Independence of linear spectral statistics and the point process at the edge of Wigner matrices

DEBAPRATIM BANERJEE
Indian Statistical Institute, Kolkata
deobratimbanerjee@gmail.com

October 17, 2022

Abstract

In the current paper we consider a Wigner matrix and consider an analytic function of polynomial growth on a set containing the support of the semicircular law in its interior. We prove that the linear spectral statistics corresponding to the function and the point process at the edge of the Wigner matrix are asymptotically independent when the entries of the Wigner matrix are sub-Gaussian. The main ingredient of the proof is based on a recent paper by Banerjee [6]. The result of this paper can be viewed as a first step to find the joint distribution of eigenvalues in the bulk and the edge.

1 Introduction

Since the groundbreaking discovery of Wigner [16], Wigner matrices have been a topic of key interest in the mathematics and physics communities. Later on these matrices proved to be important for many models in engineering, high dimensional statistics and many other branches.

In particular, a Wigner matrix is a $n \times n$ symmetric (hermitian) matrix with real (complex) entries where the entries of the upper diagonal part are i.i.d. with mean 0 and variance $\frac{1}{n}$. One is interested in the eigenvalue distribution of the matrix when the dimension grows to infinity. The study of eigenvalues of Wigner matrices started with characterizing the limiting distribution of the histogram of the eigenvalues. This is done in the seminal papers of Wigner [16] and Wigner [15]. It is known that this limiting distribution exists and coined as the famous semicircular distribution. In particular, it is given by the following density.

$$f(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4-x^2} & \text{when } |x| \leq 2 \\ 0 & \text{otherwise} \end{cases}$$ (1.1)
However after specifying the spectral distribution, there has been a remarkable advancement in this topic. Several interesting questions related to the eigenvalues have been raised and have been subsequently solved.

The main objective of this paper is to relate two such branches. The first one is the study of $\text{Tr}[g(W)]$ for an analytic function $g$ containing the support of the semicircular law and the second one is the study of eigenvalues near the support of the semicircular law. Here $W$ is the Wigner matrix. The second one is about the study of the eigenvalues near the edge of the spectrum.

The study of linear spectral statistics was initiated in the paper of Bai and Silverstein [4]. Here one considers a generalized Wishart matrix and the linear spectral statistics of a general analytic function. There have been several papers in this topic after that. For Wigner matrices the CLT of linear spectral statistics was studied in Bai et al. [3]. These results are proven using a rigorous analysis of the Stieltjes transform of the empirical distribution function of the eigenvalues of the Wigner/Wishart matrix. On the other hand a combinatorial approach for solving the similar problem was initiated in the paper by Anderson and Zeitouni [1].

On the other hand, study of the eigenvalues near the edge of the Wigner matrix is also well known. Here we look at the point processes of the eigenvalues near $\pm 2$ which is the support of the semicircular distribution. Using the explicit distribution of the eigenvalues the fluctuation of the largest eigenvalue of the Wigner matrix was first proved in the seminal works of Tracy and Widom [12], [13] in the Gaussian case. In particular it is proved that

$$\mathbb{P} \left[ n^{\frac{2}{3}} (\lambda_{1,n} - 2) \leq s \right] \to F_{\beta}(s)$$

where the Tracy-Widom distribution functions $F_{\beta}$ can be described by Painleve equations, and $\beta = 1, 2, 4$ corresponds to Orthogonal/Unitary/Symplectic ensemble, respectively. Here $\lambda_{1,n} \geq \ldots \geq \lambda_{n,n}$ are the eigenvalues of the Wigner matrix $W$. The joint distribution of $k$ largest eigenvalues can be expressed in terms of the Airy kernel, which was shown by Forrester [8]. In general the joint distribution of $(\lambda_{1,n}, \ldots, \lambda_{k,n})$ will also converge after proper rescaling and centering. In particular

$$\mathbb{P} \left[ n^{\frac{2}{3}} (\lambda_{1,n} - 2) \leq s_1, \ldots, n^{\frac{2}{3}} (\lambda_{k,n} - 2) \leq s_k \right] \to F_{\beta,k}(s_1, \ldots, s_k).$$

The $k$ dimensional distribution $F_{\beta,k}$ will also be coined as Tracy Widom distribution.

After the Gaussian case being solved, a significant amount of work has been done for general non-Gaussian entries. This problem is known as edge universality. The edge universality of the Wigner matrix was first proved in the paper Soshnikov [10] through combinatorial methods. He assumed that the distributions of the entries of the matrix are sub-Gaussian and symmetric. The universality for the non-symmetric entries was first proved by Tao and Vu [11]. Here one assumes that entries have vanishing third moment and the tail decays exponentially. Then there was a different approach by Erdős, Yau and others by analyzing the resolvent matrix. One might look at Erdős and Yau [7]. Very recently Banerjee [6] proved edge universality through a new combinatorial approach where he removed assumption of symmetry of the entries.
However to the best of our limited knowledge nothing has been proved for the joint distribution of a generic linear spectral statistics and the eigenvalues near the edge. In this paper following the combinatorial approach of Banerjee [6] and Anderson and Zeitouni [1] we prove that they are asymptotically independent. This result can be seen as a first step to prove the joint distribution of eigenvalues near a typical point in the interior of the support of the semicircular law and the eigenvalues near the edge. This is due to the following reason: we choose a point $x \in (-2, 2)$ and look at an interval of length $O\left(\frac{1}{n}\right)$ around $x$ and look at the eigenvalues falling in the interval. This type of statistics can be approximated by a function $g_{x,n}(\cdot)$ where $g_{x,n}(\cdot)$ takes $O(1)$ values in an interval of length $O\left(\frac{1}{n}\right)$ around $x$ and is very close to 0 outside the interval. However as of now, there are some technical difficulties in this scenario. This is due to the fact that the main ingredient of the proof is by approximation of an analytic function by a first few terms of it’s power series expansion. However for analytic functions just discussed, this kind of approximation is not well behaved. This is an ongoing work.

The linear spectral statistics and the largest eigenvalue are also interesting from high dimensional statistics and spin glass point of view. In spin glasses, it is known that for spherical Sherrington Kirkpatrick model the free energy at high temperature can be approximated by a linear spectral statistics of the interaction matrix and in low temperature it can be approximated by the largest eigenvalue of the interaction matrix. One might look at Baik and Lee [5] for a reference. A straight forward corollary of our result is the independence of free energies in low and high temperature of spherical SK model made out of same interaction matrix. The joint distribution of largest eigenvalue and a generic linear spectral statistics is also important in high dimensional statistics. In particular it is proved in Zhang et al. [17] for large spikes, when we consider the largest eigenvalue of a spiked Wishart matrix and a generic spectral statistics they are independent. One might also look at Li et al. [9]. However these results are for the spiked matrices where the spikes are large. In particular, the largest eigenvalue has an asymptotic normal distribution as opposed to Tracy-Widom distribution.

2 Preliminaries

2.1 Model description and related terminologies

Firstly we start with the definition of Wigner matrices.

**Definition 2.1.** We call a matrix $W = \left(\frac{x_{i,j}}{\sqrt{n}}\right)_{1 \leq i, j \leq n}$ to be a Wigner matrix if $x_{i,j} = \bar{x}_{i,j}$, $(x_{i,j})_{1 \leq i < j \leq n}$ are i.i.d., $E[x_{i,j}] = 0$ and $E[|x_{i,j}|^2] = 1$.

In this paper we deal with the real symmetric matrices and for the ease of calculation we scale the whole matrix by a factor 2. With slight abuse of notation we shall also call this matrix a Wigner matrix and denote it by $W$. Following are the assumptions of the matrices we consider.
**Assumption 2.1.** We consider a matrix \( W \) given by \( W = (x_{i,j} / \sqrt{n})_{1 \leq i,j \leq n} \) such that the following conditions are satisfied:

(i) \( x_{i,j} = x_{j,i} \) for \( i \leq j \).

(ii) \( \text{Var}(x_{i,j}) = \frac{1}{4} \)

(iii) \( (x_{i,j})_{1 \leq i,j \leq n} \) are i.i.d.

(iv) \( \mathbb{E}[x_{i,j}^2k] \leq (\text{const.})^k \forall k \in \mathbb{N} \)

Given a Wigner matrix \( W \) of dimension \( n \times n \) we denote its eigenvalues by \( \lambda_{1,n} \geq \ldots \geq \lambda_{n,n} \). It is well known that for a Wigner matrix in Definition 2.1, the measure \( \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_{i,n}} \) converge weakly to the semicircular law in the almost sure sense. The law is given by the following density:

\[
f(x) = \begin{cases} 
\frac{1}{\sqrt{4-x^2}} & \text{when } |x| \leq 2 \\
0 & \text{otherwise}
\end{cases}
\] (2.1)

When we scale the entries by a factor 2, the distribution is supported in \([-1, 1]\) and its density is given by

\[
f(x) = \begin{cases} 
\frac{2}{\sqrt{1-x^2}} & \text{when } |x| \leq 1 \\
0 & \text{otherwise}
\end{cases}
\] (2.2)

One of the main quantities of consideration in this paper is the linear spectral statistics. This is defined as follows:

**Definition 2.2.** Suppose \( g \) is an analytic function on a set containing \([-1, 1]\) in its interior. We define the linear spectral statistics corresponding to the function \( g \) as follows:

\[
\text{Tr}[g(W)] = \sum_{i=1}^{n} g(\lambda_{i,n})
\] (2.3)

Our main approach is to work with a polynomial function and use the power series representation of the polynomial function to work with general analytic function. So most of the time \( g \) will be a polynomial of finite degree.

### 2.2 Preliminaries for the combinatorial approach

The main approach of this paper is combinatorial. For the linear spectral statistics we follow the approach of Anderson and Zeitouni [1] and for the point process near the edge we follow the approach of Banerjee [6]. In this subsection we develop the required terminologies. To begin with we start with a matrix \( W \) of dimension \( n \times n \). Its \( k \) th moment is given by

\[
\text{Tr}[W^k] = \sum_{l_0, l_1, \ldots, l_{k-1}, l_0} W_{l_0, l_1} \cdots W_{l_{k-1}, l_0}
\] (2.4)
The word sentence method systematically analyzes the tuples \((i_0, \ldots, i_{k-1}, i_0)\) for some suitable \(k\). To do this we need some notations and definitions. In this part we give a very brief introduction to words, sentences and their equivalence classes essential for the combinatorial analysis of random matrices. The definitions are taken from Anderson et al. [2] and Anderson and Zeitouni [1]. For more general information, see [2, Chapter 1] and [1].

**Definition 2.3 (S words).** Given a set \(S\), an \(S\) letter \(s\) is simply an element of \(S\). An \(S\) word \(w\) is a finite sequence of letters \(s_1 \ldots s_k\), at least one letter long. An \(S\) word \(w\) is closed if its first and last letters are the same. In this paper \(S = \{1, \ldots, n\}\) where \(n\) is the number of nodes in the graph.

Two \(S\) words \(w_1, w_2\) are called equivalent, denoted \(w_1 \sim w_2\), if there is a bijection on \(S\) that maps one into the other. For any word \(w = s_1 \ldots s_k\), we use \(l(w) = k\) to denote the length of \(w\), define the weight \(wt(w)\) as the number of distinct elements of the set \(s_1, \ldots, s_k\) and the support of \(w\), denoted by \(supp(w)\), as the set of letters appearing in \(w\). With any word \(w\) we may associate an undirected graph, with \(wt(w)\) vertices and at most \(l(w) - 1\) edges, as follows.

**Definition 2.4 (Graph associated with a word).** Given a word \(w = s_1 \ldots s_k\), we let \(G_w = (V_w, E_w)\) be the graph with set of vertices \(V_w = supp(w)\) and (undirected) edges \(E_w = \{\{s_i, s_{i+1}\}, i = 1, \ldots, k - 1\}\).

The graph \(G_w\) is connected since the word \(w\) defines a path connecting all the vertices of \(G_w\), which further starts and terminates at the same vertex if the word is closed. We note that equivalent words generate the same graphs \(G_w\) (up to graph isomorphism) and the same passage-counts of the edges. Given an equivalence class \(w\), we shall sometimes denote \(\#E_w\) and \(\#V_w\) to be the common number of edges and vertices for graphs associated with all the words in this equivalence class \(w\).

**Definition 2.5 (Weak Wigner words).** Any word \(w\) will be called a weak Wigner word if the following conditions are satisfied:

1. \(w\) is closed.
2. \(w\) visits every edge in \(G_w\) at least twice.

Suppose now that \(w\) is a weak Wigner word. If \(wt(w) = (l(w) + 1)/2\), then we drop the modifier “weak” and call \(w\) a Wigner word. (Every single letter word is automatically a Wigner word.) Except for single letter words, each edge in a Wigner word is traversed exactly twice. If \(wt(w) = (l(w) - 1)/2\), then we call \(w\) a critical weak Wigner word.

It is a well known result in random matrix theory that there is a bijection from the set of the Wigner words of length \(2k + 1\) to the set of Dyck paths of length \(2k\). We now move to definitions related to sentences.
Definition 2.6 (Sentences and corresponding graphs). A sentence \( a = [w_i]_{i=1}^m = [[\alpha_{i,j}]_{j=1}^{l(w_i)}]_{i=1}^m \) is an ordered collection of \( m \) words of length \( (l(w_1), \ldots, l(w_m)) \) respectively. We define the graph \( G_a = (V_a, E_a) \) to be the graph with

\[
V_a = \text{supp}(a), \quad E_a = \{\{\alpha_{i,j}, \alpha_{i,j+1}\}|i = 1, \ldots, m; j = 1, \ldots, l(w_i) - 1\}.
\]

Notation 2.1. Given a closed word \( w = (i_0, i_1, \ldots, i_{k-1}, i_k = i_0) \) we need to consider the random variable \( \prod_{j=0}^{k-1} x_{i_j,i_{j+1}} \). We call this random variable \( X_w \).

In addition to this we also need the map from word to Dyck paths and the related results and terminologies for the analysis of the eigenvalues near the edge. For these the reader may have a look at subsection 7.2 in Banerjee [6].

3 Main result

We now state the main result of this paper.

Theorem 3.1. Suppose we have a Wigner matrix \( W \) satisfying the assumptions in Assumption 2.1. Then the following are true:

1. Let \( g \) be any polynomial of fixed degree. Then the random variable 
   \( (\text{Tr}[g(W)] - E[\text{Tr}[g(w)]] \) and the point process at the edge of the spectrum (i.e. at \( x = \pm 1 \)) are asymptotically independent.

2. More generally \( g \) can be taken to be a fixed analytic function of polynomial growth on a set containing \( [-1,1] \) in its interior.

3. As a corollary, for any fixed \( k \), the random vector
   \( (n^{\frac{k}{2}}(\lambda_1, n - 1), \ldots, n^{\frac{k}{2}}(\lambda_k, n - 1)) \)
   and \( (\text{Tr}[g(W)] - E[\text{Tr}[g(w)]] \) are independent.

4 Strategies for the proof

As mentioned at the beginning, the main technique of this current paper is combinatorial in nature. In particular for an analytic function \( g \), it is known that 
\( (\text{Tr}[g(W)] - E[\text{Tr}[g(W)]] \) converges to a centered Gaussian with explicit variance. The combinatorial approach for proving this kind of result was initiated in the paper by Anderson and Zeitouni [1]. In order the prove this the authors first consider the joint distribution of 

\( (\text{Tr}[W^{m_1}] - E[\text{Tr}[W^{m_1}]], \ldots, \text{Tr}[W^{m_k}] - E[\text{Tr}[W^{m_k}]] ) \)

for every fixed \( k \) and integers \( m_1, \ldots, m_k \in \mathbb{N} \). It is proved using the method of moments that this joint distribution is \( k \) dimensional Gaussian with explicit covariance matrix. The results are summarized in the following theorem.
Theorem 4.1. (Chapter 1 Anderson et al. [2]) Suppose $W$ is a Wigner matrix satisfying Assumption 2.1 and we want to determine the joint distribution of

$$(\text{Tr} [W^{m_1}] - E [\text{Tr} [W^{m_1}]], \ldots, \text{Tr} [W^{m_k}] - E [\text{Tr} [W^{m_k}]]).$$

Then there exists $\Sigma$ such that

1. for any $l$ and $m'_1, \ldots, m'_l \in \{m_1, \ldots, m_k\}$ the joint moment

$$E \left[ \prod_{i=1}^{l} (\text{Tr} [W^{m'_i}] - E [\text{Tr} [W^{m'_i}]]) \right] \rightarrow \begin{cases} 0 & \text{if } l \text{ is odd} \\ \sum_{\eta \in \mathcal{P}_l} \prod_{j=1}^{l} \Sigma (m'_{\eta(j,1)}, m'_{\eta(j,2)}) & \text{otherwise.} \end{cases} \tag{4.1}$$

Here for an even number $l$, $\mathcal{P}_l$ denotes all partitions of the set $\{1, 2, \ldots, l\}$ such that each block has exactly two elements. Further for any partition $\eta \in \mathcal{P}_l$, $\eta(j, 1)$ and $\eta(j, 2)$ denote the first and second element of the $j$th block in $\eta$.

2. As a corollary, we have the vector

$$(\text{Tr} [W^{m_1}] - E [\text{Tr} [W^{m_1}]], \ldots, \text{Tr} [W^{m_k}] - E [\text{Tr} [W^{m_k}]]$$

converges to a $k$ dimensional centered Gaussian with covariance matrix $(\Sigma(m_i, m_j))_{1 \leq i, j \leq k}$.

For an explicit expression of $\Sigma$ one might look at section 2.1.7 of Anderson et al. [2].

Finally for a general analytic function $g(x) = \sum_{i=0}^{\infty} g_i x^i$, one considers the following polynomial $g^{(m)}(x) = \sum_{i=0}^{m} g_i x^i$ and prove the following result:

Theorem 4.2. (Anderson and Zeitouni [1]) Suppose $W$ be a Wigner matrix satisfying Assumption 2.1. Then for any analytic function $g(x) = \sum_{i=0}^{\infty} g_i x^i$ on a set containing $[-1 - \varepsilon, 1 + \varepsilon]$ and having polynomial growth we have

$$\text{Var} [\text{Tr} [g(W)]] \leq \tilde{c} \sup_{x \in [-1 - \varepsilon, 1 + \varepsilon]} |g'(x)|^2. \tag{4.2}$$

In particular if the function is analytic on an interval containing $[-1 - \varepsilon, 1 + \varepsilon]$, for any $\eta$ we can choose large enough $i$ such that

$$\text{Var} \left[ \text{Tr} [g(W)] - \sum_{j=0}^{i} g_j x^j \right] \leq \eta. \tag{4.3}$$

Remark 4.1. One might note that for Theorem 4.2 to hold, Anderson and Zeitouni [1] require the entries to satisfy Poincare inequality. However one might look at Remark 3.12 of Van Handel [14] to see that sub-Gaussianity implies Poincare inequality.
On the other hand in the recent paper Banerjee [6] the combinatorial approach for proving edge universality with general non-symmetrically distributed entries was considered. This method relies upon proving the following results

**Theorem 4.3.** (Banerjee [6]) Consider the Wigner matrix $W$ satisfying Assumption 2.1. Then for any fixed $t \in (0, \infty)$ taking $k = \left[ \frac{m^2}{n} \right]$, we have the following results

1. $E \text{Tr} \left[ W^{2k} \right] = O(1)$ and $E \text{Tr} \left[ W^{2k+1} \right] = o(1)$.

2. If the limit of $\lim_{n \to \infty} E \text{Tr} \left[ W^{2k} \right]$ for some $t \in (0, \infty)$ exists, then the limit only depends on the first and second moment of entries.

3. As the limit exists for Gaussian entries, the limit exists and is universal for any Wigner matrix satisfying Assumption 2.1.

**Theorem 4.4.** (Banerjee [6]) Consider the Wigner matrix $W$ satisfying Assumption 2.1. Then for any fixed $t_1, \ldots, t_l \in (0, \infty)$ taking $k_i = \left[ \frac{t_i n^2}{3} \right]$, we have the following results

1. $E \left[ \prod_{i=1}^l \left( \text{Tr} \left[ W^{k_i} \right] - E \left[ \text{Tr} \left[ W^{k_i} \right] \right] \right) \right] = O(1)$. \hfill (4.4)

2. If the limit in (4.4) exists for some $t_1, \ldots, t_l$, then the limit only depends on the first and second moment of entries.

3. As the limit exists for Gaussian entries, the limit exists and is universal for any Wigner matrix satisfying Assumption 2.1.

Theorems 4.3 and 4.4 is enough to prove the edge universality as the laplace transformation of the $l$ point correlation function of the point process at the edge of the spectrum is determined by the limits in Theorems 4.4 and 4.3. As weak convergence of correlation functions of a sequence of point processes determines the weak convergence of the sequence of point processes, we get the desired result of edge universality.

So in order to prove the independence of the linear spectral statistics and the point process at the edge, we prove the following result:

**Theorem 4.5.** Suppose we fix $m_1, \ldots, m_k$. Then for any $l$, $l', m'_1, \ldots, m'_l \in \{m_1, \ldots, m_k\}$ and $t_1, \ldots, t_l \in (0, \infty)$, the following holds:

$$E \left[ \prod_{i=1}^l \left( \text{Tr} \left[ W^{m'_i} \right] - E \left[ \text{Tr} \left[ W^{m'_i} \right] \right] \right) \right] \prod_{j=1}^{l'} \left( \text{Tr} \left[ W^{t_j \frac{n^2}{3}} \right] - E \left[ \text{Tr} \left[ W^{t_j \frac{n^2}{3}} \right] \right] \right) \rightarrow \lim_{n \to \infty} E \left[ \prod_{i=1}^l \left( \text{Tr} \left[ W^{m'_i} \right] - E \left[ \text{Tr} \left[ W^{m'_i} \right] \right] \right) \right] \lim_{n \to \infty} E \left[ \prod_{j=1}^{l'} \left( \text{Tr} \left[ W^{t_j \frac{n^2}{3}} \right] - E \left[ \text{Tr} \left[ W^{t_j \frac{n^2}{3}} \right] \right] \right) \right]$$

\hfill (4.5)

Once this is proved, Theorem 3.1 readily follows with the help of Theorem 4.2.
5 Proofs of the results

To begin with we state an algorithm. This algorithm takes two closed words \( w_1 \) and \( w_2 \) of lengths \( k_1 + 1 \) and \( k_2 + 1 \) as input such that the words have at least one edge in common and gives a closed word \( w_3 \) of length \( k_1 + k_2 + 1 \) as an output which has the same edge set as the union of the edges of \( w_1 \) and \( w_2 \). This is taken from Banerjee [6]

**Algorithm 5.1.** We start with two words \( w_1 \) and \( w_2 \) such that \( w_1 \) and \( w_2 \) shares an edge. Let \( \{\alpha, \beta\} \) be the first edge in \( w_2 \) which is repeated in \( w_1 \). We consider the first appearance of \( \{\alpha, \beta\} \) in \( w_2 \). Without loss of generality we assume that the first appearance of the edge \( \{\alpha, \beta\} \) appears in the word \( w_2 \) in the order \((\alpha, \beta)\). We now consider any appearance (for concreteness say the first) of the edge \( \{\alpha, \beta\} \) in the word \( w_1 \). This appearance \( \{\alpha, \beta\} \) can be traversed in \( w_1 \) in the order \((\alpha, \beta)\) or \((\beta, \alpha)\). Considering these we have the word \( w_2 \) looks like

\[
w_2 = (\alpha_0, \alpha_1, \ldots, \alpha_{p_1}, \alpha, \beta, \ldots, \alpha_{k_1-1}, \alpha_0) \tag{5.1}
\]

and the word \( w_1 \) looks like

\[
w_1 = (\beta_0, \beta_1, \ldots, \beta_{q_1}, \alpha, \beta, \ldots, \beta_{k_2-1}, \beta_0) \tag{5.2}
\]

or

\[
w_1 = (\beta_0, \beta_1, \ldots, \beta_{q_1}, \beta, \alpha, \ldots, \beta_{k_2-1}, \beta_0). \tag{5.3}
\]

Now we output the word \( w_3 \) as follows:

1. **Suppose \( w_1 \) is of the form (5.2), then**

\[
w_3 = (\alpha_0, \alpha_1, \ldots, \alpha_{p_1}, \alpha, \beta, \beta_{q_1+3}, \ldots, \beta_{k_2-1}, \beta_0, \ldots, \beta_{q_1}, \alpha, \beta, \alpha_{p_1+3}, \ldots, \alpha_{k_1-1}, \alpha_0). \tag{5.4}
\]

2. **On the other hand when \( w_1 \) is of the form (5.3),**

\[
w_3 = (\alpha_0, \alpha_1, \ldots, \alpha_{p_1}, \alpha, \beta, \beta_{q_1}, \ldots, \beta_0, \beta_{k_2-1}, \ldots, \beta_{q_1+3}, \alpha, \beta, \alpha_{p_1+3}, \ldots, \alpha_{k_1-1}, \alpha_0). \tag{5.5}
\]

**Proof of Theorem 4.5.** As this is the most important proof in this paper, we prove it in step by step.
Step 1 (A basic observation): We begin with a very basic observation.

\[
E \left[ \prod_{i=1}^{l} \left( \text{Tr} \left[ W_{i}^{m_{i}} \right] - E \left[ \text{Tr} \left[ W_{i}^{m_{i}} \right] \right] \right) \right]
\]

\[
= \frac{1}{n} \sum_{a_{1}, a_{2}} \left( \sum_{i=1}^{l} E \left[ \prod_{i=1}^{l} \left( X_{w_{i}} - E[X_{w_{i}}] \right) \prod_{j=1}^{r} \left( X_{w_{j}} - E[X_{w_{j}}] \right) \right] \right)
\]

\[
= \frac{1}{n} \sum_{a_{1}, a_{2}} \left( \sum_{i=1}^{l} E \left[ \prod_{i=1}^{l} \left( X_{w_{i}} - E[X_{w_{i}}] \right) \prod_{j=1}^{r} \left( X_{w_{j}} - E[X_{w_{j}}] \right) \right] \right)
\]

\[
= \frac{1}{n} \sum_{a_{1}, a_{2}} \left( \sum_{i=1}^{l} E \left[ \prod_{i=1}^{l} \left( X_{w_{i}} - E[X_{w_{i}}] \right) \prod_{j=1}^{r} \left( X_{w_{j}} - E[X_{w_{j}}] \right) \right] \right) + T_{1}(\text{say})
\]

\[
= \frac{1}{n} \sum_{a_{1}, a_{2}} \left( \sum_{i=1}^{l} E \left[ \prod_{i=1}^{l} \left( X_{w_{i}} - E[X_{w_{i}}] \right) \prod_{j=1}^{r} \left( X_{w_{j}} - E[X_{w_{j}}] \right) \right] \right) \left[ a_{1} \in \mathbb{E}_{i}, a_{2} \in \mathbb{E}_{i}, a_{1} \cap a_{2} = \emptyset \right]
\]

Here \(a_{1} = \left[ w_{i} \right]_{i=1}^{l} \) and \(a_{2} = \left[ w'_{j} \right]_{j=1}^{r} \). We shall analyze \(T_{1} \) and \(T_{2} \) separately.

Step 2 (Analysis of \(T_{1} \)) \(T_{1} \) contains the cases when \(E_{a_{1}} \cap E_{a_{2}} \neq \emptyset \).

For this without loss of generality we assume \(a_{1} \) contains a single word and so does \(a_{2} \). The general case can be proved by repeated use of the arguments given next. We assume \(w_{1} \) be a word of fixed length \(m+1 \) while \(w_{2} \) is a word of length \([m^{\frac{3}{2}}] + 1 \) for some \(t \in (0, \infty) \) such that \(E_{w_{1}} \cap E_{w_{2}} \neq \emptyset \). We would like to find the following expectation

\[
\left( \frac{1}{n} \right) \sum_{w_{1}, w_{2} \mid l(w_{1}) = m+1, l(w_{2}) = [m^{\frac{3}{2}}] + 1, E_{w_{1}} \cap E_{w_{2}} \neq \emptyset} \left( E[X_{w_{1}} X_{w_{2}}] - E[X_{w_{1}}] E[X_{w_{2}}] \right)
\]

(5.7)

We now apply Algorithm 5.1 to \((w_{1}, w_{2})\) to get a word \(w \). Hence we write the sum in (5.7) in the following way:

\[
\left( \frac{1}{n} \right) \sum_{\omega} \sum_{w_{1}, w_{2} \mid f(w_{1}, w_{2}) = \omega} \left( E[X_{w_{1}} X_{w_{2}}] - E[X_{w_{1}}] E[X_{w_{2}}] \right).
\]

(5.8)

In order to have a nontrivial value of \(E[X_{w_{1}} X_{w_{2}}] - E[X_{w_{1}}] E[X_{w_{2}}] \), one needs to have each edge in \(\omega\) repeated at least twice. We call this set \(\mathcal{W}_{2,m+[m^{\frac{3}{2}}]} \). By a straight
forward application of Holder’s inequality we know \( E[|X_{w_1}X_{w_2}|] \geq E[|X_{w_1}|] E[|X_{w_2}|] \). So it is enough to bound the following quantity:

\[
\left( \frac{1}{n} \right)^{\frac{m+n^2}{2}} \sum_{w \in \mathcal{M}} 2|f^{-1}(w)| E[|X_w|].
\]  

(5.9)

We now look closely at the proof of Theorem 6.2 in Banerjee [6]. Imitating the arguments of the proof of this theorem it can be showed that all the other cases apart from case (i) have negligible contribution. Also following the proofs of Propositions 8.2 and 8.3 in Banerjee [6], we can restrict our attention to the case when every edge in \( w \) is traversed exactly twice and every type \( j \geq 2 \) instant in the skeleton word is actually a type 2 instant. Now we analyze this case in details.

Alike Banerjee [6] we at first fix the skeleton word. Let \( N \) be the number of type \( j \geq 2 \) instants in the skeleton word. Given a skeleton word, we choose a type \( j \geq 2 \) instant and find out the edge which is closed in the next instant. We choose the instant which corresponds to the left endpoint of the first traversal of this edge. So given the skeleton word there is at most \( N \) choices to choose the endpoints of \( w_1 \). Let this choice be \( l \). Let the length of the part of the skeleton word inside these two points be \( m_1 \) and the outside these two points be \( m_2 \). As a consequence, in \( w_2 \) there will be exactly \( m_2 + 1 \) Dyck paths and in \( w_1 \) there will be exactly \( m_1 \) Dyck paths. So this might appear same as the calculation in Theorem 6.2 in Banerjee [6]. However as the length of the word \( w_1 \) is fixed, there will be additional constraints. In particular, if the first endpoint of the word is the instant where this level first appears in the skeleton word, we have \( |p_1 + \ldots + p_l - r_l| \leq m \). Here we chose the \( l \) th type \( j \geq 2 \) instant. On the other hand, it is possible that the skeleton word returns to the level of first chosen point multiple times before reaching the second chosen point. In this case we have \( p_l \leq m \). As all the type \( j \geq 2 \) instants in the skeleton word are type 2, we have the following upper bound to (5.9):

\[
C \left( \frac{1}{2} \right)^{\frac{m+n^2}{2}} \sum_N \sum_{l=1}^N \sum_{p_1, \ldots, p_N \mid p_i \leq m} \sum_{q_1, \ldots, q_{N-1} \mid q_i \geq 0} \sum_{r_1, \ldots, r_N \mid r_i - p_i \geq 0} \sum n^{\frac{m+n^2}{2} - N + 1} 3^N
\]

\[
= (m+1) \frac{m_1}{m+1} \frac{(m+1)}{\left( \frac{m_1+m+1}{2} \right)} \frac{m_2 + 1}{\left( \frac{m_2+n^2}{2} + 1 \right)} \frac{m+n^2}{2} + O(1)
\]  

(5.10)

As the sum over all the unconstrained case (i.e. when both \( l(w_1) \) and \( l(w_2) \) is of \( O(n^2) \)) is of \( O(1) \), the sum in (5.10) goes to 0.
Step 3 (Analysis of $T_2$) Now we consider the case when $E_{a_1} \cap E_{a_2} = \emptyset$. Observe that in this case the random variables $\prod_{i=1}^{l} (X_{w_i} - E[X_{w_i}])$ and $\prod_{j=1}^{r} (X_{w_j'} - E[X_{w_j'}])$ are independent. As a result,

$$E \left[ \prod_{i=1}^{l} (X_{w_i} - E[X_{w_i}]) \prod_{j=1}^{r} (X_{w_j'} - E[X_{w_j'}]) \right]$$

$$= E \left[ \prod_{i=1}^{l} (X_{w_i} - E[X_{w_i}]) \right] \prod_{j=1}^{r} (X_{w_j'} - E[X_{w_j'}]) \tag{5.11}$$

By analysis of $T_1$,

$$E \left[ \prod_{i=1}^{l} \left( \text{Tr} [W_{w_i}] - E \left[ \text{Tr} [W_{w_i}] \right] \right) \prod_{j=1}^{r} \left( \text{Tr} [W_{w_j'}^{1/2}] - E \left[ \text{Tr} [W_{w_j'}^{1/2}] \right] \right) \right]$$

$$= \frac{1}{n} \sum_{a_1, a_2 | E_{a_1} \cap E_{a_2} = \emptyset} E \left[ \prod_{i=1}^{l} (X_{w_i} - E[X_{w_i}]) \right] \prod_{j=1}^{r} (X_{w_j'} - E[X_{w_j'}]) + o(1). \tag{5.12}$$

Now we know that

$$E \left[ \prod_{i=1}^{l} \left( \text{Tr} [W_{w_i}] - E \left[ \text{Tr} [W_{w_i}] \right] \right) \prod_{j=1}^{r} \left( \text{Tr} [W_{w_j'}^{1/2}] - E \left[ \text{Tr} [W_{w_j'}^{1/2}] \right] \right) \right]$$

$$= \frac{1}{n} \sum_{a_1, a_2 | E_{a_1} \cap E_{a_2} = \emptyset} E \left[ \prod_{i=1}^{l} (X_{w_i} - E[X_{w_i}]) \right] \prod_{j=1}^{r} (X_{w_j'} - E[X_{w_j'}]) \tag{5.13}$$

$$= T_2 + T'_1$$

where

$$T'_1 = \sum_{a=(a_1, a_2) \in \mathcal{A}_{2, s}} E \left[ \prod_{i=1}^{l} \left( \text{Tr} [W_{w_i}] - E \left[ \text{Tr} [W_{w_i}] \right] \right) \prod_{j=1}^{r} \left( \text{Tr} [W_{w_j'}^{1/2}] - E \left[ \text{Tr} [W_{w_j'}^{1/2}] \right] \right) \right]. \tag{5.14}$$

Here $\mathcal{A}_{2, s}$ denotes all sentences $a = (a_1, a_2)$ such that each edge in the graph $G_a$ is traversed at least twice and $E_{a_1} \cap E_{a_2} \neq \emptyset$. It is easy to see (by Holder’s inequality) that $|T'_1| \leq |T_1|$. Hence $|T'_1| \to 0$. As a consequence,

$$\left| \prod_{i=1}^{l} \left( \text{Tr} [W_{w_i}] - E \left[ \text{Tr} [W_{w_i}] \right] \right) \prod_{j=1}^{r} \left( \text{Tr} [W_{w_j'}^{1/2}] - E \left[ \text{Tr} [W_{w_j'}^{1/2}] \right] \right) \right| \rightarrow 0. \tag{5.15}$$
This completes the proof. \(\square\)

**Proof of Theorem 3.1.** We only prove part 2 as proof of part 1 is contained in the proof of Theorem 4.5 and part 3 is a straightforward corollary of part 2.

This is enough to prove that for an analytic function \(g\) of polynomial growth on a set containing \([-1-\varepsilon, 1+\varepsilon]\) and any bounded continuous function \(q : \mathbb{R} \rightarrow \mathbb{R}\) we have

\[
\lim_{n \to \infty} \mathbb{E} \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) \prod_{j=1}^{l} \left( \text{Tr} \left[ W_{i,j}^{(j,n)} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{i,j}^{(j,n)} \right] \right] \right) \right] \\
= \lim_{n \to \infty} \mathbb{E} \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) \right] \lim_{n \to \infty} \mathbb{E} \left[ \prod_{j=1}^{l} \left( \text{Tr} \left[ W_{i,j}^{(j,n)} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{i,j}^{(j,n)} \right] \right] \right) \right]
\]

(5.16)

We shall fix any \(\eta > 0\), choose \(i\) large enough and compare \(q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right)\) and \(q \left( \text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]] \right)\). Here \(g^{(i)}(x) = \sum_{j=0}^{i} g_{j} x^{j}\). Since both \((\text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]])\) is a Gaussian random variable in the limit and \(\mathbb{E} \left( \text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]] \right)^{2}\) is uniformly bounded over \(n\) and \(i\), they are tight uniformly over \(n\) and \(i\). In particular we can choose \(M\) large enough so that both \(P \left[ |\text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]]| > M \right]\) and \(P \left[ |\text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]]| > M \right]\) are less than \(\eta\) uniformly over \(n\) and \(i\). Now inside \([-M, M]\) \(q\) is uniformly continuous. So given any \(0 < \eta' < \eta\) there exists \(\delta > 0\) such that \(|x - y| < \delta\) implies \(|q(x) - q(y)| < \eta'\). Now we choose \(i\) large enough so that \(\text{Var} \left[ \text{Tr}[g(W)] - \text{Tr}[g^{(i)}(W)] \right] \leq \delta^{2}\eta\). As a consequence, by Chebyshev’s inequality

\[
P \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) - q \left( \text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]] \right) > \eta' \right] < \eta + 2\eta = 3\eta.
\]

Hence,

\[
\mathbb{E} \left[ (q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) - q \left( \text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]] \right) \right)^{2} \right] \leq 3\eta\|q\|_{\infty}^{2} + \eta^{2}
\]
. Now

\[
\limsup_n \left| \mathbb{E} \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) \prod_{j=1}^{r} \left( \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] \right] \right) \right] - \right.
\]

\[
\mathbb{E} \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) \prod_{j=1}^{r} \left( \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] \right] \right) \right] \times
\]

\[
\left| \prod_{j=1}^{r} \left( \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] \right] \right) \right| + \mathbb{E} \left| \prod_{j=1}^{r} \left( \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] \right] \right) \right| \times
\]

\[
\mathbb{E} \left[ q \left( \text{Tr}[g(W)] - \mathbb{E}[\text{Tr}[g(W)]] \right) - q \left( \text{Tr}[g^{(i)}(W)] - \mathbb{E}[\text{Tr}[g^{(i)}(W)]] \right) \right] + o(1)
\]

\[
\leq 2 \sqrt{3\eta \|q\|_{\infty}^2 + \eta^2} \limsup_n \left[ \mathbb{E} \left( \prod_{j=1}^{r} \left( \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] - \mathbb{E} \left[ \text{Tr} \left[ W_{ij,n}^{\frac{2}{3}} \right] \right] \right) \right)^{\frac{3}{2}} \right] + o(1)
\]

(5.17)

Here the last line follows from Cauchy-Schwarz inequality. As \( \eta > 0 \) was arbitrary, this proves that the first expression of (5.17) goes to 0. This proves the result. \( \square \)

**Acknowledgment:** This research is supported by an Inspire faculty fellowship. I am highly grateful to Prof. Krishna Maddaly for his interest and encouragement.

**References**

[1] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

[2] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[3] Z. Bai, X. Wang, and W. Zhou. Clt for linear spectral statistics of wigner matrices. *Electronic Journal of Probability*, 14:2391–2417, 2009.

[4] Z. D. Bai and J. W. Silverstein. Clt for linear spectral statistics of large-dimensional sample covariance matrices. In *Advances In Statistics*, pages 281–333. World Scientific, 2008.

[5] J. Baik and J. O. Lee. Fluctuations of the free energy of the spherical sherrington-kirkpatrick model. *Journal of Statistical Physics*, 165(2):185–224, 2016.

[6] D. Banerjee. A new combinatorial approach for tracy-widom law of wigner matrices. *arXiv preprint arXiv:2201.00300*, 2022.

[7] L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28. American Mathematical Soc., 2017.

[8] P. J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402(3):709–728, 1993.

[9] Z. Li, F. Han, and J. Yao. Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model. *The Annals of Statistics*, 48(6):3138–3160, 2020.

[10] A. Soshnikov. Universality at the edge of the spectrum in wigner random matrices. *Communications in mathematical physics*, 207(3):697–733, 1999.

[11] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572, 2010.

[12] C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[13] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.

[14] R. Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.

[15] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, pages 325–327, 1958.

[16] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions i. In *The Collected Works of Eugene Paul Wigner*, pages 524–540. Springer, 1993.

[17] Z. Zhang, S. Zheng, G. Pan, and P. Zhong. Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices. *arXiv preprint arXiv:2009.11010*, 2020.