C = C(\kappa)$.

Part (iii) also follows from part (i) after summing up the estimate (3.3) for all spectral intervals and for all coordinates v_j of $\mathbf{v}$ by using that the distribution of v_j is independent of j. $\square$

7 Proof of the tail of the gap distribution

Proof of Theorem 3.3. First notice that for any $K_0(\kappa)$ it is sufficient to prove the theorem for all $K \geq K_0(\kappa)$, by adjusting the prefactor $C = C(\kappa)$ in (3.6). Second, it is sufficient to consider the case of sufficient large $N \geq N_0(\kappa)$. By increasing $K_0(\kappa)$ to ensure $K_0(\kappa) \geq N_0(\kappa)^2$ if necessary, we can estimate

$$\mathbb{P}(\lambda_{\alpha+1} - E \geq K/N,\ \alpha \leq N-1) \leq \mathbb{P}(\max_\beta \lambda_\beta \geq \sqrt{K} - 2) \tag{7.1}$$

for any $K \geq K_0$ and $N \leq N_0$. We recall part i) of Lemma 7.3 of [5], i.e. that there is a constant $c > 0$ such that

$$\mathbb{P}\{\max_\beta \lambda_\beta \geq L\} \leq e^{-cL^2N} \tag{7.2}$$

for all $L \geq L_0$ sufficiently large (both c and L_0 depend on the constants in (1.3)). Thus the probability in (7.1) can be estimated by $C \exp(-c\sqrt{K})$.

Next we treat the case $K \geq CN$ with some large constant C. Since $\lambda_{\alpha+1} \geq E + K/N$ implies $\max_\beta \lambda_\beta \geq K/N - 2 \geq L_0$ for a sufficiently large C, and using (7.2), we obtain much stronger bound of the form $\exp(-cK^2N)$ for the tail probability of $\lambda_{\alpha+1}$. For the rest of the proof we can thus assume that $K \leq CN$ and both K and N are sufficiently large, depending on κ.

The event $\lambda_{\alpha+1} \geq E + K/N$ implies that there is a gap of size K/N about $E' = E + K/2N$. Fix a sufficiently large M (depending on κ) and let $z' = E' + i\eta$, with $\eta = K/(NM^2)$ and denote

$$\mathcal{N}_j = \#\{\beta\ :\ 2^{j-1}K/N \leq |\lambda_\beta - E'| \leq 2^j K/N\}\ , \quad j = 0, 1, 2, \ldots$$

On the set where $\max_\alpha |\lambda_\alpha| \leq K_0$, with some large constant K_0, we can estimate

$$\begin{aligned} \text{Im}\ m(z') =& \frac{1}{N}\sum_{\beta=1}^{N-1} \frac{\eta}{(\lambda_\beta - E')^2 + \eta^2} \\ \leq& \frac{\eta}{N}\sum_{j=0}^{C\log N} \frac{\mathcal{N}_j}{(2^{j-1}K/N)^2}\ . \end{aligned} \tag{7.3}$$

Define

$$\Omega := \max_\alpha\{|\lambda_\alpha| \leq K_0\} \cup \bigcup_{j=0}^{C\log N} \{\mathcal{N}_j \leq 2^{j+1}KM\},$$

with a sufficiently large K_0, then, similarly to the estimate (6.6), and together with $K \leq CN$, we get

$$\mathbb{P}(\Omega^c) \leq Ce^{-c\sqrt{K}}\ .$$

Then, on the set Ω, we have from (7.3)

$$\text{Im}\ m(z') \leq \frac{16}{M}\ .$$

For large M this implies that $|\text{Im}\ m(z') - \text{Im}\ m_{sc}(z')| \geq \frac{1}{2}\text{Im}\ m_{sc}(z') =: c_0 > 0$ and from Theorem 3.1 we know that

$$\mathbb{P}(|m(z') - m_{sc}(z')| \geq c_0) \leq e^{-c\sqrt{N\eta}} = e^{-c'\sqrt{K}}\ ,$$

where the constants depend on κ. Thus, recalling that α was defined to be the index of the largest eigenvalue below E, we have

$$\mathbb{P}(\lambda_{\alpha+1} - E \geq K/N,\ \alpha \leq N-1) \leq \mathbb{P}(\Omega^c) + \mathbb{P}(|m(z') - m_{sc}(z')| \geq c_0) \leq Ce^{-c\sqrt{K}} .$$

This proves Theorem 3.3. □

8 Proof of the Wegner estimate

Proof of Theorem 3.4. We can assume that $\varepsilon < 1/2$, otherwise the statement is trivial. From the basic formulae (2.7), (3.11) and using the Schwarz inequality, we obtain that

$$\begin{aligned}\mathbb{E}\,\mathcal{N}_I^2 &\leq C(N\eta)^2\,\mathbb{E}\left[\mathrm{Im}\,\frac{1}{h - z - \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\xi_\alpha}{\lambda_\alpha - z}}\right]^2 \\ &\leq C\varepsilon^2\,\mathbb{E}\left[\Big(\eta + \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\eta\xi_\alpha}{(\lambda_\alpha - E)^2+\eta^2}\Big)^2 + \Big(h - E - \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{(\lambda_\alpha - E)\xi_\alpha}{(\lambda_\alpha - E)^2+\eta^2}\Big)^2\right]^{-1},\end{aligned} \tag{8.1}$$

where $h = h_{11}$ and $\lambda_\alpha = \lambda_\alpha^{(1)}$, i.e. the eigenvalues of the minor $B = B^{(1)}$ obtained from H by removing the first row and column, and $\xi_\alpha = \xi_\alpha^{(1)} = |\mathbf{b}\cdot\mathbf{u}_\alpha^{(1)}|^2$ where $\mathbf{b} \equiv (b_1, \dots, b_{N-1}) := \sqrt{N}(h_{12}, h_{13}, \dots, h_{1N})$.

Introducing the notation

$$d_\alpha := \frac{N(\lambda_\alpha - E)}{N^2(\lambda_\alpha - E)^2 + \varepsilon^2}, \qquad c_\alpha := \frac{\varepsilon}{N^2(\lambda_\alpha - E)^2 + \varepsilon^2},$$

we have

$$\mathbb{E}\,\mathcal{N}_I^2 \leq C\varepsilon^2\,\mathbb{E}\left[\Big(\sum_{\alpha=1}^{N-1} c_\alpha\xi_\alpha\Big)^2 + \Big(h - E - \sum_{\alpha=1}^{N-1} d_\alpha\xi_\alpha\Big)^2\right]^{-1}. \tag{8.2}$$

Let γ be defined so that

$$\lambda_\gamma - E = \min\Big\{\lambda_\alpha - E\ :\ \lambda_\alpha - E \geq \frac{\varepsilon}{N}\Big\},$$

i.e. λ_γ is the first eigenvalue above $E + \varepsilon/N$. Thus $\lambda_\gamma \leq \lambda_{\gamma+1} \leq \lambda_{\gamma+2} \leq \lambda_{\gamma+3}$ are the first four eigenvalues above $E + \varepsilon/N$. If there are no four eigenvalues above $E + \varepsilon/N$, then we use the four consecutive eigenvalues below $E - \varepsilon/N$, as it will be clear from the proof, what matters is only that the signs of $d_{\gamma+j}$, $j = 0, 1, 2, 3$, are identical. At the end of the proof we will consider the exceptional case when there are less than four λ-eigenvalues both above $E + \varepsilon/N$ and below $E - \varepsilon/N$, i.e. in this case all but at most six eigenvalues are in $[E - \varepsilon/N, E + \varepsilon/N]$.

We define then

$$\Delta := N(\lambda_{\gamma+3} - E). \tag{8.3}$$

Note that, by definition,

$$\varepsilon \leq N(\lambda_\gamma - E) \leq \ldots \leq N(\lambda_{\gamma+3} - E) = \Delta\ ,$$

in particular $d_\gamma \geq d_{\gamma+1} \geq d_{\gamma+2} \geq d_{\gamma+3}$ (since the function $x \to x/(x^2 + \varepsilon^2)$ is decreasing for $x \geq \varepsilon$) and $c_\gamma \geq c_{\gamma+1} \geq c_{\gamma+2} \geq c_{\gamma+3}$ thus

$$\min_{j=0,1,2,3} d_{\gamma+j} = \frac{\Delta}{\Delta^2 + \varepsilon^2} \geq \frac{1}{2\Delta}, \qquad \min_{j=0,1,2,3} c_{\gamma+j} \geq \frac{\varepsilon}{\Delta^2}\ . \tag{8.4}$$

Next, we discard, in the first term in the denominator of (8.2), all contributions but the ones from $\alpha = \gamma, \gamma+1$. We find

$$\mathbb{E}\,\mathcal{N}_I^2 \le C\varepsilon^2\,\mathbb{E}\left[\Big(c_\gamma\xi_\gamma + c_{\gamma+1}\xi_{\gamma+1}\Big)^2 + \Big(h - E - \sum_{\alpha=1}^{N-1} d_\alpha\xi_\alpha\Big)^2\right]^{-1}.$$

Note that c_α and d_α depend on the minor B and are independent of the vector $\mathbf{b}$, so we can first take the expectation value with respect to $\mathbf{b}$. In Lemma 8.2 below we give a general estimate for such expectation values. Applying (8.11) from Lemma 8.2 with $r = p = 2$, $\beta_1 = \gamma+2$, $\beta_2 = \gamma+3$, and using the estimates (8.4), we have

$$\mathbb{E}\,\mathcal{N}_I^2 \le C\varepsilon\mathbb{E}\,\Delta^3\,.$$

To estimate the tail probability of Δ, we note that for any $K \ge \varepsilon$, the event $\Delta \ge K$ means that there must be an interval of size $(K-\varepsilon)/4N$ between $E+\varepsilon/N$ and $E+K/N$ with no λ-eigenvalue. From Theorem 3.3 we have

$$\mathbb{P}(\mathcal{N}_J^\lambda = 0) \le C\,e^{-c\sqrt{N|J|}}$$

for any interval J with length $|J| \ge 1/N$. Thus

$$\mathbb{P}(\Delta \ge t) \le C\,e^{-c\sqrt{t}}, \qquad t \ge 1.$$

Therefore $\mathbb{E}\,\Delta^3$ is finite and thus $\mathbb{E}\,\mathcal{N}_I^2 \le C\varepsilon$ is proven. The other statements in Theorem 3.4 are easy consequences of this estimate.

Finally, we have to consider the case, when all but at most six λ-eigenvalues are within $[E-\varepsilon/N, E+\varepsilon/N]$. For all these eigenvalues λ_α we have $\frac{1}{2}\varepsilon^{-1} \le c_\alpha \le \varepsilon^{-1}$. If $N-1 \ge 9$, then there are at least three eigenvalues in $[E-\varepsilon/N, E+\varepsilon/N]$, we denote them by $\lambda_{\gamma_1}, \lambda_{\gamma_2}$, and λ_{γ_3}. Then we have from (8.2) and from (8.13) of Lemma 8.2 below that

$$\mathbb{E}\,\mathcal{N}_I^2 \le \varepsilon^2\mathbb{E}\Big(c_{\gamma_1}\xi_{\gamma_1} + c_{\gamma_2}\xi_{\gamma_2} + c_{\gamma_3}\xi_{\gamma_3}\Big)^{-2} \le C\varepsilon^4\,.$$

This completes the proof for $N \ge 10$.

The case $N < 10$ requires a different argument. Let f be a smooth cutoff function supported on $[-1,1]$, $0 \le f \le 1$ and $f(x) \equiv 1$ for $|x| \le 1/2$, and let $F(s) = \int_{-\infty}^s f(x)\mathrm{d}x$ its antiderivative, clearly $0 \le F(s) \le 2$. Write

$$\mathbb{E}\mathcal{N}_I^2 \le N\sum_{\alpha=1}^N \mathbb{E}^*\Big[\mathbb{E}^{**} f\Big(\frac{\mu_\alpha - E}{\varepsilon/N}\Big)\Big]\,,$$

where $\mathbb{E}^*$ is the expectation with respect to the off-diagonal matrix elements and $\mathbb{E}^{**}$ is the expectation with respect to the diagonal elements x_{ii}, $i = 1, 2, \ldots N$. Since N is bounded, it is sufficient to show that the expectation inside the square bracket is bounded by $C\varepsilon$. Let $\mathbf{x} = (x_{11}, x_{22}, \ldots, x_{NN})$ and viewing μ_α as a function of $\mathbf{x}$, we have

$$\nabla_{\mathbf{x}}\Big[F\Big(\frac{\mu_\alpha - E}{\varepsilon/N}\Big)\Big] = N\varepsilon^{-1} f\Big(\frac{\mu_\alpha - E}{\varepsilon/N}\Big)\nabla_{\mathbf{x}}\mu_\alpha\,. \tag{8.5}$$

Simple first order perturbation shows that

$$\frac{\partial\mu_\alpha}{\partial x_{ii}} = \frac{2}{\sqrt{N}}|\mathbf{v}_\alpha(i)|^2$$

where $\mathbf{v}_\alpha$ is the eigenvector of H belonging to μ_α. Notice that the components of the gradient in (8.5) are nonnegative and their sum is $2/\sqrt{N}$. Thus, summing up each component of (8.5), we get

$$\begin{aligned}
\mathbb{E}^{**} f\Big(\frac{\mu_\alpha - E}{\varepsilon}\Big) &= \frac{\varepsilon}{2\sqrt{N}} \sum_{i=1}^N \int_{\mathbb{R}^N} \Big[\prod_{j=1}^N \mathrm{d}\widetilde{\nu}(x_{jj})\Big] \frac{\partial}{\partial x_{ii}} \Big[F\Big(\frac{\mu_\alpha - E}{\varepsilon/N}\Big)\Big] \\
&= (\text{const.}) \frac{\varepsilon\sqrt{N}}{2} \int_{\mathbb{R}} \mathrm{d}x_{11}\, e^{-\widetilde{g}(x_{11})} \frac{\partial}{\partial x_{11}} \Bigg[\int_{\mathbb{R}^{N-1}} F\Big(\frac{\mu_\alpha - E}{\varepsilon/N}\Big) \prod_{j=2}^N \mathrm{d}\widetilde{\nu}(x_{jj})\Bigg] \\
&\le C\varepsilon\sqrt{N}.
\end{aligned} \tag{8.6}$$

In the last step we used integration by parts, the boundedness of F, the fact that $\mathrm{d}\widetilde{\nu}$ is a probability measure and that $\int_{\mathbb{R}} |\widetilde{g}'(x)| \exp(-\widetilde{g}(x))\mathrm{d}x$ is finite. Thus we obtained the Wegner estimate for the small values of N as well. □

The proof actually shows the following stronger result that will be needed in Section 9. As before, let μ's be the eigenvalues of an $N \times N$ Wigner matrix, and let $\gamma = \gamma(N)$ defined as

$$\mu_\gamma - E = \min\Big\{\mu_\alpha - E \;:\; \mu_\alpha - E \ge \frac{\varepsilon}{N}\Big\}\,.$$

For any positive integer d, let

$$\Delta_d^{(\mu)} = N(\mu_{\gamma(N)+d-1} - E) \tag{8.7}$$

i.e. the rescaled distance from E to the d-th μ-eigenvalue above $E + \varepsilon/N$. If there are no d μ-eigenvalues above $E + \varepsilon/N$, then we use the eigenvalues below $E - \varepsilon/N$ to define $\gamma = \gamma(N)$ as

$$\mu_\gamma - E = \max\Big\{\mu_\alpha - E \;:\; \mu_\alpha - E \le -\frac{\varepsilon}{N}\Big\}$$

and

$$\Delta_d^{(\mu)} = N(E - \mu_{\gamma(N)-d+1})\,. \tag{8.8}$$

To unify the notation, let us introduce the symbol

$$\Delta_d^{(\mu)} = \infty \tag{8.9}$$

for the extreme case, when there are at most $d-1$ eigenvalues above $E+\varepsilon/N$ and at most $d-1$ eigenvalues below $E-\varepsilon/N$; in particular in this case all but at most $2d-2$ eigenvalues are between $E-\varepsilon/N$ and $E+\varepsilon/N$.

Corollary 8.1 *With the notation above, for any $d \ge 5$, $N \ge 10$ and $M \in \mathbb{N}$ there is a constant $C = C_{M,d}$ such that*

$$\mathbb{E}\Big[\mathbf{1}(\mathcal{N}_I \ge 1) \cdot \big[\Delta_d^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_d^{(\mu)} < \infty)\Big] \le C\,\varepsilon\,. \tag{8.10}$$

Proof. We proceed as in the proof of Theorem 3.4 above. By $\mathbf{1}(\mathcal{N}_I \ge 1) \le \mathcal{N}_I^2$ and following the estimates (8.1)–(8.2), we have

$$\mathbb{E}\Big[\mathbf{1}(\mathcal{N}_I \ge 1) \cdot \big[\Delta_d^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_d^{(\mu)} < \infty)\Big] \le C\varepsilon^2\, \mathbb{E}\, \frac{\big[\Delta_d^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_d^{(\mu)} < \infty)}{\Big(\sum_{\alpha=1}^{N-1} c_\alpha \xi_\alpha\Big)^2 + \Big(h - E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha\Big)^2}\,.$$

With the notation (8.7), the Δ in (8.3) is actually $\Delta = \Delta_4^{(\lambda)}$, where the superscript indicates that it is defined in the λ-spectrum. By the interlacing property and by $d \geq 5$, we have

$$\Delta = \Delta_4^{(\lambda)} \leq \Delta_d^{(\mu)} \leq \Delta_{d+1}^{(\lambda)},$$

thus

$$\mathbb{E}\Big[\mathbf{1}(\mathcal{N}_I \geq 1) \cdot \big[\Delta_d^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_d^{(\mu)} < \infty)\Big] \leq \varepsilon^2\, \mathbb{E}\, \frac{\big[\Delta_{d+1}^{(\lambda)}\big]^M \cdot \mathbf{1}(\Delta_{d+1}^{(\lambda)} < \infty)}{\Big(\sum_{\alpha=1}^{N-1} c_\alpha \xi_\alpha\Big)^2 + \Big(h - E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha\Big)^2}\,.$$

Now we perform the expectation with respect to the $\mathbf{b}$ variables as before; the numerator is independent of $\mathbf{b}$. We get

$$\mathbb{E}\Big[\mathbf{1}(\mathcal{N}_I \geq 1) \cdot \big[\Delta_d^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_d^{(\mu)} < \infty)\Big] \leq C\varepsilon\, \mathbb{E}\big[\Delta_4^{(\lambda)}\big]^3 \big[\Delta_{d+1}^{(\lambda)}\big]^M \cdot \mathbf{1}(\Delta_{d+1}^{(\lambda)} < \infty) \leq C\varepsilon$$

since the tail distribution of any $\Delta_d^{(\lambda)}$ decays faster than any polynomial. $\square$

Lemma 8.2 *Fix $p \in \mathbb{N}/\{0\}$ and let $N \geq p+3$. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N-1}$ be an arbitrary orthonormal basis in $\mathbb{C}^{N-1}$ and set $\xi_\alpha = |\mathbf{b} \cdot \mathbf{u}_\alpha|^2$, where the components of $\mathbf{b}$ are i.i.d complex variables with distribution ν with density h satisfying the condition* **C2)** *with an exponent $a = p+3$ in (1.4). Fix different indices $\alpha_1, \ldots, \alpha_p, \beta_1, \beta_2 \in \{1, 2, \ldots, N-1\}$. Assume that $c_j > 0$, for $j = 1, \ldots, p$. Let $d_\alpha \in \mathbb{R}$ for all $1 \leq \alpha \leq N-1$ be arbitrary numbers such that $d_{\beta_1}, d_{\beta_2} > 0$. Then, for every $1 < r < p+1$, there exists a constant $C_{r,p} < \infty$ such that*

$$\mathbb{E}_{\mathbf{b}} \left[\Big(\sum_{j=1}^p c_j \xi_{\alpha_j}\Big)^2 + \Big(E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha\Big)^2 \right]^{-\frac{r}{2}} \leq \frac{C_{p,r}}{\big(\prod_{j=1}^p c_j\big)^{\frac{r-1}{p}} \min(d_{\beta_1}, d_{\beta_2})}\,. \tag{8.11}$$

Moreover, for every $p \geq 3$, we also have the improved bound

$$\mathbb{E}_{\mathbf{b}} \left[\Big(\sum_{j=1}^p c_j \xi_{\alpha_j}\Big)^2 + \Big(E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha\Big)^2 \right]^{-\frac{p}{2}} \leq \frac{C_p}{\big(\prod_{j=1}^{p-2} c_j\big) \min(c_{p-1}, c_p) \min(d_{\beta_1}, d_{\beta_2})}\,. \tag{8.12}$$

for a constant C_p depending only on p.

Without the second term in the denominator, we have the following estimates: For all $1 \leq r < p$, there exists a constant $C_{p,r} < \infty$ such that

$$\mathbb{E}_{\mathbf{b}} \left[\sum_{j=1}^p c_j \xi_{\alpha_j} \right]^{-r} \leq \frac{C_{p,r}}{(\min c_j)^r}\,. \tag{8.13}$$

Remark. For (8.13), it is enough to assume that

$$|\widehat{h}(t,s)| \leq \frac{1}{(1 + \omega_{p+1}(t^2+s^2))^{p+1}} \tag{8.14}$$

instead of both conditions in (1.4) with exponent $a = p+3$.

Proof. To prove (8.11), we perform a change variables from $\mathbf{b} = (b_1, \ldots, b_{N-1})$ to $\mathbf{z} = (z_1, \ldots z_{N-1})$ by introducing

$$\mathbf{z} = U^* \mathbf{b}$$

where U is the unitary matrix with columns $(\mathbf{u}_1, \ldots, \mathbf{u}_{N-1})$. Notice that the Jacobian is one, thus

$$\mathrm{I} := \mathbb{E}_{\mathbf{b}} \left[\left(\sum_{j=1}^{p} c_j \xi_{\alpha_j} \right)^2 + \Big(E - \sum_{\alpha=1}^{N-1} d_\alpha \xi_\alpha \Big)^2 \right]^{-r/2} = \int \frac{\mathrm{d}\mu(\mathbf{z})}{[P(\mathbf{z})]^{r/2}} \tag{8.15}$$

with

$$\mathrm{d}\mu(\mathbf{z}) := e^{-\Phi(\mathbf{z})} \prod_{\alpha=1}^{N-1} \mathrm{d}z_\alpha \mathrm{d}\overline{z}_\alpha, \qquad \Phi(\mathbf{z}) := \sum_{\ell=1}^{N-1} g\left(\mathrm{Re}\,(U\mathbf{z})_\ell,\ \mathrm{Im}\,(U\mathbf{z})_\ell\right)$$

and

$$P(\mathbf{z}) := \Big(\sum_{j=1}^{p} c_j |z_{\alpha_j}|^2 \Big)^2 + \Big(E - \sum_{\alpha=1}^{N-1} d_\alpha |z_\alpha|^2 \Big)^2.$$

We define, for $t \in \mathbb{R}$,

$$F(t) := \int_{-\infty}^{t} \mathrm{d}s \, \left(\Big(\sum_{j=1}^{p} c_j |z_{\alpha_j}|^2 \Big)^2 + s^2 \right)^{-r/2} . \tag{8.16}$$

Note that, for every $r > 1$, there exists a constant $C_r < \infty$, such that

$$0 \le F(t) \le \frac{C_r}{\Big(\sum_{j=1}^{p} c_j |z_{\alpha_j}|^2 \Big)^{r-1}} \tag{8.17}$$

for every $t \in \mathbb{R}$. For $j = 1, 2$, we have

$$z_{\beta_j} \frac{\mathrm{d}}{\mathrm{d}z_{\beta_j}} \, F\left(E - \sum_{\alpha=1}^{N-1} d_\alpha |z_\alpha|^2 \right) = -\frac{d_{\beta_j} |z_{\beta_j}|^2}{[P(\mathbf{z})]^{r/2}} \, .$$

Introducing the first order differential operator

$$D := z_{\beta_1} \frac{\mathrm{d}}{\mathrm{d}z_{\beta_1}} + z_{\beta_2} \frac{\mathrm{d}}{\mathrm{d}z_{\beta_2}} \, ,$$

we find

$$D \left[F\Big(E - \sum_{\alpha=1}^{N-1} d_\alpha |z_\alpha|^2 \Big) \right] = -\frac{d_{\beta_1} |z_{\beta_1}|^2 + d_{\beta_2} |z_{\beta_2}|^2}{[P(\mathbf{z})]^{r/2}} \, . \tag{8.18}$$

From (8.15), we get

$$\mathrm{I} = -\int \mathrm{d}\mu(\mathbf{z}) \frac{1}{d_{\beta_1} |z_{\beta_1}|^2 + d_{\beta_2} |z_{\beta_2}|^2} \, D \left[F\Big(E - \sum_{\alpha=1}^{N-1} d_\alpha |z_\alpha|^2 \Big) \right] . \tag{8.19}$$

Integrating by parts and using the fact that

$$\frac{\mathrm{d}}{\mathrm{d}z_{\beta_1}}\,\frac{z_{\beta_1}}{d_{\beta_1}|z_{\beta_1}|^2+d_{\beta_2}|z_{\beta_2}|^2}+\frac{\mathrm{d}}{\mathrm{d}z_{\beta_2}}\,\frac{z_{\beta_2}}{d_{\beta_1}|z_{\beta_1}|^2+d_{\beta_2}|z_{\beta_2}|^2}=\frac{1}{d_{\beta_1}|z_{\beta_1}|^2+d_{\beta_2}|z_{\beta_2}|^2},$$

we find

$$\mathrm{I}=\int \mathrm{d}\mu(\mathbf{z})\frac{F\Big(E-\sum_{\alpha=1}^{N-1} d_\alpha|z_\alpha|^2\Big)}{d_{\beta_1}|z_{\beta_1}|^2+d_{\beta_2}|z_{\beta_2}|^2}\Big(1-D\Phi(\mathbf{z})\Big). \tag{8.20}$$

Clearly

$$|D\Phi(\mathbf{z})|^2\le(|z_{\beta_1}|^2+|z_{\beta_2}|^2)\left(\Big|\frac{\partial\Phi(\mathbf{z})}{\partial z_{\beta_1}}\Big|^2+\Big|\frac{\partial\Phi(\mathbf{z})}{\partial z_{\beta_2}}\Big|^2\right).$$

By a Schwarz inequality in (8.20) and using (8.17), we have

$$\mathrm{I}\le C_r\,\frac{A+B_1+B_2}{\min(d_{\beta_1},d_{\beta_2})}, \tag{8.21}$$

where

$$\begin{aligned} A &:= \int \mathrm{d}\mu(\mathbf{z})\frac{1}{\Big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\Big)^{r-1}}\frac{1}{|z_{\beta_1}|^2+|z_{\beta_2}|^2}\\ B_k &:= \int \mathrm{d}\mu(\mathbf{z})\frac{1}{\Big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\Big)^{r-1}}\Big|\frac{\partial\Phi(\mathbf{z})}{\partial z_{\beta_k}}\Big|^2. \end{aligned} \tag{8.22}$$

The integral A can be bounded as follows

$$\mathrm{A}\le A_1+A_2+A_3 \tag{8.23}$$

with

$$\begin{aligned} A_1 &:= \int \mathrm{d}\mu(\mathbf{z})\frac{\mathbf{1}\big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\le\kappa\big)}{\Big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\Big)^{r-1}}\\ A_2 &:= \frac{1}{\kappa^{r-1}}\int \mathrm{d}\mu(\mathbf{z})\frac{1}{|z_{\beta_1}|^2+|z_{\beta_2}|^2}\\ A_3 &:= \int \mathrm{d}\mu(\mathbf{z})\frac{\mathbf{1}(|z_{\beta_1}|^2+|z_{\beta_2}|^2\le 1)\cdot\mathbf{1}\Big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\le\kappa\Big)}{\Big(\sum_{j=1}^p c_j|z_{\alpha_j}|^2\Big)^{r-1}(|z_{\beta_1}|^2+|z_{\beta_2}|^2)} \end{aligned} \tag{8.24}$$

for any $\kappa>0$. We start with the estimate of A_3. Decompose $z_{\alpha_j}=x_j+iy_j$ and $z_{\beta_j}=x_{p+j}+iy_{p+j}$ into real and imaginary parts. We define the function

$$f(x_1,\dots x_{p+2},y_1,\dots,y_{p+2}):=\frac{\mathbf{1}\Big(\sum_{j=1}^p c_j(x_j^2+y_j^2)\le\kappa\Big)\cdot\mathbf{1}\Big(\sum_{j=p+1}^{p+2}(x_j^2+y_j^2)\le 1\Big)}{\Big[\sum_{j=1}^p c_j(x_j^2+y_j^2)\Big]^{r-1}\sum_{j=p+1}^{p+2}(x_j^2+y_j^2)} \tag{8.25}$$

on $\mathbb{R}^{2p+4}$. Changing variables $c_j^{1/2}x_j \to x_j$, $c_j^{1/2}y_j \to y_j$ and using that $r < p+1$, it is simple to check that

$$\|f\|_1 \le \frac{C_{r,p}\,\kappa^{p+1-r}}{\prod_{j=1}^p c_j}\,.$$

Thus, recalling that $z_\alpha = (U^*\mathbf{b})_\alpha$ and since the indices $\alpha_1,\dots,\alpha_p$, β_1,β_2 are all distinct, we find, by taking the Fourier transformation in the $x_1,\dots x_{p+2}$, $y_1,\dots,y_{p+2}$ variables, that

$$\begin{aligned}
A_3 &\le \|\widehat{f}\|_\infty \int_{\mathbb{R}^{2p+4}} \prod_{j=1}^{p+2} \mathrm{d}t_j \mathrm{d}s_j \left| \mathbb{E}_{\mathbf{b}} e^{-i\sum_{j=1}^p [t_j \mathrm{Re}(U^*\mathbf{b})_{\alpha_j} + s_j \mathrm{Im}(U^*\mathbf{b})_{\alpha_j}] - i\sum_{j=p+1}^{p+2}[t_j\mathrm{Re}(U^*\mathbf{b})_{\beta_j} + s_j \mathrm{Im}(U^*\mathbf{b})_{\beta_j}]}\right| \\
&\le \|f\|_1 \int_{\mathbb{R}^{2p+4}} \prod_{j=1}^{p+2} \mathrm{d}t_j \mathrm{d}s_j \left|\mathbb{E}_{\mathbf{b}} e^{-i[\mathrm{Re}(U\mathbf{t}')+\mathrm{Im}(U\mathbf{s}')]\cdot\mathrm{Re}\mathbf{b} - i[\mathrm{Re}(U\mathbf{t}')-\mathrm{Im}(U\mathbf{s}')]\cdot\mathrm{Im}\mathbf{b}}\right| \\
&\le \|f\|_1 \int_{\mathbb{R}^{2p+4}} \prod_{j=1}^{p+2} \mathrm{d}t_j \mathrm{d}s_j \frac{1}{(1+\omega_{p+3}\|U\mathbf{t}'\|^2 + \omega_{p+3}\|U\mathbf{s}'\|^2)^{p+3}} \\
&\le \|f\|_1 \int_{\mathbb{R}^{2p+4}} \prod_{j=1}^{p+2} \mathrm{d}t_j \mathrm{d}s_j \frac{1}{(1+\omega_{p+3}\|\mathbf{t}\|^2 + \omega_{p+3}\|\mathbf{s}\|^2)^{p+3}} \\
&\le \frac{C_{r,p}\,\kappa^{p+1-r}}{\prod_{j=1}^p c_j}
\end{aligned} \tag{8.26}$$

for an appropriate constant C_p. Here the components t_j' of the vector $\mathbf{t}' \in \mathbb{R}^{N-1}$ are defined to be all zero except $t'_{\alpha_j} := t_j$, $t'_{\beta_{p+1}} := t_{p+1}$, $t'_{\beta_{p+2}} := t_{p+2}$; the vector $\mathbf{s}'$ is defined similarly. In the last but one step we used the bound (1.4) with exponent $p+3$ for the Fourier transform of the distribution of $\mathbf{b}$. In the last step, we used that for the Euclidean norm $\|U\mathbf{t}'\| = \|\mathbf{t}'\| = \|\mathbf{t}\|$ with $\mathbf{t} = (t_1,\dots t_{p+2})$ and similarly $\|U\mathbf{s}'\| = \|\mathbf{s}\|$ with $\mathbf{s} = (s_1,\dots,s_{p+2})$. Using similar arguments to bound the terms A_1 and A_2, we conclude that

$$A \le C_{r,p}\left(\frac{1}{\kappa^{r-1}} + \frac{\kappa^{p+1-r}}{\prod_{j=1}^p c_j}\right)$$

for arbitrary $\kappa > 0$. Optimizing over κ, we find

$$A \le \frac{C_{r,p}}{\left(\prod_{j=1}^p c_j\right)^{\frac{r-1}{p}}}\,. \tag{8.27}$$

To control the integrals B_k, $k=1,2$ in (8.21), we integrate by parts and we use that $\beta_k \ne \alpha_j$:

$$B_k = -\int \prod_{\alpha=1}^{N-1} \mathrm{d}z_\alpha \mathrm{d}\bar{z}_\alpha \left(\sum_{j=1}^p c_j |z_{\alpha_j}|^2\right)^{-(r-1)} \frac{\partial \Phi(\mathbf{z})}{\partial z_{\beta_k}} \frac{\partial e^{-\Phi(\mathbf{z})}}{\partial \bar{z}_{\beta_k}} = \int \mathrm{d}\mu(\mathbf{z}) \left(\sum_{j=1}^p c_j |z_{\alpha_j}|^2\right)^{-(r-1)} \frac{\partial^2 \Phi(\mathbf{z})}{\partial z_{\beta_k}\partial \bar{z}_{\beta_k}}.$$

Simple calculation shows that

$$\frac{\partial^2\Phi(\mathbf{z})}{\partial z_\beta \partial\bar{z}_\beta} = \frac{1}{4}\sum_{\ell=1}^{N-1} |\mathbf{u}_\beta(\ell)|^2 \Delta g\left(\mathrm{Re}\,(U\mathbf{z})_\ell,\ \mathrm{Im}\,(U\mathbf{z})_\ell\right)$$

thus

$$B_k = \frac{1}{4}\sum_{\ell} |\mathbf{u}_{\beta_k}(\ell)|^2\, \mathbb{E}_{\mathbf{b}}\left[\left(\sum_{j=1}^{p} c_j |(U^*\mathbf{b})_{\alpha_j}|^2\right)^{-(r-1)} \Delta g(\operatorname{Re} b_\ell, \operatorname{Im} b_\ell)\right]. \tag{8.28}$$

For each fixed ℓ, the estimate of the expectation value is identical to that of A_1 if the density function e^{-g} for b_ℓ is replaced with $e^{-g}\Delta g$ (and all other b_m, $m \neq \ell$, are still distributed according to e^{-g}). Although $e^{-g}\Delta g$ is not a probability density, it is only the decay of its Fourier transform that is relevant to proceed similarly to the estimate (8.26). Having obtained uniform bound on the expectation in (8.28), we can perform the summation over ℓ and we obtain

$$B_k \le \frac{C_{p,r}}{\left(\prod_{j=1}^{p} c_j\right)^{\frac{r-1}{p}}}.$$

Combining this with (8.27) and (8.21), we have proved (8.11).

To prove (8.12), we proceed as before up to (8.21). This time, however, we bound the term A in (8.22), with $r = p$, by

$$A \le A_4 + A_5 + A_6 + A_7,$$

where

$$\begin{aligned}
A_4 &= \int \mathrm{d}\mu(\mathbf{z}) \frac{\mathbf{1}\left(|z_{\alpha_{p-1}}|^2 + |z_{\alpha_p}|^2 \le 1\right)}{\left(\sum_{j=1}^{p-2} c_j |z_{\alpha_j}|^2 + \widetilde{c}\left(|z_{\alpha_{p-1}}|^2 + |z_{\alpha_p}|^2\right)\right)^{p-1}}\\
A_5 &= \int \mathrm{d}\mu(\mathbf{z}) \frac{1}{\left(\sum_{j=1}^{p-2} c_j |z_{\alpha_j}|^2 + \widetilde{c}\right)^{p-1}}\\
A_6 &= \int \mathrm{d}\mu(\mathbf{z}) \frac{\mathbf{1}(|z_{\beta_1}|^2 + |z_{\beta_2}|^2 \le 1)}{\left(\sum_{j=1}^{p-2} c_j |z_{\alpha_j}|^2 + \widetilde{c}\right)^{p-1}} \frac{1}{|z_{\beta_1}|^2 + |z_{\beta_2}|^2}\\
A_7 &= \int \mathrm{d}\mu(\mathbf{z}) \frac{\mathbf{1}\left(|z_{\alpha_{p-1}}|^2 + |z_{\alpha_p}|^2 \le 1\right) \mathbf{1}(|z_{\beta_1}|^2 + |z_{\beta_2}|^2 \le 1)}{\left(\sum_{j=1}^{p-2} c_j |z_{\alpha_j}|^2 + \widetilde{c}\left(|z_{\alpha_{p-1}}|^2 + |z_{\alpha_p}|^2\right)\right)^{p-1}} \frac{1}{|z_{\beta_1}|^2 + |z_{\beta_2}|^2}
\end{aligned} \tag{8.29}$$

with $\widetilde{c} = \min(c_{p-1}, c_p)$. We consider first the term A_7. We decompose $z_{\alpha_j} = x_j + iy_j$ and $z_{\beta_j} = x_{p+j} + iy_{p+j}$ into real and imaginary parts. We define the function

$$f(x_1, \ldots x_{p+2}, y_1, \ldots, y_{p+2}) := \frac{\mathbf{1}\left(\sum_{j=p-1}^{p} x_j^2 + y_j^2 \le 1\right) \mathbf{1}\left(\sum_{j=p+1}^{p+2} (x_j^2 + y_j^2) \le 1\right)}{\left[\sum_{j=1}^{p-2} c_j (x_j^2 + y_j^2) + \widetilde{c} \sum_{j=p-1}^{p} (x_j^2 + y_j^2)\right]^{p-1} \sum_{j=p+1}^{p+2} (x_j^2 + y_j^2)} \tag{8.30}$$

on $\mathbb{R}^{2p+4}$. Changing variables $c_j^{1/2} x_j \to x_j$, $c^{1/2} y_j \to y_j$ for $j = 1, \ldots, p-2$, and then letting $r = \sum_{j=1}^{p-2}(x_j^2 + y_j^2)$ and $w = \sum_{j=p-1}^{p}(x_j^2 + y_j^2)$, we find that

$$\|f\|_1 \le \frac{C_p}{\prod_{j=1}^{p-2} c_j} \int_0^1 \mathrm{d}w\, w \int_0^\infty \mathrm{d}r \frac{r^{p-3}}{(r + \widetilde{c}\, w)^{p-1}} \le \frac{C_p}{\widetilde{c} \prod_{j=1}^{p-2} c_j} \tag{8.31}$$

for an appropriate constant C_p. Proceeding as in (8.26), we conclude that

$$A_7 \leq \frac{C_p}{\widetilde{c} \prod_{j=1}^{p-2} c_j}.$$

The terms A_4, A_5, A_6 can be controlled similarly. Hence

$$A \leq \frac{C_p}{\widetilde{c} \prod_{j=1}^{p-2} c_j} .$$

The bound for B in (8.22) can be obtained analogously as in the proof of (8.11) (with the same modifications used for the term A). The proof of (8.13) is similar (but much simpler). $\square$

9 Proof of the level repulsion

Proof of Theorem 3.5. We can assume that $\varepsilon < 1/2$ and that $k \geq 2$, the $k = 1$ case was proven in Theorem 3.4. We recall the notation $\Delta_d^{(\mu)}$ from (8.7)–(8.9) and we split

$$\mathbb{P}(\mathcal{N}_\eta \geq k) \leq (I) + (II)$$

with

$$\begin{aligned} (I) :=& \mathbb{P}(\mathcal{N}_\eta \geq k, \Delta_d^{(\mu)} = \infty) \\ (II) :=& \mathbb{P}(\mathcal{N}_\eta \geq k, \Delta_d^{(\mu)} < \infty) \end{aligned} \tag{9.1}$$

for some positive integer d. From the basic formula (2.7) we have

$$\mathcal{N}_\eta \leq \frac{C\varepsilon}{N} \sum_{j=1}^{N} \left[\left(\eta + \frac{\eta}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(j)}}{(\lambda_\alpha^{(j)} - E)^2 + \eta^2} \right)^2 + \left(E - h_{jj} + \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{(\lambda_\alpha^{(j)} - E)\, \xi_\alpha^{(j)}}{(\lambda_\alpha^{(j)} - E)^2 + \eta^2} \right)^2 \right]^{-1/2} . \tag{9.2}$$

We introduce the notations $\xi_\alpha = \xi_\alpha^{(1)}$, $\lambda_\alpha = \lambda_\alpha^{(1)}$, $h = h_{11}$, and

$$c_\alpha = \frac{\varepsilon}{N^2(\lambda_\alpha - E)^2 + \varepsilon^2}, \qquad d_\alpha = \frac{N(\lambda_\alpha - E)}{N^2(\lambda_\alpha - E)^2 + \varepsilon^2}$$

as before. Using a moment inequality, we get

$$(I) \leq C_k \varepsilon^{k^2}\, \mathbb{E}\, \frac{\mathbf{1}(\Delta_d^{(\mu)} = \infty)}{\left(\sum_{\alpha=1}^{N-1} c_\alpha\, \xi_\alpha \right)^{k^2}} .$$

This term represents the extreme case, when all but at most $2d - 2$ eigenvalues are in $[E - \varepsilon/N, E + \varepsilon/N]$. Choosing $d = 2k$ and assuming that $N \geq k^2 + 4k$, we see that for at least $k^2 + 1$ different α-indices we have $\lambda_\alpha \in [E - \varepsilon/N, E + \varepsilon/N]$, i.e. $\frac{1}{2}\varepsilon^{-1} \leq c_\alpha \leq \varepsilon^{-1}$. Using (8.13) with $r = k^2$, $p = k^2 + 1$, we get

$$(I) \leq C_k \varepsilon^{2k^2} . \tag{9.3}$$

Now we turn to the estimate of (II) and we will consider the following somewhat more general quantity:

$$I_N(M,k,\ell) := \mathbb{E}\Big[\mathbf{1}(\mathcal{N}_\eta^\mu \geq k)\cdot \big[\Delta_\ell^{(\mu)}\big]^M \cdot \mathbf{1}(\Delta_{2k-\ell+4}^{(\mu)} < \infty)\Big]$$

for any $M \in \mathbb{N}$ and $4 \leq \ell \leq k+2$. The index N refers to the fact that the μ's are the eigenvalues of an $N \times N$ Wigner matrix. The superscript μ in $\mathcal{N}_\eta^\mu$ indicates that it counts the number of μ-eigenvalues. Since by definition $\Delta_\ell^{(\mu)} \geq 1$, we know that $I_N(M,k,\ell)$ is monotone increasing in M. Moreover, with the choice $M=0$, $\ell=4$ we have

$$(II) \leq I_N(0,k,4)\,. \tag{9.4}$$

Since the existence of k μ-eigenvalues in the interval I_η implies that $\mathcal{N}_\eta^{(j)} \geq k-1$ for all $j=1,\ldots,N$ (where $\mathcal{N}_\eta^{(j)}$ denotes the number of eigenvalues $\lambda_\alpha^{(j)} \in I_\eta$), we obtain that

$$I_N(M,k,\ell) \leq C_k\,\varepsilon^{k+1}\,\mathbb{E}\,\frac{\mathbf{1}(\mathcal{N}_\eta^{(1)} \geq k-1)\cdot\big[\Delta_\ell^{(\mu)}\big]^M\cdot\mathbf{1}(\Delta_{2k-\ell+4}^{(\mu)}<\infty)}{\Big[\Big(\sum_{\alpha=1}^{N-1} c_\alpha\,\xi_\alpha\Big)^2 + \Big(E-h+\sum_{\alpha=1}^{N-1} d_\alpha\,\xi_\alpha\Big)^2\Big]^{(k+1)/2}}\,. \tag{9.5}$$

By the interlacing property we have $\Delta_{2k-\ell+3}^{(\lambda)} \leq \Delta_{2k-\ell+4}^{(\mu)}$ and $\Delta_\ell^{(\mu)} \leq \Delta_{\ell+1}^{(\lambda)}$, thus we have

$$I_N(M,k,\ell) \leq C_k\,\varepsilon^{k+1}\,\mathbb{E}\,\frac{\mathbf{1}(\mathcal{N}_\eta^{(1)} \geq k-1)\cdot\big[\Delta_{\ell+1}^{(\lambda)}\big]^M\cdot\mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)}<\infty)}{\Big[\Big(\sum_{\alpha=1}^{N-1} c_\alpha\,\xi_\alpha\Big)^2 + \Big(E-h+\sum_{\alpha=1}^{N-1} d_\alpha\,\xi_\alpha\Big)^2\Big]^{(k+1)/2}}\,. \tag{9.6}$$

We split this quantity into two terms:

$$I_N(M,k,\ell) \leq (\mathrm{A}+\mathrm{B})$$

with

$$\begin{aligned}
\mathrm{A} &:= C_k\,\varepsilon^{k+1}\,\mathbb{E}\frac{\mathbf{1}(\mathcal{N}_\eta^{(1)} \geq k+2)\cdot\big[\Delta_{\ell+1}^{(\lambda)}\big]^M\cdot\mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)}<\infty)}{\Big[\Big(\sum_{\alpha=1}^{N-1} c_\alpha\xi_\alpha\Big)^2 + \Big(E-h+\sum_{\alpha=1}^{N-1} d_\alpha\,\xi_\alpha\Big)^2\Big]^{(k+1)/2}}\\
\mathrm{B} &:= C_k\varepsilon^{k+1}\,\mathbb{E}\frac{\mathbf{1}(k-1 \leq \mathcal{N}_\eta^{(1)} < k+2)\cdot\big[\Delta_{\ell+1}^{(\lambda)}\big]^M\cdot\mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)}<\infty)}{\Big[\Big(\sum_{\alpha=1}^{N-1} c_\alpha\xi_\alpha\Big)^2 + \Big(E-h+\sum_{\alpha=1}^{N-1} d_\alpha\,\xi_\alpha\Big)^2\Big]^{(k+1)/2}}\,.
\end{aligned} \tag{9.7}$$

To control the first term, we denote by $\lambda_{\alpha_1},\ldots,\lambda_{\alpha_{k+2}}$ the first $k+2$ λ-eigenvalues in the set I_η. Then $c_{\alpha_j} \geq \frac{1}{2}\varepsilon^{-1}$, for all $j=1,\ldots,k+2$ and therefore, by (8.13),

$$\mathrm{A} \leq C_k\,\varepsilon^{k+1}\mathbb{E}\frac{\mathbf{1}(\mathcal{N}_\eta^{(1)} \geq k+2)\cdot\big[\Delta_{\ell+1}^{(\lambda)}\big]^M\cdot\mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)}<\infty)}{\Big(\varepsilon^{-1}\sum_{j=1}^{k+2}\xi_{\alpha_j}\Big)^{k+1}} \leq C_k\,\varepsilon^{2k+2} I_{N-1}(M,k-1,\ell+1) \tag{9.8}$$

using that $\Delta_m^{(\lambda)}$ is monotone increasing in m.

To control the term B in (9.7), we choose the indices $\alpha_1,\ldots,\alpha_{k-1}$ so that $\lambda_{\alpha_j} \in I_\eta$ for all $j=1,\ldots,k-1$. Since we know that there are at most $k+1$ eigenvalues in I_η, there must be, either on the right or on the left of E, λ-eigenvalues at distances larger than ε/N from E if $N \geq k+8$. Let us suppose, for example, that there are four such eigenvalues on the right of E. Then, we define the index α_k so that

$$\lambda_{\alpha_k} - E = \min\left\{\lambda_\alpha - E \;:\; \lambda_\alpha - E > \frac{\varepsilon}{N}\right\}$$

i.e. λ_{α_k} is the first eigenvalue above $E+\varepsilon/N$. Moreover, let $\alpha_{k+1} = \alpha_k + 1$, $\beta_1 = \alpha_k + 2$ and $\beta_2 = \alpha_{k+1} + 3$. Recalling the notation (8.7), we set $\Delta := \Delta_4^{(\lambda)} = N(\lambda_{\beta_2} - E)$. By definition

$$\varepsilon \leq N(\lambda_{\alpha_k} - E) \leq N(\lambda_{\alpha_{k+1}} - E) \leq N(\lambda_{\beta_1} - E) \leq N(\lambda_{\beta_2} - E) = \Delta.$$

and $\min(d_{\beta_1}, d_{\beta_2}) \geq \frac{1}{2\Delta}$. Therefore

$$\mathrm{B} \leq C_k\, \varepsilon^{k+1}\, \mathbb{E}\, \frac{\mathbf{1}\left(\mathcal{N}_\eta^{(1)} \geq k-1\right) \cdot \left[\Delta_{\ell+1}^{(\lambda)}\right]^M \cdot \mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)} < \infty)}{\left[\left(\sum_{j=1}^{k-1} \varepsilon^{-1}\xi_{\alpha_j} + \frac{\varepsilon}{\Delta^2}(\xi_{\alpha_k} + \xi_{\alpha_{k+1}})\right)^2 + \left(E - h + \sum_{\alpha=1}^{N-1} d_\alpha\, \xi_\alpha\right)^2\right]^{(k+1)/2}}\,. \tag{9.9}$$

From (9.9), we find

$$\begin{aligned} \mathrm{B} \leq C_k \varepsilon^{k+1}\, \mathbb{E}_{\lambda,h} \Bigg\{ & \left[\mathbf{1}\left(\mathcal{N}_\eta^{(1)} \geq k-1\right) \cdot \left[\Delta_{\ell+1}^{(\lambda)}\right]^M \cdot \mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)} < \infty)\right] \\ & \times \mathbb{E}_{\mathbf{b}} \left[\left(\sum_{j=1}^{k-1} \varepsilon^{-1}\xi_{\alpha_j} + \varepsilon\Delta^{-2}(\xi_{\alpha_k} + \xi_{\alpha_{k+1}})\right)^2 + \left(E - h + \sum_{\alpha=1}^{N-1} d_\alpha\, \xi_\alpha\right)^2\right]^{-\frac{k+1}{2}} \Bigg\}\,. \end{aligned} \tag{9.10}$$

Using (8.12) from Lemma 8.2 (with $p = k+1$, $c_j = \varepsilon^{-1}$ for all $j = 1, \ldots, k-1$, $c_k = c_{k+1} = \varepsilon\Delta^{-2}$), it follows that

$$\begin{aligned} \mathrm{B} &\leq C_k\, \varepsilon^{2k-1}\, \mathbb{E}_\lambda \Big[\mathbf{1}\left(\mathcal{N}_\eta^{(1)} \geq k-1\right) \left[\Delta_4^{(\lambda)}\right]^3 \cdot \left[\Delta_{\ell+1}^{(\lambda)}\right]^M \cdot \mathbf{1}(\Delta_{2k-\ell+3}^{(\lambda)} < \infty)\Big] \\ &\leq C_k\, \varepsilon^{2k-1}\, I_{N-1}(M+3, k-1, \ell+1)\,, \end{aligned} \tag{9.11}$$

where we used that $\min(d_{\beta_1}, d_{\beta_2}) \geq 1/2\Delta$ and that $\Delta_4^{(\lambda)} \leq \Delta_{\ell+1}^{(\lambda)}$.

Together with (9.8) and the monotonicity of I_{N-1} in M, we obtain that

$$I_N(M, k, \ell) \leq C_k \varepsilon^{2k-1} I_{N-1}(M+3, k-1, \ell+1)\,.$$

Iterating this inequality, we arrive at

$$I_N(M, k, \ell) \leq C_k \varepsilon^{k^2-1} I_{N-k+1}(M + 3(k-1), 1, \ell+k-1)\,.$$

Recalling (9.4), we have

$$(II) \leq I_N(0, k, 4) \leq C_k \varepsilon^{k^2-1} I_{N-k+1}(3(k-1), 1, k+3)\,.$$

Finally, $I_{N-k+1}(M, 1, d)$ was exactly the quantity that has been estimated by $C\varepsilon$ for any M and $d \geq 5$ in Corollary 8.1 (replacing N by $N-k+1$), thus we have

$$(II) \leq C_k \varepsilon^{k^2}\,.$$

Together with (9.3), this completes the proof of Theorem 3.5. $\square$

A Proof of Proposition 4.5

We first consider the case, when the real and imaginary parts of b_j are i.i.d. (first condition in (1.2)). We split a_{jk} and b_j into real and imaginary parts, $a_{jk} = p_{jk} + iq_{jk}$, $b_j = x_j + iy_j$, and form the vector $\mathbf{w} = (x_1, \dots x_N, y_1, \dots y_N) \in \mathbb{R}^{2N}$ with i.i.d. components. We write $X = X_1 + iX_2$ where $X_1 = \mathbf{w} \cdot \mathbf{P}\mathbf{w} - \mathbb{E}\mathbf{w} \cdot \mathbf{P}\mathbf{w}$, $X_2 = \mathbf{w} \cdot \mathbf{Q}\mathbf{w} - \mathbb{E}\mathbf{w} \cdot \mathbf{Q}\mathbf{w}$ with symmetric real $(2N) \times (2N)$ matrices $\mathbf{P}$ and $\mathbf{Q}$, written in a block-matrix form as

$$\mathbf{P} = \frac{1}{2}\begin{pmatrix} P + P^t & Q - Q^t \\ Q^t - Q & P + P^t \end{pmatrix}, \qquad \mathbf{Q} = \frac{1}{2}\begin{pmatrix} Q + Q^t & P^t - P \\ P - P^t & Q + Q^t \end{pmatrix},$$

where $P = (p_{jk})$ and $Q = (q_{jk})$. We define $\mathcal{P}$ to be the symmetric matrix whose entries are the absolute values of the matrix entries of $\mathbf{P}$:

$$\mathcal{P} = \frac{1}{2}\begin{pmatrix} P^\dagger & P^\# \\ P^\# & P^\dagger \end{pmatrix}, \qquad (P^\dagger)_{jk} = |p_{jk} + p_{kj}|, \qquad (P^\#)_{jk} = |q_{kj} - q_{jk}| .$$

Then

$$\operatorname{Tr} \mathcal{P}^2 = \frac{1}{2}\sum_{j,k} \Big(|p_{jk} + p_{kj}|^2 + |q_{kj} - q_{jk}|^2 \Big) \le 2 \sum_{j,k} \big[p_{jk}^2 + q_{jk}^2 \big] .$$

We apply the non-symmetric version of of the Hanson-Wright theorem [16] for X_1 and X_2 separately; note that the components of $\mathbf{w}$ are i.i.d. Together with the bound $\|\mathcal{P}\| \le \sqrt{\operatorname{Tr} \mathcal{P}^2}$ we have

$$\mathbb{P}(|X_1| \ge \delta) \le 2\exp\big(-c\min\{\delta/\sqrt{\operatorname{Tr}\mathcal{P}^2}, \delta^2/\operatorname{Tr}\mathcal{P}^2\}\big)$$

for some constant c depending on δ_0 and D from (1.3). Similar estimate holds for X_2, so we have

$$\mathbb{P}(|X| \ge \delta) \le 4\exp\big(-c\min\{\delta/A, \delta^2/A^2\}\big)$$

where $A^2 = \sum_{j,k} |a_{jk}|^2 = \sum_{j,k} \big[|p_{jk}|^2 + |q_{jk}|^2 \big]$.

In the second case in (1.2), when the distribution of the complex random variable b_j is rotationally symmetric, we can directly extend the proof [8] (note that [8] uses the notation X_j for b_j). We first symmetrize the quadratic form X by replacing a_{jk} with $\frac{1}{2}[a_{jk} + \overline{a}_{kj}]$. We then follow the proof in [8] and note that the only change is that Z used starting from Lemma 2 in [8] will be a standard complex Gaussian random variable and instead of Z^2 or Z^{2n} we consider $|Z|^2 = Z\overline{Z}$ and $|Z|^{2n}$, and similarly X^{2n} is replaced by $|X|^{2n}$, $n = 1, 2, \ldots$. With these changes, Lemma 1–6 in [8] hold true for the complex case as well. In the proof of the theorem, starting on page 1082 of [8], instead of $\prod_i \mathbb{E}X_i^{\alpha_i}(X_i^2 - \mathbb{E}X_i^2)^{\beta_i}$ the expansion will contain terms of the form $\prod_i \mathbb{E}X_i^{\alpha_i}\overline{X}_i^{\alpha_i'}(|X_i|^2 - \mathbb{E}|X_i|^2)^{\beta_i}$. Due to the rotational symmetry of the distribution, these terms are all zero (case (i) on page 1082 of [8]) unless $\alpha_i = \alpha_i'$ for all i. In the latter case, the bound $|\mathbb{E}|X_i|^{2\alpha_i}(|X_i|^2 - \mathbb{E}|X_i|^2)^{\beta_i}| \le \lambda^{2\alpha_i + 2\beta_i}\mathbb{E}|Z_i|^{2\alpha_i}(|Z_i|^2 - 1)^{\beta_i}$ holds with a sufficiently large λ (depending on δ_0 from (1.3)) exactly as in case (ii) on page 1082 of [8]. From now on the proof is unchanged and we obtain

$$\mathbb{P}\Big(\Big| \sum_{jk} a_{jk}(b_j\overline{b}_k - \mathbb{E}b_j\overline{b}_k) \Big| \ge \delta \Big) \le 2\exp\big(- c\min(\delta/A, \delta^2/A^2) \big) ,$$

where $\sum_{jk} \big| \frac{1}{2}[a_{jk} + \overline{a}_{kj}] \big|^2$ was estimated by $A^2 = \sum_{jk} |a_{jk}|^2$ from above. $\square$

References

[1] Anderson, G. W., Guionnet, A., Zeitouni, O.: Lecture notes on random matrices. Book in preparation.

[2] Bai, Z. D., Miao, B., Tsay, J.: Convergence rates of the spectral distributions of large Wigner matrices. *Int. Math. J.* **1** (2002), no. 1, 65–90.

[3] Bobkov, S. G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999), no. 1, 1–28.

[4] Deift, P.: Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. *Courant Lecture Notes in Mathematics* **3**, American Mathematical Society, Providence, RI, 1999

[5] Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. 2007, preprint. arXiv.org:0803.0542

[6] Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. 2008, preprint. arXiv.org:0711.1730

[7] Guionnet, A., Zeitouni, O.: Concentration of the spectral measure for large matrices. *Electronic Comm. in Probability* **5** (2000) Paper 14.

[8] Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Math. Stat.* **42** (1971), no.3, 1079-1083.

[9] Johansson, K.: Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* **215** (2001), no.3. 683–705.

[10] Khorunzhy, A.: On smoothed density of states for Wigner random matrices. *Random Oper. Stoch. Eq.* **5** (1997), no.2., 147–162.

[11] Ledoux, M.: The concentration of measure phenomenon. Mathematical Surveys and Monographs, **89** American Mathematical Society, Providence, RI, 2001.

[12] Mehta, M.I.: Random Matrices. New York, Academic Press, 1991.

[13] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian matrix models. J. Stat. Phys. **130** (2008), no.2., 205-250.

[14] Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999), no.3. 697-733.

[15] Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* **62** (1955), 548-564.

[16] Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** No. 6. (1973), 1068-1070.