LOCAL SEMICIRCLE LAW UNDER FOURTH MOMENT CONDITION

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

Abstract. We consider a random symmetric matrix \( X = [X_{jk}]_{j,k=1}^n \) with upper triangular entries being independent random variables with mean zero and unit variance. Assuming that \( \max_{j,k} \mathbb{E} |X_{jk}|^{4+\delta} < \infty, \delta > 0 \), it was proved in [17] that with high probability the typical distance between the Stieltjes transforms \( m_n(z), z = u + iv \), of the empirical spectral distribution (ESD) and the Stieltjes transforms \( m_{sc}(z) \) of the semicircle law is of order \( (nv)^{-1} \log n \). The aim of this paper is to remove \( \delta > 0 \) and show that this result still holds if we assume that \( \max_{j,k} \mathbb{E} |X_{jk}|^4 < \infty \). We also discuss applications to the rate of convergence of the ESD to the semicircle law in the Kolmogorov distance, rates of localization of the eigenvalues around the classical positions and rates of delocalization of eigenvectors.

1. Introduction and main result

One of the main questions in random matrix theory is to investigate the limiting behaviour of spectral statistics of eigenvalues of large dimensional random matrices, for example, the distance between neighbouring eigenvalues or \( k \)-point correlation function. It turns out that there is a universality phenomena which states that the distribution of these statistics is independent of the particular distribution of the matrix entries, but depends on some global characteristics like the existence of moments. In the recent years there was a significant progress in the analysis of universality phenomena for the Wigner ensemble of random matrices, i.e. Hermitian matrices with independent entries subject to the symmetry constraint.

We refer the interested reader for a comprehensive literature review and more details to the forthcoming book by L. Erdös and H.-T. Yau [11]. In the current paper we will not discuss the question of universality, but turn our attention to the local semicircle law which is the necessary intermediate step to the universality, but has its own important applications.

In what follows we consider a Hermitian random matrix \( X := [X_{jk}]_{j,k=1}^n \), such that \( X_{jk}, 1 \leq j \leq k \leq n \) are independent random variables (r.v.) with zero mean. We also allow the distribution of matrix entries to depend on \( n \), but omit the latter from matrix notations. Furthermore, for simplicity we will assume that \( \mathbb{E} |X_{jk}|^2 = 1 \) for all \( 1 \leq j \leq k \leq n \). As it was mentioned above we refer to such matrices as Wigner’s ensemble. Denote the eigenvalues of the normalized matrix \( W := n^{-1/2}X \) in the increasing order by \( \lambda_1 \leq \ldots \leq \lambda_n \) and introduce the empirical spectral distribution (ESD) \( \mu_n := n^{-1} \sum_{k=1}^n \delta_{\lambda_k} \). It was proved by E. Wigner [26] and further generalized by many authors (see e.g. monographs [3], [2], [23]) that with probability one \( \mu_n \) weakly converges to the deterministic limit \( \mu_{sc} \) with absolutely continuous density

\[
g_{sc}(\lambda) := \frac{1}{2\pi} \sqrt{(4 - \lambda^2)^+} \tag{1.1}
\]
where \((x)_+ := \max(x, 0)\). In particular, these results imply convergence in the macroscopic regime, i.e. for all intervals of fixed length and independent of \(n\), which contain macroscopically large number of eigenvalues. In turned out that an appropriate analytical tool is the Stieltjes transform of ESD \(\mu_n\) given by
\[
m_n(z) := \int_{-\infty}^{\infty} \frac{\mu_n(d\lambda)}{\lambda - z},
\]
where \(z = u + iv, v > 0\). Under rather general conditions one may show (see e.g. [23]) that with probability one for fixed \(v > 0\)
\[
\lim_{n \to \infty} m_n(z) = m_{sc}(z) := \int_{-\infty}^{\infty} g_{sc}(\lambda) \frac{d\lambda}{\lambda - z} = -\frac{z}{2} + \sqrt{\frac{z^2}{4} - 1}.
\]
It is of interest to investigate the microscopic regime, i.e. the case of smaller intervals, where the number of eigenvalues cease to be macroscopically large. This regime is essential for many applications as the rate of convergence of \(\mu_n\) to the limiting distribution \(\mu_{sc}\), rigidity of eigenvalues \(\lambda_j, j = 1, \ldots, n\), or delocalization of the corresponding eigenvectors \(u_j\) among others. To deal with this regime one needs to establish the convergence of \(m_n(z)\) to \(m_{sc}(z)\) in the region \(1 \geq v \geq f(n)/n\), where \(f(n) > 1\) is some function of \(n\). Significant progress in that direction was recently made in a series of results by L. Erdős, B. Schlein, H.-T. Yau, J. Yin et al [9], [8], [10], [12], [6] showing that with high probability uniformly in \(u \in \mathbb{R}\)
\[
|m_n(u + iv) - s(u + iv)| \leq \frac{\log^{\alpha(n)} n}{nv},
\]
where \(\alpha(n) := c \log \log n\) and \(c\) is some positive constant. This result was called the local semicircle law. It means that the fluctuations of \(m_n(z)\) around \(m_{sc}(z)\) are of order \((nv)^{-1}\) (up to a logarithmic factor). In the papers [9], [8], [10], [12] the inequality (1.4) has been proved assuming that the distribution of \(X_{jk}\) has sub-exponential tails for all \(1 \leq j, k \leq n\). Moreover in [6] this assumption had been relaxed to requiring \(\beta_p := \max_{j,k} \mathbb{E}|X_{jk}|^p \leq C_p\) for all \(p \geq 1\), where \(C_p\) are some constants. In the recent years the series of results appeared, where the latter assumptions were further relaxed to the condition that
\[
\beta_{4+\delta} < \infty
\]
for some \(\delta > 0\), see e.g. [7], [5], [20], [13], [14], [15], [18] and [17]. In particular, the result of [17] implies that (1.4) holds with \(\alpha(n) \equiv 1\).

The main emphasis of the current paper is to remove \(\delta\) from the condition (1.5). The main idea of the proof is motivated by the recent result of A. Aggarwal [1] who established the bulk universality for Wigner’s matrices with finite moments of order \(2 + \varepsilon, \varepsilon > 0\). He proved that (1.4) still holds true, but the factor \((nv)^{-1}\) is replaced by \((nv)^{-1/2} + n^{-c\varepsilon}\), where \(c > 0\) is some constant depending on \(\varepsilon\) and \(\beta_{2+\varepsilon}\). In the current paper we show that (1.4) still holds assuming finite fourth moment only. Taking into account the behaviour of the extreme eigenvalues of \(X\) we also believe that it is the best possible moment assumption for (1.4) to remain valid. In the section 1.3 below we briefly discuss how using technique from [1] and [17] one may achieve this aim.

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 let \(E\) be the mathematical expectation with respect to \(\mathbb{P}\). For a r.v. \(\xi\) we use notation \(E^{1/p} \xi\) to denote \((E \xi^{1/p})^{1/p}\). We denote by \(1[A]\) the indicator function of the set \(A\).
We denote by \( \mathbb{R} \) and \( \mathbb{C} \) the set of all real and complex numbers. Let \( \text{Im} \, z, \text{Re} \, z \) be the imaginary and real parts of \( z \in \mathbb{C} \). We also define \( \mathbb{C}^+ := \{ z \in \mathbb{C} : \text{Im} \, z \geq 0 \} \). Let \( T = [1, \ldots, n] \) denotes the set of the first \( n \) positive integers. For any \( J \subset T \) introduce \( T_j := T \setminus J \). To simplify all notations we will write \( T_{j,l} \) instead of \( T_{(j,l)} \) and \( T_{p,j} \) respectively.

For any matrix \( W \) together with its resolvent \( R \) and Stieltjes transform \( m_n \) we shall systematically use the corresponding notations \( W^{(j)}, R^{(j)}, m_n^{(j)} \), respectively, for the submatrix of \( W \) with entries \( X_{jk}, j, k \in T \setminus J \). For simplicity we write \( W^{(j)}, W^{(j,j)} \) instead of \( W^{(j,j)} \). The same is applies to \( R, m_n, \ldots \).

By \( C \) and \( c \) we denote some positive constants.

For an arbitrary matrix \( A \) taking values in \( \mathbb{C}^{n \times n} \) we define the operator norm by \( \| A \| := \sup_{x \in \mathbb{C}^n : \| x \|=1} \| Ax \|_2, \) where \( \| x \|_2 := (\sum_{j=1}^n |x_j|^2)^{\frac{1}{2}} \). We also define the Hilbert-Schmidt norm by \( \| A \|_2 := \text{Tr} A^* A = (\sum_{j,k=1}^n |A_{jk}|^2)^{\frac{1}{2}} \).

\[ \| A \| := \sup_{x \in \mathbb{R}^n : \| x \|=1} \| Ax \|_2, \] where \( \| x \|_2 := (\sum_{j=1}^n |x_j|^2)^{\frac{1}{2}} \). We also define the Hilbert-Schmidt norm by \( \| A \|_2 := \text{Tr} A^* A = (\sum_{j,k=1}^n |A_{jk}|^2)^{\frac{1}{2}} \).

\[ 1.2. \text{Main results.} \] Without loss of generality we will assume in what follows that \( X \) is a real symmetric matrix which satisfies the following conditions.

**Definition 1.1 (Conditions (C0)).** We say that a Hermitian random matrix \( X \) satisfies conditions \( (\text{C0}) \) if its entries in the upper triangular part are independent random variables with \( \mathbb{E} X^{(n)}_{jk} = 0, \mathbb{E} |X^{(n)}_{jk}|^2 = 1 \) and \( \max_{j,k,n} \mathbb{E} |X^{(n)}_{jk}|^4 =: \beta_4 < \infty \).

Our results proven below apply to the case of Hermitian matrices as well. Here we may additionally assume for simplicity that real and imaginary parts, \( \text{Re} \, X_{jk}, \text{Im} \, X_{jk}, \) are independent r.v. for all \( 1 \leq j < k \leq n \). Otherwise one needs to extend the moment inequalities for linear and quadratic forms in complex r.v. (see [13] [Theorem A.1-A.2]) to the case of dependent real and imaginary parts, the details of which we omit.

We will also often refer to the following condition (C1).

**Definition 1.2 (Conditions (C1)).** We say that the set of conditions (C1) holds if (C0) are satisfied and \( |X_{jk}| \leq \sqrt{n/R}, 1 \leq j, k \leq n, \) where \( R \geq \log^3 n \).

Let us introduce the following notation
\[ \Lambda_n(z) := m_n(z) - m_{sc}(z), \quad z = u + iv, \]
where \( m_n(z), m_{sc}(z) \) were defined in (1.2) and (1.3) respectively. Recall that Im \( \Lambda_n \) is the imaginary part of \( \Lambda_n \). The main result of this paper is the following theorem, which estimates the fluctuations (1.4).

**Theorem 1.3.** Assume that the conditions (C1) hold and let \( V > 0 \) be some constant.

- There exist positive constants \( A_0, A_1 \) and \( C \) depending on \( V \) and \( \beta_4 \) such that
\[ \mathbb{E} |\Lambda_n(u+iv)|^p \leq \left( \frac{C p}{nv} \right)^p, \]
for all \( 1 \leq p \leq A_1 \log n, V \geq v \geq A_0 n^{-1} \log^2 n \) and \( |u| \leq 2 + v \).
- For any \( u_0 > 0 \) there exist positive constants \( A_0, A_1 \) and \( C \) depending on \( u_0, V \) such that
\[ \mathbb{E} |\text{Im} \, \Lambda_n(u+iv)|^p \leq \left( \frac{C p}{nv} \right)^p, \]
for all \( 1 \leq p \leq A_1 \log n, V \geq v \geq A_0 n^{-1} \log^2 n \) and \( |u| \leq u_0 \).
Remark. 1. Using Markov’s inequality the bound (1.6) may be used to show that for any $Q > 0$ there exists some positive constant $C$ such that with probability at least $1 - n^{-Q}$ for all $v \geq A_0 n^{-1} \log^2 n$ and $|u| \leq 2 + v$:

$$|\Lambda_n(u + iv)| \leq C \frac{\log n}{nv}.$$  

Hence, (1.4) holds with $\alpha(n) \equiv 1$.

2. It is interesting to investigate the case of generalised matrix when for any $j = 1, \ldots, n$ \sum_{k=1}^{n} \mathbb{E}|X_{jk}|^2 = n$, but $\mathbb{E}|X_{jk}|^2$ could be different. Unfortunately, the technique of the current paper doesn’t allow to deal with such case directly. Fortunately, one may apply a combination of the multiplicative descent used in this paper (and first developed in [4]) together with the additive descent developed in the series of papers by L. Erdős, B. Schlein, H.-T. Yau, J. Yin et al; see e.g. [11]. This combination was recently used in [16]. We don’t give details here to simplify the proof.

The result of the previous theorem may be formulated under conditions (C0). In this case one may truncate and re-normalize the entries of $X$ by means of Lemmas A.1–A.3 in the appendix. We obtain the following corollary.

**Corollary 1.4.** Assume that the conditions (C0) hold and let $V > 0$ be some constant. There exist positive constants $A_0, A_1$ and $C$ depending on $V$ and $\beta_4$ such that

$$\mathbb{E}|\Lambda_n(u + iv)|^p \leq \frac{C^p \log^{12p} n}{(nv)^p},$$

for all $1 \leq p \leq A_1 \log n$, $V \geq v \geq A_0 n^{-1} \log^2 n$ and $|u| \leq 2 + v$. Similar result holds true for (1.7).

We believe that the power of the logarithm could be reduced. The main technical problem is in Lemmas A.1–A.3 in the appendix. Truncation on the level near $\sqrt{n}$ requires additional logarithmic factors.

### 1.3. Sketch of the proof of Theorem 1.3

To prove Theorem 1.3 we use the strategy from [17].

1. Applying inequality (2.6) we may estimate $\mathbb{E}|\Lambda_n|^p$ (depending on Re($z$) being near or far from the spectral interval $[-2, 2]$) by the moments $\mathbb{E}|T_n(z)|^p$. This inequality first appeared in [4][Proposition 2.2].

2. Estimation of $\mathbb{E}|T_n(z)|^p$ consists of two parts:
   a) Estimation of $\mathbb{E}|R_{jk}(z)|^p$; see Lemma 3.1. This bound requires to estimate high moments of $\varepsilon_j$ (i.e. quadratic and linear forms $\varepsilon_{2j}, \varepsilon_{3j}$); see (2.2). This step also uses the following crude bound

$$\mathbb{E}|T_n|^p \leq \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j^{1/2} |\varepsilon_j|^{2p} \mathbb{E}^{1/2} |R_{jk}|^{2p}. \quad (1.8)$$

Unfortunately, the technique from [13], [15] doesn’t work since we may truncate $X_{jk}$ on the level $\sqrt{n}/R$, where $R$ is of the logarithmic order (opposite to the case when $\mathbb{E}|X_{jk}|^{1+\delta} < \infty$. This allows to truncate on the level $n^{1/2-\delta}$ for some small $\delta$. Let us demonstrate this on the quadratic form $\varepsilon_{2j}$. Applying [15][Theorem 7] or [13][Theorem A.2] we obtain

$$\mathbb{E}|\varepsilon_{2j}|^p \leq \frac{C^p \epsilon_2^p}{(nv)^{p/2}} \mathbb{E} \log m_n^{(j)}(z) + \frac{\beta_0 p^{2p} \epsilon_2^p}{n^{p/2}} \sum_{k=1}^{n} \mathbb{E} \log m_n^{(j)}(z) + \frac{C^p \epsilon_2^p \beta_0^2}{n^{p/2}} \sum_{k,l=1}^{n} \mathbb{E}|R_{kk}|^{2p},$$
where $p$ satisfies $1 \leq p \leq (nv)^{1/4}$. There is no problem to deal with first two terms in the r.h.s. of the previous inequality. The most difficult term is the last one. In the sub-Gaussian case this term has the order $Cp^p p^{p/2} (n^2 v)^{-p}$ (see [17][Lemma 4.4]) and is small for $n^{-1} \leq v \leq 1$ (here we also use the fact that $\beta_1^{1/p} \leq C \sqrt{p}$). Under assumptions (C1) we may only guarantee that $E |R_{k,l}|^{2p} \leq Cp$. But $\beta_p \leq \beta_4 n^{p/2} - 2 |X|^{-p}, p \geq 4$.

Hence, the last term in the estimate for $\varepsilon_j z^2$ is bounded by $\beta_4^2 C p^p p^{2p}\bar{T}^{2p+8}$, which could be very large for large $p$. It is worth to mention here, that if one can truncate on the level, say, $n^{1/4}$, then there will be an additional factor $n^{p/2}$ in the denominator.

To overcome this problem we use ideas from [1]. We introduce configuration matrix $L := [L_{jk}]_{j,k=1}^n$ such that $L_{jk} = 1$ if $X_{jk} \leq n^{1/4} R$ and $L_{jk} = 0$ if $n^{1/4} R < |X_{jk}| \leq n^{1/2} R$ for some $R$ of the logarithmic order; see (3.4). One may show that with high probability this matrix has the block structure (see (3.6)). This means that with high probability in each row and in each column of $X$ there is only small (of logarithmic order) number of large entries ($L_{jk} = 0$) and large number of small entries. Fixing admissible (see Definition 3.6 below) configuration $L$ (corresponding to the block structure) one may estimate $E |R_{jk}(z)|^p |L|, z = u + iv$; see Lemma 3.7. For each subrow with small entries we use bounds from [13], [15]. For each subrow with large entries we may use crude bounds which doesn’t contain factor $p^p$; see decomposition (3.13) and corresponding estimates below. Using now total probability rule and the crude bound $|R(z)| \leq v^{-1}$ if $L$ is not admissible we estimate $E |R_{jk}(z)|^p$.

b) More accurate (than (1.8)) bounds for $E |T_n(z)|^p$; see section 4. We use Lemma 4.1 which provides a general framework for estimation of moments of statistics of independent r.v. This requires estimation of $E |\varepsilon_j|^a$ for $1 \leq a \leq 2$. The latter could be done since $X_{jk}$ has moments of order up to 4.

1.4. Applications of the main results. This section is addressed to application of Theorem 1.3 and Corollary 1.4 to different questions as the rate of convergence of the ESD $\mu_n$ to the semicircle law $\mu_{sc}$, rigidity estimates for the eigenvalues $\lambda_j, j = 1, \ldots, n$ and delocalization bounds for the corresponding eigenvectors $u_j, j = 1, \ldots, n$. Up to the power of logarithmic order these results repeat the corresponding results from [14], [17]. We formulate all results with comments, but leave the proof. The interested reader may recover the proof from the corresponding papers mentioned above. It is worth to mention that these questions has been intensively studied under stronger assumptions in many papers; see e.g. [9], [8], [10], [5], [7], [6] and [25]. We also refer to the recent monograph [11] and survey [24].

1.4.1. Rate of convergence of ESD. Our first result provides quantitative estimates for the rate of convergence of the ESD to the semicircle law in the Kolmogorov distance.

Corollary 1.5. Assume that the conditions (C0) hold. For any $Q > 0$ there exists positive constant $C$ such that with probability at least $1 - n^{-Q}$

$$\Delta_n^* := \sup_{x \in \mathbb{R}} |\mu_n((-\infty, x]) - \mu_{sc}((-\infty, x])| \leq \frac{C \log^{12} n}{n}.$$ 

For the proof see [17][Theorem 1.4]. The difference is in application of Corollary 1.4 instead of [17][Theorem 1.1]. The proof is mainly based on application of the smoothing inequality (see e.g. [17][Corollary 6.2]) and Corollary 1.4. We believe that the power of the logarithm could be reduced from 12p to $p$ or even $\frac{p}{2}$, which would be optimal due to the result of Gustavsson [19] for the Gaussian Unitary Ensembles (GUE).
Using this result on main prove the following corollary

**Corollary 1.6.** Assume that conditions (C0) hold. For any $Q > 0$ there exists positive constant $C$ such that for all $\Delta > 0$ with probability at least $1 - n^{-Q}$:

$$
\# \{ \lambda_j \in [x - \Delta/2n, x + \Delta/2n] \} - g_{sc}(x) \Delta \leq \frac{C \log^2 n}{n}.
$$

1.4.2. Rigidity. Taking into account the result of Theorem 1.5 and using Smirnov’s transform one may also get the rigidity estimates for the majority of eigenvalues $\lambda_j$. More precisely, one may control the eigenvalues on the bulk of the spectrum. To deal with the smallest (largest) eigenvalue one needs more accurate bound then in Theorem 1.7.

**Theorem 1.7.** Assume that the conditions (C1) hold and $u_0 > 2$ and $V > 0$. There exist positive constants $A_0, A_1$ and $C$ depending on $u_0, V$ and $\beta_4$ such that

$$
\mathbb{E} |\text{Im} \Lambda_n(z)|^p \leq \frac{C_p p^p}{n^p (\kappa + v)^p} + \frac{C_p p^{2p}}{(nV)^2 (\kappa + v)^p} + \frac{C_p p^2}{n^p (\kappa + v)^p} + \frac{C_p p^p}{(nV)^2 (\kappa + v)^p},
$$

for all $1 \leq p \leq A_1 \log n$, $V \geq v \geq A_0 n^{-1} \log^2 n$ and $2 \leq |u| \leq u_0$.

Let us define the quantile position of the $j$-th eigenvalue by

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

The following results give the bounds for the fluctuations of $\lambda_j$ around $\gamma_j$.

**Corollary 1.8.** Assume that the conditions (C0) hold and let $K > 0$ be an integer. Then

- (bulk) Let $j \in [\log n, n - \log n + 1]$ . For any $Q > 0$ there exists positive constant $C$ such that with probability at least $1 - n^{-Q}$:

$$
|\lambda_j - \gamma_j| \leq C_1 \log^2 n [\min(j, n - j + 1)]^{-\frac{1}{2}} n^{-\frac{3}{4}}.
$$

- (edge) Let $j \leq \log n$ or $j \geq n - \log n + 1$. There exists positive constant $C$ such that with probability at least $1 - \log^{-1} n$:

$$
|\lambda_j - \gamma_j| \leq C \log^9 n [\min(j, n - j + 1)]^{-\frac{1}{4}} n^{-\frac{3}{4}}.
$$

For the detailed proof see [14][Theorem 1.3] making minor changes. For the bulk of the spectrum we mainly use the following formula

$$
\lambda_j = G_{sc}^{-1} \left( \frac{j}{n} \right) + \mathbb{E}_r \frac{2 \pi \theta \hat{\Delta}_n^*}{\sqrt{4 - \left( G_{sc}^{-1} \left( \frac{j}{n} + \theta \hat{\Delta}_n^* \right) \right)^2}};
$$

see proof of [14][Theorem 1.3]. Here, $G_{sc}(x) := \mu_{sc}((-\infty, x])$. Taking into account that

$$
c_1 x^\frac{1}{2} \leq 2 + G_{sc}^{-1}(x) \leq c_2 x^\frac{1}{2} \quad \text{for} \ x \in [0, 1/2],
$$

$$
c_1 (1 - x)^\frac{1}{2} \leq 2 - G_{sc}^{-1}(x) \leq c_2 (1 - x)^\frac{1}{2} \quad \text{for} \ x \in [1/2, 1],
$$

and Corollary 1.5 one may obtain the estimates for the bulk of the spectrum. Clearly, the factor $\log^2 n$ comes from the bound for the $\Delta_n^*$. The proof for the edge of the spectrum requires more involved technique. In particular, following [6][Theorem 7.6] we write

$$
P(|\lambda_j - \gamma_j| \geq l) \leq P(|\lambda_j - \gamma_j| \geq l, \lambda_j > \gamma_j) + P(|\lambda_j - \gamma_j| \geq l, \lambda_j < \gamma_j),
$$
where \( l := CK_j^{-1/3}n^{-2/3} \) for some \( C > 0 \). The first case when \( \lambda_j > \gamma_j \) is trivial since in this situation \( \lambda_j > j_1 \geq -2 + c_1n^{-2/3} \) (see (1.9)) and we may repeat the calculations for the case of the bulk to get

\[
\mathbb{P}(|\lambda_j - \gamma_j| \geq l, \lambda_j > \gamma_j) \leq n^{-Q}.
\]

Applying (1.9) we obtain \( \gamma_j \leq -2 + c_2 \left( \frac{j}{n} \right)^2 \). This enables to write the estimate

\[
\mathbb{P}(|\lambda_j - \gamma_j| \geq l, \lambda_j > \gamma_j) \leq \mathbb{P}(\lambda_1 \leq -2 - (K/n)^{2/3}).
\]  (1.10)

Estimation of the r.h.s. of the previous inequality requires to use truncation technique leading to very poor probability bounds (of order \( \log^{-1} n \)). Namely, we need to replace \( W \) satisfying (C0) by the corresponding matrix \( \tilde{W} \) satisfying (C1). To estimate the r.h.s. of (1.10) one may follow [6][Theorem 7.3] and use Theorem 1.7.

1.4.3. Delocalization of eigenvectors. Let us denote by \( u_j := (u_{j1}, ..., u_{jn}) \) the eigenvectors of \( W \) corresponding to the eigenvalue \( \lambda_j \). The following theorem is the direct corollary of Lemma 3.1.

**Corollary 1.9.** Assume that conditions (C0) hold. There exist positive constant \( C \) such that with probability at least \( 1 - \log^{-1} n \):

\[
\max_{1 \leq j, k \leq n} |u_{jk}| \leq C \sqrt{\frac{\log n}{n}}.
\]

Comparison with a similar result for the Gaussian Orthogonal Ensembles (GOE) (see [2][Corollary 2.5.4]) shows that this result is optimal with respect to the power of logarithm. For the proof see [17][Theorem 1.4], replacing [17][Lemma 3.1] with Lemma 3.1( with \( \alpha = 1 \)). For the readers convenience we give an idea of the proof. We introduce the following distribution function

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

Using the eigenvalue decomposition of \( W \) it is easy to see that

\[
R_{jj}(z) = \sum_{k=1}^{n} \frac{|u_{jk}|^2}{\lambda_k(W) - z} = \int_{-\infty}^{\infty} \frac{1}{x - z} dF_{nj}(x).
\]

For any \( \lambda > 0 \) we have

\[
\max_{1 \leq k \leq n} |u_{jk}|^2 \leq \sup_x (F_{nj}(x + \lambda) - F_{nj}(x)) \leq 2 \sup_u \lambda \text{Im} R_{jj}(u + i\lambda).
\]

Estimation of the r.h.s. of the previous inequality requires again to use truncation technique leading to very poor probability bounds. Similarly to the edge case of Corollary 1.8 we replace \( W \) satisfying (C0) by the corresponding matrix \( \tilde{W} \) satisfying (C1), and apply Lemma 3.1( with \( \alpha = 1 \)). Replacing conditions (C0) by (C1) one may improve the estimate.

2. Proof of the main result

We start this section with the recursive representation of the diagonal entries of the resolvent \( R(z) := (W - zI)^{-1} \). As noted before we shall systematically use for any matrix \( W \) together with its resolvent \( R \), Stieltjes transform \( m_n \) and etc. the corresponding quantities \( W^{(j)}, R^{(j)}, m_n^{(j)} \) and etc. for the corresponding sub matrix with entries \( X_{jk}, j, k \in \mathbb{T} \setminus J \).
Here $\mathbb{T} := \{1, \ldots, n\}$ and $\mathbf{J} \subset \mathbb{T}$. We will often omit the argument $z$ from $\mathbf{R}(z)$ and write $\mathbf{R}$ instead. We may express $\mathbf{R}_{jj}$ in the following way

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

Let $\varepsilon_j := \varepsilon_{1j} + \varepsilon_{2j} + \varepsilon_{3j} + \varepsilon_{4j}$, where

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

Using these notations we may rewrite (2.1) as follows

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

Introduce

$$
\Lambda_n := m_n(z) - m_{sc}(z), \quad b(z) := z + 2m_{sc}(z), \quad b_n(z) = b(z) + \Lambda_n,
$$

and

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

Applying (2.3) we arrive at the following representation for $\Lambda_n$ in terms of $T_n$ and $b_n$

$$
\Lambda_n = \frac{T_n}{z + m_n(z) + m_{sc}(z)} = \frac{T_n}{b_n(z)}.
$$

It was proved in [4][Proposition 2.2] (see also [13][Lemma B.1]) that for all $v > 0$ and $|u| \leq 2 + v$ (using the quantities (2.4))

$$
|\Lambda_n| \leq C \min \left\{ \frac{|T_n|}{|b(z)|}, \sqrt{|T_n|} \right\}.
$$

Moreover, for all $v > 0$ and $|u| \leq u_0$

$$
|\text{Im} \Lambda_n| \leq C \min \left\{ \frac{|T_n|}{|b(z)|}, \sqrt{|T_n|} \right\}.
$$

It is easy to check that $b(z) = \sqrt{z^2 - 4}$ and moreover, there exist constants $c, C > 0$ such that $c\sqrt{\kappa + v} \leq |b(z)| \leq C \sqrt{\kappa + v}$ for all $|u| \leq u_0$, $0 < v \leq V$, where $\kappa$ is defined in the section 1.4.2; see e.g. [6][Lemma 4.3]. Hence, in order to bound $\mathbb{E} |\Lambda_n|^p$ (or $\mathbb{E} |\text{Im} \Lambda|^p$ respectively) it is enough to control $\mathbb{E} |T_n|^p$.

Let us introduce the following region in the complex plane:

$$
\mathbb{D}_2 := \{ z = u + iv \in \mathbb{C} : |u| \leq u_0, V \geq v \geq v_0 := A_0 n^{-1} \log^2 n \},
$$

where $u_0, V$ are arbitrary fixed positive real numbers and $A_0$ is some large constant defined below.

The following theorem provides a bound for $\mathbb{E} |T_n|^p$ for all $z \in \mathbb{D}_2$ in terms of diagonal resolvent entries.
Theorem 2.1. Assume that the conditions (C1) hold and $u_0 > 2$ and $V > 0$. There exist positive constants $A_0, A_1$ and $C$ depending on $u_0, V$ and $\beta_1$ such that for all $z \in \mathbb{D}_2$ we have

$$\mathbb{E}|T_n|^p \leq \frac{C^p p^p |b(z)|^p}{(nv)^p} + \frac{C^p p^{2p}}{(nv)^{2p}},$$

(2.9)

where $1 \leq p \leq A_1 \log n$.

Remark. To prove Theorem 1.7 one need more stronger bound for $\mathbb{E}|T_n|^p$ than (2.9). Minor changes in the proof of Theorem 2.1 will lead to the following estimate

$$\mathbb{E}|T_n|^p \leq \frac{C^p p^p A^p (4p)}{(nv)^p} + \frac{C^p p^{2p}}{(nv)^{2p}} + \frac{C^p p^{p/2} |b(z)|^{\frac{p}{2}} A^{\frac{p}{2}} (4p)}{(nv)^p},$$

where $A(q) := \max \{ \max_{j=1,\ldots,n} \mathbb{E}^n \text{Im}^n R_{jj}, \text{Im} m_{sc}(z) \}$. This estimate is sufficient for our purposes. The term $A(q)$ may be estimate due to Lemma 3.1. We omit the details.

The proof of Theorem 2.1 is one of the crucial steps in the proof of the main result and will be given in the next section. We finish this section with the proof of Theorems 1.3 and 1.7.

Proof of Theorem 1.3. To estimate $\mathbb{E} |\text{Im} \Lambda_n|^p$ we may choose one of the bounds (2.7), depending on whether $z$ is near the edge of the spectrum or away from it. If $|b(z)| \geq C p (nv)^{-1}$ then we may take the bound $|\text{Im} \Lambda_n| \leq C |T_n|/|b(z)|$ and obtain

$$\mathbb{E} |\text{Im} \Lambda_n|^p \leq \frac{C^p \mathbb{E} |T_n|^p}{|b(z)|^p} \leq \left( \frac{C p}{nv} \right)^p.$$ 

If the opposite inequality holds, $|b(z)| < C p (nv)^{-1}$, then we will use the bound $|\text{Im} \Lambda_n| \leq C \sqrt{|T_n|}$:

$$\mathbb{E} |\text{Im} \Lambda_n|^p \leq C^p \mathbb{E} \frac{T_n}{|b(z)|^p} \leq \left( \frac{C p}{nv} \right)^p.$$ 

Both inequalities combined yield

$$\mathbb{E} |\text{Im} \Lambda_n|^p \leq \left( \frac{C p}{nv} \right)^p.$$ 

Similar arguments are applicable to $\mathbb{E} |\Lambda_n|^p$.

Proof of Theorem 1.7. Following the remark after Theorem 2.1 we may conclude that

$$\mathbb{E}|T_n|^p \leq \frac{C^p p^p \text{Im}^n m_{sc}(z)}{(nv)^p} + \frac{C^p p^{2p}}{(nv)^{2p}} + \frac{C^p p |b(z)|^{\frac{p}{2}} \text{Im}^n m_{sc}(z)}{(nv)^p} + \frac{C^p p^{p/2} |b(z)|^{\frac{p}{2}} A^{\frac{p}{2}} (4p)}{(nv)^p}.$$ 

Using the bound $|\text{Im} \Lambda_n| \leq C |T_n|/|b(z)|$ we get

$$\mathbb{E} |\text{Im} \Lambda_n|^p \leq \frac{C^p p^p \text{Im}^n m_{sc}(z)}{(nv)^p |b(z)|^p} + \frac{C^p p^{2p}}{(nv)^{2p} |b(z)|^p} + \frac{C^p p^{p/2} \text{Im}^n m_{sc}(z)}{(nv)^p |b(z)|^{\frac{p}{2}}} + \frac{C^p p^{p/2} |b(z)|^{\frac{p}{2}}}{(nv)^{2p} |b(z)|^{\frac{p}{2}}}.$$ 

Since $c \sqrt{\kappa + v} \leq |b(z)| \leq C \sqrt{\kappa + v}$ for all $|u| \leq u_0$, $0 < v \leq V$, and

$$\frac{c v}{\sqrt{\kappa + v}} \leq \text{Im} m_{sc}(z) \leq \frac{c v}{\sqrt{\kappa + v}}$$

for all $2 \leq |u| \leq u_0$, $0 < v \leq V$;

(see e.g. [6][Lemma 4.3]) we finally get

$$\mathbb{E} |\text{Im} \Lambda_n|^p \leq \frac{C^p p^p}{n^p (\kappa + v)^p} + \frac{C^p p^{2p}}{(n^2 v^2 (\kappa + v)^2)} + \frac{C^p p^{p/2}}{(n^p v^{p/2} (\kappa + v)^{p/2})} + \frac{C^p p^{p/2}}{(n^p v^{p/2} (\kappa + v)^{p/2})}.$$
This bound concludes the proof of the theorem. \[\square\]

3. Bounds for moments of diagonal entries of the resolvent

The main result of this section is the following lemma which provides a bound for moments of the diagonal entries of the resolvent. Recall that (see the definition \((2.8)\)) for \(\alpha = 1, 2\)

\[\mathbb{D}_\alpha := \{z = u + iv \in \mathbb{C} : |u| \leq u_0, V \geq v \geq v_0 := A_0 n^{-1} \log^a n\},\]

where \(u_0, V > 0\) are any fixed real numbers and \(A_0\) is some large constant determined below.

The first value \(\alpha = 1\) is sufficient to obtain optimal bounds for delocalization of eigenvectors. The second value, \(\alpha = 2\), is necessary for the main Theorem 1.3.

**Lemma 3.1.** Assuming the conditions \((C1)\) there exist a positive constant \(H_0\) depending on \(u_0, V\) and positive constants \(A_0, A_1\) depending on \(H_0\) such that for all \(z \in \mathbb{D}_\alpha\) and \(1 \leq p \leq A_1 \log^a n\) we have

\[
\max_{1 \leq j \leq k \leq n} \mathbb{E} |R_{jk}(z)|^p \leq H_0^p, \tag{3.1}
\]

\[
\mathbb{E} \frac{1}{|z + m_n(z)|^p} \leq H_0^p, \tag{3.2}
\]

\[
\mathbb{E} \text{Im}^p R_{jj}(z) \leq H_0^p \left[ \text{Im}^p m_{sc}(z) + \frac{p^p}{(nv)^p} \right]. \tag{3.3}
\]

We provide the proof of \((3.1)\) only. The proof of \((3.2)\) and \((3.3)\) is the same and will be omitted. For the details see \([17][Lemma 3.1]\).

We start with introducing the following events

\[A_{jk} := \{ |X_{jk}| \leq n^{1/2} \bar{R} \}, \quad B_{jk} := A_{jk}^c = \{ n^{1/2} \bar{R} \leq |X_{jk}| \leq n^{1/2} / \bar{R} \},\]

where \(\bar{R} := \bar{R}_n\) is some quantity depending on \(n\). We also denote

\[p_n := \mathbb{P}(n^{1/2} \bar{R} \leq |X_{jk}| \leq n^{1/2} / \bar{R})\]

Using Markov’s inequality, it is easy to check that

\[p_n \leq n^{-1} \bar{R}^4.\]

Following \([1]\) let us introduce the following configuration matrix

\[L := L(X) := [L_{jk}]_{j,k=1}^n\]

with \(L_{jk} := 1[A_{jk}]\). Let \(\xi_{jk}\) and \(\eta_{jk}\), \(j, k = 1, \ldots, n\), be mutually independent random variables distributed as \(X_{jk}\) conditioned on \(A_{jk}\) and \(B_{jk}\) resp. Let \(X(L) := [X_{jk}(L_{jk})]_{j,k=1}^n\)

where

\[X_{jk}(L_{jk}) := \begin{cases} 
\xi_{jk}, & \text{if } L_{jk} = 1, \\
\eta_{jk}, & \text{if } L_{jk} = 0.
\end{cases} \tag{3.4}
\]

We may consider matrix \(W(L)\) as the matrix \(W\) conditioned on the configuration \(L\). We repeat some classification of configuration matrices from \([1]\).

**Definition 3.2.** Fix an \(n \times n\) configuration matrix \(L = [L_{jk}]_{j,k=1}^n\). We call \(j, k \in \mathbb{T}\) linked (w.r.t \(L\)) if \(L_{jk} = 0\); otherwise we call them unlinked.

**Definition 3.3.** The indices \(j\) and \(k\) are connected if there exists an integer \(l\) and a sequence of indices \(j = j_1, j_2, \ldots, j_l = k\) such that \(j_{\nu}\) is linked to \(j_{\nu+1}\) for each \(\nu \in [1, l - 1]\).
Definition 3.4. We call index $j$ deviant (w.r.t. to $L$) if there exist some index $k$ such that $j$ and $k$ are linked. Otherwise $j$ is called typical. Let $\mathcal{D} := \mathcal{D}_L \subset T$ denote the set of deviant indexes, and let $\mathcal{T} := \mathcal{T}_L \subset T$ denote the set of typical indexes.

Definition 3.5. We say that $L$ is:

- deviant-inadmissible if there exist at least $K \max(1, n^2 p_n)$ deviant indices, where $K$ may depend on $n$.
- $r$-connected-inadmissible if there exist distinct indices $j_1, j_2, \ldots, j_r$ that are pairwise connected.

For any $L$ define

$$r(L) := \max\{r \geq 1 : L \text{ is } r\text{-connected-inadmissible}\}.$$ 

Definition 3.6. We call configuration $L$ $r$-admissible, if it is not deviant-inadmissible and $r(L) \leq r - 1$.

Following [1] we may estimate

$$\mathbb{P}(L \text{ is } r\text{-connected-inadmissible}) \leq \binom{n}{r} \left(\frac{r^2}{r - 1}\right) p_n^{r - 1} \leq \left(\frac{ne}{r}\right)^r \left(\frac{er^2}{r - 1}\right)^{r - 1} \left(\frac{1}{nR}\right)^{r - 1} \leq e^{3nR^{-1}R^{-4(r - 1)}}.$$ 

Applying Chernoff’s inequality we may also show that

$$\mathbb{P}(L \text{ is deviant-inadmissible}) \leq e^{-cK \max(1, n^2 p_n)}.$$ 

Denote by $\mathcal{L}_r$ the set of all $L$ $r$-admissible configurations. In what follows we take $r := \log^3 n, R := \log n, K := \log^3 n$. Then

$$\mathbb{P}(L \notin \mathcal{L}_r) \leq n^{-c\log^2 n} \quad \text{(3.5)}$$

for some large $c > 0$.

Let us fix the $r$-admissible configuration $L$. By definition of $r$-admissibility we may find Hermitian matrices $A_\nu$ of order $r_\nu \leq r$, $\nu = 1, \ldots, L$ such that $r_1 + \ldots + r_L \leq K \max(1, n^2 p_n)$ and matrix $L$ may be rewritten as follows

$$L = \begin{bmatrix}
A_1 & 1 & \ldots & 1 & \ldots & 1 \\
1 & A_2 & \ldots & 1 & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
1 & \ldots & 1 & A_L & \ldots & 1 \\
1 & \ldots & 1 & \ldots & 1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\
1 & \ldots & 1 & \ldots & 1 & \ldots 
\end{bmatrix}. \quad \text{(3.6)}$$

Moreover, the zero-entries of matrix $L$ can only be inside of $A_\nu$, and in each row (column) may contain at most $r$ zero-entries.

Denote $H := X(L)$ and $G := (n^{-1/2}H - zI)^{-1}$. We also assume that $H = [h_{jk}]_{j,k=1}^n$.

Lemma 3.7. Let $L$ be $r$-admissible. 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 $C_0$ such that for all $z \in \mathbb{D}_\alpha$ and $1 \leq p \leq A_1 (nu)^{1/4}/T$ we have

$$\max_{1 \leq j,k \leq n} \mathbb{E}|G_{jk}|^p \leq C_0^p.$$
Since \( u \) is fixed and \( |u| \leq u_0 \) we shall omit \( u \) from the notation of the resolvent and denote \( G(v) := G(z) \). Sometimes in order to simplify notations we shall also omit the argument \( v \) in \( G(v) \) and just write \( G \). For any \( j \in \mathbb{T}_J \) (see section 1.1) we may express \( G^{(j)}_{jj} \) in the following way (compare with (2.1))

\[
G^{(j)}_{jj} = -\frac{1}{z + m^{(j)}_n(z) - \varepsilon^{(j)}_j},
\]

where \( m^{(j)}_n(z) := \frac{1}{n} \text{Tr} \ G^{(j)}(z) \) and \( \varepsilon^{(j)}_j := \varepsilon^{(j)}_{1j} + \ldots + \varepsilon^{(j)}_{jj} \). Here

\[
\varepsilon^{(j)}_{1j} := \frac{1}{\sqrt{n}} h_{jj}, \quad \varepsilon^{(j)}_{2j} := -\frac{1}{n} \sum_{t \neq k \in \mathbb{T}_J} h_{jk} h_{ji} G^{(j)}_{ki}, \quad \varepsilon^{(j)}_{3j} := -\frac{1}{n} \sum_{k \in \mathbb{T}_J} (h_{jk}^2 - \sigma_{jk}^2) G^{(j)}_{kk},
\]

\[
\varepsilon^{(j)}_{4j} := -\frac{1}{n} (\text{Tr} \ G^{(j)} - \text{Tr} \ G^{(\bar{j}, j)}), \quad \varepsilon^{(j)}_{5j} := \frac{1}{n} \sum_{k \in \mathbb{T}_J} (1 - \sigma_{jk}^2) G^{(j)}_{kk}.
\]

We also introduce the quantities \( \Lambda^{(j)}_n(z) := m^{(j)}_n(z) - m_{sc}(z) \) and

\[
T^{(j)}_n := \frac{1}{n} \sum_{j \in \mathbb{T}_J} \varepsilon^{(j)}_j G^{(j)}_{jj}.
\]

The following lemma allows to recursively estimate the moments of \( G^{(j)}_{jj} \).

**Lemma 3.8.** For an arbitrary set \( \mathbb{J} \subset \mathbb{T} \) and all \( j \in \mathbb{T}_\mathbb{J} \) there exist a positive constant \( c_0 \) depending on \( u_0, V \) only such that for all \( z = u + iv \) with \( V \geq v > 0 \) and \( |u| \leq u_0 \) we have

\[
|G^{(j)}_{jj}| \leq c_0 \left( 1 + |T^{(j)}_n|^{1/2} |G^{(j)}_{jj}| + |\varepsilon^{(j)}_j||G^{(j)}_{jj}| \right).
\]

**Proof.** The proof may be found in [4][Lemma 3.4] or [13][Lemma 4.2]. □

Let us take \( \tilde{v}_0 := A_0 n^{-1} \log^{4(\alpha + 1)} n \).

**Lemma 3.9.** Let \( L \) be \( r \)-admissible and assume that the conditions \( (C1) \) hold. Let \( C_0 \) and \( s_0 \) be arbitrary numbers such that \( C_0 \geq \max(1/V, 6c_0), s_0 \geq 4 \). There exist a sufficiently large constant \( A_0 \) and small constant \( A_1 \) depending on \( C_0, s_0, V \) only such that the following statement holds. Fix some \( \tilde{v} : \tilde{v}_0 s_0 \leq \tilde{v} \leq V \). Suppose that for some integer \( L > 0 \), all \( u, v, q \) such that \( \tilde{v} \leq v' \leq V, |u| \leq u_0, 1 \leq q \leq A_1(nv^{-1/4}/R) \)

\[
\max_{J:|J| \leq L} \max_{k \in \mathbb{T}_J} \mathbb{E} |G^{(j)}_{ik}(v')|^q \leq C_0^q.
\]

Then for all \( u, v, q \) such that \( \tilde{v}/s_0 \leq v \leq V, |u| \leq u_0, 1 \leq q \leq A_1(nv)^{-1/4}/R \)

\[
\max_{J:|J| \leq L-1} \max_{k \in \mathbb{T}_J} \mathbb{E} |G^{(j)}_{ik}(v)|^q \leq C_0^q.
\]

**Proof.** Let us fix an arbitrary \( s_0 \geq 4 \) and \( v \geq \tilde{v}/s_0, \mathbb{J} \subset \mathbb{T} \) such that \( |\mathbb{J}| \leq L-1 \). In the following let \( j, k \in \mathbb{T}_J \). First we note that for any \( j = 1, \ldots, n \) and \( 1 \leq q \leq A_1(nv) \) we may write

\[
\mathbb{E} |G^{(j,j)}_{jj}|^{2q} \leq s_0^{2q} \mathbb{E} |G^{(j,j)}_{jj}(s_0 v)|^{2q} \leq (s_0 C_0)^{2q}
\]

(3.8)

(3.8) (the same inequalities hold for \( G^{(j,j)} \) replaced by \( G^{(j,j)}_{jj} \)). The first inequality follows from the fact that \( G(v) \leq s_0 G(s_0 v) \) (see e.g [4][Lemma 3.4] or [13][Lemma C.1]) and the second
inequality follows from the assumption (3.7). Moreover, for any $1 \leq j < k \leq n$ we have

$$
\mathbb{E}|G_{jk}^{(J,j)}(v)|^{2q} \leq 2^{4q} \mathbb{E}|G_{jk}^{(J,j)}(s_0v)|^{2q} + 2^{2q} s_0^n 2^{q} \mathbb{E}|G_{jj}^{(J,j)}(s_0v)|^{2q} \mathbb{E}|G_{kk}^{(J,j)}(s_0v)|^{2q} \leq (3s_0C_0)^{2q}.
$$

(3.9)

Applying Lemma 3.8 we get

$$
\mathbb{E}|G_{jj}^{(J,j)}|^{q} \leq (3c_0)^{q} \left(1 + \frac{1}{n} \sum_{j \in T_j} \mathbb{E}|\varepsilon_j^{(J)}|^{2q} + \frac{1}{n} \sum_{j \in T_j} \mathbb{E}|G_{jj}^{(J,j)}(v)|^{2q}\right)^{1/2} \leq (3c_0)^{q} \left(1 + (s_0C_0)^{q} \frac{1}{n} \sum_{j \in T_j} |\varepsilon_j^{(J)}|^{2q}\right) .
$$

By the Cauchy-Schwarz inequality

$$
\mathbb{E}|T_n^{(j)}|^{q} \leq \left(\frac{1}{n} \sum_{j \in T_j} \mathbb{E}|\varepsilon_j^{(J)}|^{2q}\right)^{1/2} \left(\frac{1}{n} \sum_{j \in T_j} \mathbb{E}|G_{jj}^{(J,j)}(v)|^{2q}\right)^{1/2} \leq (s_0C_0)^q \max_{j \in T_j} \mathbb{E}|\varepsilon_j^{(J)}|^{2q}.
$$

The last two inequalities imply that

$$
\mathbb{E}|G_{jj}^{(J)}|^{q} \leq (3c_0)^{q} \left(1 + (s_0C_0)^q \left(\max_{j \in T_j} \mathbb{E}|\varepsilon_j^{(J)}|^{2q} + \mathbb{E}|\varepsilon_j^{(J)}|^{2q}\right)\right).
$$

It remains to estimate $\mathbb{E}|\varepsilon_j^{(J)}|^{2q}$. By an obvious inequality we have

$$
\mathbb{E}|\varepsilon_j^{(J)}|^{2q} \leq 5^{2q} (\mathbb{E}|\varepsilon_{1q}^{(J)}|^{2q} + \ldots + \mathbb{E}|\varepsilon_{Nq}^{(J)}|^{2q}).
$$

Let $A_j := \{k \in T : L_{jk} = 0\}$. For a $r$-admissible configuration $|A_j| \leq r$. It is easy to check that

$$
\mathbb{E}|h_{jk}|^{q} \leq \begin{cases} 
\mu_4 R^{q/4} n^{q/4 - 1}, & \text{if } k \notin A_j \\
\mu_4 R^{q/2R^{-q}}, & \text{if } k \in A_j.
\end{cases}
$$

(3.10)

Moreover, for $k \notin A_j$

$$
|\mathbb{E} h_{jk}| \leq \frac{\mu_4}{R^{q/4}}.
$$

(3.11)

The bound for $\mathbb{E}|\varepsilon_{1q}^{(J)}|^{2q}$ is the direct corollary of (3.10)

$$
\mathbb{E}|\varepsilon_{1q}^{(J)}|^{2q} \leq \mu_4 R^{q/4} n^{q/2R^{-q} - 1} + R^{2q}.
$$

(3.12)
Let us consider $\varepsilon_{2j}^{(j)}$. We may rewrite it as a sum $\varepsilon_{2j}^{(j)} = \zeta_{1j} + \zeta_{2j} + \ldots + \zeta_{6j}$, where

\[
\zeta_{1j} := -\frac{1}{n} \sum_{l \neq k \in T_{j} \cap \mathcal{A}_j} \eta_{jl} \xi_{jl} G_{kl}^{(j)},
\]

\[
\zeta_{2j} := -\frac{1}{n} \sum_{l \neq k \in T_{j} \cap \mathcal{A}_j} (\xi_{jl} - \mathbb{E} \xi_{jl})(\xi_{jl} - \mathbb{E} \xi_{jl}) G_{kl}^{(j)},
\]

\[
\zeta_{3j} := -\frac{2}{n} \sum_{l \neq k \in T_{j} \cap \mathcal{A}_j} (\xi_{jl} - \mathbb{E} \xi_{jl})(\mathbb{E} \xi_{jl}) G_{kl}^{(j)},
\]

\[
\zeta_{4j} := -\frac{2}{n} \sum_{l \neq k \in T_{j} \cap \mathcal{A}_j} (\mathbb{E} \xi_{jl})(\mathbb{E} \xi_{jl}) G_{kl}^{(j)},
\]

\[
\zeta_{5j} := -\frac{2}{n} \sum_{l \in T_{j} \cap \mathcal{A}_j} \eta_{jl} \sum_{k \in T_{j} \cap \mathcal{A}_j} (\xi_{lk} - \mathbb{E} \xi_{lk}) G_{kl}^{(j)},
\]

\[
\zeta_{6j} := -\frac{2}{n} \sum_{l \in T_{j} \cap \mathcal{A}_j} \eta_{jl} \sum_{k \in T_{j} \cap \mathcal{A}_j} (\mathbb{E} \xi_{lk}) G_{kl}^{(j)}. \tag{3.13}
\]

Applying a crude bound and (3.9) we get

\[
\mathbb{E} |\zeta_{1j}|^{2q} \leq \frac{C^{2q} q^q (s_0 C_0)^{2q}}{(nv)^q},
\]

Using Burkholder’s type inequality and (3.10) we obtain the following estimate

\[
\mathbb{E} |\zeta_{2j}|^{2q} \leq \frac{C^{2q} q^q \mathbb{E} \text{Im}^q m_n^{(j)}}{(nv)^q} + \frac{C^{2q} q^{4q} R^{2q-4}}{(n^{\beta/2} v)^q} \sum_{l \in T_{j} \cap \mathcal{A}_j} \mathbb{E} \text{Im}^q G_{ll}^{(j)}
\]

\[
+ \frac{C^{2q} q^{4q} R^{2q-8}}{n^2 (nv)^q} \sum_{l \neq k \in T_{j} \cap \mathcal{A}_j} \mathbb{E} |G_{kl}^{(j)}|^{2q}.
\]

This inequality and (3.8)–(3.9) together imply

\[
\mathbb{E} |\zeta_{2j}|^{2q} \leq \frac{C^{2q} q^q (s_0 C_0)^{2q}}{(nv)^q} + \frac{C^{2q} q^{4q} R^{2q-4}(s_0 C_0)^{2q}}{(n^{\beta/2} v)^q} + \frac{C^{2q} q^{4q} R^{2q-8}(s_0 C_0)^{2q}}{n^q}. \tag{3.14}
\]

For the term $\zeta_{3j}$ we use the Rosenthal inequality, see e.g. [21],

\[
\mathbb{E} |\zeta_{3j}|^{2q} \leq \frac{C^{2q} q^q (s_0 C_0)^{2q}}{(nv)^q}, \tag{3.15}
\]

Again the crude bound imply

\[
\mathbb{E} |\zeta_{4j}|^{2q} \leq \frac{C^{2q} q^q (s_0 C_0)^{2q}}{(nv)^q}. \tag{3.16}
\]

Similarly,

\[
\mathbb{E} |\zeta_{5j}|^{2q} \leq \frac{C^{2q} R^{2q-1}}{(nv)^{q/2}} \sum_{l \in T_{j} \cap \mathcal{A}_j} \mathbb{E} |\eta_{jl}|^{2q} \sum_{k \in T_{j} \cap \mathcal{A}_j} (\xi_{lk} - \mathbb{E} \xi_{lk}) G_{kl}^{(j)}
\]

\[
\leq \frac{C^{2q} q^{4q} (s_0 C_0)^{2q}}{(nv)^q} + \frac{C^{2q} R^{2q-4} q^{2q}(s_0 C_0)^{2q}}{n^{q/2} R^{q}}. \tag{3.17}
\]
Finally,
\[
E |ζ_{kj}|^{2q} \leq \frac{C q r^{2q} (s_0 C_0)^{2q}}{(n^{3/2} r)^q R^{2q} R^q}.
\] (3.18)

For the term \( ε_{3j}^{(j)} \) we may proceed similarly. We get that \( ε_{3j}^{(j)} = ̂{ζ_{1j}} + ̂{ζ_{2j}}, \) where
\[
\hat{ζ}_{1j} := -\frac{1}{n} \sum_{k \in T_{1,j} \cap \Lambda_j} (η_{jk} - σ_{jk}^2) G_{kk}^{(j,j)},
\]
\[
\hat{ζ}_{2j} := -\frac{1}{n} \sum_{k \in T_{1,j} \setminus \Lambda_j} (ξ_{jk} - σ_{jk}^2) G_{kk}^{(j,j)}.
\]

The crude bound implies that
\[
E |\hat{ζ}_{1j}|^{2q} \leq \frac{C q r^{2q} (s_0 C_0)^{2q}}{R^q}.
\] (3.19)

Applying the Rosenthal inequality we get
\[
E |\hat{ζ}_{2j}|^{2q} \leq \frac{C q r^{2q} (s_0 C_0)^{2q}}{n^{2q}} \left( \sum_{k \in T_{1,j} \setminus \Lambda_j} |G_{kk}^{(j,j)}|^2 \right) + \frac{C q r^{4q-4} (s_0 C_0)^{4q}}{n^q} \sum_{k \in T_{1,j} \setminus \Lambda_j} |G_{kk}^{(j,j)}|^{4q}.
\] (3.20)

It is straightforward to check that
\[
E |ε_{3j}^{(j)}|^{2q} \leq \frac{1}{(nv)^{2q}}.
\] (3.21)

Finally, for \( ε_{nj}^{(j)} \) we may write
\[
E |ε_{nj}^{(j)}|^{2q} \leq \frac{C q r^{2q} (s_0 C_0)^{2q}}{R^q} + \frac{C q (s_0 C_0)^{2q}}{n^q R^{2q}}.
\] (3.22)

The off-diagonal entries \( G_{jk}^{(j)} \) may be expressed as follows
\[
G_{jk}^{(j)} = -\sqrt{\frac{n}{r}} G_{jj}^{(j,j)} \sum_{l \in T_{1,j}} h_{jl} G_{lk}^{(j,j)}.
\]

Applying
\[
E |G_{jk}^{(j,j)}|^q \leq \frac{(s_0 C_0)^q}{n^{q/2}} \left( \sum_{l \in T_{1,j}} h_{jl} G_{lk}^{(j,j)} \right)^{2q}.
\]

We proceed similarly to the estimation of \( E |ε_{nj}^{(j)}|^{2q} \). For simplicity let us denote
\[
\tilde{ζ}_{1j} := \frac{1}{\sqrt{n}} \sum_{l \in T_{1,j} \setminus \Lambda_j} η_{jl} G_{lk}^{(j,j)},
\]
\[
\tilde{ζ}_{2j} := \frac{1}{\sqrt{n}} \sum_{l \in T_{1,j} \setminus \Lambda_j} (ξ_{jl} - E ξ_{jl}) G_{lk}^{(j,j)},
\]
\[
\tilde{ζ}_{3j} := \frac{1}{\sqrt{n}} \sum_{l \in T_{1,j} \setminus \Lambda_j} (E ξ_{jl}) G_{lk}^{(j,j)}.
\]
The crude bound implies that
\[ \mathbb{E} |\xi_{1j}|^{2q} \leq \frac{r^{2q}(s_0C_0)^{2q}}{R^{2q}}. \quad (3.23) \]

By Rosenthal’s inequality
\[ \mathbb{E} |\xi_{2j}|^{2q} \leq \frac{C_q q^{q/2}s_0^qC_0^q}{(nv)^q} + \frac{C_q q^q R^{2q-4}(s_0C_0)^{2q}}{n^{q/2}}. \quad (3.24) \]

Finally,
\[ \mathbb{E} |\xi_{3j}|^{2q} \leq \frac{C_q(s_0C_0)^{2q}}{R^{2q}}(n^{3/2}v)^q. \quad (3.25) \]

Analysing (3.12)–(3.25) it is easy to see that one may choose sufficiently large constant \( A_0 \) and small constant \( A_1 \) such that
\[ \mathbb{E} |G_{jk}|^q \leq C_0^q. \]

\[ \square \]

**Proof of Lemma 3.7.** Let us choose some sufficiently large constant \( C_0 > \max(1/V, 6c_0) \), where \( c_0 \) is defined in Lemma 3.8. We also choose \( A_0, A_1 \) as in Lemma 3.9 \( s_0 := 2^4 \). Let \( L := \lfloor \log_{s_0} V/\tilde{v}_0 \rfloor + 1 \). Since \( \|G^{(j)}(V)\| \leq V^{-1} \) we may write
\[ \max_{j : |j| \leq L} \max_{l,k \in T_j} \mathbb{E} |G^{(j)}_{lk}(V)|^p \leq C_0^p \]
for all \( u, p \) such that \( |u| < 2 \) and \( 1 \leq p \leq A_1(nv)^{1/4}/R \). Fix arbitrary \( v : V/s_0 \leq v \leq V \) and \( p : 1 \leq p \leq (nv/s_0)^{1/4}/R \). Lemma 3.9 yields that
\[ \max_{j : |j| \leq L-1} \max_{l,k \in T_j} \mathbb{E} |G^{(j)}_{lk}(v)|^p \leq C_0^p \]
for \( 1 \leq p \leq A_1(nV/s_0)^{1/4}/R \), \( v \geq V/s_0 \). We may repeat this procedure \( L \) times and finally obtain
\[ \max_{l,k \in T} \mathbb{E} |G_{lk}(v)|^p \leq C_0^p \]
for \( 1 \leq p \leq A_1(nV/s_0)^{1/4}/R \leq A_1(n\tilde{v}_0)^{1/4}/R \) and \( v \geq V/s_0^L = \tilde{v}_0 \). \[ \square \]

The previous lemma allows to obtain \( p = A_1 \log^a n \) by taking \( v = \tilde{v}_0 = A_0n^{-1}\log^{4(\alpha+1)} n \). Without loss of generality we may consider \( p = A_1 \log^a n \) only (otherwise one may apply Lyapunov’s inequality for moments). It follows from Lemma 3.7 that for any \( r \)-admissible \( L \):
\[ \max_{j,k \in T} \mathbb{E} |G_{jk}(v)|^p \leq C_0^p \]
for all \( V \geq v \geq \tilde{v}_0 \). We may descent from \( \tilde{v}_0 \) to \( v_0 \) while keeping \( p = A_1 \log^a n \). Indeed, first we may take \( s := \log^3 a n \) and show that for \( v \geq v_0 \)
\[ \max_{j,k \in T} \mathbb{E} |G_{jk}(v)|^p \leq C_0^p \log^{(3\alpha-1)p} n. \]

It remains to remove the log factor from the r.h.s. of the previous equation. To this aim we shall adopt the moment matching technique which has been successfully used recently in [20] and [17].
We consider the pairs \((j, k) : L_{jk} = 1\), and denote by \(\xi_{jk}\) random variables such that: 
\[|\xi_{jk}| \leq D, \text{ for some } D \text{ chosen later, and}\]
\[
\mathbb{E} \xi_{jk}^s = \mathbb{E} \xi_{jk}^s \quad \text{for } s = 1, ..., 4.
\]
It follows from \([20]\)[Lemma 5.2]. that such a set of random variables exists. Let us denote 
\[H^y := \{h_{jk}\}_{j,k=1}^n\] such that 
\[
h_{jk}^y = \begin{cases} 
\xi_{jk}, & \text{if } L_{jk} = 1, \\
\eta_{jk}, & \text{otherwise.}
\end{cases}
\] (3.26)

Introduce \(G^y := (n^{-1/2}H^y - z)\)^{-1} and \(m_n^y(z) := \frac{1}{n} \text{Tr} G^y(z).
Then, in Lemma A.4 we show that for all \(v \geq v_0\) and \(5 \leq p \leq A_1 \log^n n\) there exist positive constants \(C_1, C_2\) such that 
\[
\mathbb{E} |G_{jk}(v)|^p \leq C_1 + C_2 \mathbb{E}|G_{jk}^y(v)|^p.
\]
It is easy to see that \(\xi_{jk}\) are sub-Gaussian random variables. Repeating the proof of Lemma 3.9, see Lemma A.5 in the appendix, we get 
\[
\mathbb{E} |G_{jk}^y(v)|^p \leq H_0^p
\]
for some \(H_0 > C_0\). We omit the details and proceed to the proof of Lemma 3.1.

**Proof of Lemma 3.1.** From (3.5) we conclude that 
\[
\mathbb{P}(L \notin \mathcal{L}_r) \leq n^{-c \log^2 n}
\]
for some large \(c > A_1\). It is easy to see that 
\[
\mathbb{E} |R_{jk}(v)|^p = \mathbb{E} \left[ \mathbb{I}[L \in \mathcal{L}_r] \mathbb{E}(|R_{jk}(v)|^p | L) \right] + \mathbb{E} \left[ \mathbb{I}[L \notin \mathcal{L}_r] \mathbb{E}(|R_{jk}(v)|^p | L) \right] 
\leq H_0^p + \mathbb{P}^+(L \notin \mathcal{L}_r) \mathbb{E}^+(|R_{jk}(v)|^2) \leq H_0^p + v^{-p} \mathbb{P}^+(L \notin \mathcal{L}_r)
\leq H_0^p + n^{A_1 \log^n n - c \log^2 n - \log^3 n \log n} \leq H_0^p,
\]
for some \(H_1 > H_0\). \(\square\)

4. Estimate of \(T_n\)

In this section we prove Theorem 2.1. We will follow the main idea of the proof of corresponding results in \([15]\). The main technical problem is to estimate the r.h.s of (4.5). Using definition of \(\xi_j\) we come to the problem of estimation \(\mathbb{E}_j |\xi_j| \mathbb{E}_{3j}^2 |R_{jj}|\). Since \(\xi_j\) and \(R_{jj}\) are dependent we need to use the Cauchy-Schwarz inequality. Unfortunately, we can only estimate \(\mathbb{E}_j |\xi_j| \mathbb{E}_{3j}^2\) without truncation. To estimate higher moments we need to use truncation arguments, i.e. use \(|X_{jk}| \leq n^{1/2} |\mathcal{R}|\). This will lead to the non-optimal bounds. It is worth to mention that in the case when \(\beta_{1+\delta} < \infty\) we can estimate \(\mathbb{E} |\xi_j|^{2+\delta/2}\) without truncation. To overcome the problem mentioned above we split the r.h.s. of (4.5) into two terms corresponding to \(|R_{jj}(z)| \leq H_1\) or \(|R_{jj}(z)| > H_1\) for some large \(H_1\). To obtain bounds of order \(n^{-c \log n}\) for \(\mathbb{P}(|R_{jj}(z)| > H_1)\) we need to take \(p\) in Lemma 3.1 of the order \(c \log^2 n\) (\(\alpha = 2\)).

To simplify the proof of Theorem 2.1 we will formulate below a simple lemma, which provides a general framework to estimate the moments of some statistics of independent random variables.

Let us consider the following statistic 
\[
T_n^* := \sum_{j=1}^n \xi_j f_j + \mathcal{R},
\]
where $\xi_j, f_j, j = 1, \ldots, n$ and $R$ are $\mathcal{M}$-measurable r.v. for some $\sigma$-algebra $\mathcal{M}$. Assume that there exist $\sigma$-algebras $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}, \mathcal{M}^{(j)} \subset \mathcal{M}, j = 1, \ldots, n$ such that

$$E_j(\xi_j \mid \mathcal{M}^{(j)}) = 0. \quad (4.1)$$

For simplicity we denote $E_j(\cdot) := E(\cdot \mid \mathcal{M}^{(j)})$. Let $\hat{f}_j$ be arbitrary $\mathcal{M}^{(j)}$-measurable r.v. and denote

$$\tilde{T}_n^{(j)} := E_j(T_n).$$

**Lemma 4.1.** For all $p \geq 2$ there exist some absolute constant $C$ such that

$$E \left| T_n^* \right|^p \leq C^p \left( A + p^p B + p^p C + p^p D + E |R|^p \right),$$

where

$$A := E \left( \frac{1}{n} \sum_{j=1}^{n} E_j |\xi_j(f_j - \hat{f}_j)| \right)^p,$$

$$B := E \left( \frac{1}{n} \sum_{j=1}^{n} E_j (|\xi_j(T_n^* - \tilde{T}_n^{(j)}))| \hat{f}_j \right)^\frac{p}{2},$$

$$C := \frac{1}{n} \sum_{j=1}^{n} E |\xi_j||T_n^* - \tilde{T}_n^{(j)}|^{p-1} |\hat{f}_j|,$$

$$D := \frac{1}{n} \sum_{j=1}^{n} E |\xi_j||f - \hat{f}_j||T_n^* - \tilde{T}_n^{(j)}|^{p-1}.$$

**Proof.** See [16][Lemma 6.1]. \hfill \Box

**Remark.** We conclude the statement of the last lemma by several remarks.

1. It follows from the definition of $A, B, C, D$ that instead of estimation of high moments of $\xi_j$ one needs to estimate conditional expectation $E_j |\xi_j|^\alpha$ for some small $\alpha$. Typically, $\alpha \leq 4$.

2. Moreover, to get the desired bounds one needs to choose an appropriate approximation $\hat{f}_j$ of $f_j$ and estimate $T_n^* - \tilde{T}_n^{(j)}$.

**Proof of Theorem 2.1.** Recalling the definition of $T_n$ (see (2.5)) we may rewrite it in the following way

$$T_n = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{4j} R_{jj} + \frac{1}{n} \sum_{\nu=1}^{3} \sum_{j=1}^{n} \varepsilon_{\nu j} R_{jj}. $$

One may see that $T_n$ is a special case of $T_n^*$, where

$$\xi_j := \varepsilon_{1j} + \varepsilon_{2j} + \varepsilon_{3j}, \ f_j := R_{jj}, \ R := \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{4j} R_{jj}. $$

We estimate each term in Lemma 4.1. Here, $\mathcal{M} := \sigma\{X_{k}, k, l \in T\}$ and $\mathcal{M}^{(j)} := \sigma\{X_{k}, k, l \in T_j\}, j = 1, \ldots, n.$
4.1. **Bound for $\mathbb{E}|R|^p$.** Applying the Schur complement formula we get

$$\text{Tr } R - \text{Tr } R^{(j)} = (1 + \frac{1}{n} \sum_{k,l \in T_j} X_{jl}X_{jk}|(R^{(j)})^2|_{kl})R_{jj} = R_{jj}^{-1} dR_{jj}/dz.$$  \hspace{1cm} (4.2)

Since $dR_{jj}/dz = R_{jj}^2$ we rewrite

$$\sum_{j=1}^n \varepsilon_{jj}R_{jj} = \frac{1}{n} \text{Tr } R^2.$$  

Hence, using Lemma A.6 we obtain

$$|R|^p \leq \frac{1}{(nv)^p} \mathbb{E} \text{Im}^p m_n(z) \leq \frac{C^p \text{Im}^p m_{sc}(z)}{(nv)^p} + \frac{C^p \mathbb{E} \text{Im}^p \Lambda_n}{(nv)^p}.$$  \hspace{1cm} (4.3)

We may apply the bound $\text{Im} \Lambda_n \leq |T_n|^{1/2}$ (see (2.7)), Young’s inequality and get

$$|R|^p \leq \frac{1}{(nv)^p} \mathbb{E} \text{Im}^p m_n(z) \leq \frac{C^p \text{Im}^p m_{sc}(z)}{(nv)^p} + \frac{C^p}{(nv)^{2p}} + \rho \mathbb{E}|T_n|^p,$$  \hspace{1cm} (4.4)

where $0 < \rho < 1$.

4.2. **Bound for $A$.** The term $A$, may be bounded from above by the following quantity

$$\max_{j=1,\ldots,n} \mathbb{E} E_j^p |\xi_j|(|R - \hat{f}_j|).$$  

Let us fix an arbitrary $j = 1, \ldots, n$, and choose

$$\hat{f}_j := -\frac{1}{z + m_n^{(j)}(z)}.$$  

Then

$$R_{jj} - \hat{f}_j = -\xi_j \hat{f}_j R_{jj},$$  \hspace{1cm} (4.5)

This equation implies that

$$\mathbb{E} E_j^p |\xi_j|(|R - \hat{f}_j|) = \mathbb{E} |\hat{f}_j|^p \mathbb{E}_j^p |\xi_j|^2 |R_{jj}|.$$  \hspace{1cm} (4.6)

Let us take some positive constant $H_0 > C_0$ such that for $q = c \log^2 n$:

$$\mathbb{P}(|R_{jj}| \geq H_0) \leq \frac{\mathbb{E}|R_{jj}|^q}{H_0^q} \leq \left( \frac{C_0}{H_0} \right)^q \leq \frac{1}{n^{c_1 \log n}}.$$  \hspace{1cm} (4.7)

It is straightforward to check that

$$\mathbb{E} |\hat{f}_j|^p \mathbb{E}_j^p |\xi_j|^2 |R_{jj}| 1[|R_{jj}| \leq H_0] \leq \frac{C}{(nv)^{2p}} + \frac{C^p}{(nv)^p} + \rho \mathbb{E}|T_n|^p.$$  \hspace{1cm} (4.8)

Moreover, from (4.6) and negligibility of high moments of $\xi_j$ we may conclude that

$$\mathbb{E} |\hat{f}_j|^p \mathbb{E}_j^p |\xi_j|^2 |R_{jj}| 1[|R_{jj}| > H_0] \leq \frac{C}{(nv)^{2p}} \mathbb{E}|T_n|^{1/2} |R_{jj}| \leq \frac{C}{(nv)^p} \mathbb{E}|T_n|^{1/2} |R_{jj}|.$$  \hspace{1cm} (4.9)

The last two inequalities (4.7) and (4.8) imply that

$$A \leq \frac{C^p \text{Im}^p m_{sc}(z)}{(nv)^p} + \frac{C^p}{(nv)^{2p}} + \rho \mathbb{E}|T_n|^p.$$  \hspace{1cm} (4.9)
4.3. Bound for $B$. We note that
\[ \mathcal{B}_p \leq \max_{j=1, \ldots, n} \mathbb{E} \left| \hat{f}_j \right|^{p/2} \mathbb{E}^{p/2} \left| \xi_j (T_n - \tilde{T}_n^{(j)}) \right|. \] 
(4.10)

We estimate the conditional expectation. Let
\[ I := \mathbb{E}_j (|\xi_j (T_n - \tilde{T}_n^{(j)})| \zeta), \]
(4.11)
where $\zeta$ is positive r.v. with sufficiently many bounded moments. For example, to estimate the r.h.s. of (4.10) one may take $\zeta := 1$. But for further analysis it will be necessary to consider more general $\zeta$.

**Representation of $T_n - \tilde{T}_n^{(j)}$.** By definition we may write the following representation
\[ T_n - \tilde{T}_n^{(j)} = (\Lambda_n - \tilde{\Lambda}_n^{(j)}) b_n + \tilde{\Lambda}_n^{(j)} (b_n - \tilde{b}_n^{(j)}) - \mathbb{E}_j (\Lambda_n (b_n - \tilde{b}_n^{(j)})), \]
where $\tilde{\Lambda}_n^{(j)} := \mathbb{E}_j (\Lambda_n)$ and $\tilde{b}_n^{(j)} := \mathbb{E}_j (b_n)$. Hence,
\[ |T_n - \tilde{T}_n^{(j)}| \leq K^{(j)} |\Lambda_n - \tilde{\Lambda}_n^{(j)}| + \frac{2}{\mathbb{E}} \tilde{T}_n^{(j)} |^{1/2}, \]
(4.12)
where
\[ K^{(j)} := K^{(j)} (z) := |b(z)| + 2 |\tilde{T}_n^{(j)} |^{1/2} + \frac{2}{\mathbb{E}}. \]
(4.13)

The equation (4.2) and Lemma A.6[Inequality (A.11)] yield that
\[ |\Lambda_n - \tilde{\Lambda}_n^{(j)}| \leq \frac{1}{\mathbb{E}} \text{Im} \frac{R_{jj}}{|R_{jj}|}. \]

For simplicity we denote the quadratic form in (4.2) by
\[ \eta_j := \frac{1}{n} \sum_{k,l \in T_j} X_{jk} X_{kl} \left( (R_{(j)})^{2} \right)_{kl} \]
and rewrite it as a sum of the three terms $\eta_j = \eta_{0j} + \eta_{1j} + \eta_{2j}$, where
\[ \eta_{0j} := \frac{1}{n} \sum_{k \in T_j} \left( (R_{(j)})^{2} \right)_{kk} = (m_{f}^{(j)} (z))', \]
\[ \eta_{1j} := \frac{1}{n} \sum_{k \neq l \in T_j} X_{jk} X_{kl} \left( (R_{(j)})^{2} \right)_{kl}, \]
\[ \eta_{2j} := \frac{1}{n} \sum_{k \in T_j} \left[ X_{jk}^2 - 1 \right] \left( (R_{(j)})^{2} \right)_{kk}. \]

It follows from (4.2) and $\Lambda_n - \tilde{\Lambda}_n^{(j)} = \Lambda_n - \Lambda_n^{(j)} - \mathbb{E}_j (\Lambda_n - \Lambda_n^{(j)})$ that
\[ \Lambda_n - \tilde{\Lambda}_n^{(j)} = \frac{1 + \eta_{0j}}{n} R_{jj} - \mathbb{E}_j (R_{jj}) + \frac{\eta_{1j} + \eta_{2j}}{n} R_{jj} - \frac{1}{n} \mathbb{E}_j (\eta_{1j} + \eta_{2j}) R_{jj}. \]

Using representation (4.4) we estimate
\[ |R_{jj} - \mathbb{E}_j (R_{jj})| \leq \left| \hat{f}_j \right| (|\xi_j R_{jj}| + \mathbb{E}_j (|\xi_j R_{jj}|)). \]

Applying this inequality and Lemma A.6[Inequality (A.11)] we may write
\[ |\Lambda_n - \tilde{\Lambda}_n^{(j)}| \leq \frac{1}{n} \left| \hat{f}_j \right| (|\xi_j R_{jj}| + \mathbb{E}_j (|\xi_j R_{jj}|)) + \frac{|\hat{\eta}_j R_{jj}|}{n} + \frac{\mathbb{E}_j (|\hat{\eta}_j R_{jj}|)}{n}. \]

Then
\[ \mathbb{E} (\xi_j |T_n - \tilde{T}_n^{(j)}| |\zeta| |\mathcal{M}^{(j)}|) \leq |K^{(j)} |B_{1j} + \ldots + B_{kj} | + |\tilde{T}_n^{(j)} |^{1/2} B_{kj}, \]
where

\[
B_{1j} := \frac{1 + v^{-1} \text{Im} m_n^{(j)}(z)}{n} |\tilde{f}_j| \mathbb{E}_j (|\xi_j|^2 |R_{jj} \zeta|),
\]

\[
B_{2j} := \frac{1 + v^{-1} \text{Im} m_n^{(j)}(z)}{n} |\tilde{f}_j| \mathbb{E}_j (|\xi_j R_{jj}|) \mathbb{E}_j (|\xi_j \zeta|),
\]

\[
B_{3j} := \frac{1}{n} \mathbb{E}_j (|\xi_j R_{jj} \zeta|),
\]

\[
B_{4j} := \frac{1}{n} \mathbb{E}_j (|\tilde{\eta}_j R_{jj}|) \mathbb{E}_j (|\xi_j \zeta|),
\]

\[
B_{5j} := \frac{1}{nv} \mathbb{E}_j (|\xi_j \zeta|).
\]

Hence, taking \( \zeta = 1 \), we may estimate

\[
\mathcal{B} \leq C^p \max_j \left[ \sum_{j=1}^4 \mathbb{E} |\tilde{f}_j|^{p/2} |K^{(j)}|^{p/2} B_{3j}^{p/2} + \mathbb{E} |\tilde{f}_j|^{p/2} |\tilde{T}_n^{(j)}|^{p/2} B_{5j}^{p/2} \right] =: \mathcal{I}_1 + \ldots + \mathcal{I}_5.
\]

It is straightforward to check that

\[
\mathcal{I}_5 \leq \frac{\text{Im}^{p/4} m_{sc}(z)}{(nv)^{3p/4}} \mathbb{E}^{2} |T_n|^p + \frac{C^p}{(nv)^{3p/4}} \mathbb{E}^{5p/8} |T_n|^p.
\]

We proceed to estimation of \( \mathcal{I}_3 \). The arguments for all other terms are similar and will be omitted. It follows from (4.13) that

\[
\mathbb{E} |\tilde{f}_j|^{p/2} |K^{(j)}|^{p/2} B_{3j}^{p/2} \leq C^p \mathbb{E}^{2} |K^{(j)}|^{p} \mathbb{E}^{4} B_{3j}^{2p} \leq C^p |b(z)|^{p/2} \mathbb{E}^{4} B_{3j}^{2p} + \frac{C^p}{(nv)^{p/2}} \mathbb{E}^{4} B_{3j}^{2p} + C^p \mathbb{E}^{1/4} |T_n|^{p} \mathbb{E}^{1/4} B_{3j}^{2p}.
\]

It remains to estimate \( \mathbb{E} B_{3j}^{2p} \). Applying the Cauchy-Schwarz inequality and arguments similar to (4.6)–(4.8) one may write

\[
\mathbb{E} |B_{3j}|^{2p} \leq \frac{1}{n^{2p}} \mathbb{E}^{1/2} \left[ \mathbb{E}_{x}^{2p} |\tilde{\eta}_j|^2 \right] \mathbb{E}^{1/2} \left[ \mathbb{E}_{x}^{2p} |\xi_j|^2 |R_{jj}|^2 \right] \leq \frac{C^p \text{Im}^{p/2} m_{sc}(z)}{(nv)^{4p}} + \frac{C^p}{(nv)^{3p/2}} \mathbb{E} |T_n|^p.
\]

Here we also use the moment bounds for quadratic and linear forms, see [13][Lemmas A.3–A.12]. Repeating the same arguments for \( \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \) we come to the following bound

\[
\mathcal{B} \leq \frac{C^p |b(z)|^{p}}{(nv)^{p}} + \frac{C^p |b(z)|^{p/2}}{(nv)^{p}} \mathbb{E}^{4} |T_n|^p + \frac{C^p \text{Im}^{p/2} m_{sc}(z)}{(nv)^{3p/2}} + \frac{C^p}{(nv)^{3p/2}} \mathbb{E}^{4} |T_n|^p + \frac{C^p \text{Im}^{p/2} m_{sc}(z)}{(nv)^{p}} \mathbb{E}^{4} |T_n|^p + \frac{C^p}{(nv)^{p}} \mathbb{E}^{4} |T_n|^p.
\]

where we also used the crude estimate \( \text{Im} m_{sc}(z) \leq |b(z)| \). Using Young’s inequality we immediately obtain

\[
p^{p/2} \mathcal{B} \leq C^p \frac{p^{p/2} |b(z)|^{p}}{(nv)^{p}} + C^p \frac{p^{p}}{(nv)^{2p}} + \rho \mathbb{E} |T_n|^p. \tag{4.14}
\]
4.4. **Bound for** $C$ **and** $D$. It is easy to see that one may estimate $C$ and $D$ simultaneously. Indeed, it is enough to estimate

$$ C' := \max_{j=1,\ldots,n} \mathbb{E} |\xi_j||T_n - \hat{T}_n^{(j)}|^p\sqrt{\zeta}, $$

where $\zeta := \max(|f_j|, |\hat{f}_j|)$. Let us fix $j = 1, \ldots, n$. Using (4.12) we get

$$ \mathbb{E} [\xi_j||T_n - \hat{T}_n^{(j)}|^p\sqrt{\zeta}] \leq C \left[ \frac{C_p b(z)}{(nv)^{p-2}} \right] \mathbb{E} \mathbb{E}_j [\xi_j||T_n - \hat{T}_n^{(j)}||\sqrt{\zeta}] + \left[ \frac{C_p}{(nv)^{p-2}} \right] \mathbb{E} [\hat{T}_n^{(j)}]^{\frac{-2}{p}} \mathbb{E}_j [\xi_j||T_n - \hat{T}_n^{(j)}||\sqrt{\zeta}]. $$

Applying Young’s inequality we obtain

$$ p^\rho \max(C, D) \leq \frac{C_p b(z)}{(nv)^p} + \frac{C_p p_{2p}}{(nv)^{2p}} + p^{p/2} \max_{j=1,\ldots,n} \mathbb{E} \mathbb{E}_j [\xi_j||T_n - \hat{T}_n^{(j)}||\sqrt{\zeta}] + \rho \mathbb{E} |T_n|^p. $$

Similarly to (4.6)

$$ \mathbb{P} (|\zeta| \geq H_0) \leq n^{-c \log n}. $$

Repeating now all calculations above we get

$$ \max(C, D) \leq \frac{C_p b(z)}{(nv)^p} + \frac{C_p p_{2p}}{(nv)^{2p}} + \rho \mathbb{E} |T_n|^p. \quad (4.15) $$

Collecting (4.3), (4.9), (4.14) and (4.15) we conclude the claim of the Theorem 2.1. \qed

5. **Acknowledgements**

We would like to thank the Associate Editor and the Reviewer for helpful comments and suggestions.

Results have been obtained under support of the RSF grant No. 18-11-00132 (HSE University). F. Götze has been supported by DFG through the Collaborative Research Centres 1283 ”Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications”.

**Appendix A. Auxiliary results**

A.1. **Truncation.** In this section we will show that the conditions (C0) allows us to assume that for all $1 \leq j, k \leq n$ we have $|X_{jk}| \leq \sqrt{n}/R$, where $R$ is some positive constant.

Let $\tilde{X}_{jk} := X_{jk} \mathbb{I}[|X_{jk}| \leq \sqrt{n}/R]$, $\hat{X}_{jk} := X_{jk} \mathbb{I}[|X_{jk}| \geq \sqrt{n}/R] - \mathbb{E} X_{jk} \mathbb{I}[|X_{jk}| \geq \sqrt{n}/R]$ and finally $\tilde{X}_{jk} := \hat{X}_{jk} \sigma^{-1}$, where $\sigma^2 := \mathbb{E} |\tilde{X}_{11}|^2$. We denote symmetric random matrices by $\tilde{X}$, $\hat{X}$ and $\tilde{X}$ formed from $\tilde{X}_{jk}$, $\hat{X}_{jk}$ and $\tilde{X}_{jk}$ respectively. Similar notations are used for the corresponding resolvent matrices, ESD and Stieltjes transforms.

**Lemma A.1.** Assuming the conditions (C0) we have for all $1 \leq p \leq A_1 \log n$

$$ \mathbb{E} |\tilde{m}_n(z) - \hat{m}_n(z)|^p \leq \left( \frac{C R^4}{n v} \right)^p. $$

**Proof.** From Bai’s rank inequality (see [3][Theorem A.43]) we conclude that

$$ \sup_{x \in \mathbb{R}} |\mu_n((-\infty, x]) - \hat{\mu}_n((-\infty, x])| \leq \frac{1}{n} \text{Rank}(X - \tilde{X}) \leq \frac{1}{n} \sum_{j,k=1}^n \mathbb{I}[|X_{jk}| \geq \sqrt{n}/R]. $$


Integrating by parts we get
\[
E |m_n(z) - \tilde{m}_n(z)|^p \leq \frac{1}{(nv)^p} E \left( \sum_{j,k=1}^n 1_{|X_{jk}| \geq \sqrt{n/R}} \right)^p.
\]
It is easy to see that
\[
\left( \sum_{j,k=1}^n E 1_{|X_{jk}| \geq \sqrt{n/R}} \right)^p \leq C^p R^p.
\]
Applying Rosenthal’s inequality, [21], we get that
\[
E \left( \sum_{j,k=1}^n [1_{|X_{jk}| \geq \sqrt{n/R}} - E 1_{|X_{jk}| \geq \sqrt{n/R}}] \right)^p
\leq C^p \left( \frac{pR^4}{n^2} \sum_{j,k=1}^n E |X_{jk}|^4 \right) \leq (C\sqrt{pR^2})^p.
\]
From these inequalities we may conclude the statement of Lemma.

**Lemma A.2.** Assuming the conditions (C0) we have for all \(1 \leq p \leq A_1 \log n\)
\[
E |\tilde{m}_n(z) - m_n(z)|^p \leq \frac{(CR^2)^p A^p (2p)}{(nv)^p}.
\]

**Proof.** It is easy to see that
\[
\tilde{R}(z) = (\tilde{W} - zI)^{-1} = \sigma^{-1}(\tilde{W} - z\sigma^{-1}I)^{-1} = \sigma^{-1}R(z).
\]
Applying the resolvent equality we get
\[
\tilde{R}(z) - R(z) = (z - \sigma^{-1}z)R(z) - (z - \sigma^{-1}z)\sigma^{-1}R(z).
\]
From (A.1) and (A.2) we may conclude
\[
|\tilde{m}_n(z) - m_n(z)| = \frac{1}{n} |\text{Tr} \tilde{R}(z) - \text{Tr} R(z)| = \frac{1}{n} |\sigma^{-1} \text{Tr} R(z) - \text{Tr} \tilde{R}(z)|
= \frac{1}{n} |\sigma^{-1} \text{Tr} R(z) - \text{Tr} \tilde{R}(z) - (z - \sigma^{-1}z) \text{Tr} R(z) - \text{Tr} \tilde{R}(z) - \text{Tr} R(z) - \text{Tr} \tilde{R}(z)|
\leq \frac{1}{n} |\sigma^{-1} - 1| |\text{Tr} \tilde{R}(z)| + (\sigma^{-1} - 1) \frac{|z|}{n} |\text{Tr} \tilde{R}(z)| - \text{Tr} R(z) - \text{Tr} \tilde{R}(z)|.
\]
Taking the \(p\)-th power and mathematical expectation we get
\[
E |\tilde{m}_n(z) - m_n(z)|^p \leq \frac{1}{n^p} (\sigma^{-1} - 1)^p E |\text{Tr} \tilde{R}(z)|^p + (\sigma^{-1} - 1)^p \frac{C^p}{n^p} E |\text{Tr} \tilde{R}(z)|^p.
\]
Since \(\tilde{X}\) satisfies conditions (C1) we may apply Lemma 3.1 and conclude
\[
\frac{1}{n^p} E |\text{Tr} \tilde{R}(z)|^p \leq C_0^p.
\]
We also have
\[
\sigma^{-1} - 1 \leq |\sigma^{-1} (1 - \sigma)| \leq \sigma^{-1} (1 - \sigma^2) \leq \sigma^{-1} E |X_{jk}|^2 1_{|X_{jk}| \geq \sqrt{n/R}} \leq C R^2 / n.
\]

(A.3)
To finish the proof it remains to estimate the term
\[ \frac{1}{n^p} E |\text{Tr} \tilde{R}(z) \tilde{R}(\sigma^{-1} z)|^p. \]
Applying the obvious inequality \(|\text{Tr} AB| \leq \|A\|_2 \|B\|_2\) we get
\[ \frac{1}{n^p} E |\text{Tr} \tilde{R}(z) \tilde{R}(\sigma^{-1} z)|^p \leq \frac{1}{n^p} E \frac{1}{2} \|\tilde{R}(z)\|_2^2 E \frac{1}{2} \|\tilde{R}(\sigma^{-1} z)\|_2^2 \]
\[ \leq E \frac{1}{2n^p} \text{Im}^p \tilde{m}_n(z) E \frac{1}{2} \text{Im}^p \hat{m}_n(\sigma^{-1} z). \]
From this inequality and (A.3) we conclude the statement of the lemma. \(\square\)

**Lemma A.3.** Assuming the conditions (C0) we have for all \(1 \leq p \leq A_1 \log n:\)
\[ E |\tilde{m}_n(z) - \hat{m}_n(z)|^p \leq \frac{(C R^3)^p}{(n v)^{3p/2}}. \]

**Proof.** It is easy to see that
\[ \tilde{m}_n(z) - \hat{m}_n(z) = \frac{1}{n} \text{Tr}(\tilde{W} - \hat{W}) \tilde{R}. \]
Applying the obvious inequalities \(|\text{Tr} AB| \leq \|A\|_2 \|B\|_2\) and \(||AB||_2 \leq ||A||_2 ||B||_2\) we get
\[ |\tilde{m}_n(z) - \hat{m}_n(z)| \leq ||\tilde{W} - \hat{W}||_2 ||\tilde{R}||_2 ||\hat{R}|| = ||E \tilde{W}||_2 ||\hat{R}||. \]
From
\[ |E \tilde{X}_{jk}| = |E X_{jk} \mathbf{1}[|X_{jk}| \geq \sqrt{n/R}]| \leq \frac{C R^3}{n^{3/2}} \]
we obtain
\[ ||E \tilde{W}||_2 \leq \frac{C R^3}{n}. \]
By Lemma A.2 we know \(E |\tilde{m}_n(z)|^p \leq C^p.\) This implies that
\[ \frac{1}{n^p} E \|\tilde{R}\|_2^p \leq \frac{C^p}{n^p}. \]
Finally
\[ E |\tilde{m}_n(z) - \hat{m}_n(z)|^p \leq \frac{C^p R^{3p}}{(n v)^{3p/2}}. \]
\(\square\)

**A.2. Replacement.** We say that the conditions (CG) are satisfied if \(X_{jk}\) satisfies the conditions (C0) and have a sub-Gaussian distribution. It is well-known that the random variables \(\xi\) are sub-gaussian if and only if \(E^{1/p} |\xi|^p \leq C \sqrt{p}\) for some constant \(C > 0.\)

**Lemma A.4.** For all \(v \geq v_0\) and \(5 \leq p \leq \log n\) there exist positive constants \(C_1, C_2\) such that
\[ E |G_{jk}(v)|^p \leq C_1^p + C_2 E |G_{jk}^y(v)|^p, \]
where \(G_{jk}^y\) is defined in (3.26).
Proof. The method is based on the following replacement scheme, which has been used in recent results \cite{5, 20} and \cite{17}. We replace all \( h_{ab} \) by \( \overline{h}_{ab} \) for \((a, b)\) such that \( L_{ab} = 1\), thus replacing the corresponding resolvent entries \( G_{jk} \) by \( \overline{G}_{jk}^\nu \) for every pair of \((j, k)\). Let \( J, K \subset T \). Denote by \( H^{(J, K)} \) the random matrix \( H \) with all entries in the positions \((\mu, \nu), \mu \in J, \nu \in K\) replaced by \( \xi_{\mu\nu} \). Assume that we have already exchanged all entries in positions \((\mu, \nu), \mu \in J, \nu \in K\) and are going to replace an additional entry in the position \((a, b), a \in T \setminus J, b \in T \setminus K\) with \( L_{ab} = 1\). Without loss of generality we may assume that \( J = \emptyset, K = \emptyset \) (hence \( H^{(J, K)} = H \)) and then denote \( V := H^{(a, b)} \). The following additional notations will be needed.

\[
E^{(a, b)} = \begin{cases} e_a e_b^T + e_b e_a^T, & 1 \leq a < b \leq n, \\ e_a e_a^T & a = b. \end{cases}
\]

and \( U := H - E^{(a, b)} \), where \( e_j \) denotes a unit column-vector with all zeros except \( j\)-th position. In these notations we may write

\[
H = U + \xi_{ab} E^{(a, b)}, \quad V = U + \overline{\xi}_{ab} E^{(a, b)}.
\]

Recall that \( G := (n^{-1/2} H - zI)^{-1} \) and denote \( S := (V - zI)^{-1} \) and \( T := (U - zI)^{-1} \). Let us assume that we have already proved the following fact

\[
E |G_{jk}|^p = T(p) + \frac{\theta_1 C^p}{n^2} + \frac{\theta_1 E |G_{jk}|^p}{n^2},
\]

where \( T(p) \) is some quantity depending on \( p, n \) (see \((A.9)\) below for precise definition) and \( |\theta_1| \leq 1, C > 0 \) are some numbers. Similarly,

\[
E |S_{jk}|^p = T(p) + \frac{\theta_2 C^p}{n^2} + \frac{\theta_2 E |S_{jk}|^p}{n^2},
\]

where \( |\theta_2| \leq 1 \). It follows from \((A.4)\) and \((A.5)\) that

\[
\left( 1 - \frac{\theta_1}{n^2} \right) E |G_{jk}|^p \leq \left( 1 - \frac{\theta_2}{n^2} \right) E |S_{jk}|^p + \frac{2C^p}{n^2}.
\]

Let us denote \( \rho := (1 - \theta_2/n^2)^{-1} (1 - \theta_1/n^2)^{-1} \). We get

\[
E |G_{jk}|^p \leq \rho E |S_{jk}|^p + \frac{C^p}{n^2},
\]

with some positive constant \( C_1 \). Repeating \((A.6)\) recursively for \((a, b) : L_{ab} = 1\) we arrive at the following bound

\[
E |G_{jk}|^p \leq \rho^{n(n+1)/2} E |G_{jk}^V|^p + \frac{C^p}{n^2} \left( 1 + \rho_1 + \ldots + \rho_1^{M-1} \right),
\]

where \( M \leq n(n+1)/2 \). It is easy to see from the definition of \( \rho \) that for some \( \theta \), say \( |\theta| < 4 \), we have

\[
\rho \leq 1 + |\theta|/n^2.
\]

From this inequality and \((A.7)\) we deduce that

\[
E |G_{jk}|^p \leq C_2 E |G_{jk}^V|^p + C_3^p,
\]

with some positive constants \( C_2 \) and \( C_3 \). From the last inequality we may conclude the statement of the lemma. It remains to prove \((A.4)\) (resp. \((A.5)\)). Applying the resolvent equation we get for \( m \geq 0 \)

\[
G = T + \sum_{\mu=1}^m (\frac{-1}{n^2})^\mu s_{ab}^\mu (TE^{(a, b)})^\mu T + (\frac{-1}{n^2})^{m+1} \sum_{\mu=1}^{m+1} s_{ab}^{m+1} (TE^{(a, b)})^{m+1} G.
\]
The same identity holds for $S$

$$S = T + \sum_{\mu=1}^{m} \frac{(-1)^{\mu}}{n^2} \mathbb{E}_{ab}[(TE(a,b))^\mu] + \frac{(-1)^{m+1}}{n^2} e_{ab}(TE(a,b))^{m+1}S.$$  

We investigate (A.8). In order to handle arbitrary high moments of $G_{jk}$, we apply a Stein type technique similar to Theorem. Let us introduce the following function $\varphi(z) := |z|^{p-2}$ and write

$$E|G_{jk}|^p = E G_{jk} \varphi(G_{jk}).$$

Applying (A.8) we get

$$E|R_{jk}|^p = \sum_{\mu=0}^{4} \frac{(-1)^{\mu}}{n^2} E \xi_{ab}[(TE(a,b))^\mu]_{jk}\varphi(G_{jk}) + \sum_{\mu=5}^{m} \frac{(-1)^{\mu}}{n^2} E \xi_{ab}[(TE(a,b))^\mu]_{jk}\varphi(G_{jk}) + \frac{1}{n^2} E e_{ab}[(TE(a,b))^{m+1}]_{jk}\varphi(G_{jk}) =: A_0 + A_1 + A_2.$$

Repeating the arguments from [17] one may show that

$$\max(|A_1|, |A_2|) \leq C_p \rho \frac{E|G_{jk}|^p}{n^2}.$$  

For the term $A_0$ one may write down the following representation

$$A_0 = \mathcal{I}(p) + r_n(p),$$

with remainder term bounded in absolute value

$$|r_n(p)| \leq \frac{C_p}{n^2} + \frac{E|G_{jk}|^p}{n^2},$$

and

$$\mathcal{I}(p) := \sum_{\mu=0}^{4} \frac{(-1)^{\mu}}{n^2} E \xi_{ab} E[(TE(a,b))^\mu]_{jk}\varphi(T_{jk}) + \sum_{\mu=0}^{4} \sum_{l=1}^{4-\mu} \frac{(-1)^{\mu}}{l!} \mathcal{E}_{\mu l},$$

where

$$\mathcal{E}_{\mu l} := \sum_{\mu_1 + \ldots + \mu_m = l, \mu_1 + 2\mu_2 + \ldots + m\mu_m \leq 4} \frac{C_l^{\mu_1, \ldots, \mu_m}}{n^{\mu_1 + 2\mu_2 + \ldots + m\mu_m}} \mathbb{E} X_{ab}^{\mu_1 + 2\mu_2 + \ldots + m\mu_m} \mathbb{E} [TE(a,b)]_{jk}^l [TE(a,b)]_{jk}^l \ldots [TE(a,b)]_{jk}^l \varphi(T_{jk}).$$

One may see that the term $\mathcal{I}(p)$ doesn’t depend on $G$ but depends on $T$.  

\textbf{Lemma A.5.} Let $L$ be $r$-admissible and assume that the conditions (CG) hold. Let $C_0$ and $s_0$ be arbitrary numbers such that $C_0 \geq \max(1/V, 6c_0)$, $s_0 \geq 2$. There exist a sufficiently large constant $A_0$ and small constant $A_1$ depending on $C_0, s_0, V$ only such that the following statement holds. Fix some $\tilde{v} : \tilde{v}_0 s_0 \leq \tilde{v} \leq V$. Suppose that for some integer $L > 0$, all $u, v', q$ such that $\tilde{v} \leq v' \leq V$, $|u| \leq u_0$, $1 \leq q \leq A_1(nv')$

$$\max_{J: |I| \leq L} \max_{k \in \mathbb{T}_L} \mathbb{E}[G_{jk}^{(I)}(v')]^q \leq C_0^q.$$
Then for all $u, v, q$ such that $\bar{v}/s_0 \leq v \leq V, |u| \leq u_0$, $1 \leq q \leq A_1(nv)$
\[
\max_{|J| \leq L-1} \max_{l,k \in T_J} \mathbb{E}|G_{lk}^{(j)}(v)|^q \leq C_0^q.
\]

**Proof.** We first observe the fact that the factor $q$ appears only in the terms with $\bar{v}/s_0$. Let us consider only one term, for example, :
\[
\zeta_{2j} := -\frac{2}{n} \sum_{l \neq k \in T_{J,j} \setminus A_j} (\bar{\xi}_{jl} - \mathbb{E}\xi_{jl})(\bar{\xi}_{jk} - \mathbb{E}\xi_{jk})G_{kl}^{(j,j)}.
\]
Applying the Hanson-Wright inequality, see e.g. [22] we obtain that
\[
\mathbb{E}|\zeta_{2j}|^2q \leq \frac{Cq^q \mathbb{E}|\Im^q m_{n}^{(j,j)} (z)|^2}{(nv)^q} + \frac{Cq^q}{(nv)^{2q}} \leq \frac{Cq^q(C_0s_0)^q}{(nv)^q}.
\]
\[\square\]

**A.3. Inequalities for resolvent.**

**Lemma A.6.** For any $z = u + iv \in \mathbb{C}^+$ we have
\[
\frac{1}{n} \sum_{l,k \in T_J} |R_{kl}^{(j)}|^2 \leq \frac{1}{v} \Im m_{n}^{(j)} (z) \quad (A.10)
\]
For any $l \in T_J$
\[
\sum_{k \in T_J} |R_{kl}^{(j)}|^2 \leq \frac{1}{v} \Im R_{ll}^{(j)} \quad (A.11)
\]

**References**

[1] Aggarwal. A. Bulk universality for generalized Wigner matrices with few moments. *Probability Theory and Related Fields*, 173 (1-2): 375–432, 2019.

[2] 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.

[3] Z. Bai and J. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, New York, second edition, 2010.

[4] 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*, 163(1):1–59. 2015.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] L. Erdős and H.-T. Yau. A Dynamical Approach to Random Matrix Theory. Courant Lecture Notes. AMS, 2017.

[12] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices *Adv. Math.*, 229(3):1435–1515, 2012.

[13] F. Götze, A. Naumov, and A. Tikhomirov. Local semicircle law under moment conditions. Part I: The Stieltjes transfrom. arXiv:1510.07350, 2015.
[14] F. Götze, A. Naumov, and A. Tikhomirov. Local semicircle law under moment conditions. Part II: Localization and delocalization. arXiv:1511.00862, 2015.

[15] F. Götze, A. Naumov, and A. Tikhomirov. Local Semicircle Law under Moment Conditions: The Stieltjes Transform, Rigidity, and Delocalization. Theory Probab. Appl., 62 (1): 58–83, 2018.

[16] F. Götze, A. Naumov, and A. Tikhomirov. On local laws for non-Hermitian random matrices and their products. arXiv:1708.06950, 2017.

[17] F. Götze, A. Naumov, A. Tikhomirov, and A. Timushev. On the local semicircle law for Wigner ensembles. Bernoulli, 24(3): 2358–2400, 2018.

[18] F. Götze and A. Tikhomirov. Optimal bounds for convergence of expected spectral distributions to the semi-circular law. Probability Theory and Related Fields, 165(1-2):163–233, 2016.

[19] J. Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. Ann. Inst. H. Poincaré Probab. Statist., 41(2): 151–178, 2005.

[20] J. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. Duke Math. J., 163(1):117–173, 2014.

[21] H. Rosenthal. On the subspaces of $L^p$ ($p > 2$) spanned by sequences of independent random variables. Israel J. Math., 8:273–303, 1970.

[22] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. Electron. Commun. Probab., 18(82): 1–9, 2013.

[23] T. Tao. Topics in random matrix theory, volume 132 of Graduate Studies in Mathematics. American Mathematical Society, 2012.

[24] T. Tao and V. Vu. Random matrices: The universality phenomenon for Wigner ensembles. arXiv:1202.0068.

[25] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. Acta Math., 206(1):127–204, 2011.

[26] E. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. Ann. of Math. (2), 62:548–564, 1955.

FRIEDRICH GÖTZE, FACULTY OF MATHEMATICS, BIELEFELD UNIVERSITY, BIELEFELD, GERMANY

E-mail address: goetze@math.uni-bielefeld.de

ALEXEY A. NAUMOV, NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, MOSCOW, RUSSIA; AND IITP RAS, MOSCOW, RUSSIA

E-mail address: anaumov@hse.ru

ALEXANDER N. TIKHOMIROV, DEPARTMENT OF MATHEMATICS, KOMI SCIENCE CENTER OF URAL DIVISION OF RAS, SYKTYVKAR, RUSSIA; AND NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, MOSCOW, RUSSIA

E-mail address: tikhomirov@dm.komisc.ru