THE LOGARITHMIC LAW OF RANDOM DETERMINANT

ZHIGANG BAO, GUANGMING PAN, AND WANG ZHOU

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

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

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

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

1. Introduction

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

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

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 [5], [3], [13], [15] for this topic. Besides, some lower and upper bounds for the magnitudes of random determinants were obtained in [2] and [17] 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 [9] found that $\det A_n^2$ can be written as a product of $n$ independent $\chi^2$ 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 $L_n = D_n D_n^T$.

Date: August 30, 2012.

2010 Mathematics Subject Classification. 15A15, 15B52, 60F05.

Key words and phrases. Logarithmic law, Random determinant, CLT for martingale.

Z.G. Bao was partially supported by NSFC grant 11071213, NSFC grant 11101362, ZJNSF grant R6090034 and SRFDP grant 20100010110001; G.M. Pan was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11; W. Zhou was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11, and by a grant R-155-000-116-112 at the National University of Singapore.
where
\[
D_n = \begin{pmatrix}
a_n & a_{n-1} & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
b_1 & a_1 & \cdots & \cdots \\
\end{pmatrix}.
\]
Here \(\{a_n, \cdots, a_1, b_{n-1}, \cdots, b_1\}\) is a collection of independent variables such that \(a_i \sim \chi_i, b_j \sim \chi_j\) for \(i = 1, \cdots, n, j = 1, \cdots, n-1\). Such a tridiagonal form is well known for Gaussian matrix. One can refer to [4] for instance. Then apparently one has
\[
\det A_n^2 = \det A_n A_n^T \overset{d}{=} \prod_{i=1}^{n} a_i^2,
\]
which implies that \(\log |\det A_n|\) can be represented by a sum of \(n\) independent random variables. Then by the elementary properties of \(\chi^2\) distributions, the following CLT can be obtained
\[
\frac{\log(|\det A_n|) - \frac{1}{2} \log(n-1)!}{\sqrt{n}} \overset{d}{\to} N(0, 1).
\]
For details, one can see [16] or the Appendix of [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 [8] (one can also refer to Girko’s books [6] and [7] 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 \(\delta\). Girko named (1.1) as “the logarithmic law” for random determinant. In [8], 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 [8] 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 \(t > 0,\)
\[
P(|a_{ij}| \geq t) \leq C_1 \exp(-t^{C_2})
\]
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 [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 [14] inserted a lot of new ingredients. Moreover, some unrigorous steps in [8] can be fixed by the methods provided in [14]. One may find that [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 4-th 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,
\[
\frac{\log(|\det A_n|) - \frac{1}{2} \log(n-1)!}{\sqrt{ \frac{1}{2} \log n }} \xrightarrow{d} N(0,1).
\]

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 $|| \cdot ||_{op}$ 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
\[
P\{\text{all square submatrices of } A_n' \text{ are invertible} \} = 1 \quad (2.1)
\]
and
\[
P\{|\log |\det A_n| - \log |\det A_n'|| = o(\sqrt{\log n}) \} = 1 - o(1). \quad (2.2)
\]

Remark 2.2. The construction of the modified matrix $A_n'$ can be found in [14]. The strategy is to set $a_{ij}' := (1 - e^{2})^{1/2}a_{ij} + e\theta_{ij}$, 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 $e$ to be extremely small, say $n^{-Kn}$ for large enough constant $K > 0$, it was shown in [14] that (2.2) holds. The proof in [14] relies on a lower bound estimate on the smallest singular value of square random matrices provided in Theorem 2.1 of [18]. To adapt to our condition, we will use Theorem 4.1 of [10] instead (by choosing $p_n = 1$ in [10]). 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 [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, \( \sup_n \max_{1 \leq i, j \leq n} E a_{ij}^4 < \infty \). 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 \( a_k^T \) 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 \( A(1) = a_1^T \) 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 \( \gamma_{i+1} \) be the distance from \( a_{i+1}^T \) to \( V_i \) for \( 1 \leq i \leq n-1 \).

And we set \( \gamma_1 = ||a_1^T||_2 \). Then by the “base times height” formula, we can write

\[
\det A_n^2 = \prod_{i=0}^{n-1} \gamma_{i+1}^2. \tag{2.3}
\]

Observe that \( \gamma_{i+1} \) is the norm of the projection of \( a_{i+1}^T \) onto \( V_i^\perp \). Thus we also have

\[
\gamma_{i+1}^2 = a_{i+1}^T P_i a_{i+1}, \quad 1 \leq i \leq n-1. \tag{2.4}
\]

Moreover, by the definition of \( V_i \) and assumption \( C_0 \) one has that with probability one \( A(i)A_i^T \) is invertible and

\[
P_i = I_n - A(i)^T (A(i)A_i^T)^{-1} A(i). \tag{2.5}
\]

Then a direct consequence of the definition of \( \gamma_{i+1} \) is

\[
E(\gamma_{i+1}^2 | P_i) = \text{tr} P_i = n - i, \quad 0 \leq i \leq n-1.
\]

It follows from (2.3) that

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

which yields

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

Now we set

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

And we write

\[
\log \frac{\gamma_{i+1}^2}{n - i} = X_{i+1} - \frac{X_{i+1}^2}{2} + R_{i+1}, \tag{2.7}
\]

where

\[
R_{i+1} := \log(1 + X_{i+1}) - (X_{i+1} - \frac{X_{i+1}^2}{2}).
\]
Then by (2.6), one can write
\[
\log \det A_n^2 - \log(n-1)! \quad \frac{1}{\sqrt{2 \log n}} n^{1-1} \sum_{i=0}^{n-1} X_{i+1} - \frac{1}{\sqrt{2 \log n}} \sum_{i=0}^{n-1} \frac{X_{i+1}^2}{2} - \log n + \frac{1}{\sqrt{2 \log n}} i \sum_{i=0}^{n-1} R_{i+1}.
\]

(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 \( \sigma \)-algebra generated by the first \( i \) rows of \( A_n \), by definition we have
\[
\mathbb{E}\{X_{i+1}\mid \mathcal{E}_i\} = 0.
\]

Thus \( X_1, \ldots, X_n \) is a martingale difference sequence with respect to the filtration \( \emptyset \subset \mathcal{E}_1 \subset \cdots \subset \mathcal{E}_{n-1} \). 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. 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, \( \gamma_{i+1}^2 \sim \chi^2_{n-i} \), thus
\[
X_{i+1} = O((n-i)^{-1/2}), \quad R_{i+1} = O((n-i)^{-3/2})
\]
with high probability. Especially, when \( n - i \) is \( O(1) \), the main term \( X_{i+1} \) and the negligible term \( R_{i+1} \) are comparable to be \( O(1) \). 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 \( X_{i+1} = O(1) \) with high probability, especially under the assumption of the 4-th moment. Fortunately, if all the \( a_{i+1}^T \)'s are Gaussian for large \( i \geq n - s_1 \) for some positive number \( s_1 \), \( \gamma_{i+1}^2 \) are independent \( \chi^2 \) 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, 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 [13]. To pursue this idea, we set
\[ s_1 = \lfloor \log^3 a n \rfloor \]
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):
\[ \frac{\sum_{i=0}^{n-s_1} X_{i+1}^2/2 - \log n}{\sqrt{2\log n}} \xrightarrow{p} 0. \]

(iii):
\[ \frac{\sum_{i=0}^{n-s_1} R_{i+1}}{\sqrt{2\log n}} \xrightarrow{p} 0. \]

(iv): If the last \( s_1 \) rows of \( A_n \) are Gaussian, then
\[ \frac{\sum_{i=n-s_1}^{n-1} \log \frac{\gamma_{i+1}^2}{n-i}}{\sqrt{2\log n}} \xrightarrow{p} 0. \]

(v): Let \( B_n \) be a random matrix satisfying the basic assumption \( C_0 \) and differing from \( A_n \) only in the last \( s_1 \) rows. Then one has
\[ \sup_x |P\{|\det A_n| \leq x\} - P\{|\det B_n| \leq x\}| \rightarrow 0. \]

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 [11] 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
\[ \max_i |Z_{ni}| \xrightarrow{p} 0, \]
\[ \sum_i Z_{ni}^2 \xrightarrow{p} 1. \]
Moreover, $\mathbb{E}(\max_i Z^2_{ni})$ is bounded in $n$. Then we have

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

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

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

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

and

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

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 $C_0$, we have for any constant $\epsilon > 0$

$$\sum_{i=0}^{n-s_1} P\left\{ \frac{1}{\sqrt{\log n}} |X_{i+1}| \geq \epsilon \right\} \xrightarrow{n \to \infty} 0$$

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 $trQ_i = 1$ we have

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

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

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

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

$$\sum_{i=0}^{n-s_1} P\left\{ \frac{1}{\sqrt{\log n}} |U_{i+1}| \geq \frac{\epsilon}{2} \right\} \xrightarrow{n \to \infty} 0$$

and

$$\sum_{i=0}^{n-s_1} P\left\{ \frac{1}{\sqrt{\log n}} |V_{i+1}| \geq \frac{\epsilon}{2} \right\} \xrightarrow{n \to \infty} 0.$$
\[
\sum_{i=0}^{n-s_1} \mathbb{P}\left( \frac{1}{\sqrt{\log n}} |V_{i+1}| \geq s \right) \to 0,
\]
which can be implied by
\[
\frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E}U_{i+1}^2 \to 0, \tag{3.4}
\]
and
\[
\frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E}V_{i+1}^4 \to 0. \tag{3.5}
\]

First, we verify (3.4). In the sequel, we set
\[
s_2 = \lfloor n \log^{-20a_n} \rfloor.
\]
By definition, we see
\[
E_{U_{i+1}}^2 = E\left( \sum_k q_{kk}(i)(a_{i+1,k}^2 - 1) \right) \leq C \sum_k q_{kk}(i) \cdot \log^{-8a_n} n.
\]
Using the basic fact \(0 \leq q_{kk}(i) \leq 1\) one has \(0 \leq q_{kk}(i) \leq 1/(n-i)\). Taking this fact and \(trQ_i = 1\) into account we have
\[
\frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E}U_{i+1}^2 \leq C \frac{1}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_k q_{kk}(i) \\
\leq C \frac{1}{\log n} \sum_{i=0}^{n-s_1} \frac{1}{n-i} + \frac{1}{\log n} \sum_{i=n-s_2}^{n-s_1} \mathbb{E} \max_k q_{kk}(i) \\
\leq C \frac{1}{\log n} \sum_{i=n-s_2}^{n-s_1} \mathbb{E} \max_k q_{kk}(i) + O\left(\frac{\log \log n}{\log n}\right) \\
\leq C \max_{n-s_2 \leq i \leq n-s_1} \mathbb{E} \max_k q_{kk}(i) + O\left(\frac{\log \log n}{\log n}\right).
\]
We will show the following estimate which will be used repeatedly in the sequel. That is
\[
\max_{n-s_2 \leq i \leq n-1} \mathbb{E} \max_k p_{kk}(i) \leq C \log^{-8a_n} n \tag{3.6}
\]
for some positive constant \(C\). To this end, we denote the \(j\)-th column of \(A(i)\) by \(b_j(i)\) and use the notation \(A(i,j)\) to denote the matrix induced from \(A(i)\) by deleting the \(j\)-th column \(b_j(i)\). Moreover, we set the positive parameter \(\alpha =: \alpha_n = n^{-1/6}\). By (2.5), we have
\[
p_{kk}(i) = 1 - b_k(i)^T (A(i) A_i^T)^{-1} b_k(i)
\]
It is apparent that 

\[ G = \sum_{i=k}^{n} k(i) \] 

Moreover, we have 

\[ \text{ease of presentation, when there is no confusion, we will omit the parameter } \alpha. \]

Then one has 

\[ G(i) = \frac{1}{\alpha} A(i) A^T(i) + \alpha I_i, \]

and 

\[ G(i,k) = \frac{1}{\alpha} A(i,k) A^T(i,k) + \frac{\alpha}{n} I_k. \]

Let 

\[ G(i,k) = \frac{1}{\alpha} A(i,k) A^T(i,k) + \frac{\alpha}{n} I_k, \]

Then one has 

\[ p_{k,k}(i) \leq \left( 1 + \frac{1}{n} b_k(i) G(i,k) b_k(i) \right)^{-1}. \]

Hence, to verify (3.6) we only need to show 

\[ E \max_k \left( 1 + \frac{1}{n} b_k(i) G(i,k) b_k(i) \right)^{-1} \leq C \log^{-8a} n, \quad n - s_2 \leq i \leq n - 1. \]  

(3.7)

It is apparent that \( G(i) \) and \( G(i,k) \) are positive-definite and 

\[ ||G(i)\|_{op}, ||G(i,k)\|_{op} \leq \alpha^{-1}. \]

Moreover, we have 

\[ \left| trG(i) - trG(i,k) \right| \]

\[ = \left| tr\left( \frac{1}{\alpha} A(i,k) A^T(i,k) + \frac{1}{n} b_k(i)b_k(i)^T + \alpha I_i \right) - tr\left( \frac{1}{\alpha} A(i,k) A^T(i,k) + \frac{\alpha}{n} I_k \right) \right| \]

\[ = \frac{1}{\alpha} b_k(i)^T G(i,k) b_k(i) \leq \alpha^{-1}. \]  

(3.8)

Now we set 

\[ \chi(i) = 1_{\{ \frac{1}{\alpha} trG(i) \geq \log^{10a} n \}}. \]

And we denote the \((u,v)\)-th entry of \( G(i,k) \) by \( G(i,k)(u,v) \) below. Moreover, for ease of presentation, when there is no confusion, we will omit the parameter \( \alpha \) from the notation \( G(i,k) \) and \( G(i) \). Then we have for some small constant 

\[ 0 < \epsilon < 1/2, \]

\[ E \max_k \left( 1 + \frac{1}{n} b_k(i) G(i,k) b_k(i) \right)^{-1} \]

\[ \leq \mathbb{P}\left\{ \frac{1}{n} trG(i) \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left( 1 + \frac{1}{n} b_k(i) G(i,k) b_k(i) \right)^{-1} \]

\[ \leq \mathbb{P}\left\{ \frac{1}{n} trG(i) \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left( 1 + \frac{1}{n} \sum_{j=1}^{i} G(i,k)(j,j)a_{jk}^2 - \epsilon \right)^{-1} \]

\[ + \mathbb{P}\left\{ \frac{1}{n} \max_{1 \leq k \leq n} \sum_{1 \leq u \neq v \leq i} G(i,k)(u,v)a_{uk}a_{vk} \geq \epsilon \right\} \]

\[ \leq \mathbb{P}\left\{ \frac{1}{n} trG(i) \leq \log^{10a} n \right\} + \mathbb{E} \chi(i) \max_k \left( 1 + \log^{-a} n \cdot \frac{1}{n} trG(i,k) - \epsilon \right)^{-1} \]
\[
+ C \sum_{k=1}^{n} \mathbb{P}\left\{ \sum_{j=1}^{i} G_{(i,k)}(j,j)a_{jk}^2 < \log^{-a} n \cdot \text{tr} G_{(i,k)} , \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\}
\]
\[
+ \mathbb{P}\left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{1 \leq u \neq v \leq i} G_{(i,k)}(u,v)a_{uk}a_{vk} \right| \geq \epsilon \right\} \leq \mathbb{P}\left\{ \frac{1}{n} \text{tr} G_{(i)} \leq \log^{10a} n \right\} + \mathbb{E} \chi(i)(1 + \log^{-a} n \cdot \frac{1}{n} \text{tr} G_{(i)} - 2\epsilon)^{-1}
\]
\[
+ C \sum_{k=1}^{n} \mathbb{P}\left\{ \sum_{j=1}^{i} G_{(i,k)}(j,j)a_{jk}^2 < \log^{-a} n \text{tr} G_{(i,k)} , \frac{1}{n} \text{tr} G_{(i)} \geq \log^{10a} n \right\}
\]
\[
+ \mathbb{P}\left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{1 \leq u \neq v \leq i} G_{(i,k)}(u,v)a_{uk}a_{vk} \right| \geq \epsilon \right\},
\] (3.9)

where in the above last inequality, we used (3.8). Below we will estimate (3.9) term by term. To this end, we need the following two lemmas, whose proofs will be given in Appendix A.

**Lemma 3.3.** Under the assumption of Theorem 1.1, for \( n - s_2 \leq i \leq n - 1 \), we have for \( \alpha = \frac{n - 1}{6} \)
\[
\mathbb{E}\left\{ \frac{1}{n} \text{tr} G_{(i)}(\alpha) \right\} = s_i(\alpha) + \mathcal{O}(n^{-1/6})
\]
and
\[
\text{Var}\left\{ \frac{1}{n} \text{tr} G_{(i)}(\alpha) \right\} = \mathcal{O}(n^{-1/3}),
\]
where
\[
s_i(\alpha) = 2 \left( \alpha + 1 - \frac{i}{n} + \sqrt{\left[ \alpha + (1 - \frac{i}{n})^2 + 4\alpha \frac{i}{n} \right]} \right)^{-1}.
\]

The next one is a simple deviation estimate for the quadratic form, whose proof is quite elementary.

**Lemma 3.4.** Suppose \( x_i, i = 1, \cdots, n \) are independent real random variables with common mean zero and variance 1. Moreover, we assume \( \max_i \mathbb{E}|x_i|^4 \leq \nu_l \). Let \( M_n = (m_{ij})_{n,n} \) be a nonnegative definite matrix which is deterministic. Then we have
\[
\mathbb{E}\left| \sum_{i=1}^{n} m_{ii}x_i^2 - \text{tr} M_n \right|^4 \leq C(\nu_l \text{tr} M_n^4 + (\nu_l \text{tr} M_n^2)^2)
\] (3.10)
and
\[
\mathbb{E}\left| \sum_{u \neq v} m_{uv}x_u x_v \right|^4 \leq C\nu_l^2 (\text{tr} M_n^2)^2
\] (3.11)
for some positive constant \( C \).
Consequently, for the second term of (3.9), with the definition of \( s_i(\alpha) \), we have
\[
\Pr \left\{ \frac{1}{n} \text{tr} G_i \leq \log^{10a} n \right\} \leq \Pr \left\{ \left| \frac{1}{n} \text{tr} G_i - \frac{1}{n} \text{tr} G_i \right| \geq \frac{1}{20} \log^{20a} n \right\} 
\leq C \log^{-40a} n \text{Var} \left\{ \frac{1}{n} \text{tr} G_i \right\} = o(n^{-1/3}).
\]
For the second term of (3.9), with the definition of \( \chi(i) \), obviously one has
\[
\mathbb{E} \chi(i)(1 + \log^{-a} n \cdot \frac{1}{n} \text{tr} G_i - 2e)^{-1} \leq C \log^{-8a} n.
\]
Now we deal with the third term of (3.9). We set
\[
\tilde{a}_{jk} = a_{jk} 1_{\{|a_{jk}| \leq \log^a n\}}, \quad \hat{a}_{jk} = \frac{\tilde{a}_{jk} - \mathbb{E} \tilde{a}_{jk}}{\sqrt{\text{Var} \{\tilde{a}_{jk}\}}},
\]
Since \( G(i,k)(j,j) \)'s are positive and \( \hat{a}_{jk}^2 \leq a_{jk}^2 \), one has
\[
\sum_{k=1}^{n} \Pr \left\{ \sum_{j=1}^{i} G(i,k)(j,j) a_{jk}^2 \leq \log^{-a} n \cdot \text{tr} G(i,k), \frac{1}{n} \text{tr} G_i \geq \log^{10a} n \right\} 
\leq \sum_{k=1}^{n} \Pr \left\{ \sum_{j=1}^{i} G(i,k)(j,j) \hat{a}_{jk}^2 \leq 3 \log^{-a} n \cdot \text{tr} G(i,k), \frac{1}{n} \text{tr} G_i \geq \log^{10a} n \right\}.
\]
Moreover, by the assumption \( \sup_n \max_{ij} \mathbb{E} a_{ij}^4 < \infty \) it is easy to derive that
\[
\mathbb{E} \hat{a}_{jk} = O(\log^{-3a} n), \quad \text{Var} \{\hat{a}_{jk}\} = 1 + O(\log^{-2a} n).
\]
Consequently,
\[
\tilde{a}_{jk} = \hat{a}_{jk} + O(\log^{-a} n),
\]
which implies
\[
\hat{a}_{jk}^2 \leq 2\hat{a}_{jk}^2 + O(\log^{-2a} n) \leq 2\hat{a}_{jk}^2 + \log^{-a} n
\]
for sufficiently large \( n \). Therefore, we have
\[
\sum_{k=1}^{n} \Pr \left\{ \sum_{j=1}^{i} G(i,k)(j,j) a_{jk}^2 \leq \log^{-a} n \cdot \text{tr} G(i,k), \frac{1}{n} \text{tr} G_i \geq \log^{10a} n \right\} 
\leq \sum_{k=1}^{n} \Pr \left\{ \sum_{j=1}^{i} G(i,k)(j,j) \hat{a}_{jk}^2 \leq 3 \log^{-a} n \cdot \text{tr} G(i,k), \frac{1}{n} \text{tr} G_i \geq \log^{10a} n \right\}.
\]
\[ \leq C \log^{-40a} n \cdot n^{-4} \sum_{k=1}^{n} \mathbb{E}[\log^{4a} n \cdot trG_{(i,k)}^4 + (trG_{(i,k)}^2)^2] \]
\[ \leq o\left(\frac{1}{n^{\alpha}}\right). \]

In the fourth inequality we used the fact (3.8) and in the fifth inequality we used (3.10) and the fact
\[ \mathbb{E}|\tilde{a}_{ij}|^{4+t} = O(\log^{ta} n) \]
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 (3.9). Note that by (3.11)
\[ P \left\{ \frac{1}{n} \max_{1 \leq k \leq n} \left| \sum_{u \neq v} G_{(i,k)}(u,v)a_{uk}a_{vk} \right| \geq \epsilon \right\} \]
\[ \leq \sum_{k=1}^{n} P \left\{ \frac{1}{n} \sum_{u \neq v} G_{(i,k)}(u,v)a_{uk}a_{vk} \geq \epsilon \right\} \]
\[ \leq \epsilon^{-4} n^{-4} \sum_{k=1}^{n} \mathbb{E}\left( \sum_{u \neq v} G_{(i,k)}(u,v)a_{uk}a_{vk} \right)^4 \]
\[ \leq C \epsilon^{-4} n^{-4} \sum_{k=1}^{n} \mathbb{E}(trG_{(i,k)}^2)^2 \]
\[ \leq O\left(\frac{1}{n^{\alpha}}\right). \]

Therefore, (3.7) follows from the above estimates, so does (3.6). Consequently, we have (3.4).

Now we begin to show (3.5). By (3.1.1.3) one has
\[ \mathbb{E}V_{i+1}^4 \leq C \mathbb{E}(trQ_i^2)^2 = C \frac{1}{(n-i)^2}, \]
which implies that
\[ \frac{1}{4 \log^2 n} \sum_{i=0}^{n-s_1} \mathbb{E}V_{i+1}^4 = O(\log^{-2+3a} n). \]

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

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 3.5.** Under the assumption \( C_0 \), one has
\[ \frac{\sum_{i=0}^{n-s_1} X_{i+1}^2}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0, \]
(3.13)
and
\[ \frac{1}{2 \log n} \sum_{i=0}^{n-s_1} \mathbb{E}X_{i+1}^2 = 1 + o(1). \] (3.14)

**Proof.** We begin with (3.13). We split the proof of (3.13) into two steps.
\[ \sum_{i=0}^{n-s_1} \left( X_{i+1}^2 - \mathbb{E}\{X_{i+1}^2|\mathcal{E}_i\} \right) \rightarrow 0, \] (3.15)
and
\[ \sum_{i=0}^{n-s_1} \mathbb{E}\{X_{i+1}^2|\mathcal{E}_i\} - 2 \log n \rightarrow 0. \] (3.16)

Observe that
\[ \mathbb{E}\{X_{i+1}^2|\mathcal{E}_i\} = \mathbb{E}\{( \sum_{j,k=1}^n q_{jk}(i)a_{i+1,j}a_{i+1,k} - 1)^2|\mathcal{E}_i\} \]
\[ = \frac{2}{n-i} + \sum_k q_{kk}(i)^2(\mathbb{E}|a_{i+1,k}|^4 - 3) \] (3.17)

Therefore, to verify (3.16), we only need
\[ \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \sum_k q_{kk}(i)^2 \rightarrow 0. \] (3.18)

Note that from the proof of (3.4) we can get the following estimate directly.
\[ \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \sum_k q_{kk}(i)^2 = \mathcal{O}\left(\frac{\log n}{\sqrt{\log n}}\right). \] (3.19)

Thus it suffices to show (3.15). By elementary calculations, we have
\[ X_{i+1}^2 - \mathbb{E}\{X_{i+1}^2|\mathcal{E}_i\} \]
\[ = -2 \sum_u q_{uu}(i)(a_{i+1,u}^2 - 1) + 2 \sum_{u \neq v} q_{uv}(i)q_{vv}(i)(a_{i+1,u}a_{i+1,v}^2 - 1) \]
\[ + 2 \sum_{u \neq v} q_{uv}(i)^2(a_{i+1,u}^4 - 2a_{i+1,u}a_{i+1,v}^2 - 1) \]
\[ + 2 \sum_{u_1 \neq v_1, u_2 \neq v_2} q_{u_1v_1}(i)q_{u_2v_2}(i)a_{i+1,u_1}a_{i+1,v_1}a_{i+1,u_2}a_{i+1,v_2} \]
\[ + \sum_u q_{uu}(i)^2(a_{i+1,u}^4 - \mathbb{E}a_{i+1,u}^4) + 2(\sum_u q_{uu}(i)(a_{i+1,u}^2 - 1))\left(\sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v}\right) \]
\[ =: 2W_1(i) + 2W_2(i), \]
where
\[ W_1(i) = -\sum_i q_{uu}(i)(a_{i+1,u}^2 - 1) + \sum_{u \neq v} q_{uv}(i)q_{vv}(i)(a_{i+1,u}a_{i+1,v}^2 - 1) \]
By (3.19), we have that the first term of (3.21) is of the order $O_{\log}$.

First, we deal with the second statement of (3.20). Note that

$$\sum_{u_1 \neq v_1, u_2 \neq v_2} q_{u_1 v_1}(i)q_{u_2 v_2}(i)a_{i+1,u_1}a_{i+1,v_1}a_{i+1,u_2}a_{i+1,v_2},$$

and

$$W_2(i) = \frac{1}{2} \sum_u q_{uu}(i)^2(a_{i+1,u} - E a_{i+1,u}^2) + (\sum_u q_{uu}(i)(a_{i+1,u}^2 - 1))(\sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v}).$$

We split the issue to show

$$\frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_1(i) \xrightarrow{p} 0, \quad \frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_2(i) \xrightarrow{p} 0. \quad (3.20)$$

First, we deal with the second statement of (3.20). Note that

$$\frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} E|W_2(i)| \leq C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} E \sum_u q_{uu}(i)^2
\quad + C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} (E \sum_u q_{uu}(i)(a_{i+1,u}^2 - 1))^2/2(E \sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v})^{1/2}
\quad \leq C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} E \sum_u q_{uu}(i)^2 + C \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} (1 - n^{-1})^{1/2}(E \sum_u q_{uu}(i)^2)^{1/2}. \quad (3.21)$$

By (3.19), we have that the first term of (3.21) is of the order $O(\log \log n/\sqrt{\log n})$.

For the second term, by using (3.6) we have

$$\frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_1} \frac{1}{n-i}^{1/2}(E \sum_u q_{uu}(i)^2)^{1/2}
\quad \leq \frac{1}{\sqrt{\log n}} \sum_{i=0}^{n-s_2} \frac{1}{n-i} + \frac{1}{\sqrt{\log n}} \sum_{i=n-s_2}^{n-s_1} \frac{1}{n-i}E \sum_u q_{uu}(i)^2)^{1/2}
\quad \leq \frac{1}{\sqrt{\log n}} \sum_{i=n-s_2}^{n-s_1} \frac{1}{n-i}E \max_u p_{uu}(i)^{1/2} + O(\frac{\log \log n}{\sqrt{\log n}})
\quad = O(\frac{\log \log n}{\sqrt{\log n}}).$$

Now we consider the first term of (3.20). It is easy to see

$$E\{W_1(i)\} = 0, \quad E\{W_1(i)W_1(j)\} = 0, \quad i \neq j.$$
which yields

\[
E\left(\frac{1}{\sqrt{\log n}} \sum_{i=0}^{s_1} W_1(i)\right)^2 = \frac{1}{\log n} \sum_{i=0}^{s_1} EW_1(i)^2
\]

\[
\leq \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_u q_{uu}(i)^2 + \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u \neq v, u \neq w} q_{uu}(i)^2 q_{vv}(i)q_{ww}(i)
\]

\[
+ \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u \neq v, u \neq w} q_{uv}(i)^2 q_{uw}(i)^2
\]

\[
+ \frac{C}{\log n} \sum_{i=0}^{n-s_1} \mathbb{E} \sum_{u \neq v_1, u \neq v_2} |q_{u_1v_1}(i)q_{u_1v_2}(i)q_{v_2u_1}(i)q_{u_2v_2}(i)|.
\]

(3.22)

The estimation of (3.22) is elementary but somewhat tedious. In fact, one can find the estimate towards every term of (3.22) in [14] (see the estimation of \(\text{Var}(\sum_{i=0}^{s_1} Y_{i+1})\) in Section 6 of [14]). Here we omit the details and claim the following estimation

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

Then (3.15) follows, so does (3.13). Moreover, it is easy to see that (3.14) holds by combining (3.17) and (3.18). Therefore, Lemma 3.5 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 \(X_{i+1} \geq -1 + \log^{-\delta} n\), 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 \(\delta\).

**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 \(-1 + \log^{-\delta} n \leq X_{i+1} < 1 + \varepsilon\) 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 $X_{i+1} > 1 - \epsilon$, 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 $\max\{U_{i+1}, V_{i+1}\} > \frac{1}{2} - \frac{\epsilon}{2}$. Finally we deal with the last case. Note that when
\[-1 + \log^{-c} n \leq X_{i+1} < -1 + \epsilon,\]
we have
\[|R_{i+1}| \leq C \log \log n.\]
Moreover, it is obvious that we have
\[\max\{|U_{i+1}|, |V_{i+1}|\} > \frac{1}{2} - \frac{\epsilon}{2}.\]
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 \(\bigcup_{i=0}^{n-s_1} \{X_{i+1} < -1 + \log^{a/2} n\}\).

**Lemma 4.2.** Under the assumption $C_0$, we have
\[\sum_{i=0}^{n-s_1} \Pr\{X_{i+1} < -1 + \log^{-\frac{a}{2}} n\} \rightarrow 0\]
as $n$ tends to infinity.

**Proof.** Note that
\[\Pr\{X_{i+1} < -1 + \log^{-\frac{a}{2}} n\} = \Pr\{a_{i+1}^T Q a_{i+1} < \log^{-\frac{a}{2}} n\} \leq \Pr\{\sum_k q_{kk}(i)a_{i+1,k}^2 < 2 \log^{-\frac{a}{2}} n\} + \Pr\{|\sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v}| \geq \frac{1}{2} \log^{-\frac{a}{2}} n\}.\]

Now we recall the definition
\[\hat{a}_{i+1,k} = a_{i+1,k} 1_{\{|a_{i+1,k}| \leq \log n\}}, \quad \tilde{a}_{i+1,k} = \frac{\hat{a}_{i+1,k} - \mathbb{E}\hat{a}_{i+1,k}}{\sqrt{\text{Var}\{\hat{a}_{i+1,k}\}}}.\]

A similar discussion as that in the last section yields
\[\Pr\{\sum_k q_{kk}(i)a_{i+1,k}^2 < 2 \log^{-\frac{a}{2}} n\} \leq \Pr\{\sum_k q_{kk}(i)\tilde{a}_{i+1,k}^2 < C \log^{-\frac{a}{2}} n\}\]
\[\leq \Pr\{|\sum_k q_{kk}(i)\tilde{a}_{i+1,k}^2 - 1| \geq \frac{1}{2}\}\]
\[\leq C(\log^{4a} n \cdot \text{tr} Q_i^4 + (\text{tr} Q_i^2)^2)\]
\[\leq C(\log^{4a} n \cdot (n - i)^{-3} + (n - i)^{-2}).\]

Here in the third inequality we used \((3.10)\) again. Moreover, by \((3.11)\) we obtain
\[\Pr\{|\sum_{u \neq v} q_{uv}(i)a_{i+1,u}a_{i+1,v}| \geq \frac{1}{2} \log^{-\frac{a}{2}} n\} \leq C(n - i)^{-2} \log^{2a} n.\]
Thus finally we have

\[
\sum_{i=0}^{n-s_1} \mathbb{P}\{X_{i+1} < -1 + \log^{-\frac{2}{a}} n\} \\
\leq C(\log^{2a} n \sum_{i=0}^{n-s_1} (n-i)^{-3} + \log^{2a} n \sum_{i=0}^{n-s_1} (n-i)^{-2}) \\
\leq C \log^{-a} n.
\]

Therefore, we complete the proof. \qed

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

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

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

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

Note that

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

Moreover, we have

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

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 \(\gamma_{i+1}^2, i = n-s_1, \ldots, n-1\) is an independent sequence and \(\gamma_{i+1}^2 \sim \chi_n^2\). One may refer to Section 7 of [14] for instance. By using the Laplace transform trick, Nguyen and Vu [14] showed that for \(0 < c < 100\),

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

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.

**Lemma 5.1.** Let $Z_1, \cdots, Z_m$ be independent real random variables with mean zero. Assume that

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

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

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

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 $C_0$ 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 $\bar{a}_n^T$ and $\bar{\bar{a}}_n^T$ are independent. Here we use $a_n^T$ and $\bar{a}_n^T$ to denote the $n$-th row of $A_n$ and $\bar{A}_n$ respectively as above. Below we use the notation $\alpha_{ni}$ 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

$$\sup_x \left| \mathbb{P}\left\{ \frac{\det A_n}{\Delta} \leq x | \mathcal{E}_{n-1} \right\} - \Phi(x) \right| \leq C \sum_{i=1}^n \frac{|\alpha_{ni}|^3}{\Delta^3}.$$
Therefore, we have
\[
\sup_x \left| \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \right\} - \Phi(x) \right|
= \sup_x \left| \mathbb{E} \left\{ \mathbb{P} \left\{ \frac{\det A_n}{\Delta} \leq x \mid \mathcal{E}_{n-1} \right\} \right\} - \Phi(x) \right|
\leq C \mathbb{E} \sum_{k=1}^{n} \frac{|\alpha_{nk}|^3}{\Delta^3}.
\]
For simplicity, we will briefly denote \(b_k(n-1)\) and \(A_{(n-1),k}\) by \(b_k\) and \(A_{nk}\) respectively in the sequel. Then by the definitions of cofactors and the Cauchy-Binet formula, we have
\[
\left( \frac{\alpha_{nk}}{\Delta} \right)^2 = \frac{\det A_{nk}^2}{\det A_{(n-1)A_{(n-1)}}} = \det A_{nk}^T(A_{nk}A_{nk}^T + b_kb_k^T)^{-1}A_{nk}
= \det (I_{n-1} + A_{nk}^{-1}b_kb_k^T(A_{nk}^{-1})^{-1})^{-1} = (1 + b_k^T(A_{nk}A_{nk}^{-1})b_k)^{-1}.
\]
Moreover, one always has
\[
\mathbb{E} \sum_{k=1}^{n} \frac{|\alpha_{nk}|^3}{\Delta^3} \leq \mathbb{E} \max_{k=1, \ldots, n} \frac{|\alpha_{nk}|}{\Delta} = \mathbb{E} \max_{k=1, \ldots, n} \left( 1 + b_k^T(A_{nk}A_{nk}^{-1})b_k \right)^{-1/2}.
\]
Recall the definition
\[
G_{(n-1),k}(\alpha) = \left( \frac{1}{n} A_{nk}A_{nk}^T + \alpha I_{n-1} \right)^{-1}, \quad G_{(n-1)}(\alpha) = \left( \frac{1}{n} A_{(n-1)A_{(n-1)}} + \alpha I_{n-1} \right)^{-1}.
\]
By using (3.7) we obtain
\[
\mathbb{E} \max_{k=1, \ldots, n} \left( 1 + b_k^T(a_{nk}A_{nk}^{-1})b_k \right)^{-1/2}
\leq \mathbb{E} \max_{k=1, \ldots, n} \left( 1 + \frac{1}{n} b_k^T G_{(n-1),k}(\alpha) b_k \right)^{-1/2}
\leq \left( \mathbb{E} \max_{k=1, \ldots, n} \left( 1 + \frac{1}{n} b_k^T G_{(n-1),k}(\alpha) b_k \right)^{-1} \right)^{1/2}
\leq \mathcal{O}(\log^{-4a} n).
\]
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
\[
\sup_x \left| \mathbb{P}\left\{ \frac{\log \det A_n^2 - \log(n-1)!}{\sqrt{2 \log n}} \leq x \right\} - \mathbb{P}\left\{ \frac{\log \det \overline{A}_n^2 - \log(n-1)!}{\sqrt{2 \log n}} \leq x \right\} \right| \leq C \log^{-4a} n.
\]
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 3.3 and Lemma 3.4. The proof is intrinsically the same as the counterpart in [1]. For convenience of the reader, we sketch it below. For ease of notation, we represent Lemma 3.3 as follows

**Lemma 5.2.** 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 \( G(\alpha) = \left( \frac{1}{n} XX^T + \alpha \right)^{-1} \), we have for \( \alpha = n^{-1/6} \)

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

and

\[
\text{Var}\left\{ \frac{1}{n} \text{tr} G(\alpha) \right\} = \mathcal{O}(n^{-1/3})
\]

where

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

**Proof.** For convenience, we set

\[
r_p(\alpha) = \frac{1}{p} \mathbb{E} \text{tr} G(\alpha) = \frac{1}{p} \mathbb{E} \sum_{k=1}^{p} \frac{1}{n} x_k x_k^T + \alpha - \frac{1}{n} x_k X(k)^T \left( \frac{1}{n} X(k) X^T(k) + \alpha I_{p-1} \right)^{-1} X(k) x_k^T.
\]

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

\[
r_p(\alpha) = \frac{1}{p} \mathbb{E} \sum_{k=1}^{p} \frac{1}{n} \epsilon_k + 1 + \alpha - y_p + y_p \alpha r_p(\alpha) \]

\[
= \frac{1}{1 + \alpha - y_p + y_p \alpha r_p(\alpha)} + \delta,
\]

where

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

and

\[
\delta = \frac{1}{p} \mathbb{E} \left( \frac{\epsilon_k}{1 + \alpha - y_p + y_p \alpha r_p(\alpha)} \right)(1 + \alpha - y_p + y_p \alpha r_p(\alpha) + \epsilon_k).
\]
\[
\begin{align*}
&= -\frac{1}{p} \sum_{k=1}^{p} \mathbb{E} \left( \frac{\epsilon_k}{(1 + \alpha - y_p + y_p \alpha r_p(\alpha))^2} \right) \\
&\quad - \frac{1}{p} \sum_{k=1}^{p} \mathbb{E} \left( \frac{\epsilon_k^2}{(1 + \alpha - y_p + y_p \alpha r_p(\alpha))^2(1 + \alpha - y_p + y_p \alpha r_p(\alpha) + \epsilon_k)} \right).
\end{align*}
\]  
(5.2)

From (5.1) we can get
\[ r_p(\alpha) = \frac{1}{2y_p \alpha} \left( \sqrt{(1 + \alpha - y_p - y_p \alpha \delta)^2 + 4y_p \alpha} - (1 + \alpha - y_p - y_p \alpha \delta) \right). \]  
(5.3)

It is not difficult to see that
\[ \frac{1}{1 + \alpha - y_p + y_p \alpha r_p(\alpha) + \epsilon_k} \leq \alpha^{-1}. \]  
(5.4)

By (5.2) and (5.4), we can get
\[ |\delta| \leq \frac{1}{p} \sum_{k=1}^{p} (|\mathbb{E} \epsilon_k| + \alpha^{-1} \mathbb{E} \epsilon_k^2)(1 + \alpha - y_p + y_p \alpha r_p(\alpha))^2. \]  
(5.5)

First, note that
\[
|\mathbb{E} \epsilon_k| = |\mathbb{E}(y_p - y_p \alpha 1_n tr G(\alpha) - \frac{1}{n^2} tr(\frac{1}{n} X(k)X^T(k) + \alpha I_{p-1})^{-1} \frac{1}{n} X(k)X^T(k))| \\
= \alpha \frac{1}{n} |\mathbb{E}(tr(X^T + \alpha I_p)^{-1} - tr(X(k)X^T(k) + \alpha I_{p-1}))| + \frac{1}{n} \\
= \mathcal{O}(\frac{1}{n}).
\]  
(5.6)

Next, we come to estimate \( \mathbb{E} \epsilon_k^2 \). Note that
\[
\mathbb{E} \epsilon_k^2 = \text{Var}\{\epsilon_k\} + (\mathbb{E} \epsilon_k)^2 \leq \frac{C}{n} + T_1 + T_2,
\]  
(5.7)

where
\[
T_1 = \mathbb{E} \left| \frac{1}{n^2} x_k X^T(k) \left( \frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} \frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right| x_k^T \\
- \mathbb{E} \left( \frac{1}{n^2} x_k X^T(k) \left( \frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} X(k) x_k^T \right)^2,
\]
and
\[
T_2 = \frac{\alpha^2}{n^2} \mathbb{E} \left| tr \left( \frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} - tr \left( \frac{1}{n} X(k)X^T(k) + \alpha I_{p-1} \right)^{-1} \right|^2.
\]  
(5.8)

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
\[
T_1 \leq \frac{C}{n^2} \mathbb{E} tr \Gamma_k^2 \leq \frac{C}{n \alpha^2}.
\]  
(5.9)
Let $\mathbb{E}_d$ be the conditional expectation given \{\(x_{ij}, d + 1 \leq i \leq p, 1 \leq j \leq n\}\). Define
\[
\gamma_d(k) = \mathbb{E}_d - 1 \text{tr}\left( \frac{1}{n}X(k)X^T(k) + \alpha I_{p-1} \right) - 1
\]
where
\[
\sigma_d(k) = \text{tr}\left( \frac{1}{n}X(k)X^T(k) + \alpha I_{p-1} \right) - \text{tr}\left( \frac{1}{n}X(k,d)X^T(k,d) + \alpha I_{p-2} \right).
\]
Noting that $|\sigma_d(k)| \leq \alpha^{-1}$, one has
\[
T_2 \leq Cn^2 p \sum_{d=1}^{p} E|\gamma_d^2(k)| \leq C n\alpha^2.
\] (5.10)
Combining (5.7)-(5.10) we can get
\[
E \delta^2 \leq C n\alpha^2.
\] (5.11)
Substituting (5.6), (5.11) and the basic fact
\[
(1 + \alpha - y_p + yar_p(\alpha))^2 \leq \alpha^{-2}
\]
into (5.5) one has
\[
|\delta| \leq C n\alpha^5.
\]
Then by (5.3) one has
\[
\begin{align*}
\rho_p(\alpha) &= \frac{1}{2y_p\alpha}\left(\sqrt{(1 + \alpha - y_p)^2 + 4y_p\alpha - (1 + \alpha - y_p)}\right) + O(n^{-1}\alpha^{-5}) \\
&= 2\left(\sqrt{(1 + \alpha - y_p)^2 + 4y_p\alpha + (1 + \alpha - y_p)^{-1}}\right) + O(n^{-1}\alpha^{-5}).
\end{align*}
\]
Moreover, similar to the estimate towards (5.8), we can get that
\[
	ext{Var}\left\{\frac{1}{n}\text{tr}G(\alpha)\right\} \leq C \frac{n}{\alpha^4}.
\]
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
\[
\begin{align*}
\mathbb{E}\left| \sum_{i=1}^{n} m_{ii}(x_i^2 - 1) \right|^4 &= \sum_{j,k,u,v} m_{jj}m_{kk}m_{uu}m_{vv} \mathbb{E}(x_j^2 - 1)(x_k^2 - 1)(x_u^2 - 1)(x_v^2 - 1) \\
&\leq C(\nu_8 \sum_j m_{jj}^4 + \nu_4^2 \sum_{j \neq k} m_{jj}^2 m_{kk}^2) \\
&\leq C(\nu_8 \text{tr}M_n^4 + (\nu_4 \text{tr}M_n^2)^2).
\end{align*}
\]
For the off-diagonal part, we have

\[
\mathbb{E}\left| \sum_{u \neq v} m_{uv}x_u x_v \right|^4 \leq \sum_{u_i \neq v_i, i=1, \ldots, 4} \prod_{i=1}^4 m_{u_i,v_i} \mathbb{E}\prod_{i=1}^4 x_{u_i} x_{v_i}
\]

\[
\leq C\left( \nu_2^4 \sum_{u \neq v} m_{uv}^4 + \nu_2^2 \sum_{u,v,r} m_{uv} m_{ur} m_{rv} + \nu_2^4 \sum_{u,v,r,w} m_{uv} m_{ur} m_{rw} m_{wu} \right)
\]

\[
\leq C\nu_2^4 \left( (trM_n^2)^2 + trM_n^2 \cdot (trM_n^1)^{1/2} + trM_n^4 \right)
\]

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 [14]. First, we need the following lemma derived from Theorem 4.1 in [10]. Let \( s_1(W_n) \leq s_{n-1}(W_n) \leq \cdots \leq s_1(W_n) \) be the ordered singular values of an \( n \times n \) matrix \( W_n \).

**Lemma 5.3.** Under the assumptions of Theorem 1.1, there exist some positive constants \( c, C, L \) such that

\[
\mathbb{P}\{ s_n(A_n) \leq n^{-L}; s_1(A_n) \leq n \} \leq e^{-cn} + C\sqrt{\log n} \frac{1}{\sqrt{n}}.
\]

**Remark 5.4.** It is easy to get Lemma 5.3 from Theorem 4.1 of [10] by choosing \( p_n = 1 \). We also remark here Theorem 4.1 of [10] are stated for more general case under weaker moment assumption. For convenience, we just restate it under our setting.

Now by the result of [12], it is easy to see under the assumption of Theorem 1.1

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

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 5.3 we obtain

\[
\mathbb{P}\{ s_n(A_n) \geq n^{-L} \} = 1 - O\left( \frac{\sqrt{\log n}}{\sqrt{n}} \right). \tag{5.12}
\]

Now let \( \theta_0 \) 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 \( \theta_0 \). And we set \( A'_n = (a'_{ij}) \), where \( a'_{ij} = (1 - \epsilon_n^2)^{1/2} a_{ij} + \epsilon_n \theta_{ij} \). Here we choose \( \epsilon_n = n^{-((100+2L)n)} \) (say). Writing \( \Theta_n = (\theta_{ij})_{n,n} \), then by Weyl’s inequality, one has

\[
|s_i(A'_n) - (1 - \epsilon_n^2)^{1/2} s_i(A_n)| \leq \epsilon_n \|\Theta_n\|_{op} \leq Cn^{-(99+2L)n}.
\]
Therefore, by (5.12) we have with probability $1 - \frac{\sqrt{\log n}}{\sqrt{n}}$

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

which implies (2.2). Moreover, by the construction, since $\theta_0$ is a continuous variable, it is obvious that (2.1) holds. Thus we may complete the proof.

**References**

[1] Z.D. Bai, Convergence rate of expected spectral distributions of large random matrices. Part II. sample covariance matrices. Ann. Probab. 21, No. 2, 649-672. (1993)

[2] K.P. Costello, V. Vu, Concentration of random determinants and permanent estimators. SAIM J. Discrete Math. V 23, No. 3, 1356-1371. (2009)

[3] A. Dembo, On random determinants. Quart. Appl. Math. 47, No.2, 185-195. (1989)

[4] I. Dumitriu, A Edelman, Matrix models for Beta Ensembles. J. Math. Phys. 43, 5830. (2002)

[5] G.E. Forsythe, J.W. Tukey. The extent of $n$ random unit vectors. Bull. Amer. Math. Sot. 58, 502. (1952)

[6] V.L. Girko, Theory of random determinants, Kluwer Academic Publishers, (1990)

[7] V.L. Girko, Theory of linear algebraic equations with random coefficients, Allerton Press, (1996)

[8] V.L. Girko, A refinement of the central limit theorem for random determinants (Russian), translation in Theory Probab. Appl. 42, No. 1, 121-129. (1998)

[9] N.R. Goodman, Distribution of the determinant of a complex Wishart distributed matrix, Annal of Statistics, 34, 178-180. (1963)

[10] F. Götze, A. Tikhomirov, The Circular law for random matrices. Ann. of Probab. V.38, No.4, 1444-1491. (2010)

[11] P. Hall, C.C. Heyde, Martingale limit theory and its application. New York : Academic Press. (1980)

[12] R. Latała, Some estimates of norms of random matrices, Proc. Amer. Math. Soc. 133, 1273-1282. (2005)

[13] H. Nyquist, S.O. Rice and J. Riordan, The distribution of random determinants. Quart. Appl. Math. 12, 97-104. (1954)

[14] H. Nguyen, V. Vu, Random matrices: law of the determinant, accepted to Ann. of Probab., arXiv:1112.0752v2. (2011)

[15] A. Prékopa, On random determinants I, Studia Sci. Math. Hungar. 2, 125-132. (1967)

[16] A. Rouault, Asymptotic behaviour of random determinants in the laguerre, Gram and Jacobi ensembles, Latin American Journal of Probability and Mathematical Statistics (ALEA), 3, 181-230. (2007)

[17] T. Tao, V. Vu, On random ±1 matrices: singularity and determinant, Random Structure Algorithms, 28,1-23. (2006)

[18] T. Tao, V, Vu, Random matrices: The Circular law, Communication in Contemporary Mathematics, 10, 261-307. (2008)
