SEMICIRCLE LAW FOR A CLASS OF RANDOM MATRICES WITH DEPENDENT ENTRIES

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

Abstract. In this paper we study ensembles of random symmetric matrices $X_n = \{X_{ij}\}_{i,j=1}^n$ with a random field type dependence, such that $\mathbb{E} X_{ij} = 0$, $\mathbb{E} X_{ij}^2 = \sigma_{ij}^2$, where $\sigma_{ij}$ can be different numbers. Assuming that the average of the normalized sums of variances in each row converges to one and Lindeberg condition holds true we prove that the empirical spectral distribution of eigenvalues converges to Wigner’s semicircle law.

Contents

1. Introduction 2
2. Proof of Theorem 1.3 5
   2.1. Truncation of random variables 6
   2.2. Universality of the spectrum of eigenvalues 7
3. Proof of Theorem 1.4 10
References 17

Date: May 5, 2014.

Key words and phrases. Random matrices, semicircle law, Stieltjes transform.

All authors are supported by CRC 701 “Spectral Structures and Topological Methods in Mathematics”, Bielefeld. A. Tikhomirov are partially supported by RFBR, grant N 11-01-00310-a and grant N 11-01-12104-ofi-m-2011, and program of UD RAS No12-P-1-1013. A. Naumov has been supported by the German Research Foundation (DFG) through the International Research Training Group IRTG 1132.
1. Introduction

Let $X_{jk}, 1 \leq j \leq k < \infty$, be triangular array of random variables with $\mathbb{E} X_{jk} = 0$ and $\mathbb{E} X_{jk}^2 = \sigma_{jk}^2$, and let $X_{jk} = X_{kj}$ for $1 \leq j < k < \infty$. We consider the random matrix

$X_n = \{X_{jk}\}_{j,k=1}^n$. 

Denote by $\lambda_1 \leq \ldots \leq \lambda_n$ eigenvalues of matrix $n^{-1/2}X_n$ and define its spectral distribution function by

$F_{X_n}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\lambda_i \leq x),$

where $\mathbb{1}(B)$ denotes the indicator of an event $B$. We set $F_{X_n}(x) := \mathbb{E} F_{X_n}(x)$. Let $g(x)$ and $G(x)$ denote the density and the distribution function of the standard semicircle law

$g(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}(|x| \leq 2), \quad G(x) = \int_{-\infty}^x g(u)du.$

For matrices with independent identically distributed (i.i.d.) elements, which have moments of all orders, Wigner proved in [10] that $F_n$ converges to $G(x)$, later on called “Wigner’s semicircle law”. The result has been extended in various aspects, i.e. by Arnold in [2]. In the non-i.i.d. case Pastur, [9], showed that Lindeberg’s condition is sufficient for the convergence. In [7] Götte and Tikhomirov proved the semicircle law for matrices satisfying martingale-type conditions for the entries.

In the majority of previous papers it has been assumed that $\sigma_{ij}^2$ are equal for all $1 \leq i < j \leq n$. Recently Erdős, Yau and Yin and al. study ensembles of symmetric random matrices with independent elements which satisfy $n^{-1} \sum_{j=1}^n \sigma_{ij}^2 = 1$ for all $1 \leq i \leq n$. See for example the survey of results in [5].

In this paper we study a class of random matrices with dependent entries and show that limiting distribution for $F_{X_n}(x)$ is given by Wigner’s semicircle law. We do not assume that the variances are equal.

Introduce the $\sigma$-algebras

$\mathcal{F}_{(ij)} := \sigma\{X_{kl} : 1 \leq k \leq l \leq n, (k,l) \neq (i,j)\}, 1 \leq i \leq j \leq n.$

For any $\tau > 0$ we introduce Lindeberg’s ratio for random matrices as

$L_n(\tau) := \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}|X_{ij}|^2 \mathbb{1}(|X_{ij}| \geq \tau \sqrt{n}).$
We assume that the following conditions hold

\begin{align*}
(1.1) & \quad \mathbb{E}(X_{ij}|\mathcal{F}^{(i,j)}) = 0; \\
(1.2) & \quad \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbb{E} [ \mathbb{E}(X_{ij}^2|\mathcal{F}^{(i,j)}) - \sigma_{ij}^2] \to 0 \quad \text{as} \quad n \to \infty; \\
(1.3) & \quad \text{for any fixed } \tau > 0 \quad L_n(\tau) \to 0 \quad \text{as} \quad n \to \infty.
\end{align*}

Furthermore, we will use condition (1.3) not only for the matrix $X_n$, but for other matrices as well, replacing $X_{ij}$ in the definition of Lindeberg’s ratio by corresponding elements.

For all $1 \leq i \leq n$ let $B_{i}^2 := \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij}^2$. We need to impose additional conditions on the variances $\sigma_{ij}^2$ given by

\begin{align*}
(1.4) & \quad \frac{1}{n} \sum_{i=1}^{n} |B_{i}^2 - 1| \to 0 \quad \text{as} \quad n \to \infty; \\
(1.5) & \quad \max_{1 \leq i \leq n} B_{i} \leq C,
\end{align*}

where $C$ is some absolute constant.

**Remark.** It is easy to see that the conditions (1.4) and (1.5) follow from the following condition

\begin{align*}
(1.6) & \quad \max_{1 \leq i \leq n} |B_{i}^2 - 1| \to 0 \quad \text{as} \quad n \to \infty.
\end{align*}

The main result of the paper is the following theorem

**Theorem 1.1.** Let $X_n$ satisfy conditions (1.1)–(1.5). Then

$$
\sup_{x} |\mathbb{P}(X_n(x) - G(x))| \to 0 \quad \text{as} \quad n \to \infty.
$$

Let us fix $i, j$. It is easy to see that for all $(k,l) \neq (i,j)$

$$
\mathbb{E} X_{ij} X_{kl} = \mathbb{E} \mathbb{E}(X_{ij} X_{kl}|\mathcal{F}^{(i,j)}) = \mathbb{E} X_{kl} \mathbb{E}(X_{ij}|\mathcal{F}^{(i,j)}) = 0.
$$

Hence the elements of the matrix $X_n$ are uncorrelated. If we additionally assume that the elements of the matrix $X_n$ are independent random variables then conditions (1.1) and (1.2) are automatically satisfied. The following theorem follows immediately from Theorem 1.1 in the case when the matrix $X_n$ has independent entries.

**Theorem 1.2.** Assume that the elements $X_{ij}$ of the matrix $X_n$ are independent for all $1 \leq i \leq j \leq n$ and $\mathbb{E} X_{ij} = 0$, $\mathbb{E} X_{ij}^2 = \sigma_{ij}^2$. Assume that $X_n$ satisfies conditions (1.3)–(1.5). Then

$$
\sup_{x} |\mathbb{P}(X_n(x) - G(x))| \to 0 \quad \text{as} \quad n \to \infty.
$$

The following example illustrates that without condition (1.4) convergence to Wigner’s semicircle law doesn’t hold.
Figure 1. Spectrum of matrix $X_n$.

**Example.** Let $X_n$ denote a block matrix

$$X_n = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix},$$

where $A$ is $m \times m$ symmetric random matrix with Gaussian elements with zero mean and unit variance, $B$ is $m \times (n - m)$ random matrix with i.i.d. Gaussian elements with zero mean and unit variance. Furthermore, let $D$ be a $(n - m) \times (n - m)$ diagonal matrix with Gaussian random variables on the diagonal with zero mean and unit variance. If we set $m := n/2$ then it is not difficult to check that condition (1.4) doesn’t hold. We simulated the spectrum of the matrix $X_n$ and illustrated a limiting distribution on Figure 1.

**Remark.** We conjecture that Theorem 1.1 (Theorem 1.2 respectively) holds without assumption (1.5).

Define the Levy distance between the distribution functions $F_1$ and $F_2$ by

$$L(F_1, F_2) = \inf \{ \varepsilon > 0 : F_1(x - \varepsilon) - \varepsilon \leq F_2(x) \leq F_1(x + \varepsilon) + \varepsilon \}.$$ 

The following theorem formulates the Lindeberg’s universality scheme for random matrices.

**Theorem 1.3.** Let $X_n, Y_n$ denote independent symmetric random matrices with $E X_{ij} = E Y_{ij} = 0$ and $E X_{ij}^2 = E Y_{ij}^2 = \sigma_{ij}^2$. Suppose that the matrix $X_n$ satisfies conditions (1.1)–(1.4), and the matrix $Y_n$ has independent Gaussian elements. Additionally assume that for the matrix $Y_n$ conditions (1.3)–(1.4) hold. Then

$$L(F_n^X(x), F_n^Y(x)) \to 0 \text{ as } n \to \infty.$$ 

In view of Theorem 1.3 to prove Theorem 1.1 it remains to show convergence to semicircle law in the Gaussian case.
Theorem 1.4. Assume that the entries $Y_{ij}$ of the matrix $Y_n$ are independent for all $1 \leq i \leq j \leq n$ and have Gaussian distribution with $\mathbb{E}Y_{ij} = 0$, $\mathbb{E}Y_{ij}^2 = \sigma_{ij}^2$. Assume that conditions (1.3)–(1.5) are satisfied. Then
\[
\sup_x |F^{Y_n}(x) - G(x)| \to 0 \text{ as } n \to \infty.
\]

For related ensembles of random covariance matrices it is well known that spectral distribution function of eigenvalues converges to the Marchenko–Pastur law. In this case Götze and Tikhomirov in [6] received similar results to [7]. Recently Adamczak, [1], proved the Marchenko–Pastur law for matrices with martingale structure. He assumed that the matrix elements have moments of all orders and imposed conditions similar to (1.4). Another class of random matrices with dependent entries was considered in [8] by O’Rourke. In a forthcoming paper we prove analogs of these theorems for random covariance matrices.

The paper organized as follows. In Section 2 we give a proof of Theorem 1.3 using the method of Stieltjes transforms. In Section 3 we prove Theorem 1.4 by the classical moment method.

Throughout this paper we assume that all random variables are defined on common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\text{Tr}(A)$ denote the trace of a matrix $A$. For a vector $x = (x_1, \ldots, x_n)$ let $||x||_2 := \left(\sum_{i=1}^n x_i^2\right)^{1/2}$. We denote the operator norm of the matrix $A$ by $||A|| := \sup_{||x||_2=1} ||Ax||_2$. We will write $a \leq_m b$ if there is an absolute constant $C$ depending on $m$ only such that $a \leq Cb$.

From now on we shall omit the index $n$ in the notation for random matrices.

## 2. Proof of Theorem 1.3

We denote the Stieltjes transforms of $F^X$ and $F^Y$ by $S^X(z)$ and $S^Y(z)$ respectively. Due to the relations between distribution functions and Stieltjes transforms, the statement of Theorem 1.3 will follow from
\[
(2.1) \quad |S^X(z) - S^Y(z)| \to 0 \text{ as } n \to \infty.
\]

Set
\[
(2.2) \quad R^X(z) := \left( \frac{1}{\sqrt{n}} X - z1 \right)^{-1} \text{ and } R^Y(z) := \left( \frac{1}{\sqrt{n}} Y - z1 \right)^{-1}.
\]

By definition
\[
S^X(z) = \frac{1}{n} \text{Tr} \mathbb{E} R^X(z) \text{ and } S^Y(z) = \frac{1}{n} \text{Tr} \mathbb{E} R^Y(z).
\]

We divide the proof of (2.1) into the two subsections 2.1 and 2.2.
Note that we can substitute $\tau$ in (1.3) by a decreasing sequence $\tau_n$ tending to zero such that

$$L_n(\tau_n) \to 0 \text{ as } n \to \infty.$$ 

and $\lim_{n \to \infty} \tau_n \sqrt{n} = \infty$.

2.1. **Truncation of random variables.** In this section we truncate the elements of the matrices $X$ and $Y$. Let us omit the indices $X$ and $Y$ in the notations of the resolvent and the Stieltjes transforms.

Consider some symmetric $n \times n$ matrix $D$. Put $\tilde{X} = X + D$. Let

$$\tilde{R} = \left( \frac{1}{\sqrt{n}} \tilde{X} - zI \right)^{-1}.$$ 

**Lemma 2.1.**

$$|\text{Tr} R - \text{Tr} \tilde{R}| \leq \frac{1}{v^2} (\text{Tr} D^2)^{1/2}.$$ 

**Proof.** By the resolvent equation

$$R = \tilde{R} - \frac{1}{\sqrt{n}} R D \tilde{R}.$$ 

For resolvent matrices we have, for $z = u + iv, v > 0$,

$$\max\{|R|, |\tilde{R}|\} \leq \frac{1}{v}.$$ 

Using (2.4) and (2.5) it is easy to show that

$$|\text{Tr} R - \text{Tr} \tilde{R}| = \frac{1}{\sqrt{n}} |\text{Tr} R D \tilde{R}| \leq \frac{1}{v^2} (\text{Tr} D^2)^{1/2}.$$ 

We split the matrix entries as $X = \hat{X} + \check{X}$, where $\hat{X} := X 1(|X| < \tau_n \sqrt{n})$ and $\check{X} := X 1(|X| \geq \tau_n \sqrt{n})$. Define the matrix $\hat{X} = \{\hat{X}_{ij}\}_{i,j=1}^n$. Let

$$\tilde{R}(z) := \left( \frac{1}{\sqrt{n}} \hat{X} - zI \right)^{-1} \text{ and } \check{S}(z) = \frac{1}{n} E \text{Tr} \tilde{R}(z).$$

By Lemma 2.1

$$|S(z) - \check{S}(z)| \leq \frac{1}{v^2} \left( \frac{1}{n^2} \sum_{i,j=1}^n E X_{ij}^2 1(|X_{ij}| \geq \tau_n \sqrt{n}) \right)^{1/2} = v^{-2} \tilde{L}_n^{1/2} (\tau_n).$$

From (2.3) we conclude that

$$|S(z) - \check{S}(z)| \to 0 \text{ as } n \to \infty.$$
Introduce the centralized random variables $X_{ij} = \hat{X}_{ij} - \mathbb{E}(\hat{X}_{ij}|\hat{g}^{(i,j)})$ and the matrix $\mathbf{X} = \{X_{ij}\}_{i,j=1}^n$. Let 

$$\mathbf{R}(z) := \left(\frac{1}{\sqrt{n}}\mathbf{X} - z\mathbf{I}\right)^{-1}$$

and 

$$\hat{S}(z) = \frac{1}{n} \mathbb{E} \text{Tr} \mathbf{R}(z).$$

Again by Lemma 2.1,

$$|\hat{S}(z) - S(z)| \leq \frac{1}{n^2} \left(\frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E} X_{ij}^2 1(|X_{ij}| \geq \tau_n \sqrt{n})\right)^{1/2} = v^{-2} L_n^1(\tau_n).$$

In view of (2.3) the right hand side tends to zero as $n \to \infty$.

Now we show that (1.2) will hold if we replace $X$ by $\mathbf{X}$. For all $1 \leq i \leq j \leq n$

$$\mathbb{E}(\mathbf{X}_{ij}^2 | \hat{g}^{(i,j)}) - \sigma_{ij}^2 = \mathbb{E}(\mathbf{X}_{ij}^2 | \hat{g}^{(i,j)}) - \mathbb{E}X_{ij}^2 + \mathbb{E}(\hat{X}_{ij}^2 | \hat{g}^{(i,j)}) + \mathbb{E}(X_{ij}^2 | \hat{g}^{(i,j)}) - \sigma_{ij}^2.$$

By the triangle inequality and (1.2), (2.3)

$$\frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E} |\mathbb{E}(\mathbf{X}_{ij}^2 | \hat{g}^{(i,j)}) - \sigma_{ij}^2|$$

$$\leq \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E} |\mathbb{E}(X_{ij}^2 | \hat{g}^{(i,j)}) - \sigma_{ij}^2| + 2L_n(\tau_n) \to 0 \text{ as } n \to \infty.$$

It is also not very difficult to check that the condition (1.4) holds true for the matrix $\mathbf{X}$ replaced by $\mathbf{X}$.

Similarly, one may truncate the elements of the matrix $\mathbf{Y}$ and consider the matrix $\mathbf{Y}$ with the entries $Y_{ij} 1(|Y_{ij}| \leq \tau_n \sqrt{n})$. Then one may check that

$$\frac{1}{n^2} \sum_{i,j=1}^n |\mathbb{E}Y_{ij}^2 - \sigma_{ij}^2| \to 0 \text{ as } n \to \infty.$$

In what follows assume from now on that $|X_{ij}| \leq \tau_n \sqrt{n}$ and $|Y_{ij}| \leq \tau_n \sqrt{n}$. We shall write $\mathbf{X}, \mathbf{Y}$ instead of $X$ and $Y$ respectively.

2.2. Universality of the spectrum of eigenvalues. To prove (2.1) we will use a method introduced in [4]. Define the matrix $\mathbf{Z} := \mathbf{Z}(\varphi) := \mathbf{X} \cos \varphi + \mathbf{Y} \sin \varphi$. It is easy to see that $\mathbf{Z}(0) = \mathbf{X}$ and $\mathbf{Z}(\pi/2) = \mathbf{Y}$. Set $\mathbf{W} := \mathbf{W}(\varphi) := n^{-1/2} \mathbf{Z}$ and

$$\mathbf{R}(z, \varphi) := (\mathbf{W} - z\mathbf{I})^{-1}.$$

Introduce the Stieltjes transform

$$S(z, \varphi) := \frac{1}{n} \sum_{i=1}^n \mathbb{E}|\mathbf{R}(z, \varphi)|_{ii}.$$
Note that $S(z, 0)$ and $S(z, \pi/2)$ are the Stieltjes transforms $S^X(z)$ and $S^Y(z)$ respectively.

Obviously we have

$$S(z, \pi/2) - S(z, 0) = \int_0^{\pi/2} \frac{\partial S(z, \varphi)}{\partial \varphi} d\varphi. \tag{2.9}$$

To simplify the arguments we will omit arguments in the notations of matrices and Stieltjes transforms. We have

$$\frac{\partial W}{\partial \varphi} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial Z_{ij}}{\partial \varphi} \mathbf{e}_i \mathbf{e}_j^T,$$

where we denote by $\mathbf{e}_i$ the column vector with 1 in position $i$ and zeros in the other positions. We may rewrite the integrand in (2.9) in the following way

$$\frac{\partial S}{\partial \varphi} = -\frac{1}{\sqrt{n}} \mathbf{R} \frac{\partial W}{\partial \varphi} \mathbf{R} \tag{2.10}$$

$$= -\frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} \text{Tr} \mathbf{R} \frac{\partial Z_{ij}}{\partial \varphi} \mathbf{e}_i \mathbf{e}_j^T \mathbf{R}$$

$$= \frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} \frac{\partial Z_{ij}}{\partial \varphi} u_{ij},$$

where $u_{ij} = -[\mathbf{R}^2]_{ji}$.

For all $1 \leq i \leq j \leq n$ introduce the random variables

$$\xi_{ij} := Z_{ij}, \quad \dot{\xi}_{ij} := \frac{\partial Z_{ij}}{\partial \varphi} = -\sin \varphi X_{ij} + \cos \varphi Y_{ij},$$

and the sets of random variables

$$\xi^{(ij)} := \{\xi_{kl} : 1 \leq k \leq l \leq n, (k, l) \neq (i, j)\}.$$

Using Taylor’s formula one may write

$$u_{ij}(\xi_{ij}, \xi^{(ij)}) = u_{ij}(0, \xi^{(ij)}) + \xi_{ij} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) + \mathbf{E}_\theta (1 - \theta) \xi_{ij} \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta \xi_{ij}, \xi^{(ij)}),$$

where $\theta$ has a uniform distribution on $[0, 1]$ and is independent of $(\xi_{ij}, \xi^{(ij)})$.

Multiplying both sides of the last equation by $\dot{\xi}_{ij}$ and taking mathematical expectation on both sides we have

$$\mathbf{E} \dot{\xi}_{ij} u_{ij}(\xi_{ij}, \xi^{(ij)}) = \mathbf{E} \dot{\xi}_{ij} u_{ij}(0, \xi^{(ij)}) + \mathbf{E} \dot{\xi}_{ij} \xi_{ij} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)})$$

$$+ \mathbf{E} (1 - \theta) \dot{\xi}_{ij} \xi_{ij} \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta \xi_{ij}, \xi^{(ij)}). \tag{2.11}$$

By independence of $Y_{ij}$ and $\xi^{(ij)}$ we get

$$\mathbf{E} Y_{ij} u_{ij}(0, \xi^{(ij)}) = \mathbf{E} Y_{ij} \mathbf{E} u_{ij}(0, \xi^{(ij)}) = 0. \tag{2.12}$$
By the properties of conditional expectation and condition (1.1)

$$E_{X_{ij} u_{ij}}(0, \xi^{(ij)}) = E_{u_{ij}}(0, \xi^{(ij)}) E(X_{ij} | \tilde{F}^{(i,j)}) = 0.$$  

By (2.11), (2.12) and (2.13) we can rewrite (2.10) in the following way

$$\frac{\partial S}{\partial \varphi} = \frac{1}{n^2} \sum_{i,j=1}^{n} \hat{\xi}_{ij} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) + \frac{1}{n^2} \sum_{i,j=1}^{n} \theta(1-\theta) \hat{\xi}_{ij} \hat{\xi}_{ij} \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2} (\theta \xi_{ij}, \xi^{(ij)})$$

$$= A_1 + A_2.$$  

It is easy to see that

$$\hat{\xi}_{ij} \epsilon_{ij} = -\frac{1}{2} \sin 2\varphi X^2_{ij} + \cos^2 \varphi X_{ij} Y_{ij} - \sin^2 \varphi X_{ij} Y_{ij} + \frac{1}{2} \sin 2\varphi Y^2_{ij}.$$  

The random variables $Y_{ij}$ are independent of $X_{ij}$ and $\xi^{(ij)}$. Using this fact we conclude that

$$E_{X_{ij}} Y_{ij} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) = E Y_{ij} E_{X_{ij}} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) = 0,$$  

$$E_{Y_{ij}^2} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) = \sigma_{ij}^2 E \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}).$$  

By the properties of conditional mathematical expectation we get

$$E_{X_{ij}^2} \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) = E \frac{\partial u_{ij}}{\partial \xi_{ij}}(0, \xi^{(ij)}) E(X_{ij}^2 | \tilde{F}^{(i,j)}).$$  

A direct calculation shows that the derivative of $u_{ij} = -[R^2]_{ji}$ is equal to

$$\frac{\partial u_{ij}}{\partial \xi_{ij}} = \left[ \frac{R^2 \frac{\partial Z}{\partial \xi_{ij}} R}{R^2}_{ji} \right] + \left[ R \frac{\partial Z}{\partial \xi_{ij}} R^2 \right]_{ji} = \frac{1}{\sqrt{n}} [R^2 e_t e_j^T R]_{ji} + \frac{1}{\sqrt{n}} [R^2 e_t e_i^T R]_{ji} + \frac{1}{\sqrt{n}} [Re_t e_j^T R^2]_{ji} + \frac{1}{\sqrt{n}} [Re_j e_i^T R^2]_{ji} = \frac{1}{\sqrt{n}} [R^2]_{ji} [R]_{ji} + \frac{1}{\sqrt{n}} [R^2]_{jj} [R]_{ji} + \frac{1}{\sqrt{n}} [R]_{ji} [R^2]_{ji} + \frac{1}{\sqrt{n}} [R]_{jj} [R^2]_{ji}.$$  

Using the obvious bound for the spectral norm of the matrix resolvent $||R|| \leq v^{-1}$ we get

$$\frac{\partial u_{ij}}{\partial \xi_{ij}} \leq \frac{C}{\sqrt{n} v^3}.$$  

From (2.14), (2.17) and (2.6)–(2.7) we deduce

$$|A_1| \leq \frac{C}{n^2 v^3} \sum_{i,j=1}^{n} E |E(X_{ij} | \tilde{F}^{(i,j)}) - \sigma_{ij}^2| \to 0 \text{ as } n \to \infty.$$  

It remains to estimate $A_2$. We calculate the second derivative of $u_{ij}$

$$\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2} = -2 \left[ R^2 \frac{\partial W}{\partial \xi_{ij}} R \frac{\partial W}{\partial \xi_{ij}} R \right]_{ji} - 2 \left[ R \frac{\partial W}{\partial \xi_{ij}} R^2 \frac{\partial W}{\partial \xi_{ij}} R \right]_{ji} - 2 \left[ R \frac{\partial W}{\partial \xi_{ij}} R \frac{\partial W}{\partial \xi_{ij}} R^2 \right]_{ji} = T_1 + T_2 + T_3.$$
Let’s expand the term $T_1$

\begin{equation}
T_1 = -2 \left[ R^2 \frac{\partial W}{\partial \xi_{ij}} R \frac{\partial W}{\partial \xi_{ij}} \right]_{ji} = T_{11} + T_{12} + T_{13} + T_{14},
\end{equation}

where we denote

\begin{align*}
T_{11} &= -\frac{2}{n} [R^2]_{ji} [R]_{ji}, \\
T_{12} &= -\frac{2}{n} [R^2]_{ji} [R]_{jj} [R]_{ii}, \\
T_{13} &= -\frac{2}{n} [R^2]_{jj} [R]_{ii} [R]_{ji}, \\
T_{14} &= -\frac{2}{n} [R^2]_{jj} [R]_{ij} [R]_{ii}.
\end{align*}

Using again the bound $||R|| \leq v^{-1}$ we can show that

\[ \max(|T_{11}|, |T_{12}|, |T_{13}|, |T_{14}|) \leq C n v^4. \]

From the expansion (2.19) and the bounds of $T_{1i}, i = 1, 2, 3, 4$ we conclude that

\[ |T_1| \leq \frac{C}{n v^4}. \]

Repeating the above arguments one can show that

\[ \max(|T_2|, |T_3|) \leq \frac{C}{n v^4}. \]

Finally we have

\[ \left| \frac{\partial^2 u_{\theta \xi_{ij} (\xi^{(ij)})}}{\partial \xi_{ij}^2} \right| \leq \frac{C}{n v^4}. \]

Using the assumption $|\xi_{ij}| \leq \tau_n \sqrt{n}$ and the condition (1.4) we deduce the bound

\[ |A_2| \leq \frac{C \tau_n}{v^4}. \]

We may turn $\tau_n$ to zero and conclude the statement of Theorem 1.3 from (2.9), (2.10), (2.18) and (2.20).

3. Proof of Theorem 1.4

We prove the theorem using the moment method. It is easy to see that the moments of $F^Y(x)$ can be rewritten as normalized traces of powers of $Y$:

\[ \int_R x^k dF^Y(x) = \mathbb{E} \frac{1}{n} \operatorname{Tr} \left( \frac{1}{\sqrt{n}} Y \right)^k, \quad k \geq 1. \]

It is sufficient to prove that

\[ \mathbb{E} \frac{1}{n} \operatorname{Tr} \left( \frac{1}{\sqrt{n}} Y \right)^k = \int_R x^k dG(x) + o_k(1). \]

for $k \geq 1$, where $o_k(1)$ tends to zero as $n \to \infty$ for any fixed $k$.

It is well known that the moments of semicircle law are given by the Catalan numbers

\begin{align*}
\beta_k &= \int_R x^k dG(x) = \begin{cases} \\
\frac{1}{m+1} \binom{2m}{m}, & k = 2m \\
0, & k = 2m + 1.
\end{cases}
\end{align*}
Furthermore we shall use the notations and the definitions from \cite{3}. 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, \(F : E \rightarrow V \times V\). Let \(i = (i_1, \ldots, i_k)\) be a vector taking values in \(\{1, \ldots, n\}^k\). For a vector \(i\) we define a \(\Gamma\)-graph as follows. Draw a horizontal line and plot the numbers \(i_1, \ldots, i_k\) on it. Consider the distinct numbers as vertices, and draw \(k\) edges \(e_j\) from \(i_j\) to \(i_{j+1}\), \(j = 1, \ldots, k\), 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 each other by a permutation of \((1, \ldots, n)\). By this definition, all \(\Gamma\)-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 \(\{1, \ldots, t\}\);
2) Its edge set is \(\{e_1, \ldots, e_k\}\);
3) There is a function \(g\) from \(\{1, \ldots, k\}\) onto \(\{1, \ldots, t\}\) satisfying \(g(1) = 1\) and \(g(i) \leq \max\{g(1), \ldots, g(i - 1)\} + 1\) for \(1 < i \leq k\);
4) \(F(e_i) = (g(i), g(i + 1))\), for \(i = 1, \ldots, k\), with the convention \(g(k + 1) = g(1) = 1\).

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

We shall classify all canonical graphs into three categories. Category 1 consists of all canonical \(\Gamma(k, t)\)-graphs with the property that each edge is coincident with exactly one other edge of opposite direction and the graph of noncoincident edges forms a tree. It is easy to see if \(k\) is odd then there are no graphs in category 1. If \(k\) is even, i.e. \(k = 2m\), say, we denote a \(\Gamma(k, t)\)-graph by \(\Gamma_1(2m)\).

Category 2 consists of all canonical graphs that have at least one edge with odd multiplicity. We shall denote the graph from this category by \(\Gamma_2(k, t)\). Finally, category 3 consists of all other canonical graphs, which we denote by \(\Gamma_3(k, t)\).

It is known, see \cite{3} Lemma 2.4, that the number of \(\Gamma_1(2m)\)-graphs is equal to \(\frac{1}{m^2 m^{(2m)}}\).

We expand the traces of powers of \(Y\) in a sum

\[
\text{Tr} \left( \frac{1}{n} Y \right)^k = \frac{1}{n^{k/2}} \sum_{i_1, i_2, \ldots, i_k} Y_{i_1 i_2} Y_{i_2 i_3} \ldots Y_{i_k i_1},
\]

where the summation is taken over all sequences \(i = (i_1, \ldots, i_k) \in \{1, \ldots, n\}^k\).

For each vector \(i\) we construct a graph \(G(i)\) as above. We denote by \(Y(i) = Y(G(i))\).
Then we may split the moments of \( F_Y(x) \) into three terms
\[
\mathbb{E} \frac{1}{n} \text{Tr} \left( \frac{1}{\sqrt{n}} Y \right)^k = \frac{1}{n^{k/2+1}} \sum_i \mathbb{E} Y_{i_1} Y_{i_2} Y_{i_3} \ldots Y_{i_k} = S_1 + S_2 + S_3,
\]
where
\[
S_j = \frac{1}{n^{k/2+1}} \sum_{\Gamma(k,t) \in C_j} \sum_{G(i) \in \Gamma(k,t)} \mathbb{E}[Y(G(i))],
\]
and the summation \( \sum_{\Gamma(k,t) \in C_j} \) is taken over all canonical \( \Gamma(k,t) \)-graphs in category \( C_j \) and the summation \( \sum_{G(i) \in \Gamma(k,t)} \) is taken over all isomorphic graphs for a given canonical graph.

From the independence of \( Y_{ij} \) and \( E Y_{ij}^2 = s_{ij} \leq s \sigma^2 \) for \( s \geq 1 \), it follows that \( S_2 = 0 \).

For the graphs from categories \( C_1 \) and \( C_3 \) we introduce further notations. Let us consider the \( \Gamma(k,t) \)-graph \( G(i) \). Without loss of generality we assume that \( i_l, l = 1, \ldots, t \) are distinct coordinates of the vector \( i \) and define a vector \( \tilde{i} = (i_1, \ldots, i_t) \). We also set \( G(\tilde{i}) := G(i) \). Let \( \tilde{i}_q = (i_1, \ldots, i_{q-1}, i_{q+1}, \ldots, i_t) \) be vectors derived from \( \tilde{i} \) by deleting the elements in the position \( q \) and \( p, q \) respectively. We additionally assume that the coordinates of \( \tilde{i}_q \) do not coincide with \( i_p \). We denote the graph without the vertex \( i_q \) and all edges linked to it by \( G(\tilde{i}_q) \). If the vertex \( i_q \) is incident to a loop we denote by \( G'(\tilde{i}_q) \) the graph with this loop removed. By \( G'(\tilde{i}_q) \) we mean the graph derived from \( G(\tilde{i}_q) \) by deleting the edge from \( i_p \) to \( i_q \) taking into account the multiplicity.

Now we will estimate the term \( S_3 \). For a graph from category \( C_3 \) we know that \( k \) has to be even, i.e. \( k = 2m \), say. We illustrate the example of a \( \Gamma_3(k,t) \)-graph in Figure 2. This graph corresponds to the term
\[
Y(\tilde{G}(\tilde{i}_3)) = Y_{i_1}^2 Y_{i_2}^2 Y_{i_3} Y_{i_3}^2 Y_{i_3},
\]
We mention that \( \mathbb{E} Y_{i_p i_q}^{2s} \leq s \sigma^{2s}_{i_p i_q} \). Hence we may rewrite the terms which correspond to the graphs from category \( C_3 \) via variances.

Each graph \( G \) from \( C_3 \) we can decompose into two graphs \( G = G_1 \cup G_2 \) in the following way. We will paint all edges into two colours such that all coincident edges have the same colour. We choose the graph \( G_1 \) such that one of the
following cases holds true:
i) graph $G_1$ consists from only one vertex and all incident to it loops;
ii) graph $G_1$ consists from two vertices and the edge between them with the multiplicity greater than two;
iii) Each edge of the graph $G_1$ coincide with exactly one edge of the opposite direction and the graph of non coincident edges forms a simple cycle.

We may assume that the sum of multiplicities of all edges from $G_1$ is equal $2s$. It remains to consider the remaining $2(m - s)$ edges from the graph $G_2$.

Denote by $V_1, E_1$ and $V_2, E_2$ the sets of coordinates and edges of the graphs $G_1$ and $G_2$ respectively. We define $i_1^1 := (i_1, i_1 \in V_1)$. Now we may fix the graph $G_1$ and consider the graph $G_2$.

a) If the set $E_2$ is empty, then one should go to the step c). Otherwise we can consider the following opportunities:

(1) There is a loop from $G_2$ incident to the vertex $i_q \in V_2$ with the multiplicity $2a, a \geq 1$. In this case we estimate

$$
\frac{1}{n^{m+1}} \sum_{G(i) \in \Gamma(2m, \ell)} \mathbb{E}[Y(G(i))] \leq a \frac{1}{n^{m+1}} \sum_{i_i} \mathbb{E}[Y(G'(i_i))] \sigma_{i_q i_q}^{2a}.
$$

Applying $a$ times the inequality $n^{-1} \sigma_{i_q i_q}^2 \leq B^2_{i_q}$ and the condition (1.5), we delete all loops incident to this vertex;

(2) There are no loops incident to the vertex $i_q \in V_2 \setminus V_1$, but $i_q$ is connected with only one vertex $i_p \in V_2$ by an edge of the graph $G_2$ and the multiplicity of this edge is equal $2b, b \geq 1$. In this case we estimate

$$
\frac{1}{n^{m+1}} \sum_{G(i) \in \Gamma(2m, \ell)} \mathbb{E}[Y(G(i))] \leq b \frac{1}{n^{m+1}} \sum_{i_i} \mathbb{E}[Y(G'(i_i))] \sum_{i_q=1}^{n} \sigma_{i_p i_q}^{2b}.
$$

Here we may use $b - 1$ times the inequality $n^{-1} \sigma_{i_p i_q}^2 \leq B^2_{i_p}$ and condition (1.5) and consequently delete all coinciding edges except two. Then we may again apply condition (1.5) to delete $i_q$;

(3) There are no loops incident to some vertex from $V_2$ and no vertices in $V_2 \setminus V_1$ which are connected with only one vertex from $V_2$. Then we can take any two vertices, let’s say $i_p$ and $i_q$ from $V_2$ and estimate

$$
\frac{1}{n^{m+1}} \sum_{G(i) \in \Gamma(2m, \ell)} \mathbb{E}[Y(G(i))] \leq c \frac{1}{n^{m+1}} \sum_{i_i} \mathbb{E}[Y(G'(i_i))] \sigma_{i_p i_q}^{2c},
$$

where the multiplicity of the edge between $i_p$ and $i_q$ is equal $2c, c \geq 1$. Here we may use $c$ times the inequality $n^{-1} \sigma_{i_p i_q}^2 \leq B^2_{i_q}$ and the condition (1.5) and consequently delete all coinciding edges between $i_p$ and $i_q$;

b) go to step a);

c) It is easy to see that each time on the step a) we use the same bound (1.5).
Hence we will have

\begin{equation}
\frac{1}{n^{m+1}} \sum_{G(i) \in \Gamma(2m,t)} \mathbb{E}[Y(G(i))] \leq m \frac{C^{2(m-s)n^{m-s}}}{n^{m+1}} \sum_{i_1} \mathbb{E}[Y(G_1(i_1))].
\end{equation}

It remains to estimate the right hand side of (3.2). In the case i) we may estimate

\[
\frac{1}{n^{s+1}} \sum_{i_1=1}^{n} \mathbb{E}Y_{i_1}^{2s} \leq s \frac{1}{n^{s+1}} \sum_{i_1=1}^{n} \sigma_{i_1}^{2s} \leq C^{2(s-1)} \frac{n^2}{n^3} \sum_{i_1=1}^{n} \sigma_{i_1}^{2s} \leq C^{2(s-1)} \frac{n^2}{n^3} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^2 \leq C^{2(s-1)} \frac{n^2}{n^3} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^2 \leq C^{2(s-1)} \frac{n^2}{n^3} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^2.
\]

In the case ii) we will use the bound

\[
\frac{1}{n^{s+1}} \sum_{i_1,i_2=1}^{n} \mathbb{E}Y_{i_1i_2}^{2s} \leq s \frac{1}{n^{s+1}} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^{2s} \leq \frac{C^{2(s-2)}}{n^3} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^4 \leq \frac{C^{2(s-2)}}{n^3} \sum_{i_1,i_2=1}^{n} \sigma_{i_1i_2}^4.
\]

It remains to consider the case iii) only. We need to introduce further notation for this situation. We may redenote the vertices from the set $V_1$ and assume that $i_1^i = i_s := (i_1, ..., i_s)$. By $G_1(i_s)$ we denote the graph $G_1(i_1)$. Using the previous notations of $\tilde{i}_s, \bar{i}_s$ and $\tilde{G}_1(i_s)$ we set $p = 1$ and $q = 2$. Finally by $\tilde{G}_1(i_s)$ we denote the graph derived from $\tilde{G}_1(i_s)$ by deleting the vertex $i_2$ and the edge between $i_2$ and some another vertex $i_x$, $2 < x \leq s$. We may write

\begin{equation}
\frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E}Y(G_1(i_s)) \leq \frac{\tau^2}{n^s} \sum_{i_s} \mathbb{E}Y(\tilde{G}_1(i_s))
\end{equation}

\begin{equation}
+ \frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E}Y(\tilde{G}_1(i_s)) \mathbb{E}Y_{i_1i_2}^{2s} \mathbb{I}(\{|Y_{i_1i_2}| \geq \tau n \sqrt{n}\}).
\end{equation}

First we estimate the right hand side of (3.3). The graph of the non coincident edges of $\tilde{G}_1(i_s)$ forms a tree. We can sequentially delete all vertices and edges from $\tilde{G}_1(i_s)$ using the assumption (1.5) on each step. We derive the bound

\begin{equation}
\frac{\tau^2}{n^s} \sum_{i_s} \mathbb{E}Y(\tilde{G}_1(i_s)) \leq C^{2(s-1)} \frac{n^s}{n^s}.
\end{equation}
For the term (3.4) we may write

\[(3.6) \quad \frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E} Y(\tilde{G}_1(i_s)) \mathbb{E} Y_{i_1i_2}^2 \mathbb{I} (|Y_{i_1i_2}| \geq \tau_n \sqrt{n}) \]

\[\leq \frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E} Y(\tilde{G}_1(i_s)) \sigma_{i_1i_2}^2 \mathbb{E} Y_{i_1i_2}^2 \mathbb{I} (|Y_{i_1i_2}| \geq \tau_n \sqrt{n}) \]

\[\leq \frac{C^2}{n^s} \sum_{i_1, i_2 = 1}^n \mathbb{E} Y_{i_1i_2}^2 \mathbb{I} (|Y_{i_1i_2}| \geq \tau_n \sqrt{n}) \sum_{i_s} \mathbb{E} Y(\tilde{G}_1(i_s)). \]

Again using (1.5) one may show that

\[(3.7) \quad \sum_{i_s} \mathbb{E} Y(\tilde{G}_1(i_s)) \leq C^{2(s-2)} n^{s-2}. \]

By (3.6) and (3.7) we have

\[(3.8) \quad \frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E} Y(\tilde{G}_1(i_s)) \mathbb{E} Y_{i_1i_2}^2 \mathbb{I} (|Y_{i_1i_2}| \geq \tau_n \sqrt{n}) \leq C^{2(s-1)} L_n(\tau_n). \]

From (3.5) and (3.8) we derive the estimate

\[\frac{1}{n^{s+1}} \sum_{i_s} \mathbb{E} Y(G_1(i_s)) \leq C^{2(s-1)} \tau_n^2 + C^{2(s-1)} L_n(\tau_n). \]

Finally for the cases i)–iii) we will have

\[\frac{1}{n^{m+1}} \sum_{G(i) \in \Gamma(2m, t)} \mathbb{E}[Y(G(i))] \leq m C^{2(m-1)} (\tau_n^2 + L_n(\tau_n)) = o_m(1). \]

As an example we recommend to check this algorithm for the graph in Figure 2.

It is easy to see that the number of different canonical graphs in $C_3$ is of order $O_m(1)$. Finally for the term $S_3$ we get

\[S_3 = o_m(1). \]

It remains to consider the term $S_1$. For a graph from category $C_1$ we know that $k$ has to be even, i.e. $k = 2m$, say. In the category $C_1$ using the notations of $i_t$, $\tilde{i}_t$ and $\tilde{t}$ we take $t = m + 1$.

We illustrate on the left part of Figure 3 an example of the tree of noncoincident edges of a $\Gamma_1(2m)$-graph for $m = 5$. The term corresponding to this tree is $Y(G(i_6)) = Y_{i_1i_2}^2 Y_{i_2i_3}^2 Y_{i_2i_4}^2 Y_{i_1i_5}^2 Y_{i_5i_6}^2$.

We denote by $\sigma^2(i_{m+1}) = \sigma^2(G(i_{m+1}))$ the product of $m$ numbers $\sigma_{i_s i_t}^2$, where $i_s, i_t, s < t$ are vertices of the graph $G(i_{m+1})$ connected by edges of this graph. In our example, $\sigma^2(i_{m+1}) = \sigma^2(i_6) = \sigma_{i_1i_2}^2 \sigma_{i_2i_3}^2 \sigma_{i_2i_4}^2 \sigma_{i_1i_5}^2 \sigma_{i_5i_6}^2$. 


If \( m = 1 \) then \( \sigma^2(i_2) = \sigma^2_{i_1i_2} \) and

\[
\frac{1}{n^2} \sum_{i_1,i_2=1}^{n} \sigma^2_{i_1i_2} = \frac{1}{n} \sum_{i_1=1}^{n} \left[ \frac{1}{n} \sum_{i_2=1}^{n} \sigma^2_{i_1i_2} - 1 \right] + 1 + o(1),
\]

where we have used \( n^{-2} \sum_{i_1=1}^{n} \sigma^2_{i_1i_1} = o(1) \). By (1.4) the first term is of order \( o(1) \). The number of canonical graphs in \( C_1 \) for \( m = 1 \) is equal to 1. We conclude for \( m = 1 \) that

\[
S_1 = n^{-2} \sum_{\Gamma_1(2)} \sum_{i_1,i_2=1 \atop i_1 \neq i_2}^{n} \sigma^2_{i_1i_2} = 1 + o(1),
\]

Now we assume that \( m > 1 \). Let’s consider the tree of non-coincident edges of the graph. We can find a leaf in the tree, let’s say \( i_q \), and a vertex \( i_p \), which is connected to \( i_q \) by an edge of this tree. We have \( \sigma^2(i_{m+1}) = \sigma^2(i_{m+1}) \cdot \sigma^2_{i_p i_q} \), where \( \sigma^2(i_{m+1}) = \sigma^2(G(i_{m+1})) \).

In our example we can take the leaf \( i_6 \). On the right part of Figure 3 we have drawn the tree with deleted leaf \( i_6 \). We have \( \sigma^2_{i_p i_q} = \sigma^2_{i_5 i_6} \) and \( \sigma^2(i_6) = \sigma^2_{i_1i_2} \sigma^2_{i_2i_3} \sigma^2_{i_2i_4} \sigma^2_{i_1i_5} \).

\[
\frac{1}{n^{m+1}} \sum_{i_{m+1}}^{n} \sigma^2(i_{m+1}) = \frac{1}{n^{m+1}} \sum_{i_{m+1}}^{n} \sigma^2(i_{m+1}) \sum_{i_q=1}^{n} \sigma^2_{i_p i_q} + o_m(1)
\]

\[
= \frac{1}{n^m} \sum_{i_{m+1}}^{n} \sigma^2(i_{m+1}) \left[ \frac{1}{n} \sum_{i_q=1}^{n} \sigma^2_{i_p i_q} - 1 \right]
\]

\[
+ \frac{1}{n^m} \sum_{i_{m+1}}^{n} \sigma^2(i_{m+1}) + o_m(1),
\]
where we have added some graphs from category $C_3$ and use the similar bounds as for $S_3$ term. Now we will show that the term (3.10) is of order $o_m(1)$. Note that

\[
\frac{1}{n^m} \sum_{i_{m+1}} \sigma_i^2(i_{m+1}) \left| \frac{1}{n} \sum_{i_q=1}^n \sigma_{i_p}^2 - 1 \right| = \frac{1}{n} \sum_{i_p=1}^n \left| \frac{1}{n} \sum_{i_q=1}^n \sigma_{i_p}^2 - 1 \right| \frac{1}{n^{m-1}} \sum_{i_{m+1}} \sigma_i^2(i_{m+1}).
\]

We can sequentially delete leaves from the tree and using (1.5) write the bound

\[
\frac{1}{n^m} \sum_{i_{m+1}} \sigma_i^2(i_{m+1}) \leq C_2(m^2).
\]

By (3.13) and (1.4) we have shown that (3.10) is of order $o_m(1)$. For the second term (3.11) we can repeat the above procedure and stop after $m-1$ steps when we arrive at only two vertices in the tree. In the last step we can use the result (3.9). Finally we get

\[
S_1 = \frac{1}{n^{m+1}} \sum_{\Gamma_1(2m)} \sum_{i_{m+1}} \sigma_i^2(i_{m+1}) = \frac{1}{m+1} \left( \frac{2m}{m} \right) + o_m(1),
\]

which proves Theorem 1.4.

**References**

[1] R. Adamczak. On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries. *Electron. J. Probab.*, 16:no. 37, 1068–1095, 2011.

[2] L. Arnold. On Wigner’s semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 19:191–198, 1971.

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

[4] V. Bentkus. A new approach to approximations in probability theory and operator theory. *Liet. Mat. Rink.*, 43(4):444–470, 2003.

[5] L. Erdős. Universality of wigner random matrices: a survey of recent results. *arXiv:1004.0861*.

[6] F. Götze and A. Tikhomirov. Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 311(Veroyatn. i Stat. 7):92–123, 299, 2004.

[7] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. *Teor. Veroyatn. Primen.*, 51(1):171–192, 2006.

[8] S. O’Rourke. A note on the Marchenko-Pastur law for a class of random matrices with dependent entries. *arXiv:1201.3554*.

[9] L. A. Pastur. Spectra of random selfadjoint operators. *Uspehi Mat. Nauk*, 28(1(169)):3–64, 1973.

[10] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2), 67:325–327, 1958.
