

This document is downloaded from DR-NTU, Nanyang Technological University Library, Singapore.

Title	The logarithmic law of random determinant
Author(s)	Bao, Zhigang; Pan, Guangming; Zhou, Wang
Citation	Bao, Z., Pan, G., & Zhou, W. (2015). The logarithmic law of random determinant. <i>Bernoulli</i> , 21(3), 1600-1628.
Date	2015
URL	http://hdl.handle.net/10220/38998
Rights	© 2015 Bernoulli Society for Mathematical Statistics and Probability. This paper was published in <i>Bernoulli</i> and is made available as an electronic reprint (preprint) with permission of Bernoulli Society for Mathematical Statistics and Probability. The published version is available at: [http://dx.doi.org/10.3150/14-BEJ615]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Bernoulli **21**(3), 2015, 1600–1628
DOI: 10.3150/14-BEJ615

The logarithmic law of random determinant

ZHIGANG BAO¹, GUANGMING PAN² and WANG ZHOU³

¹*Department of Mathematics, Zhejiang University, Hangzhou, P.R. China.*

E-mail: maomie2007@gmail.com

²*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.*

E-mail: gmpan@ntu.edu.sg

³*Department of Statistics and Applied Probability, National University of Singapore, Singapore.*

E-mail: stazw@nus.edu.sg

Consider the square random matrix $A_n = (a_{ij})_{n,n}$, where $\{a_{ij} := a_{ij}^{(n)}, i, j = 1, \dots, n\}$ is a collection of independent real random variables with means zero and variances one. Under the additional moment condition

$$\sup_n \max_{1 \leq i, j \leq n} \mathbb{E} a_{ij}^4 < \infty,$$

we prove Girko's logarithmic law of $\det A_n$ in the sense that as $n \rightarrow \infty$

$$\frac{\log |\det A_n| - (1/2) \log(n-1)!}{\sqrt{(1/2) \log n}} \xrightarrow{d} N(0, 1).$$

Keywords: CLT for martingale; logarithmic law; random determinant

1. Introduction

Consider the square random matrix $A_n = (a_{ij})_{n,n}$, where $\{a_{ij} := a_{ij}^{(n)}, i, j = 1, \dots, n\}$ is a collection of independent real random variables with means zero and variances one. Moreover, we assume

$$\sup_n \max_{1 \leq i, j \leq n} \mathbb{E} a_{ij}^4 < \infty.$$

The main purpose of this paper is to study the determinant of A_n . As an important and fundamental function of a matrix, the random determinant has been investigated in many articles. For instance, the study of the moments of random determinants arose in the 1950s. One can refer to Dembo [3], Forsythe and Tukey [5], Nyquist, Rice and Riordan [15], Prékopa [17] for this topic. Besides, some lower and upper bounds for the magnitudes of random determinants were obtained in Costello and Vu [2] and Tao and Vu [19]

This is an electronic reprint of the original article published by the ISI/BS in *Bernoulli*, 2015, Vol. 21, No. 3, 1600–1628. This reprint differs from the original in pagination and typographic detail.

recently. A basic problem in the random determinant theory is to derive the fluctuation of the quantity $\log|\det A_n|$, which can give us an explicit description of the limiting behaviour of $|\det A_n|$. Particularly, when the entries of A_n are i.i.d. Gaussian, Goodman [10] found that $\det A_n^2$ can be written as a product of n independent χ^2 variables with different degrees of freedom. In fact, by using the Householder transform repeatedly, one can get that the joint distribution of the eigenvalues of $A_n A_n^T$ is the same as that of the tridiagonal matrix $L_n = D_n D_n^T$, where

$$D_n = \begin{pmatrix} a_n & & & \\ b_{n-1} & a_{n-1} & & \\ & \cdots & \cdots & \\ & & b_1 & a_1 \end{pmatrix}.$$

Here $\{a_n, \dots, a_1, b_{n-1}, \dots, b_1\}$ is a collection of independent variables such that $a_i \sim \chi_i, b_j \sim \chi_j$ for $i = 1, \dots, n, j = 1, \dots, n-1$. Such a tridiagonal form is well known for Gaussian matrix. One can refer to Dumitriu and Edelman [4], for instance. Then apparently one has

$$\det A_n^2 = \det A_n A_n^T \stackrel{d}{=} \prod_{i=1}^n a_i^2,$$

which implies that $\log|\det A_n|$ can be represented by a sum of n independent random variables. Then by the elementary properties of χ^2 distributions, the following CLT can be obtained

$$\frac{\log|\det A_n| - (1/2)\log(n-1)!}{\sqrt{(1/2)\log n}} \xrightarrow{d} N(0,1). \quad (1.1)$$

For details, one can see Rouault [18] or the Appendix of Costello and Vu [2] for instance. Like most of topics in the Random Matrix Theory, one may ask whether there is a ‘‘universal’’ phenomenon for the CLT of $\log|\det A_n|$ under general distribution assumption. The best result on this problem was given by Girko in [9] (one can also refer to Girko’s books [7] and [8] or his paper [6] for his former results on this topic), where the author only required the existence of the $(4 + \delta)$ th moment of the entries for some positive δ . Girko named (1.1) as ‘‘the logarithmic law’’ for random determinant. In Girko [9], using an elegant ‘‘method of perpendiculars’’ and combining the classical CLT for martingale, Girko claimed (1.1) is universal under the moment assumption mentioned above. However, though the proof route of Girko [9] is clear and quite original, it seems the proof is not complete and several parts are lack of mathematical rigour. Recently, Nguyen and Vu [14] provided a transparent proof of (1.1) for the general distribution case, under much stronger moment assumption in the sense that for all $t > 0$,

$$\mathbb{P}(|a_{ij}| \geq t) \leq C_1 \exp(-t^{C_2}) \quad (1.2)$$

with some positive constants C_1, C_2 independent of i, j, n . Obviously, (1.2) implies the existence of the moment of any order. The basic framework of the proof in Nguyen and

Vu [14] is similar to Girko's method of perpendiculars, which will be introduced in the next section. However, in order to provide a transparent proof, the authors of Nguyen and Vu [14] inserted a lot of new ingredients. Moreover, some unrigorous steps in Girko [9] can be fixed by the methods provided in Nguyen and Vu [14]. One may find that Nguyen and Vu [14] also provided a convergence rate of the logarithmic law as $\log^{-1/3+o(1)} n$, which is nearly optimal.

In this paper, also relying on the basic strategy of Girko's method of perpendiculars, we will provide a complete and rigorous proof under a weaker moment condition. More precisely, we only require the existence of 4th moment of the matrix entries. Our main result is

Theorem 1.1. *Let $A_n = (a_{ij})_{n,n}$ be a square random matrices, where $\{a_{ij}, 1 \leq i, j \leq n\}$ is a collection of independent real random variables with common mean 0 and variance 1. Moreover, we assume*

$$\sup_n \max_{1 \leq i, j \leq n} \mathbb{E} a_{ij}^4 < \infty.$$

Then we have the logarithmic law for $|\det A_n|$: as n tends to infinity,

$$\frac{\log |\det A_n| - (1/2) \log(n-1)!}{\sqrt{(1/2) \log n}} \xrightarrow{d} N(0, 1).$$

Our paper is organized as follows. In Section 2, we will sketch the main idea of Girko's method of perpendiculars and the proof route of Theorem 1.1. In Sections 3, 4 and 5, we will present the details of the proof with the aid of some additional lemmas, whose proofs will be given in the Appendix.

Throughout the paper, the notation such as C, K will be used to denote some positive constants independent of n , whose values may differ from line to line. We use $\|\cdot\|_2$ and $\|\cdot\|_{\text{op}}$ to represent the Euclidean norm of a vector and the operator norm of a matrix respectively as usual.

2. Girko's method of perpendiculars

In this section, we will sketch the main framework of Girko's method of perpendiculars, which was also pursued by Nguyen and Vu in the recent work [14]. To state this method rigorously, we need the following proposition whose proof will be given in Appendix B.

Proposition 2.1. *For the matrix A_n defined in Theorem 1.1, we can find a modified matrix $A'_n = (a'_{ij})_{n,n}$ satisfying the assumptions in Theorem 1.1 such that*

$$\mathbb{P}\{\text{all square submatrices of } A'_n \text{ are invertible}\} = 1 \quad (2.1)$$

and

$$\mathbb{P}\{|\log |\det A_n| - \log |\det A'_n|| = o(\sqrt{\log n})\} = 1 - o(1). \quad (2.2)$$

Remark 2.2. The construction of the modified matrix A'_n can be found in Nguyen and Vu [14]. The strategy is to set $a'_{ij} := (1 - \varepsilon^2)^{1/2} a_{ij} + \varepsilon \theta_{ij}$, where $\{\theta_{ij}, 1 \leq i, j \leq n\}$ is a collection of independent bounded continuous random variables with common mean zero and variance one and is independent of A_n . By choosing ε to be extremely small, say n^{-Kn} for large enough constant $K > 0$, it was shown in Nguyen and Vu [14] that (2.2) holds. The proof in Nguyen and Vu [14] relies on a lower bound estimate on the smallest singular value of square random matrices provided in Theorem 2.1 of Tao and Vu [20]. To adapt to our condition, we will use Theorem 4.1 of Götze and Tikhomirov [11] instead (by choosing $p_n = 1$ in Götze and Tikhomirov [11]). For convenience of the reader, we sketch the proof of Proposition 2.1 in Appendix B. The proof is just a slight modification of that in Nguyen and Vu [14] under our setting and assumptions.

Therefore, with the aid of Proposition 2.1, we can always work under the following assumption.

Assumption C₀. We assume that $A_n = (a_{ij})_{n,n}$ is a square random matrix, where $\{a_{ij}, 1 \leq i, j \leq n\}$ is a collection of independent real random variables with common mean zero and variance one. Besides, $\sup_n \max_{1 \leq i, j \leq n} \mathbb{E} a_{ij}^4 < \infty$. Moreover, all square submatrices of A_n are invertible with probability one.

The starting point of the method of perpendiculars is the elementary fact that the magnitude of the determinant of n real vectors in n dimensions is equal to the volume of the parallelepiped spanned by those vectors. Therefore, by the basic “base times height” formula, one can represent $|\det A_n|$ by the products of n perpendiculars. To make it more precise, we introduce some notations at first.

In the sequel, we will use \mathbf{a}_k^T to denote the k th row of A_n . And let $A_{(k)}$ be the $k \times n$ rectangular matrix formed by the first k rows of A_n . Particularly, one has $A_{(1)} = \mathbf{a}_1^T$ and $A_{(n)} = A_n$. Moreover, we use the notation V_i to denote the subspace generated by the first i rows of A_n and $P_i = (p_{jk}(i))_{n,n}$ to denote the projection matrix onto the space V_i^\perp . Let γ_{i+1} be the distance from \mathbf{a}_{i+1}^T to V_i for $1 \leq i \leq n-1$. And we set $\gamma_1 = \|\mathbf{a}_1^T\|_2$. Then by the “base times height” formula, we can write

$$\det A_n^2 = \prod_{i=0}^{n-1} \gamma_{i+1}^2. \quad (2.3)$$

Observe that γ_{i+1} is the norm of the projection of \mathbf{a}_{i+1}^T onto V_i^\perp . Thus we also have

$$\gamma_{i+1}^2 = \mathbf{a}_{i+1}^T P_i \mathbf{a}_{i+1}, \quad 1 \leq i \leq n-1. \quad (2.4)$$

Moreover, by the definition of V_i and Assumption C₀ one has that with probability one $A_{(i)} A_{(i)}^T$ is invertible and

$$P_i = I_n - A_{(i)}^T (A_{(i)} A_{(i)}^T)^{-1} A_{(i)}. \quad (2.5)$$

Then a direct consequence of the definition of γ_{i+1} is

$$\mathbb{E}\{\gamma_{i+1}^2 | P_i\} = \text{tr } P_i = n - i, \quad 0 \leq i \leq n - 1.$$

It follows from (2.3) that

$$\log \det A_n^2 = \sum_{i=0}^{n-1} \log \gamma_{i+1}^2,$$

which yields

$$\log \det A_n^2 - \log(n-1)! = \sum_{i=0}^{n-1} \log \frac{\gamma_{i+1}^2}{n-i} + \log n. \quad (2.6)$$

Now we set

$$X_{i+1} := X_{n,i+1} = \frac{\gamma_{i+1}^2 - (n-i)}{n-i}.$$

And we write

$$\log \frac{\gamma_{i+1}^2}{n-i} = X_{i+1} - \frac{X_{i+1}^2}{2} + R_{i+1}, \quad (2.7)$$

where

$$R_{i+1} := \log(1 + X_{i+1}) - \left(X_{i+1} - \frac{X_{i+1}^2}{2} \right).$$

Then by (2.6), one can write

$$\begin{aligned} & \frac{\log \det A_n^2 - \log(n-1)!}{\sqrt{2 \log n}} \\ &= \frac{1}{\sqrt{2 \log n}} \sum_{i=0}^{n-1} X_{i+1} - \frac{1}{\sqrt{2 \log n}} \left(\sum_{i=0}^{n-1} \frac{X_{i+1}^2}{2} - \log n \right) + \frac{1}{\sqrt{2 \log n}} \sum_{i=0}^{n-1} R_{i+1}. \end{aligned} \quad (2.8)$$

Crudely speaking, the main route is to prove that the first term of (2.8) weakly converges to the standard Gaussian distribution and the remaining two terms tend to zero in probability. Let \mathcal{E}_i be the σ -algebra generated by the first i rows of A_n , by definition we have

$$\mathbb{E}\{X_{i+1} | \mathcal{E}_i\} = 0.$$

Thus X_1, \dots, X_n is a martingale difference sequence with respect to the filtration $\emptyset \subset \mathcal{E}_1 \subset \dots \subset \mathcal{E}_{n-1}$. In Girko [9], under the assumption of the existence of $(4 + \delta)$ th moments

of the matrix entries for some $\delta > 0$, Girko used the CLT for martingales to show the first term of (2.8) is asymptotically Gaussian. He also showed that the last term of (2.8) is asymptotically negligible. However, some steps in the proofs of these two parts are lack of mathematical rigour. Moreover, we do not find the discussion of the second term of (2.8) in Girko's original proof in [9]. Recently, Nguyen and Vu provided a complete proof under the assumption that the distributions of the matrix entries satisfy (1.2), thus with finite moments of all orders. In the following sections, we will also adopt the representation (2.7) and the theory on the weak convergence of martingales.

It will be clear that the proof will rely on some approximations of X_{i+1} and R_{i+1} . However, these approximations are i -dependent and the large i case turns out to be badly approximated. To see this, we can take the Gaussian case for example. Note that when the entries are standard Gaussian, $\gamma_{i+1}^2 \sim \chi_{n-i}^2$, thus

$$X_{i+1} = \mathcal{O}((n-i)^{-1/2}), \quad R_{i+1} = \mathcal{O}((n-i)^{-3/2})$$

with high probability. Especially, when $n-i$ is $\mathcal{O}(1)$, the main term X_{i+1} and the negligible term R_{i+1} are comparable to be $\mathcal{O}(1)$. Such a fact will be an obstacle if we use crude estimations for X_{i+1} and R_{i+1} for general distribution case. This is explained such as follows. When we estimate the last term of (2.8), a basic strategy is to use the Taylor expansion of $\log(1 + X_{i+1})$ to gain a relatively small remainder R_{i+1} , which requires $|X_{i+1}| \leq 1 - c$ for some small positive constant c . However, as we mentioned above, when $n-i$ is too small, such a bound is hard to be guaranteed since $X_{i+1} = \mathcal{O}(1)$ with high probability, especially under the assumption of the 4th moment. Fortunately, if all the \mathbf{a}_{i+1}^T 's are Gaussian for large $i \geq n - s_1$ for some positive number s_1 , γ_{i+1}^2 are independent χ^2 variables for all $i \geq n - s_1$ even if $A_{(n-s_1)}$ is generally distributed. Such an explicit distribution information can be used to deal with the large i part. Therefore, in [9] Girko proposed to replace some rows by Gaussian ones and prove the logarithmic law for the matrix after replacement, and then recover the result to the original one by a comparison procedure. Such a strategy was also used in Nguyen and Vu [14]. To pursue this idea, we set

$$s_1 = \lfloor \log^{3a} n \rfloor$$

for some sufficiently large positive constant a . Our proof route can be split into the following five steps.

(i)

$$\frac{\sum_{i=0}^{n-s_1} X_{i+1}}{\sqrt{2 \log n}} \xrightarrow{d} N(0, 1).$$

(ii)

$$\frac{\sum_{i=0}^{n-s_1} X_{i+1}^2 / 2 - \log n}{\sqrt{2 \log n}} \xrightarrow{\mathbb{P}} 0.$$

(iii)

$$\frac{\sum_{i=0}^{n-s_1} R_{i+1}}{\sqrt{2 \log n}} \xrightarrow{\mathbb{P}} 0.$$

(iv) If the last s_1 rows of A_n are Gaussian, then

$$\frac{\sum_{i=n-s_1}^{n-1} \log(\gamma_{i+1}^2/(n-i))}{\sqrt{2 \log n}} \xrightarrow{\mathbb{P}} 0.$$

(v) Let B_n be a random matrix satisfying the basic Assumption \mathbf{C}_0 and differing from A_n only in the last s_1 rows. Then one has

$$\sup_x |\mathbb{P}\{|\det A_n| \leq x\} - \mathbb{P}\{|\det B_n| \leq x\}| \longrightarrow 0.$$

We will prove (i) and (ii) together in Section 3, and prove (iii) and (iv) in Section 4. Section 5 is devoted to the proof of (v).

3. Convergence issue on the martingale difference sequence

In this section, we will prove the statements (i) and (ii). The arguments for both two parts heavily rely on the fact that $\{X_i, 1 \leq i \leq n\}$ is a martingale difference sequence.

In order to prove (i), we will use the following classical CLT for martingales, which can be found in the book of Hall and Heyde [12], for instance.

Proposition 3.1. *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be a zero-mean, square-integrable martingale array with differences Z_{ni} . Suppose that*

$$\begin{aligned} \max_i |Z_{ni}| &\xrightarrow{\mathbb{P}} 0, \\ \sum_i Z_{ni}^2 &\xrightarrow{\mathbb{P}} 1. \end{aligned}$$

Moreover, $\mathbb{E}(\max_i Z_{ni}^2)$ is bounded in n . Then we have

$$S_{nk_n} \xrightarrow{d} N(0, 1).$$

Now we use the above proposition to prove (i). Let $k_n = n - s_1$, $Z_{ni} = X_i/\sqrt{2 \log n}$. Thus, it suffices to show that

$$\frac{1}{\sqrt{\log n}} \max_{0 \leq i \leq n-s_1} |X_{i+1}| \xrightarrow{\mathbb{P}} 0, \quad (3.1)$$

$$\frac{1}{2 \log n} \sum_{i=0}^{n-s_1} X_{i+1}^2 \xrightarrow{\mathbb{P}} 1, \quad (3.2)$$

and

$$\frac{1}{\log n} \mathbb{E} \left(\max_i X_{i+1}^2 \right) \leq \frac{1}{\log n} \mathbb{E} \sum_{i=0}^{n-s_1} X_{i+1}^2 \leq C \quad (3.3)$$

for some positive constant C independent of n . To verify (3.1) it suffices to show the following lemma.

Lemma 3.2. *Under the Assumption \mathbf{C}_0 , we have for any constant $\varepsilon > 0$*

$$\sum_{i=0}^{n-s_1} \mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |X_{i+1}| \geq \varepsilon \right\} \rightarrow 0$$

as n tends to infinity.

Proof. Below we use the notation

$$Q_i = [q_{jk}(i)]_{n,n} =: \frac{1}{n-i} P_i.$$

Thus by definition and the fact that $\text{tr} Q_i = 1$ we have

$$X_{i+1} = \mathbf{a}_{i+1}^T Q_i \mathbf{a}_{i+1} - 1 = \sum_k q_{kk}(i) (a_{i+1,k}^2 - 1) + \sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v}.$$

Now we introduce the quantities

$$U_{i+1} = \sum_k q_{kk}(i) (a_{i+1,k}^2 - 1), \quad V_{i+1} = \sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v}.$$

Obviously

$$|X_{i+1}| \leq |U_{i+1}| + |V_{i+1}|.$$

Then it is elementary to see

$$\mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |X_{i+1}| \geq \varepsilon \right\} \leq \mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |U_{i+1}| \geq \frac{\varepsilon}{2} \right\} + \mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |V_{i+1}| \geq \frac{\varepsilon}{2} \right\}.$$

Therefore, it suffices to verify the following two statements instead:

$$\sum_{i=0}^{n-s_1} \mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |U_{i+1}| \geq \frac{\varepsilon}{2} \right\} \rightarrow 0$$

and

$$\sum_{i=0}^{n-s_1} \mathbb{P} \left\{ \frac{1}{\sqrt{\log n}} |V_{i+1}| \geq \frac{\varepsilon}{2} \right\} \rightarrow 0,$$

which can be implied by

$$\frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} U_{i+1}^2 \rightarrow 0, \quad (3.4)$$

and

$$\frac{1}{\log^2 n} \sum_{i=0}^{n-s_1} \mathbb{E} V_{i+1}^4 \rightarrow 0. \quad (3.5)$$

First, we verify (3.4). In the sequel, we set

$$s_2 = \lfloor n \log^{-20a} n \rfloor.$$

By definition, we see

$$\mathbb{E} U_{i+1}^2 = \mathbb{E} \left(\sum_k q_{kk}(i) (a_{i+1,k}^2 - 1) \right)^2 \leq C \mathbb{E} \sum_k q_{kk}^2(i).$$

Using the basic fact $0 \leq p_{kk}(i) \leq 1$ one has $0 \leq q_{kk}(i) \leq 1/(n-i)$. Taking this fact and $\text{tr} Q_i = 1$ into account we have

$$\begin{aligned} \frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} U_{i+1}^2 &\leq C \frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_k q_{kk}^2(i) \\ &\leq C \left(\frac{1}{\log n} \sum_{i=0}^{n-s_2} \frac{1}{n-i} + \frac{1}{\log n} \sum_{i=n-s_2}^{n-s_1} \mathbb{E} \max_k q_{kk}(i) \right) \\ &\leq C \frac{1}{\log n} \sum_{i=n-s_2}^{n-s_1} \mathbb{E} \max_k q_{kk}(i) + \mathcal{O} \left(\frac{\log \log n}{\log n} \right) \\ &\leq C \frac{1}{\log n} \sum_{i=n-s_2}^{n-s_1} \frac{1}{n-i} \mathbb{E} \max_k p_{kk}(i) + \mathcal{O} \left(\frac{\log \log n}{\log n} \right) \\ &\leq C \max_{n-s_2 \leq i \leq n-s_1} \mathbb{E} \max_k p_{kk}(i) + \mathcal{O} \left(\frac{\log \log n}{\log n} \right). \end{aligned} \quad (3.6)$$

Now we need the following crucial technical lemma which will be used repeatedly in the sequel.

Lemma 3.3. *Under the above notation, we have*

$$\max_{n-s_2 \leq i \leq n-1} \mathbb{E} \max_k p_{kk}(i) \leq C \log^{-8a} n \quad (3.7)$$

for some positive constant C .

The proof of Lemma 3.3 will be given later. Now we proceed to the proof of (3.4) by assuming Lemma 3.3. Note that (3.4) follows from (3.6) and (3.7) immediately. Hence, it remains to show (3.5). We need the following simple deviation lemma for the quadratic form, whose proof is quite elementary and will be given in Appendix A.

Lemma 3.4. *Suppose $x_i, i = 1, \dots, n$ are independent real random variables with common mean zero and variance 1. Moreover, we assume $\max_i \mathbb{E}|x_i|^4 \leq \nu_4$. Let $M_n = (m_{ij})_{n,n}$ be a nonnegative definite matrix which is deterministic. Then we have*

$$\mathbb{E} \left| \sum_{i=1}^n m_{ii} x_i^2 - \text{tr} M_n \right|^4 \leq C (\nu_8 \text{tr} M_n^4 + (\nu_4 \text{tr} M_n^2)^2) \quad (3.8)$$

and

$$\mathbb{E} \left| \sum_{u \neq v} m_{uv} x_u x_v \right|^4 \leq C \nu_4^2 (\text{tr} M_n^2)^2 \quad (3.9)$$

for some positive constant C .

Note that by (3.9) one has

$$\mathbb{E} V_{i+1}^4 \leq C \mathbb{E} (\text{tr} Q_i^2)^2 = C \frac{1}{(n-i)^2},$$

which implies that

$$\frac{1}{4 \log^2 n} \sum_{i=0}^{n-s_1} \mathbb{E} V_{i+1}^4 = \mathcal{O}(\log^{-2-3a} n). \quad (3.10)$$

Thus (3.5) holds. Then Lemma 3.2 follows from (3.4) and (3.5) immediately. Thus (3.1) is verified. \square

Now we prove Lemma 3.3.

Proof of Lemma 3.3. We denote the j th column of $A_{(i)}$ by $\mathbf{b}_j(i)$ and use the notation $A_{(i,j)}$ to denote the matrix induced from $A_{(i)}$ by deleting the j th column $\mathbf{b}_j(i)$. Moreover,

we set the positive parameter $\alpha = \alpha_n := n^{-1/6}$. By (2.5), we have

$$\begin{aligned} p_{kk}(i) &= 1 - \mathbf{b}_k(i)^T (A_{(i)} A_{(i)}^T)^{-1} \mathbf{b}_k(i) \\ &= 1 - \mathbf{b}_k(i)^T (A_{(i,k)} A_{(i,k)}^T + \mathbf{b}_k(i) \mathbf{b}_k(i)^T)^{-1} \mathbf{b}_k(i) \\ &\leq 1 - \mathbf{b}_k(i)^T (A_{(i,k)} A_{(i,k)}^T + n\alpha I_i + \mathbf{b}_k(i) \mathbf{b}_k(i)^T)^{-1} \mathbf{b}_k(i) \\ &= (1 + \mathbf{b}_k(i)^T (A_{(i,k)} A_{(i,k)}^T + n\alpha I_i)^{-1} \mathbf{b}_k(i))^{-1}, \end{aligned}$$

where in the last step we used the Sherman–Morrison formula

$$(M + \mathbf{b}_k(i) \mathbf{b}_k(i)^T)^{-1} = M^{-1} - \frac{M^{-1} \mathbf{b}_k(i) \mathbf{b}_k(i)^T M^{-1}}{1 + \mathbf{b}_k(i)^T M^{-1} \mathbf{b}_k(i)} \quad (3.11)$$

for $k \times k$ invertible matrix M . Let

$$G_{(i,k)}(\alpha) = \left(\frac{1}{n} A_{(i,k)} A_{(i,k)}^T + \alpha I_i \right)^{-1}, \quad G_{(i)}(\alpha) = \left(\frac{1}{n} A_{(i)} A_{(i)}^T + \alpha I_i \right)^{-1}.$$

Then one has

$$p_{kk}(i) \leq \left(1 + \frac{1}{n} \mathbf{b}_k(i)^T G_{(i,k)}(\alpha) \mathbf{b}_k(i) \right)^{-1}.$$

Hence, to verify (3.7) we only need to show

$$\mathbb{E} \max_k \left(1 + \frac{1}{n} \mathbf{b}_k(i)^T G_{(i,k)}(\alpha) \mathbf{b}_k(i) \right)^{-1} \leq C \log^{-8\alpha} n, \quad n - s_2 \leq i \leq n - 1. \quad (3.12)$$

It is apparent that $G_{(i)}(\alpha)$ and $G_{(i,k)}(\alpha)$ are positive-definite and

$$\|G_{(i)}(\alpha)\|_{\text{op}}, \|G_{(i,k)}(\alpha)\|_{\text{op}} \leq \alpha^{-1}.$$

Moreover, we have

$$\begin{aligned} & |\text{tr } G_{(i)}(\alpha) - \text{tr } G_{(i,k)}(\alpha)| \\ &= \left| \text{tr} \left(\frac{1}{n} A_{(i,k)} A_{(i,k)}^T + \frac{1}{n} \mathbf{b}_k(i) \mathbf{b}_k(i)^T + \alpha I_i \right)^{-1} - \text{tr} \left(\frac{1}{n} A_{(i,k)} A_{(i,k)}^T + \alpha I_i \right)^{-1} \right| \\ &= \frac{(1/n) \mathbf{b}_k(i)^T G_{(i,k)}(\alpha)^2 \mathbf{b}_k(i)}{1 + (1/n) \mathbf{b}_k(i)^T G_{(i,k)}(\alpha) \mathbf{b}_k(i)} \leq \alpha^{-1}, \end{aligned} \quad (3.13)$$

where in the second step above we used the Sherman–Morrison formula (3.11) again. Now we set

$$\chi(i) = \mathbf{1}_{\{(1/n) \text{tr } G_{(i)}(\alpha) \geq \log^{10\alpha} n\}},$$

and we denote the (u, v) th entry of $G_{(i,k)}(\alpha)$ by $G_{(i,k)}(u, v)$ below. Moreover, for ease of presentation, when there is no confusion, we will omit the parameter α from the notation $G_{(i,k)}(\alpha)$ and $G_{(i)}(\alpha)$. Then we have for some small constant $0 < \varepsilon < 1/2$,

$$\begin{aligned}
& \mathbb{E} \max_k \left(1 + \frac{1}{n} \mathbf{b}_k(i)^T G_{(i,k)} \mathbf{b}_k(i) \right)^{-1} \\
& \leq \mathbb{P} \left\{ \frac{1}{n} \operatorname{tr} G_{(i)} \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left(1 + \frac{1}{n} \mathbf{b}_k(i)^T G_{(i,k)} \mathbf{b}_k(i) \right)^{-1} \\
& \leq \mathbb{P} \left\{ \frac{1}{n} \operatorname{tr} G_{(i)} \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left(1 + \frac{1}{n} \sum_{j=1}^i G_{(i,k)}(j, j) a_{jk}^2 - \varepsilon \right)^{-1} \\
& \quad + \mathbb{P} \left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{1 \leq u \neq v \leq i} G_{(i,k)}(u, v) a_{uk} a_{vk} \right| \geq \varepsilon \right\} \\
& \leq \mathbb{P} \left\{ \frac{1}{n} \operatorname{tr} G_{(i)} \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left(1 + \log^{-a} n \cdot \frac{1}{n} \operatorname{tr} G_{(i,k)} - \varepsilon \right)^{-1} \\
& \quad + C \sum_{k=1}^n \mathbb{P} \left\{ \sum_{j=1}^i G_{(i,k)}(j, j) a_{jk}^2 < \log^{-a} n \cdot \operatorname{tr} G_{(i,k)}, \frac{1}{n} \operatorname{tr} G_{(i)} \geq \log^{10a} n \right\} \\
& \quad + \mathbb{P} \left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{1 \leq u \neq v \leq i} G_{(i,k)}(u, v) a_{uk} a_{vk} \right| \geq \varepsilon \right\} \\
& \leq \mathbb{P} \left\{ \frac{1}{n} \operatorname{tr} G_{(i)} \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \left(1 + \log^{-a} n \cdot \frac{1}{n} \operatorname{tr} G_{(i)} - 2\varepsilon \right)^{-1} \\
& \quad + C \sum_{k=1}^n \mathbb{P} \left\{ \sum_{j=1}^i G_{(i,k)}(j, j) a_{jk}^2 < \log^{-a} n \operatorname{tr} G_{(i,k)}, \frac{1}{n} \operatorname{tr} G_{(i)} \geq \log^{10a} n \right\} \\
& \quad + \mathbb{P} \left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{1 \leq u \neq v \leq i} G_{(i,k)}(u, v) a_{uk} a_{vk} \right| \geq \varepsilon \right\},
\end{aligned} \tag{3.14}$$

where in the above last inequality, we used (3.13). Below we will estimate (3.14) term by term. To this end, we need the following two lemmas, whose proofs will be given in Appendix A.

Lemma 3.5. *Under the assumption of Theorem 1.1, for $n - s_2 \leq i \leq n - 1$, we have for $\alpha = n^{-1/6}$*

$$\mathbb{E} \left\{ \frac{1}{n} \operatorname{tr} G_{(i)}(\alpha) \right\} = s_i(\alpha) + \mathcal{O}(n^{-1/6})$$

and

$$\text{Var}\left\{\frac{1}{n}\text{tr}G_{(i)}(\alpha)\right\} = \mathcal{O}(n^{-1/3}),$$

where

$$s_i(\alpha) = 2\left(\alpha + 1 - \frac{i}{n} + \sqrt{\left[\alpha + \left(1 - \frac{i}{n}\right)\right]^2 + 4\alpha\frac{i}{n}}\right)^{-1}.$$

With the aid of Lemma 3.5, we can estimate the first term of (3.14) as follows. Note that by definition $s_i(\alpha) \geq \frac{1}{10} \log^{20a} n$ for $n - s_2 \leq i \leq n - 1$, we have

$$\begin{aligned} \mathbb{P}\left\{\frac{1}{n}\text{tr}G_{(i)} \leq \log^{10a} n\right\} &\leq \mathbb{P}\left\{\left|\frac{1}{n}\text{tr}G_{(i)} - \mathbb{E}\frac{1}{n}\text{tr}G_{(i)}\right| \geq \frac{1}{20}\log^{20a} n\right\} \\ &\leq C \log^{-40a} n \text{Var}\left\{\frac{1}{n}\text{tr}G_{(i)}\right\} = o(n^{-1/3}). \end{aligned}$$

For the second term of (3.14), with the definition of $\chi(i)$, obviously one has

$$\mathbb{E}\chi(i) \left(1 + \log^{-a} n \cdot \frac{1}{n}\text{tr}G_{(i)} - 2\varepsilon\right)^{-1} \leq C \log^{-8a} n.$$

Now we deal with the third term of (3.14). We set

$$\hat{a}_{jk} = a_{jk} \mathbf{1}_{\{|a_{jk}| \leq \log^a n\}}, \quad \tilde{a}_{jk} = \frac{\hat{a}_{jk} - \mathbb{E}\hat{a}_{jk}}{\sqrt{\text{Var}\{\hat{a}_{jk}\}}}.$$

Since $G_{(i,k)}(j, j)$'s are positive and $\hat{a}_{jk}^2 \leq a_{jk}^2$ one has

$$\begin{aligned} &\sum_{k=1}^n \mathbb{P}\left\{\sum_{j=1}^i G_{(i,k)}(j, j) a_{jk}^2 < \log^{-a} n \cdot \text{tr}G_{(i,k)}, \frac{1}{n}\text{tr}G_{(i)} \geq \log^{10a} n\right\} \\ &\leq \sum_{k=1}^n \mathbb{P}\left\{\sum_j G_{(i,k)}(j, j) \hat{a}_{jk}^2 < \log^{-a} n \cdot \text{tr}G_{(i,k)}, \frac{1}{n}\text{tr}G_{(i)} \geq \log^{10a} n\right\}. \end{aligned}$$

Moreover, by the assumption $\sup_n \max_{ij} \mathbb{E}a_{ij}^4 < \infty$ it is easy to derive that

$$\mathbb{E}\hat{a}_{jk} = \mathcal{O}(\log^{-3a} n), \quad \text{Var}\{\hat{a}_{jk}\} = 1 + \mathcal{O}(\log^{-2a} n).$$

Consequently,

$$\tilde{a}_{jk} = \hat{a}_{jk} + \mathcal{O}(\log^{-a} n),$$

which implies

$$\tilde{a}_{jk}^2 \leq 2\hat{a}_{jk}^2 + \mathcal{O}(\log^{-2a} n) \leq 2\hat{a}_{jk}^2 + \log^{-a} n$$

for sufficiently large n . Therefore, we have

$$\begin{aligned} & \sum_{k=1}^n \mathbb{P} \left\{ \sum_j G_{(i,k)}(j,j) a_{jk}^2 < \log^{-a} n \cdot \text{tr} G_{(i,k)}, \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\} \\ & \leq \sum_{k=1}^n \mathbb{P} \left\{ \sum_j G_{(i,k)}(j,j) \hat{a}_{jk}^2 < \log^{-a} n \cdot \text{tr} G_{(i,k)}, \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\} \\ & \leq \sum_{k=1}^n \mathbb{P} \left\{ \sum_j G_{(i,k)}(j,j) \tilde{a}_{jk}^2 < 3 \log^{-a} n \cdot \text{tr} G_{(i,k)}, \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\} \\ & \leq \sum_{k=1}^n \mathbb{P} \left\{ \left| \sum_j G_{(i,k)}(j,j) \tilde{a}_{jk}^2 - \text{tr} G_{(i,k)} \right| \geq \frac{1}{2} \text{tr} G_{(i,k)}, \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\} \\ & \leq \sum_{k=1}^n \mathbb{P} \left\{ \frac{1}{n} \left| \sum_j G_{(i,k)}(j,j) \tilde{a}_{jk}^2 - \text{tr} G_{(i,k)} \right| \geq \frac{1}{4} \log^{10a} n \right\} \\ & \leq C \log^{-40a} n \cdot n^{-4} \sum_{k=1}^n \mathbb{E} [\log^{4a} n \cdot \text{tr} G_{(i,k)}^4 + (\text{tr} G_{(i,k)}^2)^2] \\ & \leq o\left(\frac{1}{n\alpha^4}\right). \end{aligned}$$

In the fourth inequality we used the fact (3.13) and in the fifth inequality we used (3.8) and the fact $\mathbb{E}|\tilde{a}_{ij}|^{4+t} = \mathcal{O}(\log^{ta} n)$ for any $t \geq 0$, which is easy to see from the definition of \tilde{a}_{ij} . Now we begin to deal with the last term of (3.14). Note that by (3.9)

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{u \neq v} G_{(i,k)}(u,v) a_{uk} a_{vk} \right| \geq \varepsilon \right\} \\ & \leq \sum_{k=1}^n \mathbb{P} \left\{ \frac{1}{n} \left| \sum_{u \neq v} G_{(i,k)}(u,v) a_{uk} a_{vk} \right| \geq \varepsilon \right\} \\ & \leq \varepsilon^{-4} n^{-4} \sum_{k=1}^n \mathbb{E} \left(\sum_{u \neq v} G_{(i,k)}(u,v) a_{uk} a_{vk} \right)^4 \\ & \leq C \varepsilon^{-4} n^{-4} \sum_{k=1}^n \mathbb{E} (\text{tr} G_{(i,k)}^2)^2 \end{aligned}$$

$$\leq \mathcal{O}\left(\frac{1}{n\alpha^4}\right).$$

Therefore, (3.12) follows from the above estimates, so does (3.7). Hence, we complete the proof. \square

Now we come to deal with (3.2) and (3.3). Note that (3.2) can be implied by (ii) directly. Thus we will prove the statement (ii) and (3.3) below. We reformulate them as the following lemma and then prove it.

Lemma 3.6. *Under the Assumption C_0 , one has*

$$\frac{\sum_{i=0}^{n-s_1} X_{i+1}^2 - 2 \log n}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0, \quad (3.15)$$

and

$$\frac{1}{2 \log n} \sum_{i=0}^{n-s_1} \mathbb{E} X_{i+1}^2 = 1 + o(1). \quad (3.16)$$

Proof. We begin with (3.15). We split the proof of (3.15) into two steps.

$$\frac{\sum_{i=0}^{n-s_1} (X_{i+1}^2 - \mathbb{E}\{X_{i+1}^2 | \mathcal{E}_i\})}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0, \quad (3.17)$$

and

$$\frac{\sum_{i=0}^{n-s_1} \mathbb{E}\{X_{i+1}^2 | \mathcal{E}_i\} - 2 \log n}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0. \quad (3.18)$$

Observe that

$$\begin{aligned} \mathbb{E}\{X_{i+1}^2 | \mathcal{E}_i\} &= \mathbb{E}\left\{ \left(\sum_{j,k=1}^n q_{jk}(i) a_{i+1,j} a_{i+1,k} - 1 \right)^2 \middle| \mathcal{E}_i \right\} \\ &= \frac{2}{n-i} + \sum_k q_{kk}(i)^2 (\mathbb{E}|a_{i+1,k}|^4 - 3). \end{aligned} \quad (3.19)$$

Therefore, to verify (3.18), we only need

$$\frac{1}{\sqrt{\log n}} \mathbb{E} \sum_{i=0}^{n-s_1} \sum_k q_{kk}(i)^2 \longrightarrow 0. \quad (3.20)$$

Note that from the proof of (3.4) we can get the following estimate directly.

$$\frac{1}{\sqrt{\log n}} \mathbb{E} \sum_{i=0}^{n-s_1} \sum_k q_{kk}(i)^2 = \mathcal{O}\left(\frac{\log \log n}{\sqrt{\log n}}\right). \quad (3.21)$$

Thus it suffices to show (3.17). By elementary calculations, we have

$$\begin{aligned} & X_{i+1}^2 - \mathbb{E}\{X_{i+1}^2 | \mathcal{E}_i\} \\ &= -2 \sum_u q_{uu}(i)(a_{i+1,u}^2 - 1) + 2 \sum_{u \neq v} q_{uu}(i)q_{vv}(i)(a_{i+1,u}^2 a_{i+1,v}^2 - 1) \\ &\quad + 2 \sum_{u \neq v} q_{uv}(i)^2 (a_{i+1,u}^2 a_{i+1,v}^2 - 1) \\ &\quad + 2 \sum_{\substack{u_1 \neq v_1, u_2 \neq v_2 \\ \{u_1, v_1\} \neq \{u_2, v_2\}}} q_{u_1 v_1}(i) q_{u_2 v_2}(i) a_{i+1, u_1} a_{i+1, v_1} a_{i+1, u_2} a_{i+1, v_2} \\ &\quad + \sum_u q_{uu}(i)^2 (a_{i+1,u}^4 - \mathbb{E}a_{i+1,u}^4) + 2 \left(\sum_u q_{uu}(i)(a_{i+1,u}^2 - 1) \right) \left(\sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v} \right) \\ &=: 2W_1(i) + 2W_2(i), \end{aligned}$$

where

$$\begin{aligned} W_1(i) &= - \sum_i q_{uu}(i)(a_{i+1,u}^2 - 1) + \sum_{u \neq v} q_{uu}(i)q_{vv}(i)(a_{i+1,u}^2 a_{i+1,v}^2 - 1) \\ &\quad + \sum_{u \neq v} q_{uv}(i)^2 (a_{i+1,u}^2 a_{i+1,v}^2 - 1) \\ &\quad + \sum_{\substack{u_1 \neq v_1, u_2 \neq v_2 \\ \{u_1, v_1\} \neq \{u_2, v_2\}}} q_{u_1 v_1}(i) q_{u_2 v_2}(i) a_{i+1, u_1} a_{i+1, v_1} a_{i+1, u_2} a_{i+1, v_2}, \end{aligned}$$

and

$$\begin{aligned} W_2(i) &= \frac{1}{2} \sum_u q_{uu}(i)^2 (a_{i+1,u}^4 - \mathbb{E}a_{i+1,u}^4) \\ &\quad + \left(\sum_u q_{uu}(i)(a_{i+1,u}^2 - 1) \right) \left(\sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v} \right). \end{aligned}$$

We split the issue to show

$$\frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_1(i) \xrightarrow{\mathbb{P}} 0, \quad \frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_2(i) \xrightarrow{\mathbb{P}} 0. \quad (3.22)$$

First, we deal with the second statement of (3.22). Note that

$$\begin{aligned}
& \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \mathbb{E}|W_2(i)| \\
& \leq C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_u q_{uu}(i)^2 \\
& \quad + C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \left(\mathbb{E} \left(\sum_u q_{uu}(i)(a_{i+1,u}^2 - 1) \right)^2 \right)^{1/2} \left(\mathbb{E} \left(\sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v} \right)^2 \right)^{1/2} \\
& \leq C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_u q_{uu}(i)^2 + C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \left(\frac{1}{n-i} \right)^{1/2} \left(\mathbb{E} \sum_u q_{uu}(i)^2 \right)^{1/2}.
\end{aligned} \tag{3.23}$$

By (3.21), we have that the first term of (3.23) is of the order of $\mathcal{O}(\log \log n / \sqrt{\log n})$. For the second term, by using (3.7) we have

$$\begin{aligned}
& \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \left(\frac{1}{n-i} \right)^{1/2} \left(\mathbb{E} \sum_u q_{uu}(i)^2 \right)^{1/2} \\
& \leq \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_2} \frac{1}{n-i} + \frac{1}{\sqrt{\log n}} \sum_{i=n-s_2}^{n-s_1} \left(\frac{1}{n-i} \mathbb{E} \sum_u q_{uu}(i)^2 \right)^{1/2} \\
& \leq \frac{1}{\sqrt{\log n}} \sum_{i=n-s_2}^{n-s_1} \frac{1}{n-i} \left(\mathbb{E} \max_u p_{uu}(i) \right)^{1/2} + \mathcal{O} \left(\frac{\log \log n}{\sqrt{\log n}} \right) \\
& = \mathcal{O} \left(\frac{\log \log n}{\sqrt{\log n}} \right).
\end{aligned}$$

Now we consider the first term of (3.22). It is easy to see

$$\mathbb{E}\{W_1(i)\} = 0, \quad \mathbb{E}\{W_1(i)W_1(j)\} = 0, \quad i \neq j,$$

which yields

$$\begin{aligned}
\mathbb{E} \left(\frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_1(i) \right)^2 &= \frac{1}{\log n} \sum_{i=0}^{s_1} \mathbb{E} W_1(i)^2 \\
&\leq \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_u q_{uu}(i)^2 + \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u \neq v, u \neq w} q_{uu}(i)^2 q_{vv}(i) q_{ww}(i)
\end{aligned} \tag{3.24}$$

$$\begin{aligned}
& + \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u \neq v, u \neq w} q_{uv}(i)^2 q_{uw}(i)^2 \\
& + \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u_1 \neq v_1, u_2 \neq v_2} |q_{u_1 v_1}(i) q_{u_1 v_2}(i) q_{u_2 v_1}(i) q_{u_2 v_2}(i)|.
\end{aligned}$$

The estimation of (3.24) is elementary but somewhat tedious. In fact, one can find the estimate towards every term of (3.24) in Nguyen and Vu [14] (see the estimation of $\text{Var}(\sum_{i=0}^{n-s_1} Y_{i+1})$ in Section 6 of Nguyen and Vu [14]). Here we omit the details and claim the following estimation

$$\mathbb{E} \left(\frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_1(i) \right)^2 = \mathcal{O} \left(\frac{\log \log n}{\log n} \right).$$

Then (3.17) follows, so does (3.15). Moreover, it is easy to see that (3.16) holds by combining (3.19) and (3.20). Therefore, Lemma 3.6 is proved. \square

Thus we complete the proof of (i) and (ii).

4. Negligible parts (iii) and (iv)

In this section, we will prove the statements (iii) and (iv). We start with (iii). The following elementary but crucial lemma will be needed.

Lemma 4.1. *By the definitions above, if $X_{i+1} \geq -1 + \log^{-a/2} n$, one has*

$$|R_{i+1}| \leq C(U_{i+1}^2 + |V_{i+1}|^{2+\delta}) \log \log n$$

for any $0 \leq \delta \leq 1$. Here $C := C(a, \delta)$ is a positive constant only depends on a and δ .

Proof. We split the discussion into three cases. Choose some small constant $0 < \varepsilon < \frac{1}{10}$ (say) and consider the three cases $|X_{i+1}| \leq 1 - \varepsilon$, $X_{i+1} > 1 - \varepsilon$ and $-1 + \log^{-a/2} n \leq X_{i+1} < -1 + \varepsilon$ separately. For the first case, we can use the elementary Taylor expansion to see that

$$|R_{i+1}| \leq C|X_{i+1}|^3 \leq C|U_{i+1} + V_{i+1}|^{2+\delta}$$

for any $0 \leq \delta \leq 1$. If $|U_{i+1}| \geq 1$ or $|V_{i+1}| \geq 1$, we will immediately get

$$|R_{i+1}| \leq C(U_{i+1}^2 + |V_{i+1}|^{2+\delta})$$

since $|U_{i+1} + V_{i+1}| < 1$. If both $|U_{i+1}|$ and $|V_{i+1}|$ are less than 1, we have

$$|R_{i+1}| \leq C(|U_{i+1}|^{2+\delta} + |V_{i+1}|^{2+\delta}) \leq C(U_{i+1}^2 + |V_{i+1}|^{2+\delta}).$$

Now we come to deal with the second case. When $X_{i+1} > 1 - \varepsilon$, obviously one has

$$|R_{i+1}| \leq C(U_{i+1} + V_{i+1})^2.$$

Then it is elementary to see that we always can find some positive constant C such that

$$|R_{i+1}| \leq C(U_{i+1}^2 + |V_{i+1}|^{2+\delta})$$

since $\max\{U_{i+1}, V_{i+1}\} > \frac{1}{2} - \frac{\varepsilon}{2}$. Finally, we deal with the last case. Note that when $-1 + \log^{-c} n \leq X_{i+1} < -1 + \varepsilon$, we have

$$|R_{i+1}| \leq C \log \log n.$$

Moreover, it is obvious that we have $\max\{|U_{i+1}|, |V_{i+1}|\} > \frac{1}{2} - \frac{\varepsilon}{2}$. Consequently, one has

$$|R_{i+1}| \leq C(U_{i+1}^2 + |V_{i+1}|^{2+\delta}) \log \log n.$$

In conclusion, we completed the proof. \square

The next lemma is devoted to bounding the probability of the event $\bigcup_{i=0}^{m-s_1} \{X_{i+1} < -1 + \log^{-a} n\}$.

Lemma 4.2. *Under the Assumption \mathbf{C}_0 , we have*

$$\sum_{i=0}^{n-s_1} \mathbb{P}\{X_{i+1} < -1 + \log^{-a/2} n\} \longrightarrow 0$$

as n tends to infinity.

Proof. Note that

$$\begin{aligned} & \mathbb{P}\{X_{i+1} < -1 + \log^{-a/2} n\} \\ &= \mathbb{P}\{\mathbf{a}_{i+1}^T \mathbf{Q}_i \mathbf{a}_{i+1} < \log^{-a/2} n\} \\ &\leq \mathbb{P}\left\{\sum_k q_{kk}(i) a_{i+1,k}^2 < 2 \log^{-a/2} n\right\} + \mathbb{P}\left\{\left|\sum_{u \neq v} q_{uv}(i) a_{i+1,u} a_{i+1,v}\right| \geq \frac{1}{2} \log^{-a/2} n\right\}. \end{aligned}$$

Now we recall the definition

$$\hat{a}_{i+1,k} = a_{i+1,k} \mathbf{1}_{\{|a_{i+1,k}| \leq \log^a n\}}, \quad \tilde{a}_{i+1,k} = \frac{\hat{a}_{i+1,k} - \mathbb{E} \hat{a}_{i+1,k}}{\sqrt{\text{Var}\{\hat{a}_{i+1,k}\}}}.$$

A similar discussion as that in the last section yields

$$\mathbb{P}\left\{\sum_k q_{kk}(i) a_{i+1,k}^2 < 2 \log^{-a/2} n\right\} \leq \mathbb{P}\left\{\sum_k q_{kk}(i) \tilde{a}_{i+1,k}^2 < C \log^{-a/2} n\right\}$$

$$\begin{aligned}
&\leq \mathbb{P}\left\{\left|\sum_k q_{kk}(i)\tilde{a}_{i+1,k}^2 - 1\right| \geq \frac{1}{2}\right\} \\
&\leq C(\log^{4a} n \cdot \text{tr} Q_i^4 + (\text{tr} Q_i^2)^2) \\
&\leq C(\log^{4a} n \cdot (n-i)^{-3} + (n-i)^{-2}).
\end{aligned}$$

Here in the third inequality, we used (3.8) again. Moreover, by (3.9), we obtain

$$\mathbb{P}\left\{\left|\sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v}\right| \geq \frac{1}{2} \log^{-a/2} n\right\} \leq C(n-i)^{-2} \log^{2a} n.$$

Thus finally, we have

$$\begin{aligned}
&\sum_{i=0}^{n-s_1} \mathbb{P}\{X_{i+1} < -1 + \log^{-a/2} n\} \\
&\leq C\left(\log^{4a} n \sum_{i=0}^{n-s_1} (n-i)^{-3} + \log^{2a} n \sum_{i=0}^{n-s_1} (n-i)^{-2}\right) \\
&\leq C \log^{-a} n.
\end{aligned}$$

Therefore, we complete the proof. \square

Combining Lemmas 4.1 with 4.2, one has with probability $1 - o(1)$,

$$|R_{i+1}| \leq C(U_{i+1}^2 + V_{i+1}^{2+\delta}) \log \log n, \quad 0 \leq i \leq n - s_1.$$

Thus to show (iii), it suffices to verify that

$$\frac{\log \log n}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} (U_{i+1}^2 + |V_{i+1}|^{2+\delta}) \xrightarrow{\mathbb{P}} 0,$$

which can be implied by

$$\frac{\log \log n}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \mathbb{E}(U_{i+1}^2 + |V_{i+1}|^{2+\delta}) \longrightarrow 0. \quad (4.1)$$

Note that

$$\sum_{i=0}^{n-s_1} \mathbb{E}U_{i+1}^2 \leq C \sum_{i=0}^{n-s_1} \mathbb{E} \sum_j q_{jj}(i)^2 = \mathcal{O}(\log \log n). \quad (4.2)$$

Moreover, we have

$$\sum_{i=0}^{n-s_1} \mathbb{E}|V_{i+1}|^{2+\delta} \leq \sum_{i=0}^{n-s_1} (\mathbb{E}V_{i+1}^4)^{(2+\delta)/4} \leq C \sum_{i=0}^{n-s_1} (n-i)^{-1-\delta/2} = o(1). \quad (4.3)$$

Then (4.1) follows from (4.2) and (4.3) immediately. Thus, we completed the proof of (iii).

It remains to show (iv) in this section. The proof is quite elementary owing to the fact that $\{\gamma_{i+1}^2, i = n - s_1, \dots, n - 1\}$ is an independent sequence and $\gamma_{i+1}^2 \sim \chi_{n-i}^2$. One may refer to Section 7 of Nguyen and Vu [14] for instance. By using the Laplace transform trick, Nguyen and Vu [14] showed that for $0 < c < 100$,

$$\mathbb{P}\left\{\sum_{n-s_1 \leq i \leq n-1} \frac{\log(\gamma_{i+1}^2/(n-i))}{\sqrt{2 \log n}} < -\log^{1/2+c} n\right\} = o(\exp(-\log^{c/2} n)),$$

which implies (v) immediately.

5. A replacement issue: Proof of (v)

In this section, we present the proof for (v). In other words, we shall replace the last s_1 rows of Gaussian entries by generally distributed entries. We will need the following classical Berry–Esseen bound for sum of independent random variables. For instance, one can refer to Theorem 5.4 of Petrov [16].

Lemma 5.1. *Let Z_1, \dots, Z_m be independent real random variables such that $\mathbb{E}Z_j = 0$ and $\mathbb{E}|Z_j|^3 < \infty$, $j = 1, \dots, m$. Assume that*

$$\sigma_j^2 = \mathbb{E}Z_j^2, \quad D_m = \sum_{j=1}^m \sigma_j^2, \quad L_m = D_m^{-3/2} \sum_{j=1}^m \mathbb{E}|Z_j|^3.$$

Then there exists a constant $C > 0$ such that

$$\sup_x \left| \mathbb{P}\left(D_m^{-1/2} \sum_{j=1}^m Z_j \leq x\right) - \Phi(x) \right| \leq CL_m.$$

Our strategy is to replace one row at each step, and derive the difference between the distributions of the logarithms of the magnitudes of two adjacent determinants. Hence, it suffices to compare two matrices with only one different row. Noting that since the magnitude of a determinant is invariant under swapping of two rows, thus without loss of generality, we only need to compare two random matrices $A_n = (a_{ij})_{n,n}$ and $\bar{A}_n = (\bar{a}_{ij})_{n,n}$ satisfying Assumption **C**₀ such that they only differ in the last row. More precisely, we assume that $a_{ij} = \bar{a}_{ij}$, $1 \leq i \leq n - 1$, $1 \leq j \leq n$ and \mathbf{a}_n^T and $\bar{\mathbf{a}}_n^T$ are independent. Here we use \mathbf{a}_n^T and $\bar{\mathbf{a}}_n^T$ to denote the n th row of A_n and \bar{A}_n , respectively as above. Below we use the notation α_{ni} to denote the cofactor of a_{ni} . It is elementary that

$$\det A_n = \sum_{k=1}^n a_{nk} \alpha_{nk},$$

and

$$\det \bar{A}_n = \sum_{k=1}^n \bar{a}_{nk} \alpha_{nk}.$$

Now we set

$$\Delta = \sqrt{\alpha_{n1}^2 + \cdots + \alpha_{nn}^2}.$$

Consider the quantities

$$\frac{\det A_n}{\Delta} = \sum_{k=1}^n a_{nk} \frac{\alpha_{nk}}{\Delta}, \quad \frac{\det \bar{A}_n}{\Delta} = \sum_{k=1}^n \bar{a}_{nk} \frac{\alpha_{nk}}{\Delta}.$$

By using Lemma 5.1, we obtain

$$\sup_x \left| \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \mid \mathcal{E}_{n-1} \right\} - \Phi(x) \right| \leq C \sum_{i=1}^n \frac{|\alpha_{ni}|^3}{\Delta^3}.$$

Therefore, we have

$$\begin{aligned} & \sup_x \left| \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \right\} - \Phi(x) \right| \\ &= \sup_x \left| \mathbb{E} \left\{ \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \mid \mathcal{E}_{n-1} \right\} \right\} - \Phi(x) \right| \\ &\leq C \mathbb{E} \sum_{k=1}^n \frac{|\alpha_{nk}|^3}{\Delta^3}. \end{aligned}$$

For simplicity, we will briefly denote $\mathbf{b}_k(n-1)$ and $A_{(n-1,k)}$ by \mathbf{b}_k and A_{nk} respectively in the sequel. Then by the definitions of cofactors and the Cauchy–Binet formula, we have

$$\begin{aligned} \left(\frac{\alpha_{nk}}{\Delta} \right)^2 &= \frac{\det A_{nk}^2}{\det A_{(n-1)} A_{(n-1)}^T} = \det A_{nk}^T (A_{nk} A_{nk}^T + \mathbf{b}_k \mathbf{b}_k^T)^{-1} A_{nk} \\ &= \det (I_{n-1} + A_{nk}^{-1} \mathbf{b}_k \mathbf{b}_k^T (A_{nk}^T)^{-1})^{-1} = (1 + \mathbf{b}_k^T (A_{nk} A_{nk}^T)^{-1} \mathbf{b}_k)^{-1}. \end{aligned}$$

Moreover, one always has

$$\mathbb{E} \sum_{k=1}^n \frac{|\alpha_{nk}|^3}{\Delta^3} \leq \mathbb{E} \max_{k=1, \dots, n} \frac{|\alpha_{nk}|}{\Delta} = \mathbb{E} \max_{k=1, \dots, n} (1 + \mathbf{b}_k^T (A_{nk} A_{nk}^T)^{-1} \mathbf{b}_k)^{-1/2}.$$

Recall the definition

$$G_{(n-1,k)}(\alpha) = \left(\frac{1}{n} A_{nk} A_{nk}^T + \alpha I_{n-1} \right)^{-1}, \quad G_{(n-1)}(\alpha) = \left(\frac{1}{n} A_{(n-1)} A_{(n-1)}^T + \alpha I_{n-1} \right)^{-1}.$$

By using (3.12), we obtain

$$\begin{aligned}
& \mathbb{E} \max_{k=1, \dots, n} (1 + \mathbf{b}_k^T (A_{nk} A_{nk}^T)^{-1} \mathbf{b}_k)^{-1/2} \\
& \leq \mathbb{E} \max_{k=1, \dots, n} \left(1 + \frac{1}{n} \mathbf{b}_k^T G_{(n-1, k)}(\alpha) \mathbf{b}_k \right)^{-1/2} \\
& \leq \left(\mathbb{E} \max_{k=1, \dots, n} \left(1 + \frac{1}{n} \mathbf{b}_k^T G_{(n-1, k)}(\alpha) \mathbf{b}_k \right)^{-1} \right)^{1/2} \\
& \leq \mathcal{O}(\log^{-4a} n).
\end{aligned}$$

Thus we have

$$\sup_x \left| \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \right\} - \Phi(x) \right| \leq C \log^{-4a} n,$$

and

$$\sup_x \left| \mathbb{P} \left\{ \frac{\det \bar{A}_n}{\Delta} \leq x \right\} - \Phi(x) \right| \leq C \log^{-4a} n.$$

Consequently, we have

$$\sup_x \left| \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \right\} - \mathbb{P} \left\{ \frac{\det \bar{A}_n}{\Delta} \leq x \right\} \right| \leq C \log^{-4a} n,$$

which implies

$$\sup_x \left| \mathbb{P} \left\{ \frac{\log \det A_n^2 - \log(n-1)!}{\sqrt{2 \log n}} \leq x \right\} - \mathbb{P} \left\{ \frac{\log \det \bar{A}_n^2 - \log(n-1)!}{\sqrt{2 \log n}} \leq x \right\} \right| \leq C \log^{-4a} n.$$

Then after $s_1 = \lfloor \log^{3a} n \rfloor$ steps of replacing, we can finally recover the logarithmic law to general distribution case. Thus we completed the proof.

Appendix A

In this appendix, we provide the proof of Lemma 3.5 and Lemma 3.4. The proof is intrinsically the same as the counterpart in Bai [1]. For convenience of the reader, we sketch it below. For ease of notation, we represent Lemma 3.5 as follows.

Lemma A.1. *Let $X = (x_{ij})_{p \times n}$ be a random matrix, where $n - s_1 \leq p \leq n$ and $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$ is a collection of real independent random variables with means zero and variances 1. Moreover, we assume $\sup_n \max_{i,j} \mathbb{E} x_{ij}^4 < \infty$. Let $G(\alpha) = (\frac{1}{n} X X^T + \alpha)^{-1}$,*

we have for $\alpha = n^{-1/6}$

$$\mathbb{E}\left\{\frac{1}{n} \operatorname{tr} G(\alpha)\right\} = s_p(\alpha) + \mathcal{O}(n^{-1/6})$$

and

$$\operatorname{Var}\left\{\frac{1}{n} \operatorname{tr} G(\alpha)\right\} = \mathcal{O}(n^{-1/3}),$$

where

$$s_p(\alpha) = 2\left(\alpha + 1 - \frac{i}{n} + \sqrt{\left[\alpha + \left(1 - \frac{i}{n}\right)\right]^2 + 4\alpha\frac{i}{n}}\right)^{-1}.$$

Proof. For convenience, we set

$$\begin{aligned} r_p(\alpha) &= \frac{1}{p} \mathbb{E} \operatorname{tr} G(\alpha) \\ &= \frac{1}{p} \mathbb{E} \sum_{k=1}^p \frac{1}{(1/n)x_k x_k^T + \alpha - (1/n^2)x_k X(k)X^T(k) + (1/n)X(k)X^T(k) + \alpha I_{p-1}}^{-1} X(k)x_k^T. \end{aligned}$$

Here x_k is the k th row of X and $X(k)$ is the submatrix of X by deleting its k th row. Set $y_p = p/n$. Then we write

$$\begin{aligned} r_p(\alpha) &= \frac{1}{p} \sum_{k=1}^p \mathbb{E} \frac{1}{\varepsilon_k + 1 + \alpha - y_p + y_p \alpha r_p(\alpha)} \\ &= \frac{1}{1 + \alpha - y_p + y_p \alpha r_p(\alpha)} + \delta, \end{aligned} \tag{A.1}$$

where

$$\varepsilon_k = \frac{1}{n} \sum_{j=1}^n (x_{kj}^2 - 1) + y_p - y_p \alpha r_p(\alpha) - \frac{1}{n^2} x_k X^T(k) \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} X(k)x_k^T,$$

and

$$\begin{aligned} \delta &= \delta_p = -\frac{1}{p} \sum_{k=1}^p \mathbb{E} \frac{\varepsilon_k}{(1 + \alpha - y_p + y_p \alpha r_p(\alpha))(1 + \alpha - y_p + y_p \alpha r_p(\alpha) + \varepsilon_k)} \\ &= -\frac{1}{p} \sum_{k=1}^p \mathbb{E} \frac{\varepsilon_k}{(1 + \alpha - y_p + y_p \alpha r_p(\alpha))^2} \\ &\quad - \frac{1}{p} \sum_{k=1}^p \mathbb{E} \frac{\varepsilon_k^2}{(1 + \alpha - y_p + y_p \alpha r_p(\alpha))^2 (1 + \alpha - y_p + y_p \alpha r_p(\alpha) + \varepsilon_k)}. \end{aligned} \tag{A.2}$$

From (A.1), we can get

$$r_p(\alpha) = \frac{1}{2y_p\alpha} \left(\sqrt{(1 + \alpha - y_p - y_p\alpha\delta)^2 + 4y_p\alpha} - (1 + \alpha - y_p - y_p\alpha\delta) \right). \quad (\text{A.3})$$

It is not difficult to see that

$$\frac{1}{|1 + \alpha - y_p + y_p\alpha r_p(\alpha) + \varepsilon_k|} \leq \alpha^{-1}. \quad (\text{A.4})$$

By (A.2) and (A.4), we can get

$$|\delta| \leq \frac{1}{p} \sum_{k=1}^p (|\mathbb{E}\varepsilon_k| + \alpha^{-1}\mathbb{E}\varepsilon_k^2)(1 + \alpha - y_p + y_p\alpha r_p(\alpha))^2. \quad (\text{A.5})$$

First, note that

$$\begin{aligned} |\mathbb{E}\varepsilon_k| &= \left| \mathbb{E} \left(y_p - y_p\alpha \frac{1}{n} \text{tr} G(\alpha) - \frac{1}{n^2} \text{tr} \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} \frac{1}{n} X(k)X^T(k) \right) \right| \\ &= \alpha \frac{1}{n} |\mathbb{E}(\text{tr}(XX^T + \alpha I_p)^{-1} - \text{tr}(X(k)X^T(k) + \alpha I_{p-1}))| + \frac{1}{n} \\ &= \mathcal{O}\left(\frac{1}{n}\right). \end{aligned} \quad (\text{A.6})$$

Next, we come to estimate $\mathbb{E}\varepsilon_k^2$. Note that

$$\mathbb{E}\varepsilon_k^2 = \text{Var}\{\varepsilon_k\} + (\mathbb{E}\varepsilon_k)^2 \leq \frac{C}{n} + T_1 + T_2, \quad (\text{A.7})$$

where

$$\begin{aligned} T_1 &= \mathbb{E} \left| \frac{1}{n^2} x_k X^T(k) \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} X(k)x_k^T \right. \\ &\quad \left. - \mathbb{E}^{(k)} \frac{1}{n^2} x_k X^T(k) \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} X(k)x_k^T \right|^2, \end{aligned}$$

and

$$T_2 = \frac{\alpha^2}{n^2} \mathbb{E} \left| \text{tr} \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} - \mathbb{E} \text{tr} \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} \right|^2. \quad (\text{A.8})$$

Here $\mathbb{E}^{(k)}$ represents the conditional expectation given $\{x_{ij}, i \neq k\}$. Let

$$\Gamma_k = (\gamma_{ij}(k)) = \frac{1}{n} X^T(k) \left(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} X(k).$$

Then one has

$$T_1 \leq \frac{C}{n^2} \mathbb{E} \operatorname{tr} \Gamma_k^2 \leq \frac{C}{n\alpha^2}. \quad (\text{A.9})$$

Let \mathbb{E}_d be the conditional expectation given $\{x_{ij}, d+1 \leq i \leq p, 1 \leq j \leq n\}$. Define

$$\begin{aligned} \gamma_d(k) &= \mathbb{E}_{d-1} \operatorname{tr} \left(\frac{1}{n} X(k) X^T(k) + \alpha I_{p-1} \right)^{-1} - \mathbb{E}_d \operatorname{tr} \left(\frac{1}{n} X(k) X^T(k) + \alpha I_{p-1} \right)^{-1} \\ &= \mathbb{E}_{d-1} \sigma_d(k) - \mathbb{E}_d \sigma_d(k), \quad d = 1, 2, \dots, p, \end{aligned}$$

where

$$\sigma_d(k) = \operatorname{tr} \left(\frac{1}{n} X(k) X^T(k) + \alpha I_{p-1} \right)^{-1} - \operatorname{tr} \left(\frac{1}{n} X(k, d) X^T(k, d) + \alpha I_{p-2} \right)^{-1}.$$

Noting that

$$|\sigma_d(k)| \leq \alpha^{-1},$$

one has

$$T_2 \leq \frac{C}{n^2} \sum_{d=1}^p \mathbb{E} |\gamma_d^2(k)| \leq \frac{C}{n\alpha^2}. \quad (\text{A.10})$$

Combining (A.7)–(A.10) we can get

$$\mathbb{E} \varepsilon_k^2 \leq \frac{C}{n\alpha^2}. \quad (\text{A.11})$$

Substituting (A.6), (A.11) and the basic fact

$$(1 + \alpha - y_p + y_p r_p(\alpha))^2 \leq \alpha^{-2}$$

into (A.5) one has

$$|\delta| \leq \frac{C}{n\alpha^5}.$$

Then by (A.3) one has

$$\begin{aligned} r_p(\alpha) &= \frac{1}{2y_p\alpha} \left(\sqrt{(1 + \alpha - y_p)^2 + 4y_p\alpha} - (1 + \alpha - y_p) \right) + \mathcal{O}(n^{-1}\alpha^{-5}) \\ &= 2 \left(\sqrt{(1 + \alpha - y_p)^2 + 4y_p\alpha} + (1 + \alpha - y_p) \right)^{-1} + \mathcal{O}(n^{-1}\alpha^{-5}). \end{aligned}$$

Moreover, similar to the estimate towards (A.8), we can get that

$$\text{Var}\left\{\frac{1}{n}\text{tr}G(\alpha)\right\}\leq\frac{C}{n\alpha^4}.$$

Therefore, we can complete the proof. \square

Now let us prove Lemma 3.4.

Proof of Lemma 3.4. For the diagonal part, we have

$$\begin{aligned}\mathbb{E}\left|\sum_{i=1}^nm_{ii}(x_i^2-1)\right|^4&=\sum_{j,k,u,v}m_{jj}m_{kk}m_{uu}m_{vv}\mathbb{E}(x_{jj}^2-1)(x_{kk}^2-1)(x_{uu}^2-1)(x_{vv}^2-1) \\ &\leq C\left(\nu_8\sum_jm_{jj}^4+\nu_4^2\sum_{j\neq k}m_{jj}^2m_{kk}^2\right) \\ &\leq C(\nu_8\text{tr}M_n^4+(\nu_4\text{tr}M_n^2)^2).\end{aligned}$$

For the off-diagonal part, we have

$$\begin{aligned}\mathbb{E}\left|\sum_{u\neq v}m_{uv}x_u x_v\right|^4&=\sum_{u_i\neq v_i,i=1,\dots,4}\prod_{i=1}^4m_{u_i v_i}\mathbb{E}\prod_{i=1}^4x_{u_i}x_{v_i} \\ &\leq C\left(\nu_4^2\sum_{u\neq v}m_{uv}^4+\nu_3^2\nu_2\sum_{u,v,r}m_{uv}^2m_{ur}m_{rv}+\nu_2^4\sum_{u,v,r,w}m_{uv}m_{vr}m_{rw}m_{wu}\right) \\ &\leq C\nu_4^2((\text{tr}M_n^2)^2+\text{tr}M_n^2\cdot(\text{tr}M_n^4)^{1/2}+\text{tr}M_n^4) \\ &\leq C\nu_4^2(\text{tr}M_n^2)^2.\end{aligned}$$

Thus, we may complete the proof of Lemma 3.4. \square

Appendix B

In this appendix, we state the proof of Proposition 2.1. We use the idea in Nguyen and Vu [14]. First, we need the following lemma derived from Theorem 4.1 in Götze and Tikhomirov [11]. Let $s_n(W_n)\leq s_{n-1}(W_n)\leq\cdots\leq s_1(W_n)$ be the ordered singular values of an $n\times n$ matrix W_n .

Lemma B.1. *Under the assumptions of Theorem 1.1, there exist some positive constants c, C, L such that*

$$\mathbb{P}\{s_n(A_n)\leq n^{-L}; s_1(A_n)\leq n\}\leq e^{-cn}+\frac{C\sqrt{\log n}}{\sqrt{n}}.$$

Remark B.2. It is easy to get Lemma B.1 from Theorem 4.1 of Götze and Tikhomirov [11] by choosing $p_n = 1$. We also remark here Theorem 4.1 of Götze and Tikhomirov [11] are stated for more general case under weaker moment assumption. For convenience, we just restate it under our setting.

Now by the result of Latała [13], it is easy to see under the assumption of Theorem 1.1,

$$\mathbb{E}s_1(A_n) \leq C \left(\max_i \sqrt{\sum_j \mathbb{E}a_{ij}^2} + \max_j \sqrt{\sum_i \mathbb{E}a_{ij}^2} + \sqrt[4]{\sum_{ij} \mathbb{E}a_{ij}^4} \right) \leq C\sqrt{n}.$$

Therefore, one has

$$\mathbb{P}\{s_1(A_n) \geq n\} \leq \frac{\mathbb{E}s_1(A_n)}{n} = Cn^{-1/2}.$$

Together with Lemma B.1 we obtain

$$\mathbb{P}\{s_n(A_n) \geq n^{-L}\} = 1 - \mathcal{O}\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right). \quad (\text{B.1})$$

Now let θ_0 follow the uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$ independent of A_n . Let $\theta_{ij}, 1 \leq i, j \leq n$ be independent copies of θ_0 . And we set $A'_n = (a'_{ij})$, where $a'_{ij} = (1 - \varepsilon_n^2)^{1/2} a_{ij} + \varepsilon_n \theta_{ij}$. Here we choose $\varepsilon_n = n^{-(100+2L)n}$ (say). Writing $\Theta_n = (\theta_{ij})_{n,n}$, then by Weyl's inequality, one has

$$|s_i(A'_n) - (1 - \varepsilon_n^2)^{1/2} s_i(A_n)| \leq \varepsilon_n \|\Theta_n\|_{\text{op}} \leq Cn^{-(99+2L)n}.$$

Therefore, by (B.1) we have with probability $1 - \frac{\sqrt{\log n}}{\sqrt{n}}$

$$|\det A'_n| = \prod_{i=1}^n s_i(A'_n) = (1 - \varepsilon_n^2)^{n/2} (1 + \mathcal{O}(n^{-(99+L)n}))^n \prod_{i=1}^n s_i(A_n) = (1 + o(1)) |\det A_n|,$$

which implies (2.2). Moreover, by the construction, since θ_0 is a continuous variable, it is obvious that (2.1) holds. Thus, we may complete the proof.

Acknowledgements

Z.G. Bao was supported in part by NSFC Grant 11371317, NSFC Grant 11101362, ZJNSF Grant R6090034 and SRFDP Grant 20100101110001; G.M. Pan was supported in part by the Ministry of Education, Singapore, under Grant # ARC 14/11; W. Zhou was supported in part by the Ministry of Education, Singapore, under Grant # ARC 14/11, and by a Grant R-155-000-131-112 at the National University of Singapore.

References

- [1] BAI, Z.D. (1993). Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.* **21** 649–672. [MR1217560](#)
- [2] COSTELLO, K.P. and VU, V. (2009). Concentration of random determinants and permanent estimators. *SIAM J. Discrete Math.* **23** 1356–1371. [MR2556534](#)
- [3] DEMBO, A. (1989). On random determinants. *Quart. Appl. Math.* **47** 185–195. [MR0998095](#)
- [4] DUMITRIU, I. and EDELMAN, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. [MR1936554](#)
- [5] FORSYTHE, G.E. and TUKEY, J.W. (1952). The extent of n random unit vectors. *Bull. Amer. Math. Soc.* **58** 502.
- [6] GIRKO, V.L. (1979). A central limit theorem for random determinants. *Teor. Veroyatnost. i Primenen.* **24** 728–740. [MR0550529](#)
- [7] GIRKO, V.L. (1990). *Theory of Random Determinants. Mathematics and Its Applications (Soviet Series)* **45**. Dordrecht: Kluwer Academic. Translated from the Russian. [MR1080966](#)
- [8] GIRKO, V.L. (1996). *Theory of Linear Algebraic Equations with Random Coefficients*. New York: Allerton Press. [MR1424467](#)
- [9] GIRKO, V.L. (1997). A refinement of the central limit theorem for random determinants. *Teor. Veroyatnost. i Primenen.* **42** 63–73. [MR1453330](#)
- [10] GOODMAN, N.R. (1963). The distribution of the determinant of a complex Wishart distributed matrix. *Ann. Math. Statist.* **34** 178–180. [MR0145619](#)
- [11] GÖTZE, F. and TIKHOMIROV, A. (2010). The circular law for random matrices. *Ann. Probab.* **38** 1444–1491. [MR2663633](#)
- [12] HALL, P. and HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press. [MR0624435](#)
- [13] LATAŁA, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** 1273–1282 (electronic). [MR2111932](#)
- [14] NGUYEN, H.H. and VU, V. (2014). Random matrices: Law of the determinant. *Ann. Probab.* **42** 146–167. [MR3161483](#)
- [15] NYQUIST, H., RICE, S.O. and RIORDAN, J. (1954). The distribution of random determinants. *Quart. Appl. Math.* **12** 97–104. [MR0063591](#)
- [16] PETROV, V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Oxford Studies in Probability* **4**. New York: Clarendon Press. [MR1353441](#)
- [17] PRÉKOPA, A. (1967). On random determinants. I. *Studia Sci. Math. Hungar.* **2** 125–132. [MR0211439](#)
- [18] ROUAULT, A. (2007). Asymptotic behavior of random determinants in the Laguerre, Gram and Jacobi ensembles. *ALEA Lat. Am. J. Probab. Math. Stat.* **3** 181–230. [MR2365642](#)
- [19] TAO, T. and VU, V. (2006). On random ± 1 matrices: Singularity and determinant. *Random Structures Algorithms* **28** 1–23. [MR2187480](#)
- [20] TAO, T. and VU, V. (2008). Random matrices: The circular law. *Commun. Contemp. Math.* **10** 261–307. [MR2409368](#)

Received May 2013 and revised January 2014