LOCAL SEMICIRCLE LAW UNDER MOMENT CONDITIONS. PART II: LOCALIZATION AND DELOCALIZATION.

F. GÖTZE, A. NAUMOV, AND A. N. TIKHOMIROV

Abstract. We consider a random symmetric matrix $\mathbf{X} = [X_{jk}]_{j,k=1}^n$ with upper triangular entries being independent identically distributed random variables with mean zero and unit variance. We additionally suppose that $\mathbb{E}|X_{11}|^{4+\delta} =: \mu_{4+\delta} < C$ for some $\delta > 0$ and some absolute constant C. Under these conditions we show that the typical Kolmogorov distance between the empirical spectral distribution function of eigenvalues of $n^{-1/2}\mathbf{X}$ and Wigner's semicircle law is of order $1/n$ up to some logarithmic correction factor. As a direct consequence of this result we establish that the semicircle law holds on a short scale. Furthermore, we show for this finite moment ensemble rigidity of eigenvalues and delocalization properties of the eigenvectors. Some numerical experiments are included illustrating the influence of the tail behavior of the matrix entries when only a small number of moments exist.

1. Introduction and main result

This paper is the second part of the project aimed to establish local semicircle law under moment conditions. For the readers convenience we shortly recall the most important notions of our setup in the first part [15] and give a very short survey of recent results. We consider a random symmetric matrix $\mathbf{X} = [X_{jk}]_{j,k=1}^n$ with upper triangular entries being independent random variables with mean zero and unit variance. Denote the n eigenvalues of the symmetric matrix $\mathbf{W} := \frac{1}{\sqrt{n}}\mathbf{X}$ in the increasing order by

$$\lambda_1(\mathbf{W}) \le ... \le \lambda_n(\mathbf{W})$$

and introduce the eigenvalue counting function

$$N_I(\mathbf{W}) := |\{1 \le k \le n : \lambda_k(\mathbf{W}) \in I\}|$$

for any interval $I \subset \mathbb{R}$, where $|A|$ denotes the number of elements in the set A. Note that sometimes we shall omit $\mathbf{W}$ from the notation of $\lambda_j(\mathbf{W})$.

Date: December 7, 2015.

Key words and phrases. Local semicircle law, rate of convergence to the semicircle law, rigidity of eigenvalues, delocalization of eigenvectors, moment conditions.

All authors were supported by CRC 701 "Spectral Structures and Topological Methods in Mathematics". A. Naumov and A. Tikhomirov were supported by RFBR N 14-01-00500. A. Tikhomirov was also supported by Program of UD RAS, project No 15-16-1-3. A. Naumov was also supported by Hong Kong RGC GRF 403513.

It is well known since the pioneering work of E. Wigner [31] that for any interval $I \subset \mathbb{R}$ of fixed length and independent of n

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\, N_I(\mathbf{W}) = \int_I g_{sc}(\lambda)\, d\lambda, \tag{1.1}$$

where

$$g_{sc}(\lambda) := \frac{1}{2\pi}\sqrt{4-\lambda^2}\, \mathbb{1}[|\lambda| \le 2]$$

is the density function of Wigner's semicircle law. Here and in what follows we denote by $\mathbb{1}[A]$ the indicator function of the set A. Wigner considered the special case when all X_{jk} take only two values ± 1 with equal probabilities.Wigner's semicircle law has been extended in various aspects, see, for example, [2], [25], [14], [20], [24] and [16] and etc. For an extensive list of references we refer to the monographs [1], [4] and [26].

All these results hold for intervals I of fixed length, independent of n, which typically contain a macroscopically large number of eigenvalues, which means a number of order n. It is of the great interest to investigate the case of smaller intervals where the number of eigenvalues cease to be macroscopically large. Here an appropriate analytical for asymptotic approximations is the Stieltjes transform of the empirical spectral distribution function F_n, which is is given by

$$m_n(z) := \int_{-\infty}^{\infty} \frac{dF_n(\lambda)}{\lambda - z} = \frac{1}{n}\operatorname{Tr}(\mathbf{W} - z\mathbf{I})^{-1} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j(\mathbf{W}) - z},$$

where $z = u + iv, v \ge 0$. Taking the imaginary part of $m_n(z)$ we get

$$\operatorname{Im} m_n(u+iv) = \int_{-\infty}^{\infty} \frac{v}{(\lambda - u)^2 + v^2}\, dF_n(\lambda) = \frac{1}{v}\int_{-\infty}^{\infty} K\left(\frac{u-\lambda}{v}\right) dF_n(\lambda)$$

which is the kernel density estimator with Poisson's kernel K and bandwidth v. For a meaningful estimator of the spectral density we cannot allow the distance v to the real line, that is the bandwidth of the kernel density estimator, to be smaller than the typical $\frac{1}{n}$ -distance between eigenvalues. Hence, in what follows we shall be mostly interested in the situations when $v \gg \frac{1}{n}$.

Under rather general conditions for fixed $v > 0$ one may establish the convergence of $m_n(z)$ to the the Stieltjes transform of Wigner's semicircle law which is given by

$$s(z) = \int_{-\infty}^{\infty} \frac{g_{sc}(\lambda)\, d\lambda}{\lambda - z} = -\frac{z}{2} + \sqrt{\frac{z^2}{4} - 1}.$$

It is much more difficult to establish the convergence in the region $1 \gg v \gg \frac{1}{n}$. Significant progress in that direction was recently made in a series of results by L. Erdös, B. Schlein, H.-T. Yau and et al., [12], [11] , [13], [9], showing that with high probability uniformly in $u \in \mathbb{R}$

$$|m_n(u+iv) - s(u+iv)| \le \frac{\log^c n}{nv}, \tag{1.2}$$

which they called *local semicircle law*. It means that the fluctuations of $m_n(z)$ around $s(z)$ are of order $(nv)^{-1}$ (up to a logarithmic factor). To prove (1.2) in those papers [12], [11], [13] it was assumed that the distribution of X_{jk} for all $1 \le j, k \le n$ has sub-exponential tails. Moreover in [9] this assumption had been relaxed to requiring $\mathbb{E}\,|X_{jk}|^p \le \mu_p$ for all $p \ge 1$, where μ_p are some constants. Since there is meanwhile an extensive literature on the local semicircle law we refrain from providing a complete list here and refer the reader to the surveys of L. Erdös [7] and T. Tao, V. Vu, [27].

Our main goal in [15] was to show that (1.2) holds assuming that $\mathbb{E}\,|X_{jk}|^{4+\delta} =: \mu_{4+\delta} < \infty$. The first proof of a result of this type follows from a combination of arguments in a series of papers [10], [8], [23] (we sketched the underlying main ideas in the introduction of [15]). In [15] we gave a self-contained proof based on the method from [21], [18] while at the same time reducing the power of $\log n$ from $c = \log\log n$ to some constant with precise dependence on δ and independent of n. Our work and some crucial bounds of our proof were motivated by the methods used in a recent paper of C. Cacciapuoti, A. Maltsev and B. Schlein, [5], where the authors improved the log-factor dependence in (1.2) in the sub-Gaussian case.

For a detailed statement of our result recall that the conditions (**C0**) hold if $X_{jk}, 1 \le j \le k \le n$ are i.i.d. with zero mean, unit variance and $\mathbb{E}\,|X_{11}|^{4+\delta} := \mu_{4+\delta} < \infty$ for some $\delta > 0$. We also introduce the following quantity

$$\alpha = \frac{2}{4+\delta},$$

which will control the level of truncation of the matrix entries. It was proved in the paper [15][Theorem 1.1] that under conditions (**C0**) and any fixed $V > 0$ there exist positive constants A_0, A_1 and C depending on α and V such that

$$\mathbb{E}\,|m_n(z) - s(z)|^p \le \left(\frac{Cp^2}{nv}\right)^p, \tag{1.3}$$

for all $1 \le p \le A_1(nv)^{\frac{1-2\alpha}{2}}$, $v_1 \ge v \ge A_0 n^{-1}$ and $|u| \le 2 + v$. Moreover, for any $u_0 > 0$ there exist positive constants A_0, A_1 and C depending on u_0, V and α such that

$$\mathbb{E}\,|\operatorname{Im} m_n(z) - \operatorname{Im} s(z)|^p \le \left(\frac{Cp^2}{nv}\right)^p, \tag{1.4}$$

for all $1 \le p \le A_1(nv)^{\frac{1-2\alpha}{2}}$, $v_1 \ge v \ge A_0 n^{-1}$ and $|u| \le u_0$.

In the current paper we apply (1.4) and establish an estimate for the rate of convergence in probability of F_n to $G_{sc}(x) := \int_{-\infty}^x g_{sc}(\lambda)\,d\lambda$, the rigidity of eigenvalues and delocalization properties of the eigenvectors. We will formulate these results in the sequel and discuss them.

Let us denote

$$\Delta_n^* := \sup_{x\in\mathbb{R}} |F_n(x) - G_{sc}(x)|.$$

F. Götze and A. Tikhomirov in [19] proved that assuming $\mathbb{E}\,|X_{11}|^{12} =: \mu_{12} < \infty$, one may obtain the following estimate

$$\mathbb{E}\,\Delta_n^* \le \mu_{12}^{\frac{1}{6}} n^{-\frac{1}{2}}.$$

Particularly this estimate implies by Markov's inequality that

$$\mathbb{P}\,(\Delta_n^* \ge K) \le \frac{\mu_{12}^{\frac{1}{6}}}{K n^{\frac{1}{2}}}. \tag{1.5}$$

It is easy to see from the previous bound that one may take $K \gg n^{-\frac{1}{2}}$. This result was extended by Bai and et al., see [3], where it was shown that instead of existence of the 12th moment it suffices to assume existence of the 6th moment. Applying (1.3) we may obtain a much stronger bound.

Theorem 1.1. *Assume that the condition* (**C0**) *holds. Then there exist positive constants* c *and* C *depending on* α *only such that for all* $1 \le p \le c \log n$

$$\mathbb{P}\,(\Delta_n^* \ge K) \le \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p n^p}$$

for all $K > 0$.

As a consequence we may choose $K \gg n^{-1}$ which is optimal. In particular, taking $K = n^{-1} \log^{\kappa} n$, where $\kappa := 1 + \frac{2}{1-2\alpha}$, we get that

$$\mathbb{P}\left(\Delta_n^* \ge \frac{\log^{\kappa} n}{n}\right) \le \frac{1}{n^{c \log\log n}}. \tag{1.6}$$

Under additional assumptions (1.6) was proved in [17], [28] and [18]. Comparing our result with [23]Theorem 3.6] note that we reduced the logarithmic factor and give explicit dependence on δ. Using out technique it is possible to reduce the power of logarithm in the stochastic size of Δ_n^* to 1 assuming that the distribution of X_{11} has sub-Gaussian decay, for details see Tikhomirov and Timushev (in preparation). The optimal power of logarithm is $\frac{1}{2}$ due to a result of Gustavsson [22]. In Section 5 we provide some numerical experiments to illustrate the bounds of Theorem 1.1.

Let $N[x - \frac{\xi}{2n}, x + \frac{\xi}{2n}] := N_I(\mathbf{W})$ for $I = [x - \frac{\xi}{2n}, x + \frac{\xi}{2n}], \xi > 0$. The following result is the direct corollary of Theorem 1.1.

Corollary 1.2. *Assume that condition* (**C0**) *holds. Then there exist positive constants* c *and* C *depending on* α *such that for all* $1 \le p \le c \log n$ *and all* $\xi > 0, K > 0$

$$\mathbb{P}\left(\left|\frac{N[x - \frac{\xi}{2n}, x + \frac{\xi}{2n}]}{\xi} - g_{sc}(x)\right| \ge \frac{K}{\xi}\right) \le \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p n^p}.$$

Another application of (1.3) is the following result which shows the rigidity of the eigenvalues. Let us define the quantile position of the j-th eigenvalue by

$$\gamma_j : \quad \int_{-\infty}^{\gamma_j} g_{sc}(\lambda)\, d\lambda = \frac{j}{n}, \quad 1 \le j \le N.$$

We will prove the following theorem.

Theorem 1.3. *Assume that the conditions* (**C0**) *hold. Then*
(i). For all $j \in [K, n-K+1]$ *there exist constants* c *and* C, C_1 *depending on* α *such that such that for all* $1 \le p \le c\log n$ *we have*

$$\mathbb{P}(|\lambda_j - \gamma_j| \ge C_1 K[\min(j, n-j+1)]^{-\frac{1}{3}} n^{-\frac{2}{3}}) \le \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p}.$$

(ii). Assume that $\delta = 4$. *For all* $j \le K$ *or* $j \ge n-K+1$ *there exist constants* c *and* C, C_1 *such that for* $5 \le p \le c\log n$ *and any* $\phi > 0$

$$\mathbb{P}(|\lambda_j - \gamma_j| \ge C_1 K[\min(j, n-j+1)]^{-\frac{1}{3}} n^{-\frac{2}{3}}) \le \frac{C}{n^{2-\phi}} + \frac{C^p \log^{18p} n}{K^p}.$$

Let us complement the results of this theorem by the following remarks. First we refer the interested reader to relevant results [22] (Gaussian case), [9][Theorem 7.6], [10][Theorem 2.13], [17][Remark 1.2], [23][Theorem 3.6] and [5][Theorem 4]. In particular, the result under comparable moment conditions in [23][Theorem 3.6] has an additional factor $\log^{c\log\log n} n$ which in our case may be reduced to $\log^{\kappa} n$.

The bound in the bulk of the limit spectrum, that is part (i), holds for all $\delta > 0$. Since the proof of this part is based on Theorem 1.1 we expect that it should be valid for $\delta = 0$ as well. It is shown in the proof that with high probability $n - n\Delta_n^*$ eigenvalues lie in the support of the semicircle law. Applying this fact we may use the well-known Smirnov transform from mathematical statistics together with the bound from Theorem 1.1. Concerning the edges of the limit spectrum, that is part (ii), we have to assume in addition that there exist a moment of order eight (corresponding to $\delta = 4$) to prove part (i). In this step we use ideas from [5][Lemma 8.1] and [9][Theorem 7.6]. It is still possible to get a bound for smaller δ, $0 < \delta < 4$, but here our methods allow to prove that the estimate in part (ii) holds with small probability of order $n^{-\varepsilon}$ only, where $\varepsilon := \varepsilon(\delta) > 0$. In order to improve this error to $O(n^{-2+\phi})$ we have to assume the existence of eight moments. The main problem here is to estimate the distance between $\max_{1\le k\le n} |\lambda_k(\mathbf{W})|$ and $\max_{1\le k\le n} |\lambda_k(\hat{\mathbf{W}})|$, where $\hat{\mathbf{W}}$ is the random matrix with entries from $\mathbf{W}$, but truncated on the level of order n^{α} (see the definition in the proof of Theorem 1.3). This dependence on the tails of the distribution of entries is illustrated in Section 5 with numerical experiments, where we try to explain the role of matrix truncation.

To prove Theorem 1.3 we need to apply stronger bounds for the distance between Stieltjes transforms then (1.3). Let us denote

$$\gamma := \gamma(u) := ||u| - 2|. \tag{1.7}$$

We say that the conditions (**C1**) hold if (**C0**) are satisfied and $|X_{jk}| \le Dn^{\alpha}, 1 \le j, k \le n$, where $D := D(\alpha)$ is some positive constant. We also denote

$$\mathcal{E}_p := \frac{C^p p^p}{n^p(\gamma+v)^p} + \frac{C^p p^{3p}}{(nv)^{2p}(\gamma+v)^{\frac{p}{2}}} + \frac{C^p}{n^p v^{\frac{p}{2}}(\gamma+v)^{\frac{p}{2}}} + \frac{C^p p^p}{(nv)^{\frac{3p}{2}}(\gamma+v)^{\frac{p}{4}}}.$$

It was shown in [15][Theorem 1.2] that assuming the conditions (**C1**) hold, $u_0 > 2$ and $V > 0$, there exist positive constants A_0, A_1 and C depending on u_0, V and α such that

$$\mathbb{E}\,|\operatorname{Im} m_n(z) - \operatorname{Im} s(z)|^p \le \mathcal{E}_p, \tag{1.8}$$

for all $1 \le p \le A_1 (nv)^{\frac{1-2\alpha}{2}}$, $v_1 \ge v \ge A_0 n^{-1}$ and $2 \le |u| \le u_0$. See Theorem 1.2 in [15] for details.

We conclude this paper by showing delocalization of eigenvectors. This question has been intensively studied in many papers, for example, in [12] [17], [10] and [8]. Let us denote by $u_j := (u_{j1}, ..., u_{jn})$ the eigenvectors of $\mathbf{W}$ corresponding to the eigenvalue $\lambda_j(\mathbf{W})$.

Theorem 1.4. *Assume that conditions* (**C0**) *hold with* $\delta = 4$. *Then there exist positive constants* C *and* C_1 *such that*

$$\mathbb{P}\left(\max_{1\le j,k\le n} |u_{jk}|^2 \ge \frac{C_1 \log^8 n}{n}\right) \le \frac{C}{n}.$$

Similarly as in Theorem 1.3 we restrict ourselves here to the case $\delta = 4$ only.

We mention here that it is possible to extend the result for $0 < \delta < 4$ but reducing the power in the bound in probability from 1 to some positive constant ε depending on δ only. In the case $\delta = 4$ our methods yield the following bound

$$\mathbb{P}\left(\max_{1\le j,k\le n} |u_{jk}|^2 \ge \frac{C_1 \log^{4+\varepsilon'} n}{n}\right) \le \frac{C}{n^{c(\varepsilon')}},$$

for any $\varepsilon > 0$ and some positive constant $c(\varepsilon')$ depending on ε'. We omit the details. See Section 5 for numerical experiments illustrating this remark.

We finally remark that applying a moment matching technique as used in [8][Inequality 7.12], and [10][Remark 2.18] one may prove the following bound assuming the conditions of Theorem 1.4

$$\mathbb{P}\left(\max_{1\le j,k\le n} |u_{jk}|^2 \ge \frac{C_1 \log^{8+\varepsilon} n}{n}\right) \le \frac{C}{n^{2-\varepsilon}},$$

for any $\varepsilon > 0$.

1.1. **Notations.** Throughout the paper we will use the following notations. We assume that all random variables are defined on common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $\mathbb{E}$ the mathematical expectation with respect to $\mathbb{P}$.

We denote by $\mathbb{R}$ and $\mathbb{C}$ the set of all real and complex numbers. We also define $\mathbb{C}^+ := \{z \in \mathbb{C} : \operatorname{Im} z \ge 0\}$. Let $\mathbb{T} = [1, ..., n]$ denotes the set of the first n positive integers. For any $\mathbb{J} \subset \mathbb{T}$ introduce $\mathbb{T}_{\mathbb{J}} := \mathbb{T} \setminus \mathbb{J}$.

For any matrix $\mathbf{W}$ together with its resolvent $\mathbf{R}$ and Stieltjes transform m_n we shall systematically use the corresponding notions $\mathbf{W}^{(\mathbb{J})}, \mathbf{R}^{(\mathbb{J})}, m_n^{(\mathbb{J})}$, respectively, for the sub-matrix of $\mathbf{W}$ with entries $X_{jk}, j, k \in \mathbb{T} \setminus \mathbb{J}$.

By C and c we denote some absolute positive constants.

For an arbitrary matrix $\mathbf{A}$ taking values in $\mathbb{C}^{n\times n}$ we define the operator norm by $\|\mathbf{A}\| := \sup_{x\in\mathbb{R}^n:\|x\|=1}\|\mathbf{A}x\|_2$, where $\|x\|_2 := \sum_{j=1}^n |x_j|^2$. We also define the Hilbert-Schmidt norm by $\|\mathbf{A}\|_2 := \operatorname{Tr}\mathbf{A}\mathbf{A}^* = \sum_{j,k=1}^n |\mathbf{A}_{jk}|^2$. By $\binom{2m}{m}$ we denote the binomial number $\frac{(2m)!}{m!m!}$.

1.2. **Acknowledgment.** We would like to thank L. Erdös and H.-T. Yau for drawing our attention to relevant previous results and papers in connection with the results of this paper, in particular, [8], [9], [10] and [23].

2. Rate of convergence in probability

In this section we prove Theorem 1.1. We estimate the difference between F_n and G_{sc} in Kolmogorov's metric via the distance between corresponding Stieltjes transforms. For this purpose we formulate the following smoothing inequality from [19][Corollary 2.3], which allows to relate distribution functions to their Stieltjes transforms. For all $x \in [-2,2]$ let us define $\gamma(x) := 2-|x|$. Given $\frac{1}{2} > \varepsilon > 0$ we introduce the following intervals $\mathbb{J}_\varepsilon := \{x \in [-2,2] : \gamma(x) \ge \varepsilon\}$ and $\mathbb{J}_\varepsilon' := \mathbb{J}_{\varepsilon/2}$.

Lemma 2.1. *Let $v_0 > 0, a > 0$ and $\frac{1}{2} > \varepsilon > 0$ be positive numbers such that*

$$\frac{1}{\pi}\int_{|u|\le a}\frac{1}{u^2+1}\,du = \frac{3}{4} =: \beta,$$

and

$$2v_0 a \le \varepsilon^{\frac{3}{2}}.$$

Then for any $V > 0$ and $v' := v'(x) := v_0/\sqrt{\gamma(x)}, x \in \mathbb{J}_\varepsilon'$, there exist positive constants C_1 and C_2 such that the following inequality holds

$$\Delta_n^* \le \int_{-\infty}^{\infty} |m_n(u+iV) - s(u+iV)|\,du + C_1 v_0 + C_2\varepsilon^{\frac{3}{2}} + \sup_{x\in\mathbb{J}_\varepsilon'}\left|\int_{v'}^{V}(m_n(x+iv) - s(x+iv))\,dv\right|.$$

Proof. See [19][Corollary 2.3] or [18][Proposition 2.1]. □

It what follows we will need the following version of this lemma.

Corollary 2.2. *Assuming the conditions of Lemma 2.1 we have*

$$\mathbb{E}^{\frac{1}{p}}[\Delta_n^*]^p \le \int_{-\infty}^{\infty}\mathbb{E}^{\frac{1}{p}}\,|m_n(u+iV) - s(u+iV)|^p\,du + C_1 v_0 + C_2\varepsilon^{\frac{3}{2}} + \mathbb{E}^{\frac{1}{p}}\sup_{x\in\mathbb{J}_\varepsilon'}\left|\int_{v'}^{V}(m_n(x+iv) - s(x+iv))\,dv\right|^p. \tag{2.1}$$

Proof. The proof is the direct consequence of the previous lemma and we omit it. For details the interested reader is referred to [18][Corollary 2.1]. □

Proof of Theorem 1.1. We proceed as in the proof of Theorem 1.1 in [18]. We choose in Corollary 2.2 the following values for the parameters v_0, ε and V. Let us take $v_0 := A_0 n^{-1} \log^{\frac{2}{1-2\alpha}} n$, $\varepsilon := (2v_0 a)^{\frac{2}{3}}$ and $V := 4$. We may partition $\mathbb{J}'_\varepsilon$ into $k_n := n^4$ disjoint subintervals of equal length. Let us denote the endpoints of these intervals by $x_k, k = 0, ..., k_n$. We get $-2+\varepsilon = x_0 < x_1 < ... < x_{k_n} = 2-\varepsilon$. For simplicity we denote $\Lambda_n(u+iv) := m_n(u+iv) - s(u+iv)$ but we will not omit the argument. We start to estimate the second integral in the r.h.s. of (2.1). It is easy to see that

$$\sup_{x\in\mathbb{J}'_\varepsilon} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right| \le \max_{1\le k\le k_n} \sup_{x_{k-1}\le x\le x_k} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right|. \tag{2.2}$$

Applying the Newton-Leibniz formula we may write

$$\begin{aligned}\sup_{x_{k-1}\le x\le x_k} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right| \le & \left| \int_{v'}^V \Lambda_n(x_{k-1}+iv)\,dv \right| \\ &+ \int_{x_{k-1}}^{x_k} \int_{v'}^V |\Lambda'_n(x+iv)|\,dv\,dx. \end{aligned}\tag{2.3}$$

It follows from Cauchy's integral formula that for all $z = x+iv$ with $v \ge v_0$ we have

$$|\Lambda'_n(x+iv)| \le \frac{C}{v^2} \le Cn^2. \tag{2.4}$$

We may conclude from (2.3) and (2.4) that

$$\sup_{x_{k-1}\le x\le x_k} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right| \le \left| \int_{v'}^V \Lambda_n(x_{k-1}+iv)\,dv \right| + \frac{C}{n}.$$

Applying this inequality to (2.2) and taking the mathematical expectation we obtain

$$\begin{aligned}\mathbb{E}\sup_{x\in\mathbb{J}'_\varepsilon} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right|^p &\le \mathbb{E}\max_{1\le k\le k_n} \left| \int_{v'}^V \Lambda_n(x_{k-1}+iv)\,dv \right|^p + \frac{C^p}{n^p} \\ &\le \sum_{k=1}^{k_n} \left| \int_{v'}^V \mathbb{E}^{\frac{1}{p}} \left|\Lambda_n(x_{k-1}+iv)\right|^p dv \right|^p + \frac{C^p}{n^p}. \end{aligned}\tag{2.5}$$

Since $x \in \mathbb{J}'_\varepsilon$ it follows from (1.3) that

$$\mathbb{E}\,|\Lambda_n(x+iv)|^p \le \left(\frac{Cp^2}{nv}\right)^p. \tag{2.6}$$

Choosing $p = A_1(nv_0)^{\frac{1-2\alpha}{2}} = c\log n$ we finally get from (2.5) and (2.6) that

$$\mathbb{E}^{\frac{1}{p}} \sup_{x\in\mathbb{J}'_\varepsilon} \left| \int_{v'}^V \Lambda_n(x+iv)\,dv \right|^p \le \frac{Ck_n^{\frac{1}{p}}\log^3 n}{n} + \frac{C}{n} \le \frac{C\log^3 n}{n}. \tag{2.7}$$

It remains to estimate the first of the integrals in (2.1). Let us suppose that we have already shown the following bound

$$\mathbb{E}^{\frac{1}{p}}\,|\Lambda_n(u+iV)|^p \le \frac{Cp|s(z)|^{\frac{p+1}{p}}}{n}, \tag{2.8}$$

valid for all $z = u + iV, u \in \mathbb{R}$. Hence,

$$\int_{-\infty}^{\infty} \mathbb{E}^{\frac{1}{p}}\,|\Lambda_n(u+iV)|^p\,du \le \frac{Cp}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{du\,dG_{sc}(x)}{((x-u)^2+V^2)^{\frac{p+1}{2p}}} \le \frac{C\log^2 n}{n}. \tag{2.9}$$

Combining now (2.1), (2.7) and (2.9) we get

$$\mathbb{E}^{\frac{1}{p}}[\Delta_n^*]^p \le \frac{C\log^{\frac{2}{1-2\alpha}} n}{n}.$$

Since $\mathbb{E}^{\frac{1}{p}}[\Delta_n^*]^p$ is non-decreasing function of p, the last inequality remains valid for all $1 \le p \le c\log n$. To finish the proof of Theorem 1.1 it remains to apply Markov's inequality

$$\mathbb{P}\left(\Delta_n^* \ge K\right) \le \frac{\mathbb{E}[\Delta_n^*]^p}{K^p} \le \frac{C^p\log^{\frac{2p}{1-2\alpha}} n}{K^p n^p}.$$

We conclude the proof by (2.8). To derive this bound we will proceed in the same way as in the proof of Theorem 2.1 in [15]. The main difference is that we don't need to estimate $\mathbb{E}\,|\mathbf{R}_{jj}|^p$, but we have to establish (2.8) for all $u \in \mathbb{R}$. Since means repeating the arguments in the proof of Theorem 2.1 in [15] we shall omit many details and routine calculations here.

Firstly is easy to show that one can assume that the entries of $\mathbf{X}$ satisfy the conditions (**C1**). We omit the details.

We start with a recursive representation for the diagonal entries $\mathbf{R}_{jj} = (\mathbf{W} - z\mathbf{I})^{-1}$ of the resolvent. We may express $\mathbf{R}_{jj}$ in the following way

$$\mathbf{R}_{jj} = \frac{1}{-z + \frac{X_{jj}}{\sqrt{n}} - \frac{1}{n}\sum_{l,k\in T_j} X_{jk}X_{jl}\mathbf{R}_{kl}^{(j)}}, \tag{2.10}$$

where $\mathbf{R}^{(j)}$ is defined in Section 1.1. Let $\varepsilon := \varepsilon_{1j} + \varepsilon_{2j} + \varepsilon_{3j} + \varepsilon_{4j}$, where

$$\varepsilon_{1j} = \frac{1}{\sqrt{n}}X_{jj}, \quad \varepsilon_{2j} = -\frac{1}{n}\sum_{l\ne k\in T_j} X_{jk}X_{jl}\mathbf{R}_{kl}^{(j)}, \quad \varepsilon_{3j} = -\frac{1}{n}\sum_{k\in T_j}(X_{jk}^2 - 1)\mathbf{R}_{kk}^{(j)},$$
$$\varepsilon_{4j} = \frac{1}{n}(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(j)}).$$

In these notations we may rewrite (2.10) as follows

$$\mathbf{R}_{jj} = -\frac{1}{z+m_n(z)} + \frac{1}{z+m_n(z)}\varepsilon_j\mathbf{R}_{jj}. \tag{2.11}$$

Let us denote $b(z) := z + 2s(z)$, $b_n(z) = b(z) + \Lambda_n(z)$ and

$$T_n := \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \mathbf{R}_{jj}. \tag{2.12}$$

Applying (2.11) we arrive at the following representation for Λ_n in terms of T_n and b_n

$$\Lambda_n = \frac{T_n}{z + m_n(z) + s(z)} = \frac{T_n}{b_n(z)}. \tag{2.13}$$

Now we show that for $V = 4$ and all $u \in \mathbb{R}$ one may estimate the denominator in (2.13). It is easy to check that

$$|m_n(z)| \le \frac{1}{4} \le \frac{1}{2}|z + s(z)| \quad \text{and} \quad |s(z) - m_n(z)| \le \frac{1}{2}. \tag{2.14}$$

These inequalities imply

$$|b_n(z)| \ge \frac{1}{2}|z + s(z)| \quad \text{and} \quad |z + m_n(z)| \ge \frac{1}{2}|s(z) + z|. \tag{2.15}$$

Moreover, since $1 + zs(z) + s^2(z) = 0$ we get

$$\frac{1}{|b_n(z)|} \le 2|s(z)| \quad \text{and} \quad |m_n(z)| \le |s(z)|(1 + 2|T_n|). \tag{2.16}$$

We rewrite (2.13) in the following way

$$\Lambda_n = \frac{1}{n}\sum_{j=1}^{n} \frac{\varepsilon_{4j}\mathbf{R}_{jj}}{b_n(z)} + \frac{1}{n}\sum_{\nu=1}^{3}\sum_{j=1}^{n} \frac{\varepsilon_{\nu j}\mathbf{R}_{jj}}{b_n(z)}.$$

Since

$$\sum_{j=1}^{n} \varepsilon_{4j}\mathbf{R}_{jj} = \frac{1}{n}\operatorname{Tr}\mathbf{R}^2 = m_n'(z)$$

we get that

$$\Lambda_n = \frac{1}{n}\frac{m_n'(z)}{b_n(z)} + \frac{1}{n}\sum_{\nu=1}^{3}\sum_{j=1}^{n} \frac{\varepsilon_{\nu j}\mathbf{R}_{jj}}{b_n(z)} = \frac{1}{n}\frac{m_n'(z)}{b_n(z)} + \widehat{\Lambda}_n,$$

where we denoted

$$\widehat{\Lambda}_n := \frac{1}{n}\sum_{\nu=1}^{3}\sum_{j=1}^{n} \frac{\varepsilon_{\nu j}\mathbf{R}_{jj}}{b_n(z)}.$$

Let us introduce the function $\varphi(z) = \overline{z}|z|^{p-2}$. Then

$$\mathbb{E}\,|\Lambda_n|^p = \mathbb{E}\,\Lambda_n\varphi(\Lambda_n) = \frac{1}{n}\,\mathbb{E}\,\frac{m_n'(z)}{b_n(z)}\varphi(\Lambda_n) + \mathbb{E}\,\widehat{\Lambda}_n\varphi(\Lambda_n).$$

Applying the result of Lemma C.11 we obtain a bound for the first term of the r.h.s. of the previous equation

$$\frac{1}{n}\left|\mathbb{E}\,\frac{m_n'(z)}{b_n(z)}\varphi(\Lambda_n)\right| \le \frac{C|s(z)|^2}{n}(1+\mathbb{E}^{\frac{1}{p}}\,|T_n|^p)\,\mathbb{E}^{\frac{p-1}{p}}\,|\Lambda_n|^p. \tag{2.17}$$

Cauchy-Schwartz inequality, Lemmas (C.1)– (C.5) and $\max_j|\mathbf{R}_{jj}|\le V^{-1}$ together imply that for all $p\le c\log n$ that

$$\mathbb{E}\,|T_n|^p \le \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n|\varepsilon_j|^2\right)^{\frac{p}{2}}\left(\frac{1}{n}\sum_{j=1}^n|\mathbf{R}_{jj}|^2\right)^{\frac{p}{2}} \le C.$$

From this inequality and (2.17) it follows that

$$\mathbb{E}\,|\Lambda_n|^p \le |\,\mathbb{E}\,\widehat{\Lambda}_n\varphi(\Lambda_n)| + \frac{|s(z)|^2}{n}\,\mathbb{E}^{\frac{p-1}{p}}\,|\Lambda_n|^p. \tag{2.18}$$

Now we consider the term $\mathbb{E}\,\widehat{\Lambda}_n\varphi(\Lambda_n)$. We split it into three parts with respect to $\varepsilon_{\nu j},\nu=1,2,3$ obtaining

$$\mathbb{E}\,\widehat{\Lambda}_n\varphi(\Lambda_n) = \frac{1}{n}\sum_{\nu=1}^3\sum_{j=1}^n\mathbb{E}\,\frac{\varepsilon_{\nu j}\mathbf{R}_{jj}}{b_n(z)}\varphi(\Lambda_n) = \mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3.$$

We rewrite $\mathcal{A}_\nu$ as a sum of two terms as follows

$$\mathcal{A}_{\nu 1} := \frac{s(z)}{n}\,\mathbb{E}\sum_{j=1}^n\frac{\varepsilon_{\nu j}}{b_n(z)}\varphi(\Lambda_n),$$

$$\mathcal{A}_{\nu 2} := \frac{1}{n}\sum_{j=1}^n\mathbb{E}\,\frac{\varepsilon_{\nu j}[\mathbf{R}_{jj}-s(z)]}{b_n(z)}\varphi(\Lambda_n).$$

From Hölder's inequality and Lemma C.2 with $q=1$ it follows that

$$|\mathcal{A}_{11}| \le |s(z)|^2\,\mathbb{E}^{\frac{1}{p}}\left|\frac{1}{n}\sum_{j=1}^n\varepsilon_{1j}\right|^p\mathbb{E}^{\frac{p-1}{p}}\,|\Lambda_n|^p \le \frac{Cp|s(z)|^2}{n}\,\mathbb{E}^{\frac{p-1}{p}}\,|\Lambda_n|^p. \tag{2.19}$$

To estimate $\mathcal{A}_{21}$ and $\mathcal{A}_{31}$ let us introduce the following notation

$$\widetilde{\Lambda}_n^{(j)} := \mathbb{E}(\Lambda_n\big|\mathfrak{M}^{(j)}),$$

where $\mathfrak{M}^{(j)} := \sigma\{X_{lk},l,k\in\mathbb{T}_j\}$. Since $\mathbb{E}(\varepsilon_{\nu j}\big|\mathfrak{M}^{(j)})=0$ for $\nu=2,3$ it is easy to see that $\mathcal{A}_{\nu 1}=\mathcal{B}_{\nu 1}+\mathcal{B}_{\nu 2}$, where

$$\mathcal{B}_{\nu 1} = \frac{s(z)}{n}\sum_{j=1}^n\mathbb{E}\,\frac{\varepsilon_{\nu j}}{b_n^{(j)}(z)}[\varphi(\Lambda_n)-\varphi(\widetilde{\Lambda}_n^{(j)})],$$

$$\mathcal{B}_{\nu 2} = \frac{s(z)}{n}\sum_{j=1}^n\mathbb{E}\,\frac{\varepsilon_{\nu j}\varepsilon_{4j}}{b_n(z)b_n^{(j)}(z)}\varphi(\Lambda_n).$$

Applying Lemma C.7 and (2.15) one may show that

$$|\mathcal{B}_{\nu 2}| \le \frac{C|s(z)|^2}{n} \mathbb{E}^{\frac{p-1}{p}} |\Lambda|^p. \tag{2.20}$$

From Newton-Leibniz formula (see Lemma C.12 for details) and the simple inequality $(x+y)^p \le ex^p + (p+1)^p y^p, x, y > 0, p \ge 1$ we get

$$|\mathcal{B}_{\nu 1}| \le \mathcal{C}_{\nu 1} + \mathcal{C}_{\nu 2},$$

where

$$\mathcal{C}_{\nu 1} := \frac{ep|s(z)|^2}{n} \sum_{j=1}^n \mathbb{E}\, |\varepsilon_{\nu j}||\Lambda_n - \widetilde{\Lambda}_n^{(j)}||\widetilde{\Lambda}_n^{(j)}|^{p-2},$$

$$\mathcal{C}_{\nu 2} := \frac{p^{p-2}|s(z)|^2}{n} \sum_{j=1}^n \mathbb{E}\, |\varepsilon_{\nu j}||\Lambda_n - \widetilde{\Lambda}_n^{(j)}|^{p-1}.$$

Applying the Schur complement formula (see for details [18][Lemma 7.23] or [19][Lemma 3.3]) we get

$$\operatorname{Tr} \mathbf{R} - \operatorname{Tr} \mathbf{R}^{(j)} = (1+\eta_j)\mathbf{R}_{jj}, \tag{2.21}$$

where $\eta_j := \eta_{0j} + \eta_{1j} + \eta_{2j}$ and

$$\eta_{0j} := \frac{1}{n} \sum_{k \in T_j} [(\mathbf{R}^{(j)})^2]_{kk} = (m_n^{(j)}(z))', \quad \eta_{1j} := \frac{1}{n} \sum_{k \ne l \in \mathbb{T}_j} X_{jl} X_{jk} [(\mathbf{R}^{(j)})^2]_{kl},$$

$$\eta_{2j} := \frac{1}{n} \sum_{k \in \mathbb{T}_j} [X_{jk}^2 - 1][(\mathbf{R}^{(j)})^2]_{kk}.$$

It follows from (2.21) and $\Lambda_n - \widetilde{\Lambda}_n^{(j)} = \Lambda_n - \Lambda_n^{(j)} - \mathbb{E}(\Lambda_n - \Lambda_n^{(j)} \big| \mathfrak{M}^{(j)})$ that

$$\Lambda_n - \widetilde{\Lambda}_n^{(j)} = \frac{1+\eta_{j0}}{n} [\mathbf{R}_{jj} - \mathbb{E}(\mathbf{R}_{jj} \big| \mathfrak{M}^{(j)})] + \frac{\eta_{1j} + \eta_{2j}}{n} \mathbf{R}_{jj} - \frac{1}{n} \mathbb{E}((\eta_{j1} + \eta_{j2}) \mathbf{R}_{jj} \big| \mathfrak{M}^{(j)}).$$

Let us introduce additional notations. We define $\hat{\varepsilon}_j := \varepsilon_{1j} + \varepsilon_{2j} + \varepsilon_{3j}$ and

$$a_n^{(j)} := \frac{1}{z + m_n^{(j)}(z)}.$$

It is easy to check that

$$\begin{aligned} |\mathbf{R}_{jj} - \mathbb{E}(\mathbf{R}_{jj} | \mathfrak{M}^{(j)})| &\le |a_n^{(j)}|(|\hat{\varepsilon}_j \mathbf{R}_{jj}| + \mathbb{E}(|\hat{\varepsilon}_j \mathbf{R}_{jj}| \big| \mathfrak{M}^{(j)})) \le C|s(z)|(|\hat{\varepsilon}_j| + \mathbb{E}(|\hat{\varepsilon}_j| \big| \mathfrak{M}^{(j)})) \\ &\le C|s(z)|(|\hat{\varepsilon}_j| + Cn^{-\frac{1}{2}}), \end{aligned}$$

and similarly,

$$|\eta_{1j} + \eta_{2j}||\mathbf{R}_{jj}| \le C|s(z)||\eta_{1j} + \eta_{2j}|(1 + |\varepsilon_j|).$$

Applying these inequalities we estimate $|\Lambda_n - \widetilde{\Lambda}_n^{(j)}|$ as follows

$$|\Lambda_n - \widetilde{\Lambda}_n^{(j)}| \le \frac{C|s(z)|}{n}(|\hat{\varepsilon}_j| + Cn^{\frac{1}{2}}) + \frac{C|s(z)|}{n}|\eta_{1j} + \eta_{2j}|(1 + |\varepsilon_j|) \tag{2.22}$$

Let us introduce the following quantity $\beta := \frac{1}{2\alpha}, \beta > 1$. Denote by ζ an arbitrary random variable such that the expectation $\mathbb{E}\,|\zeta|^{\frac{4\beta}{\beta-1}}$ exists. Then

$$\mathbb{E}(\varepsilon_{\nu j}|\Lambda_n - \widetilde{\Lambda}_n^{(j)}||\zeta|\big|\mathfrak{M}^{(j)}) \le B_1 + ... + B_6,$$

where

$$B_1 := \frac{C|s(z)|}{n}\,\mathbb{E}(|\varepsilon_{\nu j}\hat{\varepsilon}_j\zeta|\big|\mathfrak{M}^{(j)}), \qquad B_2 := \frac{C|s(z)|}{n^{\frac{3}{2}}}\,\mathbb{E}(|\varepsilon_{\nu j}\zeta|\big|\mathfrak{M}^{(j)}),$$
$$B_3 := \frac{C|s(z)|}{n}\,\mathbb{E}(|\varepsilon_{\nu j}\eta_{1j}||\zeta|\big|\mathfrak{M}^{(j)}), \qquad B_4 := \frac{C|s(z)|}{n}\,\mathbb{E}(|\varepsilon_{\nu j}\eta_{2j}||\zeta|\big|\mathfrak{M}^{(j)}),$$
$$B_5 := \frac{C|s(z)|}{n}\,\mathbb{E}(|\varepsilon_{\nu j}\eta_{1j}||\varepsilon_j||\zeta|\big|\mathfrak{M}^{(j)}), \qquad B_6 := \frac{C|s(z)|}{n}\,\mathbb{E}(|\varepsilon_{\nu j}\eta_{2j}||\varepsilon_j||\zeta|\big|\mathfrak{M}^{(j)}).$$

Applying Lemmas C.1– C.9 one may check that

$$\max_{k=1,...6} B_k \le \frac{C|s(z)|}{n^2}\,\mathbb{E}^{\frac{\beta-1}{2\beta}}(|\zeta|^{\frac{2\beta}{\beta-1}}\big|\mathfrak{M}^{(j)}).$$

Hence,

$$\mathbb{E}(\varepsilon_{\nu j}|\Lambda_n - \widetilde{\Lambda}_n^{(j)}||\zeta|\big|\mathfrak{M}^{(j)}) \le \frac{C|s(z)|}{n^2}\,\mathbb{E}^{\frac{\beta-1}{4\beta}}(|\zeta|^{\frac{4\beta}{\beta-1}}\big|\mathfrak{M}^{(j)}). \tag{2.23}$$

Taking $\zeta = 1$ in (2.23) we get

$$\begin{aligned}\mathcal{C}_{\nu 1} &= \frac{ep}{n}\sum_{j=1}^{n}\mathbb{E}\,|\widetilde{\Lambda}_n^{(j)}|^{p-2}\,\mathbb{E}(|\varepsilon_{\nu j}||\Lambda_n - \widetilde{\Lambda}_n^{(j)}|\big|\mathfrak{M}^{(j)}) \\ &\le \frac{Cp|s(z)|^3}{n^3}\sum_{j=1}^{n}\mathbb{E}\,|\widetilde{\Lambda}_n^{(j)}|^{p-2} \le \frac{Cp|s(z)|^3}{n^2}\,\mathbb{E}^{\frac{p-2}{p}}\,|\Lambda_n|^p. \end{aligned}\tag{2.24}$$

Similarly, applying (2.23) with $\zeta = |\Lambda_n - \widetilde{\Lambda}_n^{(j)}|^{p-2}$ we obtain

$$\mathcal{C}_{\nu 2} \le \frac{Cp^{p-2}|s(z)|^3}{n^3}\sum_{j=1}^{n}\mathbb{E}^{\frac{\beta-1}{4\beta}}\,|\Lambda - \widetilde{\Lambda}_n^{(j)}|^{\frac{4\beta(p-2)}{\beta-1}}. \tag{2.25}$$

It is easy to check (see (2.22)) that for all $q \ge 1$

$$\mathbb{E}\,|\Lambda - \widetilde{\Lambda}_n^{(j)}|^q \le \frac{C^q|s(z)|^q}{n^q}. \tag{2.26}$$

This inequality and (2.25) together imply

$$\mathcal{C}_{\nu 2} \le \frac{C^p p^{p-2}|s(z)|^{p+1}}{n^p}. \tag{2.27}$$

The estimates (2.19),(2.20), (2.24) and (2.27) yield

$$\sum_{\nu=1}^{3}\mathcal{A}_{\nu 1} \le \frac{Cp|s(z)|^2}{n}\,\mathbb{E}^{\frac{p-1}{p}}\,|\Lambda_n|^p + \frac{Cp|s(z)|^3}{n^2}\,\mathbb{E}^{\frac{p-2}{p}}\,|\Lambda_n|^p + \frac{C^p p^{p-2}|s(z)|^{p+1}}{n^p}. \tag{2.28}$$

It remains to estimate $\mathcal{A}_{\nu 2}, \nu = 1,2,3$. Recall that

$$\mathcal{A}_{\nu 2} := \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\, \frac{\varepsilon_{\nu j}[\mathbf{R}_{jj} - s(z)]}{b_n(z)} \varphi(\Lambda_n).$$

From the representation $\mathbf{R}_{jj} - s(z) = s(z)(\Lambda_n - \varepsilon_j)\mathbf{R}_{jj}$ it follows that

$$\mathcal{A}_{\nu 2} = \frac{s(z)}{n}\sum_{j=1}^{n} \mathbb{E}\, \frac{\varepsilon_{\nu j}(\Lambda_n - \varepsilon_j)\mathbf{R}_{jj}}{b_n(z)} \varphi(\Lambda_n).$$

We may bound $\mathcal{A}_{\nu 2}, \nu = 1,2,3$, by the sum of two terms (up to some constant) $\mathcal{N}_{\nu,1}$ and $\mathcal{N}_{\nu,2}, \nu = 1,2,3$, where

$$\mathcal{N}_{\nu 1} := \frac{e|s(z)|^2}{n}\sum_{j=1}^{n} \mathbb{E}\, |\varepsilon_{\nu j}||\Lambda_n - \varepsilon_j||\widetilde{\Lambda}_n^{(j)}|^{p-1},$$
$$\mathcal{N}_{\nu 2} := \frac{p^{p-1}|s(z)|^2}{n}\sum_{j=1}^{n} \mathbb{E}\, |\varepsilon_{\nu j}||\Lambda_n - \varepsilon_j||\Lambda_n - \widetilde{\Lambda}_n^{(j)}|^{p-1}.$$

Let us consider $\mathcal{N}_{\nu 1}$. Applying Lemmas C.1– C.7 we obtain

$$\mathcal{N}_{\nu 1} \le \frac{C|s(z)|^2}{n}\sum_{j=1}^{n} \mathbb{E}\, |\widetilde{\Lambda}_n^{(j)}|^{p-1}\, \mathbb{E}^{\frac12}(|\varepsilon_{\nu j}|^2 \big|\mathfrak{M}^{(j)})\, \mathbb{E}^{\frac12}(|\Lambda_n - \varepsilon_j|^2 \big|\mathfrak{M}^{(j)}) \le \frac{C|s(z)|^2}{n}\, \mathbb{E}^{\frac{p-1}{p}} |\Lambda_n|^p.$$

Similarly, in view of (2.26) we conclude

$$\mathcal{N}_{\nu 2} \le \frac{C^p p^{p-1}|s(z)|^{p+1}}{n^p}.$$

Finally we get the following inequality for the sum of the $\mathcal{A}_{\nu 2}, \nu = 1,2,3$

$$\sum_{\nu=1}^{3} \mathcal{A}_{\nu 2} \le \frac{C|s(z)|^2}{n}\, \mathbb{E}^{\frac{p-1}{p}} |\Lambda_n|^p + \frac{C^p p^{p-1}|s(z)|^{p+1}}{n^p}. \tag{2.29}$$

Combining(2.28) and (2.29) we get

$$\mathbb{E}\, |\Lambda_n|^p \le \frac{Cp|s(z)|^2}{n}\, \mathbb{E}^{\frac{p-1}{p}} |\Lambda_n|^p + \frac{Cp|s(z)|^3}{n^2}\, \mathbb{E}^{\frac{p-2}{p}} |\Lambda_n|^p + \frac{C^p p^{p-1}|s(z)|^{p+1}}{n^p}$$

Applying Lemma C.14 we obtain the following estimate

$$\mathbb{E}\, |\Lambda_n|^p \le \frac{C^p p^p|s(z)|^{p+1}}{n^p},$$

which conclude s the proof. $\square$

3. Rigidity of eigenvalues

In this section we prove Theorem 1.3. We start with a lemma which shows that with high probability all eigenvalues lie in the interval $[-2-Kn^{-\frac{2}{3}}, 2+Kn^{\frac{2}{3}}]$ for some large $K > 0$. Here we shall use methods similar to those in [5][Lemma 8.1] and [9][Theorem 7.6] adapting them to our setup.

Lemma 3.1. *Assume the conditions* (**C0**) *hold with* $\delta = 4$*. Then exist positive constants* c, C *such that for any* $\phi > 0$

$$\mathbb{P}\left(\|\mathbf{W}\| \geq 2 + \frac{K}{n^{\frac{2}{3}}}\right) \leq \frac{(Cp^{12})^p}{K^p} + \frac{C}{n^{2-\phi}}. \tag{3.1}$$

for all $4 < p \leq cn^{\frac{1}{24}}$ *and* $K > 0$*.*

Remark. We remark here that for case $0 < \delta < 4$ we are not getting a reasonable bound yet. We can only guarantee the existence of some $\varepsilon := \varepsilon(\delta) > 0$ such that (3.1) holds with probability less then $Cn^{-\varepsilon}$ for some C depending on δ only. The main problem here is to estimate the distance between $\max_{1\leq k\leq n}|\lambda_k(\mathbf{W})|$ and $\max_{1\leq k\leq n}|\lambda_k(\mathbf{W})|$. So far we can estimate the probability of the event $\mathbf{W} \neq \hat{\mathbf{W}}$ only which holds with very small probability depending on the level of truncation and hence, on δ. We omit the details, but refer instead to Section 5 with numerical results illustrating this behavior.

Proof of Lemma 3.1. Recall that $\lambda_1(\mathbf{W}) \leq ... \leq \lambda_n(\mathbf{W})$ and

$$\|\mathbf{W}\| = \max_{1\leq j\leq n} |\lambda_j(\mathbf{W})|$$

Hence, it is enough to prove that

$$\mathbb{P}\left(\lambda_1 \leq -2 - \frac{K}{n^{\frac{2}{3}}}\right) \leq \frac{(Cp^3)^p}{K^p} + \frac{C}{n^{2-\phi}} \quad \text{or} \quad \mathbb{P}\left(\lambda_n \geq 2 + \frac{K}{n^{\frac{2}{3}}}\right) \leq \frac{(Cp^3)^p}{K^p} + \frac{C}{n^{2-\phi}}.$$

Without loss of generality we consider only the bound for λ_1, since the same proof is valid for λ_n. Now we need to truncate the entries of $\mathbf{X}$. We introduce the usual notations. Let $\hat{X}_{jk} := X_{jk}\,\mathbb{1}[|X_{jk}| \leq Dn^{\frac{1}{2}-\phi}]$, $\tilde{X}_{jk} := X_{jk}\,\mathbb{1}[|X_{jk}| \geq Dn^{\frac{1}{2}-\phi}] - \mathbb{E}\,X_{jk}\,\mathbb{1}[|X_{jk}| \geq Dn^{\frac{1}{2}-\phi}]$ and finally $\breve{X}_{jk} := \tilde{X}_{jk}\sigma^{-1}$, where $\sigma^2 := \mathbb{E}\,|\tilde{X}_{11}|^2$. By $\hat{\mathbf{X}}, \tilde{\mathbf{X}}$ and $\breve{\mathbf{X}}$ we denote the symmetric random matrices with entries $\hat{X}_{jk}, \tilde{X}_{jk}$ and $\breve{X}_{jk}$ respectively. In a similar way we denote the resolvent matrices and corresponding Stieltjes transforms. In this case we have

$$\mathbb{P}(\mathbf{W} \neq \hat{\mathbf{W}}) \leq \frac{C}{n^{2-\phi}}.$$

It what follows we may assume $\mathbf{W} = \hat{\mathbf{W}}$. Let us fix some large positive constant, say u_0. Then by Lemma A.1 and A.2 in the Appendix we obtain that there exist some positive constants c, C and C_1 depending on u_0 such that

$$\mathbb{P}(\|\mathbf{W}\| \geq u_0) \leq e^{-n^c \log u_0} + \frac{C}{n^{2-\phi}} \leq \frac{C_1}{n^{2-\phi}}.$$

We conclude that

$$\mathbb{P}\left(\lambda_1(\mathbf{W}) \le -2 - \frac{K}{n^{\frac{2}{3}}}\right) \le \mathbb{P}\left(\lambda_1(\mathbf{W}) \le -u_0\right) + \mathbb{P}\left(-u_0 \le \lambda_1(\mathbf{W}) \le -2 - \frac{K}{n^{\frac{2}{3}}}\right)$$
$$\le \mathbb{P}\left(-u_0 \le \lambda_1(\mathbf{W}) \le -2 - \frac{K}{n^{\frac{2}{3}}}\right) + \frac{C_1}{n^{2-\phi}}.$$

In order to estimate the probability of $\lambda_1(\mathbf{W})$ to lie in the interval $-u_0$ to $-2 - Kn^{-\frac{2}{3}}$ let us divide this interval into sub intervals. We denote

$$\kappa_j := \frac{K + j}{n^{\frac{2}{3}}} \quad \text{and} \quad v_j := \frac{(K + j)^{\frac{1}{5}}}{n^{\frac{2}{3}}}.$$

Then we define the following intervals $I_j := [-2 - \kappa_{j+1}, -2 - \kappa_j]$ for $j = 0, ..., j_N$, where N is the smallest integer such that $2 + \kappa_{j+1} \ge u_0$. Denote $x_j := -2 - \kappa_j$. By a union bound we may write

$$\mathbb{P}\left(-u_0 \le \lambda_1(\mathbf{W}) \le -2 - \frac{K}{n^{\frac{2}{3}}}\right) \le \sum_{j=0}^{N} \mathbb{P}(\lambda_1(\mathbf{W} \in I_j).$$

By definition the intervals I_j are of length $|I_j| \le n^{-\frac{2}{3}}$ and the event $\lambda_1(\mathbf{W}) \in I_j$ involves $|\lambda_1(\mathbf{W}) - x_j| \le |I_j| \le v_j$. We may take $z_j := x_j + iv_j$ and note the following fact. Suppose that $\lambda_1(\mathbf{W}) \in I_j$ then

$$\operatorname{Im} m_n(z_j) = \frac{1}{n} \sum_{k=1}^{n} \frac{v_k}{(\lambda_k(\mathbf{W}) - x_k)^2 + v_k^2} \ge \frac{1}{2nv_j}. \tag{3.2}$$

For the imaginary part of $s(z)$ and $|u| \ge 2$ we have the following bound

$$c\frac{v}{|b(z)|} \le \operatorname{Im} s(z) \le C\frac{v}{|b(z)|},$$

where $b(z) := z + 2s(z)$. Moreover, $c_1\sqrt{\gamma + v} \le |b(z)| \le C_1\sqrt{\gamma + v}$, recalling that $\gamma := \gamma(u) := ||u| - 2|$ (see the definition (1.7)). Taking $z := z_j$ we write

$$\operatorname{Im} s(z_j) \le \frac{Cv_j}{\sqrt{\kappa_j}}. \tag{3.3}$$

It is easy to see that

$$\frac{Cv_j}{\sqrt{\kappa_j}} \le \frac{Cv_j^2 n}{nv_j\sqrt{\kappa_j}} \le \frac{C(K + j)^{\frac{2}{5}}}{nv_j(K + j)^{\frac{1}{2}}} \le \frac{1}{4nv_j}$$

for K large enough. Hence, applying (3.2) and (3.3) we get

$$\operatorname{Im} m_n(z_j) - \operatorname{Im} s(z_j) \ge \frac{1}{2nv_j} - \frac{Cv_j}{\sqrt{\kappa_j}} \ge \frac{1}{4nv_j}.$$

Applying the definition of κ_j and v_j we write

$$\frac{1}{nv_j} \geq \frac{(K+j)^{\frac{4}{5}}}{n\kappa_j} \geq \frac{(K+j)^{\frac{1}{4}}}{n(\kappa_j+v_j)},$$
$$\frac{1}{nv_j} \geq \frac{(K+j)^{\frac{7}{10}}}{(nv_j)^2\sqrt{\kappa_j}} \geq \frac{(K+j)^{\frac{1}{4}}}{(nv_j)^2\sqrt{\kappa_j+v_j}},$$
$$\frac{1}{nv_j} \geq \frac{(K+j)^{\frac{2}{5}}}{n\sqrt{v_j}\sqrt{\kappa_j}} \geq \frac{(K+j)^{\frac{1}{4}}}{n\sqrt{v_j}\sqrt{\kappa_j+v_j}},$$
$$\frac{1}{nv_j} \geq \frac{(K+j)^{\frac{7}{20}}}{(nv_j)^{\frac{3}{2}}\kappa_j^{\frac{1}{4}}} \geq \frac{(K+j)^{\frac{1}{4}}}{(nv_j)^{\frac{3}{2}}(\kappa_j+v_j)^{\frac{1}{4}}}.$$

Let us introduce the following quantity, which is the sum of four terms on the r.h.s. of the previous inequalities,

$$\Psi_j := \frac{1}{n(\kappa_j+v_j)} + \frac{1}{(nv_j)^2\sqrt{\kappa_j+v_j}} + \frac{1}{n\sqrt{v_j}\sqrt{\kappa_j+v_j}} + \frac{1}{(nv_j)^{\frac{3}{2}}(\kappa_j+v_j)^{\frac{1}{4}}}.$$

Therefore, if $\lambda_1(\mathbf{W}) \in I_j$ then

$$\operatorname{Im}\Lambda_n(z_j) = \operatorname{Im} m_n(z_j) - \operatorname{Im} s(z_j) \geq C(K+j)^{\frac{1}{4}}\Psi_j. \tag{3.4}$$

Applying now Lemmas B.2, B.3 in the Appendix, (1.8) (see Theorem 2.2 in [15]) and (3.4) we get

$$\begin{aligned}\mathbb{P}\left(-u_0 \leq \lambda_1(\mathbf{W}) \leq -2 - \frac{K}{n^{\frac{2}{3}}}\right) &\leq \sum_{j=0}^{N} \mathbb{P}(\lambda_1(\mathbf{W}) \in I_j) \\ &\leq \sum_{j=0}^{N} \mathbb{P}(|\operatorname{Im}\Lambda(z_j)| \geq C(K+j)^{\frac{1}{4}}\Psi_j) \\ &\leq \sum_{j=0}^{N} \frac{(Cp^3)^p}{(K+j)^{\frac{p}{4}}} \leq \frac{(Cp^3)^p}{K^{\frac{p}{4}-1}}.\end{aligned}$$

The last inequality concludes the proof of the lemma. □

Proof of Theorem 1.3. We first investigate the case (i) when $j \in [K, n-K+1]$. Without loss of generality we may assume that in this case $\lambda_j \in [-2, 2]$ since otherwise

$$\mathbb{P}(\lambda_j \leq -2) \leq \mathbb{P}\left(F_n(-2) \geq \frac{j}{n}\right) \leq \mathbb{P}\left(\Delta_n^* \geq \frac{K}{n}\right) \leq \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p}$$

and

$$\mathbb{P}(\lambda_j \geq 2) \leq \mathbb{P}\left(F_n(2) \leq \frac{j}{n}\right) \leq \mathbb{P}\left(\Delta_n^* \geq \frac{K}{n}\right) \leq \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p},$$

where $1 \le p \le c \log n$ and we applied the fact that $G_{sc}(-2) = 0$ and $G_{sc}(2) = 1$. It was proved in [17] (see Section 9 in the Appendix), that there exist constants c_1 and c_2 such that

$$c_1 x^{\frac{2}{3}} \le 2 + G_{sc}^{-1}(x) \le c_2 x^{\frac{2}{3}} \quad \text{for } x \in \left[0, \frac{1}{2}\right] \quad \text{and} \tag{3.5}$$

$$c_1 (1-x)^{\frac{2}{3}} \le 2 - G_{sc}^{-1}(x) \le c_2 (1-x)^{\frac{2}{3}} \quad \text{for } x \in \left[\frac{1}{2}, 1\right]. \tag{3.6}$$

Obviously, the maximum in Δ_n^* is reached at the jump points of F_n, i.e.

$$\Delta_n^* = \max_{1 \le k \le n} |F_n(\lambda_k) - G_{sc}(\lambda_k)| = \max_{1 \le k \le n} \left| \frac{k}{n} - G_{sc}(\lambda_k) \right|.$$

This fact implies that for every j there exists $\theta, |\theta| \le 1$ such that

$$\lambda_j = G_{sc}^{-1}\left(\frac{j}{n} + \theta \Delta_n^*\right).$$

By Taylor's formula we get

$$\lambda_j = G_{sc}^{-1}\left(\frac{j}{n}\right) + \mathbb{E}_\tau \frac{2\pi\theta\Delta_n^*}{\sqrt{4 - \left(G_{sc}^{-1}\left(\frac{j}{n} + \theta\Delta_n^*\right)\right)^2}}. \tag{3.7}$$

Again applying Theorem 1.1 we obtain that

$$\mathbb{P}\left(\Delta_n^* \le \frac{K}{2n}\right) \ge 1 - \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p}.$$

This means that without loss of generality we may assume that $\Delta_n^* \le \frac{K}{2n}$. It remains to consider two cases. In the first, $2\Delta_n^* \le \frac{j}{n} \le \frac{1}{2} - \theta\Delta_n^*$ we may apply (3.5) and conclude

$$\sqrt{4 - \left(G_{sc}^{-1}\left(\frac{j}{n} + \theta\Delta_n^*\right)\right)^2} \ge c_1 \left|\frac{j}{n} + \theta\Delta_n^*\right|^{\frac{1}{3}} \ge c_1' \left(\frac{j}{n}\right)^{\frac{1}{3}},$$

for some positive constant c_1'. This inequality together with (3.7) yield that

$$|\lambda_j - \gamma_j| \le C_1 \Delta_n^* \left(\frac{n}{j}\right)^{\frac{1}{3}}.$$

In the opposite case we apply (3.6) and obtain

$$|\lambda_j - \gamma_j| \le C_1 \Delta_n^* \left(\frac{n}{n-j+1}\right)^{\frac{1}{3}}.$$

Combining the last two inequalities we get

$$\mathbb{P}\left(|\lambda_j - \gamma_j| \le C_1 K [\min(j, n-j+1)]^{-\frac{1}{3}} n^{-\frac{2}{3}}\right) \ge 1 - \frac{C^p \log^{\frac{2p}{1-2\alpha}} n}{K^p}.$$

(ii). We now turn our attention to the case $j \le K$ or $j \ge n-K+1$ and without loss of generality we restrict ourselves to the first one. Let us denote

$$l := \frac{C_1 K}{n^{\frac{2}{3}}} \left(\frac{1}{j}\right)^{\frac{1}{3}}.$$

In the opposite case we take $l := C_1 K (n-j+1)^{\frac{1}{3}} n^{-\frac{2}{3}}$. It is easy to see that

$$\mathbb{P}(|\lambda_j - \gamma_j| \ge l) \le \mathbb{P}(|\lambda_j - \gamma_j| \ge l, \lambda_j > \gamma_j) + \mathbb{P}(|\lambda_j - \gamma_j| \ge l, \lambda_j < \gamma_j).$$

The first case when $\lambda_j > \gamma_j$ is trivial since in this situation $\lambda_j > j_1 \ge -2 + c_1 n^{-2/3}$ (see (3.5)) and we may repeat the calculations above and get

$$\mathbb{P}(|\lambda_j - \gamma_j| \ge l, \lambda_j > \gamma_j) \le \frac{C^p \log^{4p} n}{K^p},$$

for $1 \le p \le c \log n$. It remains to bound $\mathbb{P}(|\lambda_j - \gamma_j| \ge l, \lambda_j < \gamma_j)$. Again applying (3.5) we get $\gamma_j \le -2 + c_2 \left(\frac{j}{n}\right)^{\frac{1}{3}}$. Hence, choosing an appropriate constant C_1 we obtain

$$\begin{aligned}
\mathbb{P}(|\lambda_j - \gamma_j| \ge l, \lambda_j < \gamma_j) &= \mathbb{P}(\lambda_j \le \gamma_j - l, \lambda_j < \gamma_j) \\
&\le \mathbb{P}\left(\lambda_1 \le -2 + c_2 \left(\frac{j}{n}\right)^{\frac{2}{3}} - \frac{C_1 K}{n^{\frac{2}{3}}} \left(\frac{1}{j}\right)^{\frac{1}{3}}\right) \\
&\le \mathbb{P}\left(\lambda_1 \le -2 - \left(\frac{K}{n}\right)^{\frac{2}{3}}\right) \le \frac{(Cp^{12})^p}{K^{\frac{2p}{3}}} + \frac{C}{n^{2-\phi}}.
\end{aligned}$$

□

4. Delocalization of eigenvectors

In this section we prove Theorem 1.4. Here we shall apply the following result from [15][Lemma 4.1]. Let us denote

$$\mathbb{D} := \{z = u + iv \in \mathbb{C} : |u| \le u_0, V \ge v \ge v_0 := A_0 n^{-1}\},$$

where $u_0, V > 0$ are any fixed real numbers and A_0 is some large constant to be determined below. Then assuming the conditions (**C1**) there exist a positive constant C_0 depending on u_0, V and positive constants A_0, A_1 depending on C_0, α such that for all $z \in \mathbb{D}$ and $1 \le p \le A_1 (nv)^{\frac{1-2\alpha}{2}}$ we have

$$\max_{1 \le j \le n} \mathbb{E} |\mathbf{R}_{jj}(z)|^p \le C_0^p. \tag{4.1}$$

Proof of Theorem 1.4. Let us introduce the following distribution function

$$F_{nj}(x) := \sum_{k=1}^{n} |u_{jk}|^2 \, \mathbb{1}[\lambda_k(\mathbf{W}) \le x].$$

Using the eigenvalue decomposition of $\mathbf{W}$ it is easy to see that

$$\mathbf{R}_{jj}(z) = \sum_{k=1}^n \frac{|u_{jk}|^2}{\lambda_k(\mathbf{W}) - z} = \int_{-\infty}^{\infty} \frac{1}{x-z}\, dF_{nj}(x),$$

which means that $\mathbf{R}_{jj}(z)$ is the Stieltjes transform of $F_{nj}(x)$. For any $\lambda > 0$ we have

$$\max_{1 \le k \le n} |u_{jk}|^2 \le \sup_x (F_{nj}(x+\lambda) - F_{nj}(x)) =: Q_{nj}(\lambda). \tag{4.2}$$

Furthermore, it is easy to check that

$$Q_{nj}(\lambda) \le 2 \sup_u \lambda \operatorname{Im} \mathbf{R}_{jj}(u + i\lambda). \tag{4.3}$$

Indeed,

$$\begin{aligned}\sup_u \lambda \operatorname{Im} \mathbf{R}_{jj}(u+i\lambda) &= \sup_u \sum_{k=1}^n \frac{\lambda^2 |u_{jk}|^2}{\lambda^2 + (\lambda_k - u)^2} \\ &\ge \sup_u \sum_{k=1}^n \frac{\lambda^2 |u_{jk}|^2}{\lambda^2 + (\lambda_k - u)^2} \mathbb{1}[u \le \lambda_k \le u + \lambda] = \frac{1}{2} Q_{nj}(\lambda).\end{aligned}$$

To finish the proof we need to show that with high probability the r.h.s. of (4.3) is bounded by $n^{-1} \log^8 n$. Let us recall the following notations. Let $\hat{X}_{jk} := X_{jk} \mathbb{1}[|X_{jk}| \le Dn^{\frac{3}{8}}]$, $\tilde{X}_{jk} := X_{jk} \mathbb{1}[|X_{jk}| \ge Dn^{\frac{3}{8}}] - \mathbb{E}\, X_{jk} \mathbb{1}[|X_{jk}| \ge Dn^{\frac{3}{8}}]$ and finally $\breve{X}_{jk} := \tilde{X}_{jk} \sigma^{-1}$, where $\sigma^2 := \mathbb{E}\,|\tilde{X}_{11}|^2$. Let $\hat{\mathbf{X}}, \tilde{\mathbf{X}}$ and $\breve{\mathbf{X}}$ denote symmetric random matrices with entries $\hat{X}_{jk}, \tilde{X}_{jk}$ and $\breve{X}_{jk}$ respectively. In a similar way we denote the resolvent matrices by $\hat{\mathbf{R}}, \tilde{\mathbf{R}}$ and $\breve{\mathbf{R}}$. In this case we have

$$\mathbb{P}(\mathbf{W} \ne \hat{\mathbf{W}}) \le \frac{C}{n}.$$

Let $u_0 > 0$ denote a large constant, whose exact value will be chosen later. Applying Lemmas A.2 and A.1 in the Appendix it follows that

$$\mathbb{P}(\|\mathbf{W}\| \ge u_0) \le \frac{C}{n}. \tag{4.4}$$

It what follows we may assume that $\|\mathbf{W}\| \le u_0$ and $\mathbf{W} = \hat{\mathbf{W}}$. Then for $|u| \ge 2u_0$ and $v > 0$ we get

$$|\mathbf{R}_{jj}(u+iv)| \le \int_{-u_0}^{u_0} \frac{1}{\sqrt{(x-u)^2 + v^2}}\, dF_{nj}(x) \le \frac{1}{u_0} \le C, \tag{4.5}$$

where C is some large positive constant which will be chosen later. It remains to estimate $|\mathbf{R}_{jj}(u+iv)|$ for all $-2u_0 \le u \le 2u_0$. For simplicity let us denote this interval by $\mathcal{U}_0$, i.e. $\mathcal{U}_0 := [-2u_0, 2u_0]$. Be the triangular inequality we may write $|\mathbf{R}_{jj}| = |\hat{\mathbf{R}}_{jj}| \le |\tilde{\mathbf{R}}_{jj}| + |\hat{\mathbf{R}}_{jj} - \tilde{\mathbf{R}}_{jj}|$. Applying the simple equation

$$\hat{\mathbf{R}}_{jj} - \tilde{\mathbf{R}}_{jj} = [\hat{\mathbf{R}}(\hat{\mathbf{W}} - \tilde{\mathbf{W}})\tilde{\mathbf{R}}]_{jj}$$

we get

$$|\hat{\mathbf{R}}_{jj} - \tilde{\mathbf{R}}_{jj}| \le \|\hat{\mathbf{W}} - \tilde{\mathbf{W}}\| \|\mathbf{e}_j^T \hat{\mathbf{R}}\|_2 \|\tilde{\mathbf{R}}\mathbf{e}_j\|_2,$$

where $\mathbf{e}_j$ is a unit column-vector with all entries zero except for an entry one at the position j. Using Lemma C.11 in the Appendix we conclude that

$$|\hat{\mathbf{R}}_{jj}| \le |\tilde{\mathbf{R}}_{jj}| + \frac{1}{v}\|\hat{\mathbf{W}} - \tilde{\mathbf{W}}\| \sqrt{|\hat{\mathbf{R}}_{jj}||\tilde{\mathbf{R}}_{jj}|}.$$

It is easy to see that

$$\|\hat{\mathbf{W}} - \tilde{\mathbf{W}}\|_2^2 = \frac{1}{n}\sum_{j,k} [\mathbb{E}\,|X_{jk}|\,\mathbb{1}[|X_{jk}| \ge Dn^{\frac{3}{8}}]]^2 \le \frac{C}{n^4},$$

We may take $v = v_0 := C_1 n^{-1}\log^8 n$, with $C_1 \ge A_0$. Applying the simple inequality $2|ab| \le a^2 + b^2$ we get

$$\sup_{u \in \mathcal{U}_0} |\mathbf{R}_{jj}| \le 3 \sup_{u \in \mathcal{U}_0} |\tilde{\mathbf{R}}_{jj}|. \tag{4.6}$$

It remains to estimate $\sup_{u \in \mathcal{U}_0} |\tilde{\mathbf{R}}_{jj}(u + iv_0)|$. It is easy to see that

$$\tilde{\mathbf{R}}(z) = (\tilde{\mathbf{W}} - z\mathbf{I})^{-1} = \sigma^{-1}(\breve{\mathbf{W}} - z\sigma^{-1}\mathbf{I})^{-1} = \sigma^{-1}\breve{\mathbf{R}}(\sigma^{-1}z). \tag{4.7}$$

Applying the resolvent equality we get

$$\breve{\mathbf{R}}(z) - \breve{\mathbf{R}}(\sigma^{-1}z) = (z - \sigma^{-1}z)\breve{\mathbf{R}}(z)\breve{\mathbf{R}}(\sigma^{-1}z). \tag{4.8}$$

Combining (4.7) and (4.8) we obtain

$$|\tilde{\mathbf{R}}_{jj}(z) - \breve{\mathbf{R}}_{jj}(z)| \le (\sigma^{-1} - 1)|\breve{\mathbf{R}}_{jj}(\sigma^{-1}z)| + \frac{|z|(\sigma^{-1} - 1)}{v}\sqrt{|\breve{\mathbf{R}}_{jj}(z)||\breve{\mathbf{R}}_{jj}(\sigma^{-1}z)|}.$$

It is easy to check that $(\sigma^{-1} - 1) \le Cn^{-\frac{3}{2}}$ and $\max(|z\breve{\mathbf{R}}_{jj}(z)|, |z\breve{\mathbf{R}}_{jj}(\sigma^{-1}z)|) \le C$ for some constant C. Similarly to the previous calculations we get that

$$\sup_{u \in \mathcal{U}_0} |\tilde{\mathbf{R}}_{jj}| \le 3 \sup_{u \in \mathcal{U}_0} |\breve{\mathbf{R}}_{jj}|. \tag{4.9}$$

Note, that the matrix $\breve{\mathbf{W}}$ satisfies the conditions (**C1**). Applying (4.1) with $p = c\log n$ we obtain

$$\mathbb{P}(|\breve{\mathbf{R}}_{jj}(u + iv_0)| \ge C_0 e^{\frac{5}{c}}) \le \frac{\mathbb{E}\,|\breve{\mathbf{R}}_{jj}(u + iv_0)|^p}{(C_0 e^{\frac{5}{c}})^p} \le \frac{1}{n^5}.$$

We may partition interval $\mathcal{U}_0$ into $k_n := n^4$ disjoint subintervals of equal length, i.e $-2u_0 = x_0 \le x_1 \le ... \le x_{k_n} = 2u_0$. Then by the Newton-Leibniz formula

$$\begin{aligned}\sup_{u \in \mathcal{U}_0} |\breve{\mathbf{R}}_{jj}(u + iv_0)| &\le \max_{1 \le k \le k_n} \sup_{x_{k-1} \le x \le x_k} |\breve{\mathbf{R}}_{jj}(x + iv_0)| \\ &\le \max_{1 \le k \le k_n} |\breve{\mathbf{R}}_{jj}(x_{k-1} + iv_0)| + \max_{1 \le k \le k_n} \int_{x_{k-1}}^{x_k} |\breve{\mathbf{R}}'_{jj}(u + iv_0)|\,du.\end{aligned}$$

We may write

$$\max_{1 \le k \le k_n} \int_{x_{k-1}}^{x_k} |\breve{\mathbf{R}}'_{jj}(u + iv_0)|\,du \le \frac{C}{n}.$$

Thus we arrive at

$$\mathbb{P}\left(\sup_{u\in\mathcal{U}_0}|\breve{\mathbf{R}}_{jj}(u+iv_0)|\geq 2C_0e^{\frac{5}{c}}\right) \\ \leq \sum_{k=1}^{k_n}\mathbb{P}\left(|\breve{\mathbf{R}}_{jj}(x_{k-1}+iv_0)|\geq C_0e^{\frac{5}{c}}\right)\leq\frac{C}{n}. \tag{4.10}$$

We choose now $\lambda := v_0$. In view of (4.2), (4.3), (4.6), (4.9) and (4.10) we get that there exist C and C_1 such that

$$\mathbb{P}\left(\max_{1\leq j,k\leq n}|u_{jk}|^2\leq\frac{C_1\log^8 n}{n}\right)\geq 1-\frac{C}{n},$$

which concludes the proof. □

5. Numerical simulations

The aim of this section is to illustrate by numerical experiments some effects arising in cases where a only small number of moments of matrix entries are finite. We restrict ourselves to those statistics which correspond to the main results of the current paper.

We start by choosing an appropriate distribution for the matrix entries. To this end consider a random variable ξ which has the following density and distribution function depending on a parameter μ

$$f_\mu(x)=\frac{\mu-1}{x^\mu}\,\mathbb{1}[x\geq 1] \quad\text{and}\quad F_\mu(x)=\left(1-\frac{1}{x^{\mu-1}}\right)\mathbb{1}[x\geq 1].$$

This choice guarantees a non zero skewness i.e. the moment of order three that differs from the standard Gaussian distribution. To ensure existence of m finite moments requires to choose $\mu > m+1$. In what follows we shall take $\mu = m+1.1$. Let ξ_{jk} denote i.i.d copies of ξ. Then we consider

$$X_{jk}:=\frac{\xi_{jk}-\mathbb{E}\,\xi}{\sqrt{\mathbb{D}\,\xi}}$$

which are combined in the random matrix $\mathbf{X} := [X_{jk}]_{j,k=1}^n$ with $\mathbb{E}\,X_{jk}=0$ and $\mathbb{E}\,X_{jk}^2=1$. As usual we also introduce the truncated (also normalized) random matrix $\breve{\mathbf{X}}$.

In Figure 1 we plotted the normalized frequency histogram of the eigenvalues of $\mathbf{W}$ for different μ and $n = 2000$. We use the simplest procedure dividing the range $[\lambda_1(\mathbf{W}), \lambda_n(\mathbf{W})]$ into m intervals of equal size. In our case we take $m = 70$. We know from [25] that to guarantee convergence to Wigner's semicircle law it is enough to have finite second moments only. It is visible that for $\mu = 3.1$ (this case corresponds to a finite second moment only) convergence is rather poor. But starting from $\mu = 4.1$ one observes a rather fast convergence. It is easy from the picture that the width of histogram's bars depends on number of finite moments, indicating the fact that with growing number of finite moments the number of eigenvalues outside of the suppport of the semicircle law becomes smaller.

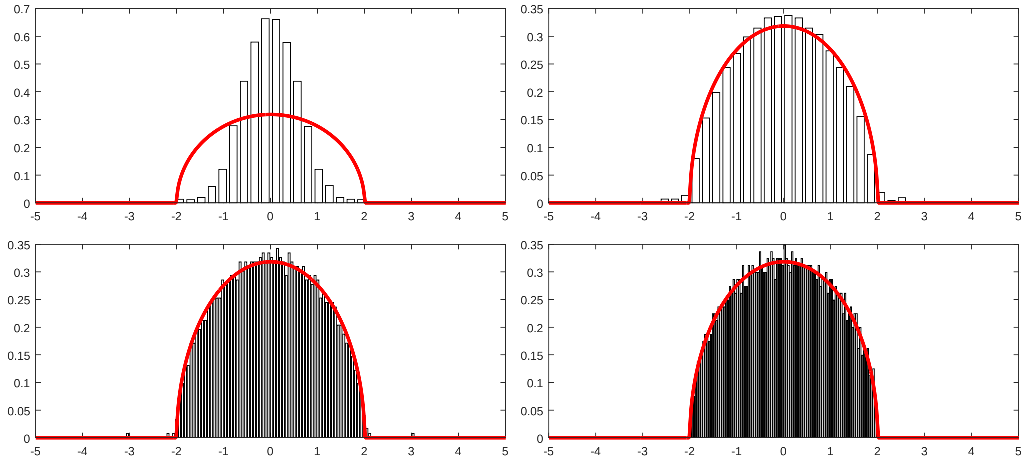

FIGURE 1. Empirical spectral density of the eigenvalues of $\mathbf{W}$ for different μ and $n = 2000$. In the top row $\mu = 3.1$ (on the left) and $\mu = 4.1$ (on the right). In the bottom row $\mu = 5.1$ (on the left) and $\mu = 9.1$ (on the right). Red line – Wigner's semicircle law density function g_{sc}.

Let us consider the following statistics (motivated by the minimum error size, see [22])

$$T_n := \frac{n\Delta_n^*}{\sqrt{\log n}}.$$

In Figure 2 we plotted $\mathbb{E}\,T_n$ (red line) with ± 1 standard deviation around $\mathbb{E}\,T_n$ (black lines) for n from 100 to 5000 with step 100. We take the following values for μ: 5.1 (top left), 7.1 (top right), 9.1 (bottom left) and Gaussian case (bottom right).

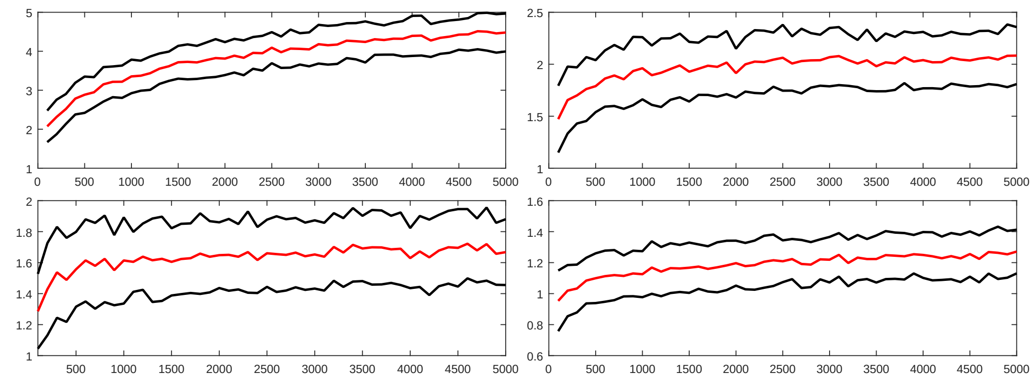

FIGURE 2. The plot of $\mathbb{E}\,T_n$ with ± 1 standard deviation around the mean for n from 100 to 5000 with step 100. The values for μ here are $5.1; 7.1, 9.1$ and Gaussian distribution.

It is interesting to investigate the dependence of the largest eigenvalues on the tail behavior. For example, we consider $\lambda_n(\mathbf{W})$ and study the following statistic

$$\zeta_n = n^{\frac{2}{3}}(\lambda_n(\mathbf{W}) - 2).$$

In Figure 3 we plotted on the left the distribution of $\zeta_n, n = 2000$ for the values $\mu = 5.1; 6.1; 7.1$ and 9.1. On the right the distribution of truncated versions $\breve{\zeta}_n = n^{\frac{2}{3}}(\lambda_n(\breve{\mathbf{W}}) - 2)$ for the corresponding values of μ. Here the red line is the Tracy–Widom density function with parameter $\beta = 1$, see [29]. To plot the Tracy–Widom density function we applied the method of [6] where the Tracy–Widom distribution has been approximated by a gamma distribution with specific values. The impact of truncation is obvious from this graph. These figures motivate the remarks following Theorems 1.3 and Lemma 3.1.

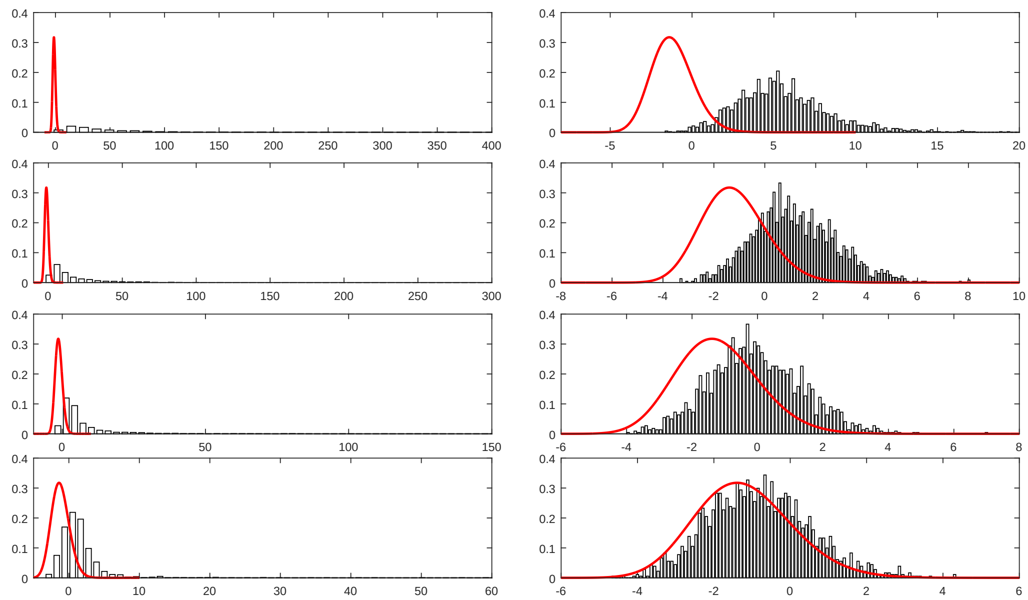

FIGURE 3. On the left the distribution of $\zeta_n, n = 2000$ for the following values of $\mu = 5.1; 6.1; 7.1$ and 9.1. On the right the distribution of $\breve{\zeta}_n$ for the corresponding values of μ. Red line – Tracy–Widom density function with parameter $\beta = 1$.

Finally, we consider simulations of the empirical distribution of the following delocalization statistics

$$V_n := n \max_{1 \le j,k \le n} |u_{jk}|^2,$$

where $u_j := (u_{j1}, ..., u_{jn})$ are the eigenvectors of $\mathbf{W}$ corresponding to the eigenvalue $\lambda_j(\mathbf{W})$. In Figure 4 we plotted in the top row V_n (on the left) and $\breve{V}_n$ (on the right), where $\breve{V}_n$ is V_n with $\mathbf{W}$ replaced by $\breve{\mathbf{W}}$, for $\mu = 5.1$ and $n = 2000$. The middle row shows the same statistics for $\mu = 9.1$. Finally, in the bottom row we compare $\breve{V}_n$ for $\mu = 9.1$ with V_n in the Gaussian case. It seems evident that for the truncation $\breve{V}_n$ in

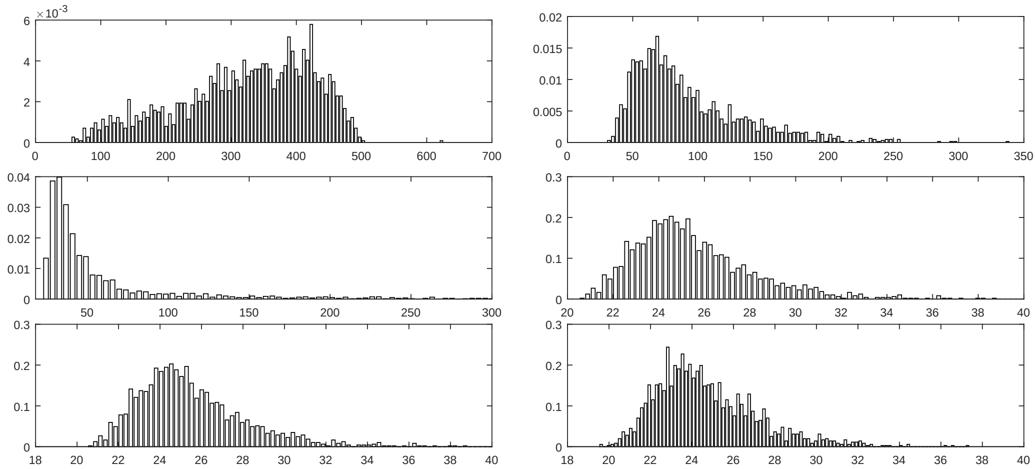

FIGURE 4. In the top row V_n (on the left) and $\breve{V}_n$ (on the right) for $\mu = 5.1$ and $n = 2000$. In the middle row the same statistics for $\mu = 9.1$. Finally, in the bottom row we compare $\breve{V}_n$ for $\mu = 9.1$ with V_n in the Gaussian case

the case $\mu = 9.1$ there is a good correspondence to Gaussian case. Even in case of high moments, $\mu = 9.1$, The histogram of V_n shows some deviation from the Gaussian case, which indicates a bad convergence rate.

Appendix A. Spectral norm of random matrices

Lemma A.1. *Assume that the conditions* (**C1**) *hold and let* $K \geq 4$. *Then there exists a constant* $c > 0$ *depending on* α *such that*

$$\mathbb{P}(\|\mathbf{W}\| \geq K) \leq e^{-n^c \log K}.$$

Proof. It is common practice to control the extreme eigenvalues by the moment method, estimating $\mathbb{E}\operatorname{Tr}\mathbf{X}^k$ for large k applying graph representation. The list of references is extensive, we only mention here some selected results. More details can be found in Chapter 2 of the monograph of T.Tao [26]. In this paper we shall adopt a method due tp V. Vu from [30]. Recall that

$$\|\mathbf{W}\| = \max_{1\leq j\leq n} |\lambda_j(\mathbf{W})|$$

and we obtain for even k

$$\mathbb{E}\max_{1\leq j\leq n} |\lambda_j(\mathbf{W})|^k \leq \sum_{j=1}^{n} \mathbb{E}\,\lambda_j(\mathbf{W})^k = \mathbb{E}\operatorname{Tr}\mathbf{W}^k.$$

In the following we shall use notations and definitions used in [4]. A graph is a triple (E, V, F), where E is the set of edges, V is the set of vertices, and F is a function, $F : E \to V \times V$. Let $\boldsymbol{i} = (i_1, ..., i_k)$ be a vector taking values in $\{1, ..., n\}^k$. For a vector $\boldsymbol{i}$ we define a Γ-graph as follows. Draw a horizontal line and plot the numbers $i_1, ..., i_k$ on it. Consider the distinct numbers as vertices, and draw k edges e_j from i_j to $i_{j+1}, j = 1, ..., k$, using $i_{k+1} = i_1$ by convention. Denote the number of distinct i_j's by t. Such a graph is called a $\Gamma(k,t)$-graph.

Two $\Gamma(k,t)$-graphs are said to be *isomorphic if they can be converted into each other by a permutation of* $(1, ..., n)$. By this definition, all Γ-graphs are classified into isomorphism classes. We shall call the $\Gamma(k,t)$-graph canonical if it has the following properties:
1) Its vertex set is $\{1, ..., t\}$;
2) Its edge set is $\{e_1, ..., e_k\}$;
3) There is a function g from $\{1, ..., k\}$ onto $\{1, ..., t\}$ satisfying $g(1) = 1$ and $g(i) \le \max\{g(1), ..., g(i-1)\} + 1$ for $1 < i \le k$;
4) $F(e_i) = (g(i), g(i+1))$, for $i = 1, ..., k$, with the convention $g(k+1) = g(1) = 1$.

It is easy to see that each isomorphism class contains one and only one canonical Γ-graph that is associated with a function g, and a general graph in this class can be defined by $F(e_j) = (i_{g(j)}, i_{g(j+1)})$. Obviously, each isomorphism class contains $n(n-1)...(n-t+1)$ $\Gamma(k,t)$-graphs.

We expand the traces of powers of $\mathbf{W}$ in a sum

$$\operatorname{Tr} \mathbf{W}^k = \frac{1}{n^{\frac{k}{2}}} \sum_{i_1, i_2, ..., i_k} X_{i_1 i_2} X_{i_2 i_3} ... X_{i_k i_1} = \frac{1}{n^{\frac{k}{2}}} \sum_{i_1, i_2, ..., i_k} X(\boldsymbol{i}), \tag{A.1}$$

where the summation is taken over all sequences $\boldsymbol{i} = (i_1, ..., i_k) \in \{1, ..., n\}^k$. For each vector $\boldsymbol{i}$ we construct a graph $G(\boldsymbol{i})$ as above and set $X(G(\boldsymbol{i})) := X(\boldsymbol{i})$. Let us denote

$$E(n,k,t) := \sum_{\Gamma(k,t)} \sum_{G(\boldsymbol{i}) \in \Gamma(k,t)} \mathbb{E}[X(G(\boldsymbol{i}))], \tag{A.2}$$

where $\sum_{\Gamma(k,t)}$ is taken over all canonical $\Gamma(k,t)$-graphs with t vertices and k edges; and the summation $\sum_{G(\boldsymbol{i}) \in \Gamma(k,t)}$ is taken over all isomorphic graphs for a given canonical graph. It is easy to check that if $t \ge \frac{k}{2} + 1$ then $E(n,k,t) = 0$. Since $\mathbb{E}\, X_{i_1,i_2} = 0$ for all $1 \le i_1 \le i_2 \le n$ and all X_{i_1,i_2} are independent we may restrict ourself to the canonical graphs where each edge appears at least twice.

Let us also denote by $W(n,k,t)$ the number of these canonical graphs using k edges and t distinct vertices where each edge is used at least twice. It was proved in [30] that

$$W(n,k,t) \le \binom{2t-2}{k} t^{2(k-2t+2)} 2^{2t-2}. \tag{A.3}$$

If a graph $G(\boldsymbol{i})$ has k edges and t vertices then

$$\mathbb{E}\, X(G(\boldsymbol{i})) \le D^{k-2(t-1)} n^{\alpha(k-2(t-1))}. \tag{A.4}$$

Thus applying (A.1)– (A.4) we obtain

$$
\begin{aligned}
\mathbb{E}\operatorname{Tr}\mathbf{W}^k &= \frac{1}{n^{\frac{k}{2}}}\sum_{t=1}^{\frac{k}{2}+1} E(n,k,t) \le \sum_{t=1}^{\frac{k}{2}+1} D^{k-2(t-1)} n^{\alpha(k-2(t-1))} n(n-1)...(n-t+1) W(n,k,t) \\
&\le \frac{1}{n^{\frac{k}{2}}}\sum_{t=1}^{\frac{k}{2}+1} D^{k-2(t-1)} n^{\alpha(k-2(t-1))} n(n-1)...(n-t+1) \binom{2t-2}{k} t^{2(k-2t+2)} 2^{2t-2} \\
&\le \frac{1}{n^{\frac{k}{2}}}\sum_{t=1}^{\frac{k}{2}+1} S(n,k,t).
\end{aligned}
$$

It is easy to check that

$$
S(n,k,t-1) \le \frac{D^2 n^{2\alpha} k^6}{4n} S(n,k,t).
$$

We may take $k = D^{-\frac{1}{3}} n^{\frac{1-2\alpha}{6}}$ and get $S(n,k,t-1) \le \frac{1}{2} S(n,k,t)$. It follows that

$$
\begin{aligned}
\mathbb{E}\operatorname{Tr}\mathbf{W}^k &\le \frac{1}{n^{\frac{k}{2}}}\sum_{t=1}^{\frac{k}{2}+1} S(n,k,t) \le \frac{2}{n^{\frac{k}{2}}} S(n,k,k/2+1) \\
&= \frac{2n(n-1)...(n-k/2)2^k}{n^{\frac{k}{2}}} \le n 2^{k+1}.
\end{aligned} \tag{A.5}
$$

Since $K \ge 4$, applying Markov's inequality for even k and (A.5) we obtain

$$
\mathbb{P}(\|\mathbf{W}\| \ge K) \le \frac{\mathbb{E}\operatorname{Tr}\mathbf{W}^k}{K^k} \le 2n\left(\frac{2}{K}\right)^k \le e^{-n^c \log K}.
$$

$\square$

We denote by $\hat{X}_{jk} := X_{jk}\,\mathbb{1}[|X_{jk}| \ge Dn^\alpha]$, $\tilde{X}_{jk} := \hat{X}_{jk} - \mathbb{E}\,\hat{X}_{jk}$ and finally $\breve{X}_{jk} := \sigma^{-1}\tilde{X}_{jk}$, where $\sigma^2 := \mathbb{E}\,|\tilde{X}_{jk}|^2$. By $\hat{\mathbf{W}}$, $\tilde{\mathbf{W}}$ and $\breve{\mathbf{W}}$ we denote the symmetric random matrices with these entries.

Lemma A.2. *Under the conditions* (**C0**) *for* $K > 0$ *we have*

$$
\mathbb{P}(\|\mathbf{W}\| \ge K) \le 2\,\mathbb{P}\left(\|\breve{\mathbf{W}}\| \ge \frac{K}{4}\right) + \mathbb{P}\left(\|\mathbf{W} - \hat{\mathbf{W}}\| \ge \frac{K}{4}\right).
$$

Proof. We start the proof with the triangular inequality which yield the following estimate of $\|\mathbf{W}\|$

$$
\|\mathbf{W}\| \le \|\mathbf{W} - \hat{\mathbf{W}}\| + \|\hat{\mathbf{W}} - \tilde{\mathbf{W}}\| + \|\tilde{\mathbf{W}} - \breve{\mathbf{W}}\| + \|\breve{\mathbf{W}}\|. \tag{A.6}
$$

It is easy to see that

$$
\|\hat{\mathbf{W}} - \tilde{\mathbf{W}}\|^2 \le \frac{1}{n}\sum_{j,k}[\mathbb{E}\,X_{jk}\,\mathbb{1}[|X_{jk}| \ge Dn^\alpha]]^2 \le \frac{\mu_{4+\delta}^2}{D^{3+\delta} n^{2\alpha(3+\delta)-1}} \le \frac{C}{n^2}. \tag{A.7}
$$

Since $\breve{\mathbf{W}}$ differs from $\tilde{\mathbf{W}}$ by a global change of variance we may write

$$\|\tilde{\mathbf{W}} - \breve{\mathbf{W}}\| = (1-\sigma)\|\breve{\mathbf{W}}\| \le (1-\sigma^2)\|\breve{\mathbf{W}}\|.$$

By definition of σ we obtain

$$(1-\sigma^2) = \mathbb{E}\,|X_{jk}|^2\,\mathbb{1}[|X_{jk}| \ge Dn^\alpha] \le \frac{\mu_{4+\delta}}{D^{2+\delta}n^{\alpha(2+\delta)}}.$$

The last two inequalities together imply

$$\|\tilde{\mathbf{W}} - \breve{\mathbf{W}}\| \le \frac{C}{n^{\alpha(2+\delta)}}\|\breve{\mathbf{W}}\|. \tag{A.8}$$

Collecting the bounds (A.6)– (A.8) we get the desired bound. □

Appendix B. Truncation of matrix entries

In this section we will show that the conditions (**C0**) allow to assume that for all $1 \le j, k \le n$ we have $|X_{jk}| \le Dn^\alpha$, where D is some positive constant and

$$\alpha = \frac{2}{4+\delta}.$$

Let $\hat{X}_{jk} := X_{jk}\,\mathbb{1}[|X_{jk}| \le Dn^\alpha]$, $\tilde{X}_{jk} := X_{jk}\,\mathbb{1}[|X_{jk}| \ge Dn^\alpha] - \mathbb{E}\,X_{jk}\,\mathbb{1}[|X_{jk}| \ge Dn^\alpha]$ and finally $\breve{X}_{jk} := \tilde{X}_{jk}\sigma^{-1}$, where $\sigma^2 := \mathbb{E}\,|\tilde{X}_{11}|^2$. Let again $\hat{\mathbf{X}}, \tilde{\mathbf{X}}$ and $\breve{\mathbf{X}}$ denote the symmetric random matrices with entries $\hat{X}_{jk}, \tilde{X}_{jk}$ and $\breve{X}_{jk}$ respectively. In a similar way we denote the corresponding empirical spectral distribution functions, resolvent matrices and corresponding Stieltjes transforms.

Lemma B.1. *Under conditions* (**C0**) *we have*

$$\mathbb{E}^{\frac{1}{p}} \sup_{x\in\mathbb{R}} |F_n(x) - \hat{F}(x)|^p \le \frac{Cp}{n}.$$

Moreover,

$$\mathbb{E}\,|m_n(z) - \hat{m}_n(z)|^p \le \left(\frac{Cp}{nv}\right)^p.$$

Proof. See in [15][Lemma D.1]. □

Lemma B.2. *Under conditions* (**C0**) *we have*

$$\mathbb{E}\,|\tilde{m}_n(z) - \breve{m}_n(z)|^p \le \frac{C^p p^p \operatorname{Im}^p s(z)}{(nv)^p} + \frac{C^p p^{3p}}{(nv)^{2p}}.$$

Proof. See in [15][Lemma D.2]. □

Lemma B.3. *Under conditions* (**C0**) *we have*

$$\mathbb{E}\,|\tilde{m}_n(z) - \hat{m}_n(z)|^p \le \left(\frac{C}{nv}\right)^{\frac{3p}{2}}.$$

Proof. See in [15][Lemma D.3]. □

Appendix C. Auxiliary Lemmas

We start this section with several lemmas providing inequalities for moments of linear and quadratic forms. Recall that

$$\varepsilon_{1j} = \frac{1}{\sqrt{n}}X_{jj}, \quad \varepsilon_{2j} = -\frac{1}{n}\sum_{l\neq k\in T_j} X_{jk}X_{jl}\mathbf{R}^{(j)}_{kl}, \quad \varepsilon_{3j} = -\frac{1}{n}\sum_{k\in T_j}(X_{jk}^2-1)\mathbf{R}^{(j)}_{kk},$$
$$\varepsilon_{4j} = \frac{1}{n}(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(j)}).$$

The following result is obvious but will be needed in the proof of Theorem 1.1

Lemma C.1. *Under conditions* (**C1**) *for* $p \geq 1$ *we have*

$$\mathbb{E}\,|\varepsilon_{1j}|^{2p} \leq \frac{\mu_4 D^{2p-4}}{n^{p(1-2\alpha)+4\alpha}}.$$

Proof. The proof follows directly from the definition of $\varepsilon_{1j} := \frac{1}{\sqrt{n}}X_{jj}$. □

Lemma C.2. *Under conditions* (**C1**) *for* $p \geq 1$ *and* $q = 1, 2$ *there exists a positive constant* C *depending on* α *such that*

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_{1j}^{q}\right|^{p} \leq \frac{(Cp)^p}{n^p}. \tag{C.1}$$

Proof. See in [15][Lemma A.4]. □

The following Lemmas C.3– C.10 were proved in [15]. For completeness we state them here again but for the special case of v being a fixed constant denoted by V. In this case all inequalities obviously hold.

Lemma C.3. *Under conditions* (**C1**) *for* $p \geq 2$ *and* $z = u + iV$ *with some fixed* $V > 0$ *we have*

$$\mathbb{E}\,|\varepsilon_{2j}|^p \leq C^p\left(\frac{p^{\frac{3p}{2}}}{n^{\frac{p}{2}}} + \frac{p^{2p}}{n^{p(1-2\alpha)}}\right),$$

where C *depends on* V *and* α.

Proof. See [15][Lemma A.5]. □

For $p = 2$ and 4 we may give the better bound for quadratic form ε_{2j}. Let $\mathfrak{M}^{(j)} := \sigma\{X_{lk}, l, k \in \mathbb{T}_j\}$.

Lemma C.4. *Under conditions* (**C0**) *for* $q = 2$ *and* 4 *we have*

$$\mathbb{E}(|\varepsilon_{2j}|^q \big| \mathfrak{M}^{(j)}) \leq \frac{C}{n^{\frac{q}{2}}}\operatorname{Im}^{\frac{q}{2}} m_n^{(j)}(z),$$

where $z = u + iV$ *with some fixed* $V > 0$ *and* C *depending on* V.

Proof. See [15][Lemma A.6]. □

Lemma C.5. *Under conditions* (**C0**) *for $p \geq 2$ and $z = u + iV$ with some fixed $V > 0$ we have*

$$\mathbb{E}\,|\varepsilon_{3j}|^p \leq C^p \left(\frac{p^{\frac{p}{2}}}{n^{\frac{p}{2}}} + \frac{p^p}{n^{p(1-2\alpha)}} \right),$$

where C depends on V and α.

Proof. See [15][Lemma A.7]. □

For small $p \leq \frac{1}{\alpha}$ we may write a better bound for ε_{3j}.

Lemma C.6. *Under conditions* (**C1**) *for $2 \leq p \leq \frac{1}{\alpha}$ and $z = u + iV$ with fixed $V > 0$ we have*

$$\mathbb{E}(|\varepsilon_{3j}|^p \big| \mathfrak{M}^{(j)}) \leq \frac{C}{n^{\frac{p}{2}}},$$

where C depends on V and α

Proof. See [15][Lemma A.8]. □

Lemma C.7. *For $p \geq 2$ and $z = u + iV$ with fixed $V > 0$ we have*

$$\mathbb{E}\,|\varepsilon_{4j}|^p \leq \frac{1}{n^p}.$$

Proof. See [15][Lemma A.9]. □

Recall the definition of $\eta_{\nu j}, \nu = 0, 1, 2$

$$\eta_{0j} := \frac{1}{n} \sum_{k \in T_j} [(\mathbf{R}^{(j)})^2]_{kk}, \qquad \eta_{1j} := \frac{1}{n} \sum_{k \neq l \in \mathbb{T}_j} X_{jl} X_{jk} [(\mathbf{R}^{(j)})^2]_{kl},$$
$$\eta_{2j} := \frac{1}{n} \sum_{k \in \mathbb{T}_j} [X_{jk}^2 - 1] [(\mathbf{R}^{(j)})^2]_{kk}.$$

Lemma C.8. *Under conditions* (**C0**) *for $2 \leq p \leq 4$ and $z = u + iV$ with fixed $V > 0$ we have*

$$\mathbb{E}(|\eta_{1j}|^p \big| \mathfrak{M}^{(j)}) \leq \frac{C}{n^{\frac{p}{2}}},$$

where C depends on V.

Proof. See [15][Lemma A.10]. □

Lemma C.9. *Under conditions* (**C1**) *for $2 \leq p \leq \frac{1}{\alpha}$ and $z = u + iV$ with fixed $V > 0$ we have*

$$\mathbb{E}(|\eta_{2j}|^p \big| \mathfrak{M}^{(j)}) \leq \frac{C}{n^{\frac{p}{2}}}.$$

where C depends on V and α.

Proof. See [15][Lemma A.11]. □

Lemma C.10. *For $p \geq 2$ and $z = u + iV$ with fixed $V > 0$ we have*

$$|\eta_{0j}|^p \leq \frac{\operatorname{Im}^p m_n^{(j)}(z)}{n^p}.$$

Proof. See [15][Lemma A.12]. $\square$

The following lemma provides estimates for norms of vectors and matrices in terms of the resolvent.

Lemma C.11. *For any $z = u + iv \in \mathbb{C}^+$ we have*

$$\frac{1}{n}\sum_{l,k\in\mathbb{T}_j} |\mathbf{R}_{kl}^{(j)}|^2 \leq \frac{1}{v}\operatorname{Im} m_n^{(j)}(z). \tag{C.2}$$

For any $l \in \mathbb{T}_j$

$$\sum_{k\in\mathbb{T}_j} |\mathbf{R}_{kl}^{(j)}|^2 \leq \frac{1}{v}\operatorname{Im} \mathbf{R}_{ll}^{(j)}. \tag{C.3}$$

Moreover,

$$\frac{1}{n}\big|\operatorname{Tr}(\mathbf{R}^{(j)})^2\big| \leq \frac{1}{v}\operatorname{Im} m_n^{(j)}(z). \tag{C.4}$$

Proof. See [15][Lemma C.4, Lemma C.5]. $\square$

Recall that $\varphi(z) = \overline{z}|z|^{p-1}$. In the following lemma we show how to estimate the difference between $\varphi(\Lambda_n)$ and $\varphi(\widetilde{\Lambda}_n^{(j)})$.

Lemma C.12. *For $p \geq 2$ and arbitrary $j \in \mathbb{T}$ we have*

$$|\varphi(\Lambda_n) - \varphi(\widetilde{\Lambda}_n^{(j)})| \leq p\,\mathbb{E}_\tau\,|\Lambda_n - \widetilde{\Lambda}_n^{(j)}||\widetilde{\Lambda}_n^{(j)} + \tau(\Lambda_n - \widetilde{\Lambda}_n^{(j)})|^{p-2},$$

where $\mathbb{E}_\tau$ denotes expectation with respect to a random variable τ which is uniformly distributed on $[0, 1]$.

Proof. The proof follows from the Newton-Leibniz formula applied to

$$\hat{\varphi}(x) = \varphi(\widetilde{\Lambda}_n^{(j)} + x(\Lambda_n - \widetilde{\Lambda}_n^{(j)})), \quad x \in [0, 1],$$

and

$$|\hat{\varphi}'(x)| \leq p|\widetilde{\Lambda}_n^{(j)} + x(\Lambda_n - \widetilde{\Lambda}_n^{(j)})|^{p-2}.$$

$\square$

The following two lemmas are simple, but will used many times in the proof of Theorem 1.1.

Lemma C.13. *Assume that for all $p > q \geq 1$ and $a, b > 0$ the following inequality holds*

$$x^p \leq a + bx^q. \tag{C.5}$$

Then

$$x^p \leq 2^{\frac{p}{p-q}}(a + b^{\frac{p}{p-q}}).$$

Proof. See [15][Lemma B.3]. $\square$

Lemma C.14. *Let $0 < q_1 \le q_2 \le ... \le q_k < p$ and $c_j, j = 0, ..., k$ be positive numbers such that*

$$x^p \le c_0 + c_1 x^{q_1} + c_2 x^{q_2} + ... + c_k x^{q_k}.$$

Then

$$x^p \le \beta \left[c_0 + c_1^{\frac{p}{p-q_1}} + c_2^{\frac{p}{p-q_2}} + ... + c_k^{\frac{p}{p-q_k}} \right],$$

where

$$\beta := \prod_{\nu=1}^{k} 2^{\frac{p}{p-q_\nu}} \le 2^{\frac{kp}{p-q_k}}.$$

Proof. See [15][Lemma B.4]. □

References

[1] G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[2] L. Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20:262–268, 1967.
[3] Z. Bai, J. Hu, G. Pan, and W. Zhou. A note on rate of convergence in probability to semicircular law. *Electron. J. Probab.*, 16:no. 88, 2439–2451, 2011.
[4] Z. Bai and J. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer, New York, second edition, 2010.
[5] C. Cacciapuoti, A. Maltsev, and B. Schlein. Bounds for the stieltjes transform and the density of states of wigner matrices. *Probability Theory and Related Fields*, pages 1–59, 2014.
[6] M. Chiani. Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution. *J. Multivariate Anal.*, 129:69–81, 2014.
[7] L. Erdős. Universality of Wigner random matrices: a survey of recent results. *Uspekhi Mat. Nauk*, 66(3(399)):67–198, 2011.
[8] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314(3):587–640, 2012.
[9] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18:no. 59, 58, 2013.
[10] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.
[11] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.
[12] L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.
[13] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN*, (3):436–479, 2010.
[14] V. Girko. Spectral theory of random matrices. *Uspekhi Mat. Nauk*, 40(1(241)):67–106, 256, 1985.
[15] F. Götze, A. Naumov, and A. Tikhomirov. Local semicircle law under moment condtions. Part I: The Stieltjes transfrom. *arXiv:1510.07350.*
[16] F. Götze, A. Naumov, and A. Tikhomirov. Limit theorems for two classes of random matrices with dependent entries. *Theory Probab. Appl.*, 59 (114)(1):23–39, 2015.
[17] F. Götze and A. Tikhomirov. On the rate of convergence to the semi-circular law. *arXiv:1109.0611.*

[18] F. Götze and A. Tikhomirov. Rate of convergence of the empirical spectral distribution function to the semi-circular law. *arXiv:1407.2780*.
[19] F. Götze and A. Tikhomirov. Rate of convergence to the semi-circular law. *Probab. Theory Related Fields*, 127(2):228–276, 2003.
[20] F. Götze and A. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. *Teor. Veroyatn. Primen.*, 51(1):171–192, 2006.
[21] F. Götze and A. Tikhomirov. Optimal bounds for convergence of expected spectral distributions to the semi-circular law. *Probability Theory and Related Fields*, pages 1–71, 2015.
[22] J. Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(2):151–178, 2005.
[23] J. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163(1):117–173, 2014.
[24] A. Naumov. Limit theorems for two classes of random matrices with Gaussian elements. *Journal of Mathematical Sciences*, 204(1):140–147, 2014.
[25] L. Pastur. Spectra of random selfadjoint operators. *Uspehi Mat. Nauk*, 28(1(169)):3–64, 1973.
[26] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.
[27] T. Tao and V. Vu. Random matrices: The universality phenomenon for Wigner ensembles. *arXiv:1202.0068*.
[28] T. Tao and V. Vu. Random matrices: sharp concentration of eigenvalues. *Random Matrices Theory Appl.*, 2(3):1350007, 31, 2013.
[29] C. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.
[30] V. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.
[31] E. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

Friedrich Götze, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
E-mail address: goetze@math.uni-bielefeld.de

Alexey A. Naumov, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia, and Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia, and Chinese University of Hong Kong, Department of Statistics, Hong Kong
E-mail address: anaumov@cs.msu.su

Alexander N. Tikhomirov, Department of Mathematics, Komi Research Center of Ural Division of RAS, Syktyvkar, Russia
E-mail address: tikhomirov@dm.komisc.ru