Several Applications of the Moment Method in Band Random Matrix Model

December 12, 2017

1 Introduction

Random matrix theory (RMT) has wide applications in various areas of science. It can be traced back to sample covariance matrices studied by J. Wishart in data analysis in 1920s-1930s. In 1951, E. Wigner associated the energy levels of heavy-nuclei atoms with Hermitian matrices with i.i.d entries. To simplify the complex Hamiltonians of heavy-nuclei atoms, Wigner introduced the ensemble of real symmetric random matrices in 1950. The ensemble is now known as real Wigner matrices:

**Definition 1.** Real Wigner matrix $A_N$ is a symmetric matrix $\in \mathbb{R}^{N \times N}$ such that all entries are (real) independent random variables with mean zero. The off diagonal entries has variance one while the diagonal components has finite second moment. For some technical reasons, assume that all the random variables have finite moments.

To study the energy levels of heavy-nuclei atoms, which is conjectured to be statistically similar to the eigenvalues of Wigner matrices, it’s natural to study:

**Definition 2.** The empirical spectral distribution (ESD) of the scaled Wigner matrix $A_N \sqrt{N}$ is defined as $\mu_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\frac{\lambda_i}{\sqrt{N}}}$ where $\lambda_i$ represents the $i^{th}$ eigenvalue of $A$ in increasing order. Thus,

$$\int f d\mu_N = \frac{1}{N} \sum_{i=1}^{N} f \left( \frac{\lambda_i}{\sqrt{N}} \right).$$

Here $f$ is any continuous function compactly supported in $\mathbb{R}$, i.e. $f \in C_c(\mathbb{R})$. $f$ is called test function. We can also define the expected empirical spectral distribution $E\mu_N$ as

$$\int f dE\mu_N = \frac{1}{N} E \sum_{i=1}^{N} f \left( \frac{\lambda_i}{\sqrt{N}} \right).$$
Definition 3. We say a sequence of random probability measure $\mu_N$ converge in probability(or almost sure) to a deterministic probability measure $\mu$, $\mu_N \to \mu$, if for every test function $f \in C_c(\mathbb{R})$, as $N \to \infty$, we have in probability (or almost sure),

$$\int_{\mathbb{R}} f d\mu_N \to \int_{\mathbb{R}} f d\mu.$$ 

Wigner advanced the moment method to prove the following Semicircle Law in his paper [33].

Theorem 4. (Wigner’s Semicircle Law) For Wigner matrices $A_N$ defined in Definition 1, $\mu_N(A_N) \to \mu_{sc}$ almost sure as $N \to \infty$, where

$$\mu_{sc} := \frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}_{|x| \leq 2} dx.$$ 

In 1988, Bai used the moment method to find the necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix $A_N = \{A_{ij}\}_{1 \leq i,j \leq N}$ whose off-diagonal and diagonal entries are two independent family of i.i.d. random variables.

Theorem 5. (Bai) The largest eigenvalue of $\frac{A_N}{\sqrt{N}}$ tends to a finite constant $2\sigma$ as $N \to \infty$ with probability 1, if and only if

$$\mathbb{E}(A_{11})^2 < \infty; \quad \mathbb{E}(A_{12}) \leq 0;$$

$$\mathbb{E}(A_{12})^2 = \sigma^2; \quad \mathbb{E}(A_{12})^4 < \infty.$$ 

In addition to studying the first moment of ESD, it’s natural for us to consider the Central Limit Theorem (CLT) for the fluctuation of linear eigenvalue statistics of random matrices. This question was initiated in 1982 by D. Jonsson, who used the moment method to study Wishart matrices. For Wigner random matrices, CLT for linear eigenvalue statistics has been studied in several papers.

In 1998, A. Soshnikov and Y. Sinai studied the largest and smallest eigenvalues of Wigner matrices in [19] and [20] by looking at $\text{Tr} \left( \frac{A_N}{\sqrt{N}} \right)^{b_N}$, where $b_N$ grows with $N$. They also proved CLT for the linear eigenvalue statistics when the degree of the test function $x^n$ is growing with the dimension $N$.

Theorem 6. (Sinai, Soshnikov) Given a sequence of Wigner matrices $A_N$ in Definition 1, assume that the distributions of all entries are symmetric, and

$$\text{Var}(A_{ij}) = \frac{1}{4}, \quad \mathbb{E} A_{ij}^{2k} \leq (Ck)^k, \quad i \neq j, C > 0.$$ 

If $1 \ll n \ll N^{\frac{2}{3}}$, then

$$\mathbb{E} \text{Tr} \left( \frac{A_N}{\sqrt{N}} \right)^{2n} \sim \frac{1}{\sqrt{\pi} (2n)^{\frac{3}{2}}} \frac{N}{2^{n}}.$$ 

In addition, $\text{Tr} \left( \frac{A_N}{\sqrt{N}} \right)^n - \mathbb{E} \text{Tr} \left( \frac{A_N}{\sqrt{N}} \right)^n$ converges in distribution to normal law with expectation zero and variance $\frac{1}{\pi}$.
Their technique from [19] and [20] was later used to prove the Tracy-Widom law of edge spectrum of Wigner matrices in [25]. This approach can also be applied to the largest eigenvalue of sample covariance matrices in [26]. However, it does not work for the smallest eigenvalues, since they make vanishing contribution to moments of ESD. Recently S. Sodin considered Chebyshev polynomials and used the modified moment method to prove the universality at both edges of spectrum of sample covariance matrices and band random matrices, which will be defined in Definition 8.

**Theorem 7.** (Sodin) Let \( \{B^{(N)}\}_{N} \) be a sequence of \( M \times N \) matrices, \( M \leq N \), \( M \to \infty \), and \( \frac{M}{N} \to c \leq 1 \). We also assume that \( \{B_{uv}^{(N)}, 1 \leq u \leq M, 1 \leq v \leq N\} \) are independent random variables, whose distributions are symmetric and have subgaussian tails, with variance 1. Let \( \lambda_1^{(N)} \) be the smallest eigenvalue of the sample covariance matrix \( B^{(N)}B^{(N)*} \), then the distribution of the smallest rescaled eigenvalue 

\[
\frac{\lambda_1^{(N)} - (\sqrt{M} - \sqrt{N})^2}{(\sqrt{M} - \sqrt{N})(M^{-\frac{1}{2}} - N^{-\frac{1}{2}})^{\frac{1}{2}}}
\]

converges to the Tracy-Widom distribution \( \beta = 1 \).

Recently, a modification of Wigner matrix, named band random matrix model, has attracted a lot of interest.

**Definition 8.** We call \( H_N^W \in \mathbb{R}^{N \times N} \) a band random matrix with bandwidth \( W \) and dimension \( N \) if

1. \( H_{uv} = 0 \) if \( |u - v|_N = \min(|u - v|, N - |u - v|) > W_N \);
2. \( \{H_{uv}, u \leq v, |u - v|_N \leq W_N\} \) are independent random variables, with mean zero;
3. \( H_{uv} = H_{vu} \) if \( u > v \) and \( |u - v|_N \leq W_N \).

Band random matrix model was first introduced by physicists since it has associations with Schrodinger operators when \( W \) is at constant scale. Recently it has been related to Quantum Chaos. As the dimension \( W \) goes from constant to the full size, there has been conjectured to be a crossover at \( N^{\frac{1}{2}} \) between a strongly disordered regime, with localized eigenvectors and weak eigenvalue correlation, and a weakly disordered regime, with extended eigenvectors and strong eigenvalue repulsion. Specifically, if \( W \ll \sqrt{N} \) the local eigenvalue statistics follows a Poisson distribution, while as \( W \gg \sqrt{N} \), it appears to be similar with GOE or GUE (see reference [5], [6]).

In 2008, J. Schenker proved the localization for the Gaussian case as \( W \ll N^{\frac{1}{2}} \). In 2013, L. Erdos, H.T. Yau and others proved that eigenvectors corresponding to the bulk of the spectrum are delocalized if \( W \gg N^{\frac{1}{2}} \). Moreover, S. Sodin proved the phase transition for the largest eigenvalue statistics.
Theorem 9. (Theorem 1.2, [23]) Given a band random matrix ensemble in Definition 8, make additional assumption that all diagonal entries are zero and all off-diagonal entries follow symmetric Bernoulli distribution, i.e.

\[
P(H_{uv} = 1) = P(H_{uv} = -1) = \frac{1}{2}.
\]

If \( W_N \gg N^{\frac{5}{6}} \), set \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \) be the extreme eigenvalues of the rescaled matrix \( \frac{H_N}{2\sqrt{2W_N}} \), then

\[2N^{\frac{3}{4}}(\lambda_{\text{max}} - 1), \quad \text{and} \quad -2N^{\frac{3}{4}}(\lambda_{\text{min}} + 1)\]

converge in distribution to the Tracy-Widom distribution for \( \beta = 1 \).

Theorem 10. (Theorem 1.3, [23]) If \( W \log N \to \infty \), then \( \| H/(2\sqrt{2W_M}) \| \to 1 \) in distribution.

Recently, L. Li and A. Soshnikov used the Stein Method to prove the CLT for linear eigenvalue statistics of band random matrices assuming that \( \sqrt{N} \ll W \ll N \).

Theorem 11. (Li, Soshnikov) Under some technical conditions on smoothness of \( f \), if \( \sqrt{N} \ll W \ll N \),

\[
\sqrt{\frac{W}{N}} \left( \sum_{i=1}^{N} f \left( \frac{\lambda_i}{\sqrt{W}} \right) \right) - \mathbb{E} \sum_{i=1}^{N} f \left( \frac{\lambda_i}{\sqrt{W}} \right)
\]

converges in distribution to normal law of expectation zero and variance written in an integral formula(see reference [11]).

2 Outline and notations

There are two parts of the rest of this paper. Section 3 is devoted to use the moment method to study the limiting operator norm of band random matrices under different moment conditions and the corresponding growth rates of \( W \) depending on \( N \). In Section 4, I use modified moment method introduced in [23] to study the CLT for the linear eigenvalue statistics of Wigner matrices and then generalize it to band random matrices.

Remark 12. Throughout this paper, the letter \( C \) and \( c \) denote positive constants whose values might change in different positions, but are always independent of other relevant parameters. We say \( A \ll B \) if \( \frac{A}{B} \to 0 \) as \( N \to \infty \). To simplify our statements, we often neglect logarithmic factors, by introducing the notation \( \succ \). More precisely, \( A \succ B \) implies \( A \gg (\log N)^c B \) for some positive constant \( c \). In addition, \( A \sim B \) means that \( \frac{A}{B} \to 1 \) as \( N \to \infty \).

3 Moment method

The moment method plays an important role in studying the limit of the largest eigenvalue, denoted as \( \lambda_{\text{max}} \). Assuming \( p \) is even, we have an inequality

\[
\lambda_{\text{max}}^p(A_N) \leq \text{Tr} A_N^p \leq N\lambda_{\text{max}}^p(A_N).
\]
Roughly speaking, $(\text{Tr } A^p)^{\frac{1}{2}}$ controls the largest eigenvalue up to a multiplicative factor of $N^{\frac{1}{2}}$. If we choose $p \gg \log N$, then as $N \to \infty$, $\lambda_{\text{max}} \sim (\text{Tr } A^p)^{\frac{1}{2}}$. Thus we need to study the higher moments of $\text{Tr } A^p$, where $p$ is at least of the order of $\log N$.

In this section, we apply the moment method to generalize Theorem 5 to band random matrices. We are given different conditions on the distributions of the matrix entries.

**Definition 13.** We study the $N \times N$ dimensional ensemble of band random matrix with width $W_N$, denoted as $(A)_{i,j} \in \mathbb{R}^{N\times N}$, which satisfies the following conditions:

1. If $\min\{|i - j|, N - |i - j|\} > W_N$, $A_{ij} = A_{ji} = 0$;
2. If $\min\{|i - j|, N - |i - j|\} \leq W_N$, $A_{ij} = A_{ji}$, and they are independent random variables with symmetric distribution laws;
3. $\text{Var} A_{ij} = 1$ for $i \neq j$ and $\min\{|i - j|, N - |i - j|\} \leq W_N$, and $\text{Var} A_{ii}$ are finite;
4a) $E(A_{ij})^{2k}$ are finite for $k \leq a$, $a$ is a constant. Or,
4b) All moments are finite, $E(A_{ij})^{2k} \leq (Ck)^{2k}$ for some constant $C$. Or,
4c) $E(A_{ij}^{(N)})^{2k} \leq (Ck)^{2k}$ for $1 \leq k \leq m_N$, where $1 \ll m_N \ll \sqrt{\log N}$.

### 3.1 Under the condition (4a)

#### 3.1.1 Statement of theorem and its proof

We start with the proof of the limiting operator norm under the condition (4a).

**Theorem 14.** Given a matrix ensemble $A_N$ satisfying the conditions (1), (2), (3) and (4a), with the additional condition $W \succ (or \gg) N^{\frac{1}{a-1}}$, then the operator norm of the rescaled matrix $\frac{A_N}{2\sqrt{2W}}$ converges to 1 almost sure, i.e.

$$\left\| \frac{A_N}{2\sqrt{2W}} \right\|_{op} = \lambda_{\text{max}} \left( \frac{A_N}{2\sqrt{2W}} \right) \to 1.$$ 

**Proof.**

Since the distributions of all entries are symmetric, we note that:

$$E \text{ Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^p = \frac{1}{(2\sqrt{2W})^p} \sum_{P(k)} E A_{u_0u_1} A_{u_1u_2} \ldots A_{u_{p-1}u_0}, \quad (2)$$

where $P(k)$ is the set of all closed paths of length $k$ in the associated graph of $A$ whose edges are passed even number of times. Therefore, all the odd moments vanish. Next, we study the case of $p = 2s$. 

It’s not hard to find the lower bound. We only count the tree-like paths, whose edges are passed exactly twice. We consider a random walk on the positive half-line, where \( x(0) = 0 \). If the \( k \)-th step of the path is to create a new edge, then \( x(k) - x(k - 1) = 1 \). Otherwise, \( x(k) - x(k - 1) = -1 \). Obviously, \( x(t) \geq 0 \), and \( x(2s) = 0 \). Therefore, the shape of tree paths (up to isomorphism) is in one-to-one correspondence with Dyck paths. Once we fix the shape of the path, we start with any initial vertex, with \( N \) choices. For the first time moving towards a new vertex, we have \( 2W \) choices; for the second time moving towards a new vertex, we have \( 2W - 1 \) choices (exclude the initial vertex and itself); for the third time moving towards a new vertex, we have at least \( 2W - 2 \) choices (if the second new vertex can be reached by the initial vertex directly, then we have \( 2W - 2 \) choices). Continue till the \( s \)-th new vertex, we have

\[
\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \geq N \frac{(2s)!}{s!(s + 1)!} 2W(2W - 1) \cdots (2W - s + 1) \left( \frac{1}{2\sqrt{2W}} \right)^{2s}.
\]

If we further assume \( s \ll W^{\frac{1}{2}} \) and use the Stirling approximation,

\[
\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \geq \frac{N}{\sqrt{\pi s^3}} (1 - o(1)). \tag{3}
\]

Next, we need to obtain a suitable upper bound. We use the following definition.

**Definition 15.** The \( k \)-th step \( u_{k-1}u_k \) is called marked if during the first \( k \) steps, the edge \( u_{k-1}u_k \) (or \( u_ku_{k-1} \)) appeared an odd number of times. Otherwise, it’s unmarked.

Since each edge is passed even number of times, there will be \( s \) unmarked steps and \( s \) marked steps. At any time, the number of marked steps should be no less than the number of unmarked steps. We consider a random walk on the positive half-line, where \( x(0) = 0 \). If the \( k \)-th step of the path is marked, then \( x(k) - x(k - 1) = 1 \). Otherwise, \( x(k) - x(k - 1) = -1 \). Obviously, \( x(t) \geq 0 \), and \( x(2s) = 0 \). So there is an one-to-one correspondence between the choices of marked and unmarked steps and the Dyck paths. Therefore, we have \( \frac{(2s)!}{s!(s + 1)!} \) ways to choose the positions for marked and unmarked steps.

For each marked steps, it can create a new edge or it’s a multiple edge. For the marked steps that create new edges, there are still two situations: it can create a new vertex or it’s an intersection edge. And we use following notations to record them:

- \( J \): the set of marked steps that create new edges, \#\( J = j \) (\( 1 \leq j \leq s \));
- \( M \): the set of marked steps that create new edges, but point to the old vertices, \#\( M = m \) (\( 0 \leq m \leq j - 1 \));
- \( J/M \): the set of marked steps that create new vertices, \#\( J/M = j - m \);
- \( S/J \): the set of marked steps that coincident with old edges(create multiple edges), \#\( S/J = s - j \).
We have $N$ choices for the initial vertex to start with. Once we fix the set $J$ and $M$, we have at most $(2W)^{j-m}$ choices for the set of $J/M$, and at most $(j-m)^m$ choices for the set of $M$. To consider the choices for the set of $S/J$, we introduce a lemma, see details in the Appendix of [5].

**Lemma 16.** The maximum number of vertices that can be visited at marked instants from a given vertex is denoted as $v_s(\mathcal{P})$. For any $\gamma$ with $0 < \gamma < 1$, the sum of $\sum_{\mathcal{P}(k)} \mathbb{E}A_{u_0u_1}A_{u_1u_2} \ldots A_{u_{2s-1}u_0}$ over paths $\mathcal{P}$ with $v_s(\mathcal{P}) > s^\gamma$ is negligible compared to the total sum.

Thus, we have at most $(2W)^{j-m}(j-m)^m(s^\gamma)^{s-j}$ choices for all the marked steps. Next, we consider all the unmarked steps. If each edge is passed exactly twice, then the unmarked steps are determined uniquely. Since we have intersection edges and multiple edges in the general case, according to the above lemma, we have at most $2^{s-j}(2s\gamma)^m$ choices. Therefore, (2) has an upper bound as

$$N \frac{(2s)!}{s!(s+1)!} \sum_{1 \leq j \leq s, 0 \leq m \leq j-1} \binom{s}{j} \binom{m}{j} (2W)^{j-m}(j-m)^m(2s\gamma)^{s-j+m} \left( \frac{1}{2\sqrt{2W}} \right)^{2s} \mathbb{E}A_{u_0u_1} \ldots A_{u_{2s-1}u_0}$$

$$\leq N \frac{(2s)!}{s!(s+1)!} \frac{1}{4^s} \sum_{1 \leq j \leq s, 0 \leq m \leq j-1} \frac{s!}{(s-j)!j!(j-m)!m!(CW)^{s-j+m}} \mathbb{E}A_{u_0u_1} \ldots A_{u_{2s-1}u_0}$$

$$\leq N \frac{(2s)!}{s!(s+1)!} \frac{1}{4^s} \sum_{1 \leq j \leq s, 0 \leq m \leq j-1} \frac{1}{m!} \left( \frac{Cs^{2+\gamma}}{W} \right)^m \frac{1}{(s-j)!} \left( \frac{Cs^{1+\gamma}}{W} \right)^{s-j} \mathbb{E}A_{u_0u_1} \ldots A_{u_{2s-1}u_0},$$

(4)

since $s! \leq (j-m)!s^{s-j+m}$. We only need to obtain an upper bound of $\mathbb{E}A_{u_0u_1} \ldots A_{u_{2s-1}u_0}$. Let us first consider the simplest case, where each edge is visited exactly twice. Since the second moment of each entry is 1, then $\mathbb{E}A_{u_0u_1} \ldots A_{u_{2s-1}u_0} = \prod_{i<j} \mathbb{E}A_{ij}^2 = 1$.

Next, we note that if the random variable has an upper bound $M$ almost sure, then we can bound higher moments as following,

$$\mathbb{E}|X|^p \leq M^{p-2} \mathbb{E}X^2 = M^{p-2}.$$  \hspace{1cm} (5)

In general, $A_{ij}$ is not bounded almost sure, but it has finite moments up to $2a$. By the Markov Inequality, we have

$$\mathbb{P} \left( |A_{ij}^{(N)}| > \lambda \right) \leq \frac{C}{\lambda^{2a}}.$$ \hspace{1cm} (6)

Thus it has an upper bound with high probability. And the method we used next is named truncation method.

Let

$$\tilde{A}_{ij}^{(N)} = \begin{cases} A_{ij}^{(N)} & \text{if } |A_{ij}^{(N)}| \leq (WN)^{\frac{a}{2}} \\ 0 & \text{if } |A_{ij}^{(N)}| > (WN)^{\frac{a}{2}} \end{cases}$$ \hspace{1cm} (7)
Lemma 17. If $\mathbb{E}|A_{ij}^{(N)}|^{2a} < \infty$, then $A^{(N)} = \tilde{A}^{(N)}$ almost sure as $N \to 0$.

Proof.

$$
P(A_N \neq \tilde{A}_N) \leq \sum_{i,j} \mathbb{P}\left(|A_{ij}^{(N)}| > (WN)^{\frac{1}{2a}}\right) = WN \cdot \mathbb{P}\left(|A_{ij}^{(N)}| > (WN)^{\frac{1}{2a}}\right).
$$

Since $\mathbb{E}|A_{ij}^{(N)}|^{2a} < \infty$, we have $\sum_{m=1}^{\infty} 2^m \cdot \mathbb{P}(|A_{ij}^{(N)}| > (2^m)^{\frac{1}{2a}}) < \infty$.

Choosing $2^{m-1} < WN \leq 2^m$, then $\sum_{N=1}^{\infty} P(A_N \neq \tilde{A}_N)$ is finite.

By Borel Cantelli Lemma, we conclude that

$$
P(\tilde{A}_N \neq A_N \text{ i.o.}) = 0.
$$

Thus, Lemma 6 is proved.

Therefore, we could replace $A_{ij}$ with $\tilde{A}_{ij}$ almost sure. Thus, given a even path of length $2s$, the number of marked steps that coincident with old edges is $s - j$, then by (5) and (7),

$$
\mathbb{E}A_{u_0 u_1} A_{u_1 u_2} \ldots A_{u_{2s-1} u_0} \leq \left(\mathbb{E}X^2\right)^{2j} (\max_{i,j} |A_{ij}|)^{2(s-j)} \leq \left((WN)^{\frac{1}{2a}}\right)^{2(s-j)}.
$$

(8)

In addition, we can obtain a stronger lemma for the upper bound.

Lemma 18. If $\mathbb{E}|A_{ij}^{(N)}|^{2a} < \infty$, then $A_{ij}^{(N)} \leq \delta_N (WN)^{\frac{1}{2a}}$ for all $i,j$, where $\delta_N \to 0$ arbitrary slowly, say slower than $\frac{1}{\log N}$.

Proof. Since $\mathbb{E}|A_{ij}^{(N)}|^{2a} < \infty$, for any $k > 0$,

$$
\sum_{N=1}^{\infty} WN \cdot \mathbb{P}(|A_{ij}^{(N)}| > \frac{1}{k} (WN)^{\frac{1}{2a}}) < \infty.
$$

Consider a sequence $\{c_k\}_{k \geq 1}$, such that $\sum_{k=1}^{\infty} c_k < \infty$.

Then there exists $\{m_k\}_{k \geq 1}$, such that

$$
\sum_{N=m_k}^{\infty} WN \cdot \mathbb{P}(|A_{ij}^{(N)}| > \frac{1}{k} (WN)^{\frac{1}{2a}}) < c_k.
$$

Consider a new multi-sequence $\tilde{m}_N$ of $1, \frac{1}{2}, \frac{1}{3}, \ldots$ in non-increasing order, where $\frac{1}{k}$ appears $m_k$ times. Then

$$
\sum_{N=1}^{\infty} WN \cdot \mathbb{P}(|A_{ij}^{(N)}| > \tilde{m}_N (WN)^{\frac{1}{2a}}) < \infty.
$$

Let $\delta_N = \tilde{m}_N$, there exists a sequence $\{\delta_n\}_{n \geq 1}$ decaying to zero, whose decay rate can be made arbitrary slow, say slower than $\frac{1}{\log N}$. By Borel Cantelli Lemma, we finish the proof of Lemma 7. \qed
By the above lemma, we induce that

$$E A_{u_0u_1} A_{u_1u_2} \cdots A_{u_{2s-1}u_0} \leq \left( \delta_N \cdot (WN)^{\frac{1}{2}} \right)^{s-j}. \quad (9)$$

Then, combining (4) and (9),

$$\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \leq N \frac{(2s)!}{s!(s+1)!} \frac{1}{4^s} \left( \sum_{s-j \geq 0} \frac{1}{(s-j)!} \left( \frac{C\delta s^{1+\gamma}(WN)^{\frac{1}{2}}}{W} \right)^{s-j} \right) \left( \sum_{m} \frac{1}{m!} \left( \frac{CS^{2+\gamma}}{W} \right)^m \right).$$

If we further assume that $s^{2+\gamma} \ll W$, and $\delta s^{(1+\gamma)}(WN)^{\frac{1}{2}} \ll W$, then

$$\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \leq \frac{N}{\sqrt{\pi}s^3} (1 + o(1)). \quad (10)$$

Therefore, by (3), if $s^{1+\gamma} \ll \min\{ \frac{W^{1-\gamma}}{N^2}, \sqrt{W} \}$,

$$\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} = \frac{N}{\sqrt{\pi}s^3} (1 + o(1)). \quad (11)$$

By the Markov Inequality, we have

$$\mathbb{P} \left( \lambda_{\max}^{(N)} > 1 + \epsilon \right) \leq \frac{\mathbb{E} \lambda_{\max}^{2s}}{(1 + \epsilon)^{2s}} \leq \frac{\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s}}{(1 + \epsilon)^{2s}} = \frac{N}{\sqrt{\pi}s^3} (1 + o(1)). \quad (12)$$

Let $s$ be large enough, as $\log^c N$, where $c > 1$. Then the RHS of the above inequality goes to 0, and the decay rate depends on $\epsilon$. Since $s \gg \log N$, it implies that $W \gg (\log N)^c N^{\frac{1}{\pi-1}}$. Thus we prove that the operator norm converges to 1 in probability. Furthermore,

$$\sum_N \mathbb{P} \left( \lambda_{\max}^{(N)} > 1 + \epsilon \right) \leq \frac{N}{\sqrt{\pi}s^3(1 + \epsilon)^{2s}} (1 + o(1)). \quad (13)$$

Choosing $s \gg \log^2 N$, which implies that $W \gg (\log N)^c N^{\frac{1}{\pi-1}}$ then the RHS of the above inequality will be summable (of the form of $N^{-c} \log^{-3} N$, where $c > 1$). Then by Borel Cantelli Lemma, we have $\mathbb{P} \left( \lambda_{\max}^{(N)} > 1 \ i.o. \right) = 0$, and we thus prove Theorem 3.

$$\square$$

### 3.1.2 Removing the logarithmic factor

To remove the logarithmic factor in the condition of Theorem 14, we could modify the truncation procedure a little bit.

Fix a row or column of $A - \~A$. Since the entries are independent and by (13), the probability that there are at least $k$ non-zeros entries is at most $O(\frac{W^k}{N^{2k}})$. The probability that there
exists at least one row or column of $A - \tilde{A}$ with at least $k$ non-zero entries is bounded by 

$$O\left(\frac{\sqrt{N}k}{\lambda}\right) = o(1)$$

provided $\lambda \gg W^{\frac{1}{2a}}N^{\frac{1}{2ak}}$.

Fix some constant $k \geq 2$, let

$$\tilde{A}^{(N)}_{ij} = \begin{cases} 
A^{(N)}_{ij} & \text{if } |A^{(N)}_{ij}| \leq W^{\frac{1}{2a}}N^{\frac{1}{2ak}} \\
0 & \text{if } |A^{(N)}_{ij}| > W^{\frac{1}{2a}}N^{\frac{1}{2ak}} 
\end{cases} \quad (14)$$

Now we know that $A - \tilde{A}$ can only have $k$ non-zero entry for each row and column in probability. Then $\|A - \tilde{A}\|_{op} = \max_{\|x\|=1} \left\{ \sum_{i,j} (A_{ij} - \tilde{A}_{ij})x_i x_j \right\} \leq k^2 \max |A_{ij}|$. From Lemma 7 we know that $\max |A_{i,j}| \leq \delta(WN)^{\frac{1}{2a}}$ almost sure. If we further assume that $\delta(WN)^{\frac{1}{2a}} \sqrt{W} \to 0$, which implies that $W \gg N^{\frac{1}{2a}-1}$, then we obtain

$$\left\| \frac{A - \tilde{A}}{2\sqrt{2W}} \right\|_{op} \leq \frac{k^2\delta(WN)^{\frac{1}{2a}}}{2\sqrt{2W}} \to 0. \quad (15)$$

Since the distribution laws of the entries are symmetric, the mean of the truncated entries is still 0 and the variance should be less than 1. Thus we combine (4) and (14), then

$$\mathbb{E} \text{Tr} \left( \frac{\tilde{A}}{2\sqrt{2W}} \right)^{2s} \leq N^{(2s)!} \frac{1}{s!(s+1)!} 4^s \left( \sum_{s-j \geq 0} \frac{1}{(s-j)!} \left( C\delta s^{1+\gamma}N^\frac{1}{2a}W^\frac{1}{2} \right)^{s-j} \right) \left( \sum_{m} \frac{1}{m!} \left( Cs^{2+\gamma} \right)^{m} \right).$$

Then, if $s^{2+\gamma} \ll W$, and $s^{1+\gamma}N^\frac{1}{2a}W^\frac{1}{2} \ll W$, which implies that $s^{1+\gamma} \ll \min\{W^{1-\frac{1}{2a}}, \sqrt{W}\}$, then

$$\mathbb{E} \text{Tr} \left( \frac{\tilde{A}}{2\sqrt{2W}} \right)^{2s} \leq \frac{N}{\sqrt{\pi}s^3}(1 + o(1)). \quad (16)$$

Let $s \ll \log N$, which implies that $W \gg (\log N)^\alpha W^{\frac{1}{\alpha(a-1)}}$ we combine (3) and (1) thus obtain that in probability,

$$\left\| \frac{\tilde{A}}{2\sqrt{2W}} \right\|_{op} = 1 + o(1). \quad (17)$$

And by (15)

$$\left\| \frac{A}{2\sqrt{2W}} \right\|_{op} \leq \left\| \frac{\tilde{A}}{2\sqrt{2W}} \right\|_{op} + \left\| \frac{A - \tilde{A}}{2\sqrt{2W}} \right\|_{op} = 1 + o(1). \quad (18)$$

To conclude, if $W \gg N^{\frac{1}{2a-1}}$, then in probability

$$\left\| \frac{A}{2\sqrt{2W}} \right\|_{op} = 1 + o(1). \quad (19)$$
3.1.3 Stronger version of Theorem 14

To get a stronger result, we need to get a more detailed estimate on the upper bound of $\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)$. Here we use the combinatorial technique introduced in [4] and [5] to get a stronger version of Theorem 14.

**Theorem 19.** Given a matrix ensemble $A_N$ satisfying the conditions (1), (2), (3) and (4a) in the Definition 1, with the additional condition $W \succ N^{\frac{1}{a-1}}$, then the operator norm of the rescaled matrix $\frac{A_N}{2\sqrt{2W}}$ converges to 1 almost sure.

First, we split the set of $N$ vertices into $(s+1)$ subsets, denoted as $N_k, 0 \leq k \leq s$.

**Definition 20.** A vertex is said to belong to $N_k, 1 \leq i \leq N$, if this vertex was arrived at $k$ marked steps. $n_k = \#N_k$. We say that the path is of type $(n_0, n_1, \cdots, n_s)$.

We have the following equations:

$$
\sum_{k=0}^{s} n_k = N, \quad \sum_{k=1}^{s} kn_k = s. \quad (20)
$$

In particular, paths of type $(n - s, s, 0, \cdots, 0)$ are just the tree-like paths that we used for the lower bound, in addition to the circle paths.

**Proof.** To estimate the upper bound, we start with any initial vertex, with $N$ choices. Except the initial vertex, there are at most $n_1 + n_2 + \cdots + n_s$ vertices involved in this path. Thus we have at most $(2W)^{n_1 + \cdots + n_s}$ different ways to choose the vertices for the path, ordered by the first appearance of the vertex in the path.

Once the positions for marked steps are fixed, for paths of the type $(n_0, n_1, \cdots, n_s)$, we need to divide the position set of $s$ marked steps into subsets of cardinality $kn_k$. For the $k$th position subset, we again need to divide into $n_k$ unordered subsets of cardinality $k$. In total, we have at most

$$
\frac{s!}{\prod_{k=1}^{s}(kn_k)!} \left( \prod_{k=1}^{s} n_k!(k!)^{n_k} \right) (2W)^{n_1 + \cdots + n_s} = \frac{s!}{\prod_{k=1}^{s} n_k!(k!)^{n_k}} (2W)^{n_1 + \cdots + n_s}. \quad (21)
$$

ways to choose the marked steps.

For the unmarked steps, if the initial vertex belongs to $N_k$, then we have at most $2k$ possibilities to choose the end points. In particular, if the vertex belongs to $N_2$, we have the unique way to return. Thus the number of ways to choose marked steps is $\prod_{k=2}^{s} (2k)^{kn_k}$.

For paths of the type $(n_0, n_1, \cdots, n_s)$, given the unit variance and condition (4b), we have

$$
\mathbb{E} A_{u_0 u_1} A_{u_1 u_2} \cdots A_{u_{2s-1} u_0} \leq \prod_{k=2}^{s} C_{n_k} \prod_{k=a+1}^{s} \left( \delta(WN)^{\frac{1}{2}} \right)^{(k-a)n_k}. \quad (22)
$$
Combining (2), (21) and (22) we obtain an upper bound
\[
\sum N \frac{(2s)!}{(s+1)!s! \prod_{k=1}^{s} n_k!(k!)^{n_k}} \frac{s!}{(2\sqrt{2W})^{2s}} \prod_{k=2}^{s} (2k)^{kn_k} \prod_{k=2}^{s} \left( \delta(WN)^{\frac{1}{2}} \right)^{(k-a)n_k} \\
\leq N \frac{(2s)!}{(s+1)!s! 4^s} \left( \sum \frac{s!}{\prod_{k=2}^{s} (Ck)^{kn_k} \prod_{k=2}^{s} \left( \delta(WN)^{\frac{1}{2}} \right)^{(k-a)n_k}} \prod_{k=2}^{s} \frac{1}{n_k!} \left( \frac{(Cs)^k \delta^{k-a}(WN)^{\frac{k-a}{a}}}{W^{k-1}} \right)^{n_k} \right) \\
\leq N \frac{(2s)!}{(s+1)!s! 4^s} \left( \exp \left( \sum_{k=2}^{s} \frac{(Cs)^k}{W^{k-1}} \right) \right) \left( \exp \left( \sum_{k=2}^{s} \frac{(Cs)^k \delta^{k-a}(WN)^{\frac{k-a}{a}}}{W^{k-1}} \right) \right), \quad (23)
\]
since \( s! \leq n_1! s^{s-n_1} \) and \( N - n_0 - s = \sum_k (k-1)n_k \).
If we further assume that \( \frac{s^2}{W} = o(1) \) and \( \delta \frac{s+1}{a(WN)^{\frac{1}{a}}} = o(1) \), then
\[
\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \leq N \frac{(2s)!}{(s+1)!s! 4^s} \left( 1 + o(1) \right) = \frac{N}{\sqrt{\pi} s^3} \left( 1 + o(1) \right). \quad (24)
\]
Since we require \( s \gg \log N \), it implies that \( W \gg (\log N)^{\frac{a}{a-1}} N \frac{1}{\alpha-1} \), which is stronger than the technical condition in Theorem 14.

3.2 Under condition (4b) and (4c)

Next, we use the method in subsection 3.1.3 to prove similar results under condition (4b) and (4c).

3.2.1 Under the condition (4b)

\textbf{Theorem 21.} Given a matrix ensemble \( A_N \) satisfying the conditions (1), (2), (3) and (4b) in the Definition 1, with the additional condition \( 1 < W \leq N \), then the operator norm of the rescaled matrix \( \frac{A_N}{2\sqrt{2W}} \) converges to 1 almost sure.

\textbf{Proof.} For paths of the type \( (n_0, n_1, \cdots, n_s) \), given the unit variance and condition (4b), we have
\[
\mathbb{E} A_{u_0u_1} A_{u_1u_2} \cdots A_{u_{2s-1}u_0} \leq \prod_{k=2}^{s} (Ck)^{2kn_k}. \quad (25)
\]
Then,
\[
\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \leq \sum N \frac{(2s)!}{(s+1)!s! \prod_{k=1}^{s} n_k!(k!)^{n_k}} \frac{s!}{(2\sqrt{2W})^{2s}} \prod_{k=2}^{s} (2k)^{kn_k} \prod_{k=2}^{s} \left( \delta(WN)^{\frac{1}{2}} \right)^{(k-a)n_k} \\
\leq N \frac{(2s)!}{(s+1)!s! 4^s} \left( \sum \frac{s!}{\prod_{k=2}^{s} (Ck)^{kn_k} \prod_{k=2}^{s} \left( \delta(WN)^{\frac{1}{2}} \right)^{(k-a)n_k}} \prod_{k=2}^{s} \frac{1}{n_k!} \left( \frac{(Cs)^k \delta^{k-a}(WN)^{\frac{k-a}{a}}}{W^{k-1}} \right)^{n_k} \right) \\
\leq N \frac{(2s)!}{(s+1)!s! 4^s} \left( \exp \left( \sum_{k=2}^{s} \frac{(Cs)^k}{W^{k-1}} \right) \right) \left( \exp \left( \sum_{k=2}^{s} \frac{(Cs)^k \delta^{k-a}(WN)^{\frac{k-a}{a}}}{W^{k-1}} \right) \right), \quad (23)
\]

Since we require \( s \gg \log N \), it implies that \( W \gg (\log N)^{\frac{a}{a-1}} N \frac{1}{\alpha-1} \), which is stronger than the technical condition in Theorem 14.

\[\square\]
\[ \leq N \frac{(2s)!}{(s+1)!s!4^s} \left( \sum s! \prod_{k=2}^{s} \frac{(C_k^2 s^k)_{n_k}}{n_k! (k!)^{n_k}} (2W)^{N-n_0-s} \right) \]
\[ \leq N \frac{(2s)!}{(s+1)!s!4^s} \left( \sum \prod_{k=2}^{s} \frac{1}{n_k!} \left( \frac{(C_k^2 s^k)^{n_k}}{W^{k-1}} \right) \right) \]
\[ \leq N \frac{(2s)!}{(s+1)!s!4^s} \left( \exp \left\{ \sum_{k=2}^{s} \frac{(C_k^2 s^k)^{n_k}}{W^{k-1}} \right\} \right). \quad (26) \]

If \( s^3 \ll W \), then the leading order of the sum above should be \( \frac{s^2}{W} \). Thus if we further assume that \( s \ll W^\frac{1}{3} \), then
\[ \mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} \leq \frac{N}{\sqrt{\pi} s^3} (1 + o(1)). \quad (27) \]

Therefore, by (3), if \( s \ll W^\frac{1}{2} \),
\[ \mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s} = \frac{N}{\sqrt{\pi} s^3} (1 + o(1)). \quad (28) \]

By the Markov Inequality, we have
\[ \mathbb{P} \left( \lambda_{\max}^{(N)} > 1 + \epsilon \right) \leq \frac{\mathbb{E} \lambda_{\max}^{2s}}{(1+\epsilon)^{2s}} \leq \frac{\mathbb{E} \text{Tr} \left( \frac{A_N}{2\sqrt{2W}} \right)^{2s}}{(1+\epsilon)^{2s}} = \frac{N}{\sqrt{\pi} s^3} (1 + o(1)) \to 0. \quad (29) \]

Let \( s \) be large enough, like \( \log^c N \), where \( c > 1 \). Then the RHS in the above inequality goes to 0, and the decay rate depends on \( \epsilon \). Since \( s \gg \log N \), it implies that \( W \gg \log^3 N \). Thus we prove that the operator norm converges to 1 in probability. Finally, using the similar arguments as proving Theorem 14, we use Borel Cantelli Lemma to prove the statement in the almost sure case. The only difference is that we need to choose \( s \gg \log^2 N \), which implies that \( W \gg \log^6 N \).

**Remark 22.** If we assume all moments are finite, and the moments satisfy \( \mathbb{E} X^k \leq (Ck^r)^k \), \( r \) is some constant, then we can get similar statement as Theorem 21. After modifying the above proof, there exists some constant \( \gamma > 0 \), such that if \( W \gg \log^7 N \), then the largest eigenvalue of the band random matrix goes to 1 almost sure.

**3.2.2 Under the condition (4c)**

Furthermore, we consider the random matrix ensemble in a general situation, where we allow the distributions of the entries depend on \( N \), the dimension of the random matrices. We can assume that the higher moments are not necessarily finite, but grow with the dimension of the matrices. For example, assuming \( \mathbb{E} (A_{ij}^{(N)})^{2k} \leq (Ck \log N)^k \), we can obtain similar
conclusion as above. Or one can relax the moments \( \mathbb{E}(A_{ij}^{(N)})^{2k} \leq (CkN^x)^k \), then we need a stronger technique condition \( W \succ N^{2x} \).

To generalize in another direction, we can assume that the exponential decay condition is satisfied up to \( 2m_N \), where \( 1 \ll m_N \ll \sqrt{\log N} \) is growing with \( N \).

**Theorem 23.** Given a matrix ensemble \( A_N \) satisfying the conditions (1), (2), (3) and (4c) in the Definition 1, with the additional condition \( W \succ N^{\frac{1}{2m_N}} \), then the operator norm of the rescaled matrix \( \frac{A_N}{\sqrt{2W}} \) converges to 1 in probability. Moreover, if under a stronger condition \( W \succ N^{\frac{1}{m_N}} \), then the operator norm converges almost sure.

**Proof.**

\[
\sum_{i,j} \mathbb{P} \left( A_{ij}^{(N)} > \lambda \right) = WN \frac{\mathbb{E}^{2m_N} A_{ij}^{(N)}}{\lambda^{2m_N}} \leq WN \frac{(Cm_N)^{2m_N}}{\lambda^{2m_N}} \to 0. \tag{30}
\]

If we choose \( \lambda \gg (WN)^{\frac{1}{2m_N}} m_N \). Then the probability that all entries of \( A_{ij} = O((WN)^{\frac{1}{2m_N}} m_N) \) approaches 1 as \( N \) is sufficient large. For paths of the type \((n_0, n_1, \cdots, n_s)\), given the unit variance and condition (4c), let \( m = m_N \),

\[
\mathbb{E} A_{u_0 u_1} A_{u_1 u_2} \cdots A_{u_{2s} u_0} \leq \prod_{k=2}^{m} (Ck)^{2kn} \prod_{k=m+1}^{s} ((WN)^{\frac{1}{2m}} m)^{2(k-m)n_k} (Cm)^{2mn_k}. \tag{31}
\]

Then, \( \mathbb{E} \text{Tr} \left( \frac{A_N}{\sqrt{2W}} \right)^{2s} \leq \frac{2s!}{(s+1)!s!} N^{s} \frac{(2W)^{n_1 + \cdots + n_s}}{(2\sqrt{2W})^{2s}} \prod_{k=2}^{m} (Ck)^{kn} \prod_{k=m+1}^{s} (Ck)^{kn} \prod_{k=1}^{s} \left( Cm^2 k^2 (WN)^{\frac{k-m}{m}} \right)^{nk} \]

\[
\leq N^{(2s)!} \frac{1}{(s+1)!s! 4^s} \left( \sum_{n_k} \prod_{k=1}^{s} \frac{1}{n_k!} \frac{(Ck^2)^{kn}}{(Wk-1)^{k}} \prod_{k=m+1}^{s} \frac{1}{n_k!} \frac{(Cm^2 k^2 )^{kn}}{(Wk-1)^{k}} \right) \]

\[
\leq N^{(2s)!} \frac{1}{(s+1)!s! 4^s} \left( \exp \left\{ \sum_{k=2}^{m} \frac{(Ck^2)^{k}}{(Wk-1)^{k}} \right\} \right) \left( \exp \left\{ \sum_{k=m+1}^{s} \frac{(Cm^2)^{k} (WN)^{\frac{k-m}{m}}}{(Wk-1)^{k}} \right\} \right). \tag{32}
\]

If \( s^2 \ll W, m^2 s \ll W \), and \( (Cm^2)^{m+1} (WN)^{\frac{1}{m}} \ll 1 \), that is \( s \ll \frac{W^{m-1}}{m^2 N^{m(m+1)}} \), then

\[
\mathbb{E} \text{Tr} \left( \frac{A_N}{\sqrt{2W}} \right)^{2s} = \frac{N}{\sqrt{\pi} s^3} (1 + o(1)). \tag{33}
\]
By the Markov Inequality, we have
\[ P(\lambda_{\text{max}}^{(N)} > 1 + \epsilon) \leq \frac{E\lambda_{\text{max}}^{2s}}{(1 + \epsilon)^{2s}} \leq \frac{E \text{Tr} \left( \frac{A_n}{2\sqrt{2W}} \right)^{2s}}{(1 + \epsilon)^{2s}} = \frac{N}{\sqrt{\pi s^3}} (1 + o(1)) \rightarrow 0. \] (34)

Let \( s \) be large enough, like \( \log^c N \), where \( c > 1 \). Then the RHS in the above inequality goes to 0, and the decay rate depends on \( \epsilon \). Since \( s \gg \log N \) and \( 1 \ll m \ll \sqrt{\log N} \), it implies that \( W \gg \log^2 N \cdot N^{\frac{1}{m^2}} \gg 1 \). Thus we prove that the operator norm converges to 1 in probability.

To prove the statement in the almost sure situation, we need to modify the choice of \( \lambda \) at the beginning of the proof. We need \( WN \frac{(Cm^2 s)^m N}{\lambda^{mN}} \ll \frac{1}{NW} \), that is \( \lambda \gg (WN)^{\frac{1}{mN}} m_N \). Then we can use Borel Cantelli lemma to prove that all entries can be bounded from above by \( \lambda \) almost sure. After similar argument from above, we can show that if \( (Cm^2 s)^{m+1} (WN)^{\frac{1}{W}} \ll 1 \), then
\[ E \text{Tr} \left( \frac{A_n}{2\sqrt{2W}} \right)^{2s} = \frac{N}{\sqrt{\pi s^3}} (1 + o(1)). \] (35)

By the Markov Inequality, we have
\[ P(\lambda_{\text{max}}^{(N)} > 1 + \epsilon) \leq \frac{E\lambda_{\text{max}}^{2s}}{(1 + \epsilon)^{2s}} \leq \frac{E \text{Tr} \left( \frac{A_n}{2\sqrt{2W}} \right)^{2s}}{(1 + \epsilon)^{2s}} = \frac{N}{\sqrt{\pi s^3}} (1 + o(1)) \rightarrow 0. \] (36)

Let \( s \gg \log^2 N \), which implies that \( W \gg \log^2 N \cdot N^{\frac{1}{m^2}} \). The RHS of the above inequality is summable with respect to \( N \), thus we can use Borel Cantelli Lemma again to prove almost sure convergence. Since \( 1 \ll m \ll \sqrt{\log N} \), we need \( W \gg \log^2 N \cdot N^{\frac{1}{m^2}} \gg 1 \).

\[ \square \]

3.3 Generalization to non-symmetric distributions

Moreover, we can also extend above theorems to the random matrix ensemble where the entries of the matrix are not symmetric in distribution laws.

**Theorem 24.** Given a matrix ensemble \( A_N \) satisfying the conditions (1), (2), (3) and (4a) in the Definition 1, but the distributions of the entries are not necessarily symmetric, with the additional condition \( W \gg N^{\frac{1}{m^2}} \), then the operator norm of the rescaled matrix \( \frac{A_N}{2\sqrt{2W}} \) converges to 1 almost sure.

**Proof.** We still consider \( E \text{Tr} \left( \frac{A_n}{2\sqrt{2W}} \right)^{2s} \). Since the distributions of the entries are not necessarily symmetric, we need to count the paths where each edge is passed more than twice. Note that the number of the edges that are passed an odd number of times must be even. First, we introduce a new variable \( 2l \) to record the number of marked steps that will never be returned, i.e. the last step of odd edges. For the rest \( 2s - 2l \) steps, there should be \( s - l \)
marked steps and \( s - l \) unmarked steps. Then for these \( s - l \) marked steps, we split the set of \( N \) vertices into \( s - l + 1 \) subsets, as in Definition 9:

\[
\sum_{k=0}^{s-l} n_k = N, \quad \sum_{k=1}^{s-l} kn_k = s - l.
\]  

(37)

Using the similar argument in Theorem 10, we have at most

\[
N \frac{(2s - 2l)!}{(s - l + 1)! (s - l)! \prod_{k=1}^{s-l} n_k! (k!)^{a_k}} (2W)^{N - n_o} \prod_{k=2}^{s-l} (2k)^{kn_k}.
\]

(38)

choices for \( s - l \) marked steps and \( s - l \) unmarked steps. For these extra \( 2l \) marked steps, we have at most \( \left( \frac{2s - 2l}{2l} \right) \) choices for the positions, and at most \( (s - l)^{2l} \) choices for the edges. Thus we obtain an upper bound of \( \mathbb{E} \text{Tr} \left( \frac{A}{2\sqrt{2W}} \right)^{2s} \) as following:

\[
\sum_{t=0}^{s} \left( \frac{2s - 2l}{2l} \right) (s - l)^{2l} \sum_{\mathcal{P}} N \frac{(2s - 2l)!}{(s - l + 1)! (s - l)! \prod_{k=1}^{s-l} n_k! (k!)^{a_k}} (2W)^{N - n_o} \prod_{k=2}^{s-l} (2k)^{kn_k} \\
\prod_{k=2}^{s-l} (2k)^{kn_k} \prod_{k=2}^{s-l} C_{\mathcal{P}}^{n_k} \prod_{k=a+1}^{s} \left( (WN)^{\frac{1}{2}} \right)^{(k-a)n_k} \left( (WN)^{\frac{1}{2}} \right)! 
\]

\[
\leq N \frac{(2s)!}{(s + 1)! 4^a} \sum_{t=0}^{s} \frac{1}{(2l)!} \left( \frac{C s^4 (WN)^{\frac{1}{2}}}{W} \right)^{t} \left( \sum_{k=2}^{a} \frac{1}{n_k!} \left( \frac{(Cs)^{k}}{W^{k-1}} \right) \prod_{k=a+1}^{s-l} \frac{1}{n_k!} \left( \frac{(Cs)^{k}}{W^{k-1}} \right) \right)^{n_k} 
\]

\[
\leq N \frac{(2s)!}{(s + 1)! 4^a} \exp \left( \frac{C s^2 (WN)^{\frac{1}{2}}}{\sqrt{W}} \right) \left( \exp \left\{ \sum_{k=2}^{a} \left( \frac{(Cs)^{k}}{W^{k-1}} \right) \right\} \left( \exp \left\{ \sum_{k=a+1}^{s-l} \left( \frac{(Cs)^{k}}{W^{k-1}} \right) \right\} \right) .
\]

(39)

Therefore, if \( s^4(WN)^{\frac{1}{2}} \ll W \) and \( s^{a+1}(WN)^{\frac{1}{2}} \ll W^a \),

\[
\mathbb{E} \text{Tr} \left( \frac{A}{2\sqrt{2W}} \right)^{2s} \leq N \frac{(2s)!}{(s + 1)! 4^a} (1 + o(1)) = N \frac{N}{\sqrt{\pi s^4}} (1 + o(1)).
\]

(40)

By the Markov Inequality, we have

\[
\mathbb{P} (\lambda_{\max} > 1 + \epsilon) \leq \frac{\mathbb{E} \lambda_{\max}^{2s}}{(1 + \epsilon)^{2s}} \leq \frac{\mathbb{E} \text{Tr} \left( \frac{A}{2\sqrt{2W}} \right)^{2s}}{(1 + \epsilon)^{2s}} = \frac{N^{\frac{1}{4s}} (1 + o(1))}{(1 + \epsilon)^{2s}}.
\]

(41)

Let \( s \gg \log^2 N \), which implies that \( W \gg N^{\frac{1}{4s-1}} \), then we can use Borel Cantelli Lemma to prove the above theorem.

**Remark 25.** We can use similar argument above to prove the non-symmetric versions for Theorem 10 and Theorem 12.
4 Modified moment method

The modified moment method is originally introduced by S. Sodin (see reference [21], [8], [24]). Though linear eigenvalue statistics with monomial test functions have the combinatorial meaning as paths in the associated graph, the monomials are ill-suited to study local statistics of eigenvalues in the bulk and at the hard edge. So a natural question arises if there exist other polynomial test functions that also have combinatorial representations. Wigner’s Semicircle Law suggests to think about Chebyshev polynomials, since they are orthogonal with respect to the semicircle distribution.

Let us briefly introduce the Chebyshev polynomials of the second kind. They form a complete orthogonal (w.r.t semicircle distribution) basis, which is defined as

\[ U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x), \quad U_0 = 1, \quad U_1 = 2x. \]  

Sodin found that they are indeed related to non-backtracking paths, defined as follows:

**Definition 26.** Given a 0-1 matrix \( G \in \{0,1\}^{N \times N} \), we create the vertex set \( V \) consisting of \( N \) elements labeled as \( 1,2,\ldots,N \) and the edge set \( E = \{e_{ij}|A_{ij} \neq 0, i, j \in V\} \). Conversely, given a graph \( G = \{V,E\} \), we can label the vertices as \( 1,2,\ldots,N \) with \( N = \#V \) and define the associated adjacency matrix \( G_{ij} = 1 \) if \( e_{ij} \in E, 1 \leq i, j \leq N \). Otherwise, \( G_{ij} = 0 \).

**Definition 27.** A path \((u_1,u_2,\ldots,u_{k-1},u_k)\) is called non-backtracking if \( u_j \neq u_{j+2} \) for \( j = 1,2,\ldots,k-2 \) and \( u_j \neq u_{j+1} \) for \( j = 1,2,\ldots,k-1 \). We denote the set of closed non-backtracking paths of length \( k \) as \( Q(k) \).

**Definition 28.** The closed non-backtracking paths of length \( 2n \) whose edges are passed even number of times are called even paths, denoted by \( Q_{even}(2n) \). If edges are passed exactly twice, they are called simple paths, denoted by \( Q_{sim}(2n) \).

Using mathematical induction, we have

**Lemma 29.** Assume \( G = (V,E) \) is a simple \( d \)-regular graph with \( N \) vertices, and \((G)^{N \times N}\) is the associated adjacency matrix. If a sequence of polynomials satisfies: \( Q_1 = x, Q_2 = x^2 - d, \) and for \( k \geq 2 \)

\[ Q_{k+1}(x) = xQ_k(x) - (d-1)Q_{k-1}(x), \]  

then \( Q_k(G)_{uv} \) represents the number of non-backtracking paths of length \( k \) from \( u \) to \( v \) in the graph \( G \), denoted as \( Q_{uv}(k) \).

**Proof.** If there is an edge between \( w \) and \( u \), then \( G_{uw} = 1 \) and otherwise 0. It’s easy to prove for \( k = 1 \).

Since each vertex is not self-looped and has exactly \( d \) neighbour vertices, thus \( G \) is a symmetric 0-1 matrix with diagonal entries 0 and exactly \( d \) 1’s in each row. Recall that \((G^2)_{uv}\) represents the number of all paths of kind \( \{uw,wv\} \) in \( G \). The subset of non-backtracking paths can be written as \( \{uw,wv\}\{u \neq v\} \). So if \( u = v \), \#\( Q_{uv}(2) = 0 = d - d = (G^2)_{uv} - d \).

Otherwise, \#\( Q_{uv}(2) = (G^2)_{uv} = d \). Thus we prove for \( k = 2 \).
Use mathematical induction and assume that for some fixed \( k \geq 2 \) and for any \( j \leq k \), \((Q_j(G))_{uv} = \#Q_{uv}(j)\). Note that each non-backtracking path of length \( k+1 \), \(\{\cdots, x, y, z, w\}\) can be created by adding a proper edge, say \(\{z, w\}\) on the non-backtracking path of length \( k \), say \(\{\cdots, x, y, z\}\). But we need this new edge still satisfies the condition of non-backtracking, i.e. \( y \neq w \). If \( y = w \), then \(\{\cdots, x, y, z, y\}\) is formed with a non-backtracking path \(\{\cdots, x, y\}\) of length \( k - 1 \) and an additional edge \(\{y, z\}\). Note that \( z \neq x \), so we have \( d - 1 \) choices of \( z \). Thus

\[
\#Q_{uv}(k + 1) = \sum_{w \in V} G_{uw}(Q_k(G))_{uw} - (d - 1)(Q_{k-1}(G))_{uv} \\
= (G \cdot Q_k(G))_{uv} - (N - 2)(Q_{k-1}(G))_{uv} = (Q_{k+1})_{uv}
\]

Comparing (43) with (42), choose

\[
Q_n(x) = (N - 2)^\frac{n}{2} U_n \left( \frac{x}{2\sqrt{N - 2}} \right) - (N - 2)^\frac{n-2}{2} U_{n-2} \left( \frac{x}{2\sqrt{N - 2}} \right). \tag{44}
\]

### 4.1 Estimating the number of non-backtracking paths

For simplicity, let us consider a Bernoulli Wigner matrix \( A^b \), whose diagonal elements are zeros and off-diagonal variables are i.i.d. symmetric Bernoulli random variables, i.e \( \mathbb{P}(A^b_{ij} = 1) = \mathbb{P}(A^b_{ij} = -1) = \frac{1}{2} \). We have the important property that

\[
(A^b)_{ij}^2 = 1, \mathbb{E}(A^b)_{ij}^{2k} = 1, \mathbb{E}(A^b)_{ij}^{2k+1} = 0. \tag{45}
\]

Even though \( A^b \) is no longer 0-1 matrix, since \( (A^b)^2_{ij} = 1 \), we still have the similar representation as in Lemma 29. Since each off-diagonal entry is non-zero with probability 1, the graph associated with \( A^b \) is the simple complete graph, and \( d = N - 1 \). Given the same conditions in Lemma 29, \( d = N - 1 \), if \( n \in \mathbb{N} \), then

\[
(Q_n(A^b))_{uv} = \sum_{Q_{uv}(n)} A^b_{uu_1}A^b_{u_1u_2} \cdots A^b_{u_{n-1}v}. \tag{46}
\]

where \( Q_{uv}(n) \) represents the set of non-backtracking paths of length \( n \) from \( u \) to \( v \) in the graph \( G \). By (46) and (44), we have

\[
(N - 2)^\frac{n}{2} \text{Tr} U_n \left( \frac{A^b}{2\sqrt{N - 2}} \right) - (N - 2)^\frac{n-2}{2} \text{Tr} U_{n-2} \left( \frac{A^b}{2\sqrt{N - 2}} \right) = \sum_{u \in V} \left( \sum_{Q_u(n)} A^b_{uu_1}A^b_{u_1u_2} \cdots A^b_{u_{n-1}u} \right), \tag{47}
\]

where \( Q_u(n) \) represents the set of non-backtracking paths of length \( n \) starting from \( u \) and ending with \( u \) in the graph \( G \).
Summing up (47) with respect to \( n \) and taking expectation, because of (45), then we get

\[
(N - 2)^n \mathbb{E} \text{Tr} U_{2n} \left( \frac{A^b}{2\sqrt{N-2}} \right) = \sum_{k=1}^{n} \#Q_{\text{even}}(2k) + N, \tag{48}
\]

where \( Q_{\text{even}} \) is defined in Definition 28.

To count \( \#Q_{\text{even}}(2n) \), the definition of matched paths was introduced in [8]. For the convenience of the reader, this section is mostly copied, up to change of notation, from [8].

**Definition 30.** A matching of any even non-backtracking path \( q_{2n} \) is an involution of \( \{0, 1, 2, \cdots, 2n - 1\} \) without fixed points, such that every edge \( (u, v) \) is matched either to a coincident edge \( (u, v) \) or to \( (v, u) \). A path together with a matching is matched path. The set of matched paths of length \( 2n \) is defined as \( Q_{\text{mat}}(2n) \).

**Remark 31.** For simple non-backtracking paths, each edge is passed exactly twice, so the matching is unique.

It’s obvious that

\[
\#Q_{\text{sim}}(2n) \leq \#Q_{\text{even}}(2n) \leq \#Q_{\text{mat}}(2n). \tag{49}
\]

To count \( \#Q_{\text{mat}}(2n) \), we can classify matched paths by using (weighted) diagrams.

**Definition 32.** A diagram is an undirected multigraph \((V, E)\) together with a non-backtracking circuit \((u_0, u_1, \cdots, u_0)\), and the degree of \( u_0 \) is 1, the degrees of all other vertices are equal to 3. A weighted diagram is a diagram together with a weight function \( w(E) \), taking value in \( \{-1, 0, 1, 2, \cdots\} \).

Next, several steps were introduced in [8] to transform the matched paths to the corresponding weighted diagrams.

**Step 1** Starting from the non-backtracking path \((u_0, u_1, \cdots, u_0)\), view each pair of matched edges as a single undirected edge;

**Step 2** If \( u_0 \) has degree \( \geq 2 \), then we create a fake start vertex \( r \), and assign \( w(ru_0) = -1 \);

**Step 3** For all vertices of degree \( > 3 \), we can create fake edges to separate the corresponding intersections as shown in the figure below. This step will be used repeatedly until all degrees are less than or equal to 3;

**Step 4** We assign -1 as the weight of the fake edges created in step 3.

**Step 5** Erase all vertices(not including the associated vertices of fake edges) of degree 2 and assign the number of erased vertices on the new edge as its weight.

19
In addition, the automation to create diagrams will consist of three basic operations.

Operation 1 Creation of a new loop;

Operation 2 Annihilation of the j-th loop;

Or,

Operation 3 Creation and annihilation.
We use $D(s)$ to denote the set of diagrams that need $s$ steps to generate.

**Lemma 33.** Each diagram in $D(s)$ has $3s - 1$ edges and $2s$ vertices, and has a total weight of $n - 3s + 1$. and for some constant $C > 1$,

$$\left(\frac{s}{C}\right)^s \leq \#D(s) \leq (Cs)^s. \tag{50}$$

For example, for $s=1$ and $s=2$, we have following diagrams.

![Diagram](image)

Figure 1: Some 1-diagrams: $s = 1$ (left), $s = 2$ (center, right)

Next, we can use diagrams to count $Q_{\text{even}}$.

**Lemma 34.** If $1 \ll n \ll N^{1/3}$,

$$\#Q_{\text{even}}(2n) \sim \#Q_{\text{sim}}(2n) \sim \#Q_{\text{mat}}(2n) \sim nN^n. \tag{51}$$

**Proof.** By Lemma 33, the number of diagrams in $D(s)$ has an upper bound $(Cs)^s$. Also, each diagram contains $\#V = 2s$ vertices and $\#E = 3s - 1$ edges. Denote the number of vertices in the original matched path as $\#\tilde{V}$ and number of edge as $\#\tilde{E} = n$. Then we have $\#\tilde{E} - \#E = \#\tilde{V} - \#V = \sum_{e \in E} w(e)$. Thus $\sum w(e) = n - \#E = n - 3s + 1$. Since $w(e) \geq -1$, there are $\binom{n+3s-1}{3s-2}$ different ways to allocate total weight among $3s - 1$ edges. Choose $s_0 \ll n^{1/2}$. For $s < s_0$, $\binom{n+3s-1}{3s-2} = \left(\frac{n^{3s-2}}{(3s-2)!}\right)(1 + o(1))$. Once the weighted diagram is fixed, it corresponds to at most $N^{\#\tilde{V} + \sum w(e)} = N^{n-s+1}$ paths. If $1 \ll n \ll N^{1/3}$, we have

$$\#Q_{\text{mat}} \leq \sum_{1 \leq s \leq s_0} \#D(s)N^{n-s+1}\frac{n^{3s-2}}{(3s-2)!}(1+o(1)) \leq nN^n \left(\#D(1) + \sum_{1 \leq s' \leq n^{1/2}} \frac{1}{(2s')!}\left(\frac{Cn^3}{N}\right)^{s'}\right)$$

$$\leq nN^n \exp\left(\frac{Cn^3}{N^{1/2}}\right)(1 + o(1)) = nN^n(1 + o(1)). \tag{52}$$

For the lower bound, it’s easy to count the number of diagrams with positive weight. There are $\binom{n-3s+1}{3s-2} = \left(\frac{n^{3s-2}}{(3s-2)!}\right)(1 + o(1))$ different ways to allocate total weight and each diagram corresponds to exactly $N(N-1) \cdots (N-(\#\tilde{V} + \sum w(e)) + 1) = N^{n-s+1}(1 + o(1))$ different paths. Therefore,

$$\#Q_{\text{sim}} \geq \sum_{1 \leq s \leq s_0} \#D(s)\frac{n^{3s-2}}{(3s-2)!}N^{n-s+1}(1 + o(1)) = nN^n(1 + o(1)). \tag{53}$$
Applying Lemma 34 to (37), we have

**Proposition 35.** If $1 \leq n \ll N^{1/3}$,

$$\mathbb{E} \text{Tr} \left( U_{2n} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) \leq Cn, \quad \mathbb{E} \text{Tr} \left( U_{2n+1} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) = o_N(1). \tag{54}$$

If we further assume that $n \gg 1$, then

$$\mathbb{E} \text{Tr} \left( U_{2n} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) \sim n. \tag{55}$$

Proposition 35 implies Wigner Semicircle Law since

$$\mathbb{E} \int_{-\infty}^{\infty} U_n \left( \frac{x}{N} \right) d\mu_{ESD}(A^b) = 1 \quad \text{and} \quad \mathbb{E} \int_{-\infty}^{\infty} U_n \left( \frac{x}{N} \right) d\mu_{sc}. \tag{56}$$

Next, we study the product of traces in the form of

$$\mathbb{E} \text{Tr} \left( U_{n_1} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) \cdot \text{Tr} \left( U_{n_2} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) \cdots \text{Tr} \left( U_{n_k} \left( \frac{A^b}{2\sqrt{N-2}} \right) \right).$$

From (44), we notice that $\text{Tr} \left( U_n \left( \frac{A^b}{2\sqrt{N-2}} \right) \right) \sim Q_n \left( \frac{A^b}{\sqrt{N}} \right)$. And by (46), we have

$$\text{Tr} \left( Q_{n_1} \left( \frac{A^b}{\sqrt{N}} \right) \right) \cdot \text{Tr} \left( Q_{n_2} \left( \frac{A^b}{\sqrt{N}} \right) \right) \cdots \text{Tr} \left( Q_{n_k} \left( \frac{A^b}{\sqrt{N}} \right) \right) = \sum_{\mathcal{T}(n_1, n_2, \ldots, n_k)} A^{b_{u_1u_1}} A^{b_{u_2u_2}} \cdots A^{b_{u_{n_k-1}u_{n_k}}},$$

where $\mathcal{T}(n_1, n_2, \ldots, n_k)$ is the set of non-backtracking k-paths $(Q_1, Q_2, \ldots, Q_k)$ defined as below, where $Q_i$ is a non-backtracking circuit of length $n_i$, $1 \leq i \leq k$.

**Definition 36.** $(Q_1, Q_2, \ldots, Q_k)$ is k-path of type $(n_1, n_2, \ldots, n_k)$, if $Q_i \in Q(n_i)$. We denote the set of k-paths as $\mathcal{T}(n_1, n_2, \ldots, n_k)$ and $\mathcal{T}^{(k)}(n)$ for short if $n_1 = n_2 = \cdots = n_k = n$. We also denote the set of even k-paths where each edge is passed even number of times by $\mathcal{T}_{\text{even}}$. And denote the set of simple k-paths where each edge is passed exactly twice by $\mathcal{T}_{\text{sim}}$.

After rescaling by $N^{-\frac{1}{2}}$ and taking expectation of Proposition 57, by (45), we have

**Lemma 37.** Let $n_1 + n_2 + \cdots + n_k = 2n$,

$$\mathbb{E} \text{Tr} Q_{n_1} \left( \frac{A^b}{\sqrt{N}} \right) Q_{n_2} \left( \frac{A^b}{\sqrt{N}} \right) \cdots Q_{n_k} \left( \frac{A^b}{\sqrt{N}} \right) = \frac{1}{N^n} \# \mathcal{T}_{\text{even}}(n_1, n_2, \ldots, n_k),$$

If $n_1 + n_2 + \cdots + n_k$ is odd, the expectation is 0.
By analogy with Definition 30 we define the matched k-path \( T_{\text{mat}}(n_1, n_2, \cdots, n_k) \). And
\[
\# T_{\text{sim}}(n_1, n_2, \cdots, n_k) \leq \# T_{\text{even}}(n_1, n_2, \cdots, n_k) \leq \# T_{\text{mat}}(n_1, n_2, \cdots, n_k).
\]

To count the number of elements in \( T_{\text{mat}} \), we generalize the Definition 33 to k-diagram. When we generate k-diagram using three basic operations on Page[] if we need \( s_i \) steps to generate the \( i \)-th loop in the diagram, \( 1 \leq i \leq k \), then it belong to the set \( D(s_1, s_2, \cdots, s_k) \), and define
\[
\# D_k(s) = \sum_{s=s_1+s_2+\cdots+s_k,s_i \geq 1} \# D(s_1, s_2, \cdots, s_k).
\]

Similar as Lemma 37, we have

**Lemma 38.** Each diagram in \( D(s_1, s_2, \cdots, s_k) \) has \( 3s - k \) edges and \( 2s \) vertices. And
\[
\frac{(s)^{s+k-1}}{(k-1)!} \leq \# D_k(s) \leq \frac{(Cs)^{s+k-1}}{(k-1)!}
\]
for some universal \( C > 1 \).

![Diagram](image)

**Figure 2:** Some 2-diagrams: \( s = 2 \) (left, center), \( s = 3 \) (right)

We use coefficient \( c_i(e), i = 1, 2, \cdots, k \) to indicate the number of appearances of the edge of the diagram appearing in \( i \)-th subpath \( Q_i(n_i) \). It takes values in \( \{0, 1, 2\} \).

\[
\sum_{e \in E} c_i(e)w(e) = n_i - 2(3s_i - 1), i = 1, 2, \cdots, k
\]
\[
c_1(e) + c_2(e) + \cdots + c_k(e) = 2
\]

Therefore, let \( n_1 + n_2 + \cdots + n_k = 2n \), we have \( \sum_{e \in E} w(e) = n - 3s + k \). Note that \( c_k(c_k) = 2 \). Since \( w(e) \geq -1, \sum'_{e \in E} w(e) \leq n - 3s + 2k \), where \( \sum' \) is summing over all edges in the diagram except the first one in each subpath. For \( w(e) \geq -1 \), if \( s \ll n^{\frac{2}{3}} \), there are at most
\[
\binom{n - 3s + 2k + 2(3s - 2k)}{3s - 2k} = \binom{n + 3s - 2k}{3s - 2k} \sim \frac{n^{3s-2k}}{(3s-2k)!}
\]
different ways to allocate total weight. By Proposition 38, once the weighted diagram is fixed, it corresponds to at most \( N^\#V + \sum_{e \in E} w(e) = N^{n-s+k} \) paths. Choose \( s_0 \ll n^{\frac{2}{3}} \). If \( 1 \ll n \ll N^{\frac{1}{4}} \),
then

\[
\#\mathcal{T}_{\text{mat}} \leq \sum_{k \leq s \leq s_0} \#D_k(s)N^{n-s+k} \frac{n^{3s-2k}}{(3s-2k)!}(1+o(1)) \leq \frac{n^k N^n}{k!} \left( \#D_k(k) + \sum_{1 \leq s' \ll n^{\frac{1}{2}}} \frac{1}{(2s')!} \left( \frac{C n^3}{N^\frac{3}{2}} \right)^{s'} \right)
\]

\[
\leq n^k N^n \frac{\#D_k(k)}{k!} \exp \left( \frac{C n^3}{N^\frac{3}{2}} \right) (1 + o(1)) \sim n^k N^n \frac{\#D_k(k)}{k!} \leq (Cn)^k N^n.
\]

Similarly, we can prove that

**Proposition 39.** Let \( n_1 + n_2 + \cdots + n_k = 2n \), and \( 1 \leq n \ll N^{\frac{1}{3}} \). Then

\[
\mathbb{E} \text{Tr} Q_{n_1} \left( \frac{A^b}{\sqrt{N}} \right) Q_{n_2} \left( \frac{A^b}{\sqrt{N}} \right) \cdots Q_{n_k} \left( \frac{A^b}{\sqrt{N}} \right) \leq (Cn)^k.
\]

*If we further assume \( n \gg 1 \), then

\[
\mathcal{T}_{\text{even}}(n_1, \cdots, n_k) \sim \mathcal{T}_{\text{sim}}(n_1, \cdots, n_k),
\]

and the main contribution comes from the simple paths corresponding to diagrams in \( D_k(k) \).

### 4.2 CLT for the linear eigenvalue statistics of Wigner matrices

Below, we extend these ideas to prove the CLT for the fluctuation.

**Theorem 40.** If \( 1 \ll n \ll N^{1/3} \), then

\[
\text{Var} \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) \right) \sim n^2. \tag{58}
\]

*Proof.*

\[
\text{Var} \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) \right) = \mathbb{E} \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) - \mathbb{E} \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) \mathbb{E} \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right)
\]

\[
= \frac{1}{N^n} \sum \mathbb{E} A^b_{u_1v_1} A^b_{u_1u_2} \cdots A^b_{u_{n-1}u_n} A^b_{v_1v_2} \cdots A^b_{v_{n-1}v_n} - \mathbb{E} A^b_{u_1v_1} A^b_{u_1u_2} \cdots A^b_{u_{n-1}u_n} \mathbb{E} A^b_{v_1v_2} \cdots A^b_{v_{n-1}v_n}
\]

where is the sum is over the set of non-backtracking 2-path \((Q_1, Q_2) \in \mathcal{T}(n, n)\) satisfying the following conditions:

(A1) \( Q_1 \) and \( Q_2 \) have at least one edge in common;

(A2) Each edge appears in \((Q_1, Q_2)\) an even number of times.
The set of \((Q_1, Q_2) \in \mathcal{T}(n, n)\) satisfying conditions (A1) and (A2) is denoted \(\mathcal{T}(n, n)\). Note that if we consider diagrams corresponding to \(\mathcal{T}(n, n)\), denoted \(D^c(s)\), then \(#D^c(2) = 2\). See the leftmost diagram in Figure 2, either in same direction or different directions. When we take \(k = 2\) and \(s \ll n^{\frac{1}{3}}\),

\[
\left(\frac{n + 3s - 4}{3s - 4}\right) \sim \frac{n^{3s-4}}{(3s-4)!} (1 + o(1))
\]

\[
\#\mathcal{T}^c_{\text{mat}} \leq \sum_{2 \leq s \leq s_0} D^c_2(s) N^{n-s+2} \frac{n^{3s-4}}{(3s-4)!} (1 + o(1)) \sim D^c_2(2) N^n \frac{n^2}{2!} = n^2 N^n
\]

On the other hand, we consider the non-backtracking 2-path corresponding to some diagram of \(D(2)\), i.e. the leftmost one in Figure 2. Thus,

\[
\#\mathcal{T}^c_{\text{sim}} \geq D^c_2(2) \left(\frac{n}{2}\right) N^n (1 + o(1)) \sim n^2 N^n.
\]

On the other hand, we consider the non-backtracking 2-path corresponding to some diagram of \(D(2)\), i.e. the leftmost one in Figure 2. Thus,

\[
\#\mathcal{T}^c_{\text{sim}} \geq D^c_2(2) \left(\frac{n}{2}\right) N^n (1 + o(1)) \sim n^2 N^n.
\]

\(\mathbb{E} \left( \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N^2}} \right) \right) - \mathbb{E} \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N^2}} \right) \right) \right)^k\) is the sum over the set of the non-backtracking k-path \((Q_1, Q_2, \cdots, Q_k) \in \mathcal{T}^c(n)\) satisfying the following conditions:

(B1) Each subpath \(Q_i\) must share at least one common edge with some other subpath \(Q_j\);
(B2) Each edge appears in \((Q_1, Q_2, \cdots, Q_k)\) an even number of times.

The set of non-backtracking k-path satisfying condition (B1) and (B2) is named \(\mathcal{T}^c(n)\). We can use the same machinery to show that if \(k\) is odd, \(\mathcal{T}^c_{\text{mat}}(n) = o(N^{\frac{2k}{k+1}})\). If \(k\) is even, \(D^c_k(k)\) consists of \(\frac{k}{2}\) pairs of the diagrams in \(D(2)\), and \(#D^c_k(k) = 2^k(k-1)!!\). If we further assume \(1 \ll n \ll N^{\frac{1}{3}}\), then diagrams in \(D_k(k)\) make the main contribution of \(\mathcal{T}^c_{\text{mat}}(n)\)

\[
\mathcal{T}^c_{\text{even}}(n) \sim \mathcal{T}^c_{\text{sim}}(n) \sim 2^k (k-1)!! \left(\frac{n}{2}\right) \frac{k}{2} N^{\frac{n^k}{k}} = (k-1)!! n^k N^{\frac{n^k}{k}}
\]

Therefore, we can show that,

\[
\mathbb{E} \left( \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N^2}} \right) \right) - \mathbb{E} \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N^2}} \right) \right)^k \rightarrow \begin{cases} (k-1)!! & \text{k is even} \\ 0 & \text{k is odd} \end{cases} \text{ } (59)
\]

This implies the CLT for the linear statistics.

**Theorem 41.** If \(1 \ll n \ll N^{\frac{1}{3}}\), then \(\frac{1}{n} \left( \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) - \mathbb{E} \text{Tr} Q_n \left( \frac{A^b}{\sqrt{N}} \right) \right)\) converges in distribution to normal law with expectation zero and variance 1.
Corollary 42. For $1 \ll n_1 \ll N^{1/3}$, and $n_1 = \gamma n_2$, where $0 < \gamma < 1$

$$\# \mathcal{T}_{\text{sim}}(n_1, n_2) \sim \# \mathcal{T}_{\text{even}}(n_1, n_2) \sim n_1^2 N^{n_1 + n_2}$$

Corollary 43. Let $1 \ll n_1 \ll N^{1/3}$, and $n_1 = \gamma n_2$, where $0 < \gamma < 1$.

If $n_2 - n_1$ is even, then

$$\text{Cov} \left( \text{Tr} Q_{n_1} \left( \frac{A^b}{\sqrt{N}} \right), \text{Tr} Q_{n_2} \left( \frac{A^b}{\sqrt{N}} \right) \right) \sim n_1^2.$$  \hfill (60)

If $n_2 - n_1$ is odd, then

$$\text{Cov} \left( \text{Tr} Q_{n_1} \left( \frac{A^b}{\sqrt{N}} \right), \text{Tr} Q_{n_2} \left( \frac{A^b}{\sqrt{N}} \right) \right) = o_N(1).$$  \hfill (61)

Remark 44. Choose $1 \ll n_0 \ll N^{1/3}$ and $n_i = c_i n_0 (i = 1, 2, \cdots, k)$, where $c_i$ are positive real numbers. Then the joint distribution of $\text{Tr} Q_{n_i} \left( \frac{A^b}{\sqrt{N}} \right) - \mathbb{E} \text{Tr} Q_{n_i} \left( \frac{A^b}{\sqrt{N}} \right)$ converges to multivariate normal distribution. If further assuming the difference of $n_i$ and $n_j$ are odd, then they are asymptotically independent.

Remark 45. We can generalize the above results to random matrices with non-Bernoulli entries, see part III in [8].

4.3 Generalization to band random matrices

For simplicity, we consider a Bernoulli band random matrix model, denoted as $H^b$. Given a band random matrix ensemble in Definition 8, and set all diagonal entries zero and all off-diagonal entries follow symmetric Bernoulli distribution, i.e.

$$\mathbb{P} (H^b_{uv} = 1) = \mathbb{P} (H^b_{uv} = -1) = \frac{1}{2}.$$  \hfill (46)

Since the Wigner matrix model is a special case of band random matrix model, we can extend Sodin’s technique to prove CLT for band random matrices. The associated graph for band Wigner matrices is $2W$ regular. Comparing with (44) in Wigner model, we choose

$$Q_n(x) = (2W - 1)^{\frac{n}{2}} U_n \left( \frac{x}{2\sqrt{2W - 1}} \right) - (2W - 1)^{\frac{n-2}{2}} U_{n-2} \left( \frac{x}{2\sqrt{2W - 1}} \right).$$  \hfill (62)

By (46), $\mathbb{E} \text{Tr} Q_n(A^b)$ is related to the number of closed non-backtrack even paths of length $n$ in the $2W$ regular associated graph. Following is our main theorem. We can also generalize to non-Bernoulli band random matrices, see part III in [8].

Theorem 46. Let $1 \ll W \ll N$ and $1 \ll n \ll \min \{ \frac{N^2}{W}, \frac{W^{1/3}}{3} \}$,

$$\mathbb{E} \text{Tr} \left( \frac{W}{N} U_{2n} \left( \frac{H^b}{2\sqrt{2W}} \right) \right) \sim \sqrt{\frac{6}{\pi} n^{3/2}}.$$  \hfill (63)
Moreover,
\[
\frac{1}{n^3} \left( \text{Tr} \left( \sqrt{\frac{W}{N}} Q_n \left( \frac{H^b}{\sqrt{2W}} \right) \right) - \mathbb{E} \text{Tr} \left( \sqrt{\frac{W}{N}} Q_n \left( \frac{H^b}{\sqrt{2W}} \right) \right) \right)
\]
converges in distribution to normal law with expectation zero and variance \(\sqrt{\frac{2}{3\pi}}\).

Before we prove the above theorem, we first introduce some preliminary details. The associated graph of \(H^b\) is \(2W + 1\) regular and all vertices as starting point are equivalent. Assuming without loss of generality that \(N\) is odd and starting point is 0, we can label the rows and columns of \(H^W\) by \(\{-\frac{N-1}{2}, -\frac{N-1}{2} + 1, \ldots, 0, 1, \ldots, \frac{N-1}{2}\}\). Let \(x_i\) be the jump at each step, taking value from \(-W\) to \(W\). All paths of length \(n\) starting at 0 and ending at \(R\) without reaching the corner of the matrix can be written as:

\[
-\frac{N-1}{2} \leq x_1 + x_2 + \cdots + x_k \leq \frac{N-1}{2}; \\
-W \leq x_k \leq W, \quad k = 1, 2, \ldots, n; \\
x_1 + x_2 + \cdots + x_n = R \\
x_i \neq 0
\]

If we assume that \(\sqrt{n} \ll \frac{N}{W}\) and \(n \ll W\), then the polytope defined by the solutions of the second inequality in (64) is mostly contained in that of the first one. Then the inequalities can be viewed as

\[
-W \leq x_k \leq W, \quad k = 1, 2, \ldots, n - 1; \\
R - W \leq x_1 + x_2 + \cdots + x_{n-1} \leq R + W. \tag{65}
\]

We use \(S_n(R)\) to denote all paths of length \(n\) with distance \(R\), which are in one-to-one correspondence with the integer solutions of (65). We first consider closed paths \((R = 0)\) and use the Monte Carlo method. Let \(x_i^W = \frac{x_i}{W}\), taking value uniformly at \(\{-1, -1 + \frac{1}{W}, \ldots, 1 - \frac{1}{W}, 1\}\). It converges to uniform distribution on \([-1, 1]\) as \(W \to \infty\). The mean is 0 and the variance is \(\frac{1}{3}\).

\[
\frac{\#\{S_n(0)\}}{(2W + 1)^{n-1}} = \mathbb{P}(-1 \leq x_1^W + x_2^W + \cdots + x_{n-1}^W \leq 1) + o_W(1). \tag{66}
\]

Recall the Local Central Limit Theorem, the proof can be found in [7].

**Proposition 47.** Let \(\{X_i\}_{i \geq 1}\) be i.i.d with \(\mathbb{E}X_i = 0\), \(\text{Var}X_i = \sigma^2\), and have a common characteristic function \(\phi(t)\) that \(|\phi(t)| < 1\) for all \(t \neq 0\). Let \(S_n = X_1 + X_2 + \cdots + X_n\), if \(\frac{x_n}{\sqrt{n}} \to x\) and \(a \leq b\),

\[
\sqrt{n}\mathbb{P}(S_n \in (x_n + a, x_n + b)) \to (b - a) \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{67}
\]
Applying Proposition 47 to (66), we have

\[
\frac{\#\{Q_n(0)\}}{(2W)^n} = \sqrt{\frac{3}{2n\pi}} \frac{1}{W} + o\left(\frac{1}{W}\right).
\]  

(68)

Actually, for all \(R \ll \sqrt{n}W\), \(\frac{\#\{Q_n(R)\}}{(2W)^n} \sim \sqrt{\frac{3}{2n\pi}} \frac{1}{W}\). If \(R = O\left(\sqrt{n}W\right)\), then \(\frac{\#\{Q_n(R)\}}{(2W)^n} = O\left(\frac{1}{W} \exp\left\{\frac{-R^2}{nW^2}\right\}\right)\). If \(R = \O\left(\sqrt{n}W\right)\), then

\[
\#\{Q_n(R)\} \leq \frac{\#\{Q_n(0)\}}{(2W)^n}.
\]

(69)

For closed paths without intersection, denoted as \(S_n\), we notice that, for some \(C\),

\[
\#\{S_n(0)\} \geq \#\{S_n\} \geq (1 - C\frac{n^2}{W})\#\{S_n(0)\}.
\]

(69)

If we further assume \(n \ll W^{\frac{1}{2}}\), then

\[
\#\{S_n\} = (1 - o(1))\#\{S_n(0)\}.
\]

(70)

**Proof.** (Theorem 46) We use the same definitions of matched paths and weighted diagrams in [8] and define standard (simple) diagram if it has a binary tree as the branch and circles as the leaves. Let us denote the number of leaves as \(s_0\). Then we have \(2s_0\) vertices and \(3s_0 - 1\) edges.

The lower bound is not hard if we only consider standard diagrams with positive weights. Since \(1 \ll n \ll W^{\frac{1}{2}}\),

\[
Q_{sim}(2n) \geq \sum_{n_1+n_2=n;n_1\geq 0;n_2\geq 3} N(2W)^n \sqrt{\frac{3}{2n_2\pi}} W^{-1}(1 + o(1))
\]

\[
Q_{sim}(2n) = \frac{N(2W)^n}{W} \sum_{k=3}^{n} \frac{1}{\sqrt{k}}(1 + o(1)).
\]

(71)

Then we have:

\[
\mathbb{E} \text{Tr}\left(\frac{W}{N}U_{2n}\left(\frac{H^b}{2\sqrt{2W}}\right)\right) \geq \sqrt{\frac{6}{\pi}} n^{\frac{3}{2}}(1 + o(1)).
\]

(72)

To get the upper bound, we introduce a procedure of deleting edges to deform every diagram to standard (simple) diagram. One example is shown as figure below. And the number of deleted edges is denoted as \(s_1\). Every time we delete one edge, we will lose 2 vertices and 3 edges. It’s easy to check that the number of steps needed to generate the diagram is \(s_1 + s_0 = s\), \(#V = 2s\), \(#E = 3s - 1\), and the total weight \(n - 3s + 1\).
Next, we separate weight in 3 parts: \( w_t \) for trees, \( w_c \) for circles, and \( w_d \) for deleted edges.

\[
Q_{mat} \leq \sum_{s \geq 1} \sum_{s_0 + s_1 = s} \sum_{w_t + w_c + w_d = n - 3s + 1} \sum_{i \geq 1} N C_{s_0 - 1} \left( \frac{w_t + 4s_0 - 2}{2s_0 - 2} \right) (2W)^{w_t + 2s_0 - 1}.
\]

\[
2^{s_0} \left( \frac{3}{2\pi} \right)^{s_0} (2W)^{w_c + s_0 + 2s_1} \frac{1}{n^{s_0 - 1} W^{s_0}} \left( \frac{w_d + 2s_1}{s_1 - 1} \right) n^{2s_1} s_1^{s_1} (2W)^w_d
\]

\[
\leq \sum_{s \geq 1} \sum_{s_0 + s_1 = s} \frac{N (2W)^n}{W} C_{s_0 - 1} s_1^{s_1} \frac{4^{s_0}}{2^{s_0 + 1}} \left( \frac{3}{2\pi} \right)^{s_0} n^{3s - \frac{5}{2}} (2W)^{s_1 - 1}.
\]  

(73)

Thus the diagram with \( s_0 = 1, s_1 = 0 \) makes the main contribution:

\[
\mathbb{E} Tr \left( \frac{W}{N} U_{2n} \left( \frac{H^W}{2\sqrt{2W}} \right) \right) \leq \sqrt{\frac{6}{\pi n^2}} (1 + o(1)).
\]  

(74)

In the same way, the leftmost diagram of Figure 2 makes the main contribution of 2-paths satisfying (A1) and (A2). \( k \) pairs of this diagram make the main contribution of the \( 2k \)-paths satisfying (B1) and (B2).

\[
T_{even}^c = T_{sim}^c (1 + o(1)) = \sum_{n_1 + n_2 = n; n_1 \geq 0; n_2 \geq 3} 2N (2W)^n \frac{n_2}{2} \sqrt{\frac{3}{2n_2 \pi}} W^{-1} (1 + o(1))
\]

\[
= \frac{N (2W)^n}{W} \sqrt{\frac{3}{2\pi}} \left( \sum_{k=3}^{n} \sqrt{k} \right) (1 + o(1)).
\]  

(75)

Thus we have:

\[
\text{Var} Tr \left( \sqrt{\frac{W}{N}} Q_n \left( \frac{H^b}{\sqrt{2W}} \right) \right) = \sqrt{\frac{2}{3\pi}} n^3 (1 + o(1)).
\]  

(76)
References


