LIMIT THEOREMS FOR LINEAR EIGENVALUE STATISTICS OF
OVERLAPPING MATRICES

VLADISLAV KARGIN

Abstract
The paper proves several limit theorems for linear eigenvalue statistics of
overlapping Wigner and sample covariance matrices. It is shown that the
covariance of the limiting multivariate Gaussian distribution is diagonalized
by choosing the Chebyshev polynomials of the first kind as the basis for the
test function space. The covariance is explicitly computed and it is shown that
for the Chebyshev polynomials of sufficiently high degree the covariance of
linear statistics depends only on the variance of matrix entries.

1. INTRODUCTION

Recently, there was a surge of interest in the spectral properties of overlapping submatrices
of large random matrices. A seminal study was done by Baryshnikov [3], which derived
the joint eigenvalue distribution for principal minors of Gaussian Unitary Ensemble (GUE)
and related this distribution to the last passage percolation problem. Later Johansson and
Nordenstam [14] established the determinantal structure of this joint distribution, and their
results were generalized in [10], [11] and [17] to other unitarily invariant ensembles of random
matrices. Recently, Borodin in [4] and Reed in [18] obtained limit theorems for eigenvalue
statistics of overlapping real Gaussian matrices. We extend these results further to Wigner and
sample covariance matrices which lack the rotation invariance structure.

![Figure 1. Overlapping symmetric matrices](image-url)

Date: July 2014.
e-mail: vladislav.kargin@gmail.com.
1.1. **Wigner matrices.** In our Wigner matrix model, \( A_N \) and \( B_N \) are two Hermitian matrices which are submatrices of an infinite Hermitian matrix \( X \). This matrix \( X \) has independent identically distributed entries above the main diagonal and independent entries (with a possibly different distribution) on the main diagonal. We impose the following assumptions about the matrix entries. Let the off-diagonal entries of \( X \) be real random variables with all moments finite and assume that \( \mathbb{E}(X_{ij}) = 0 \), \( \mathbb{E}(X_{ij}^2) = 1 \), and \( \mathbb{E}(X_{ij}^4) = m_4 \). In addition, assume that the diagonal terms have all finite moments and in particular, \( \mathbb{E}(X_{ii}) = 0 \) and \( \mathbb{E}(X_{ii}^2) = d_2 \).

Let \( A_N \) have \( a(N) \) rows and columns, and let \( B_N \) have \( b(N) \) rows and columns. In addition, \( A_N \) and \( B_N \) have \( \Delta(N) \) rows and columns in common. (See Figure 1.1.)

We think about \( A_N \) and \( B_N \) as sequences of matrices of increasing size and we assume that there is a parameter \( t_N \) that approaches infinity as \( N \to \infty \), and that quantities \( a(N)/t_N \), \( b(N)/t_N \), and \( \Delta(N)/t_N \) approach some positive limits \( a, b, \) and \( \Delta \), respectively.

We define the normalized matrices

\[
\tilde{A}_N := \frac{1}{2\sqrt{a(N)}} A_N, \quad \text{and} \quad \tilde{B}_N := \frac{1}{2\sqrt{b(N)}} B_N.
\]

The normalization is chosen in such a way that the empirical distribution of eigenvalues of \( \tilde{A}_N \) and \( \tilde{B}_N \) converges to a distribution supported on the interval \([-1, 1]\).

If \( f : \mathbb{R} \to \mathbb{R} \) is a test function, then we define the corresponding linear statistic for matrix \( A_N \) as

\[
N(f, A_N) := \text{Tr} \left[ f \left( \tilde{A}_N \right) \right] = \sum f(\lambda_i),
\]

where \( \lambda_i \) are the eigenvalues of the matrix \( \tilde{A}_N \). It is the linear statistic of the eigenvalues of \( A_N \) after rescaling. We define linear statistics for matrix \( B_N \) similarly. Let us also define the centered statistics,

\[
N^0(f, A_N) := N(f, A_N) - \mathbb{E}N(f, A_N), \quad \text{and} \quad N^0(g, B_N) := N(g, B_N) - \mathbb{E}N(g, B_N).
\]

We are interested in the joint distribution of these linear statistics when \( N \) is large. Recall that the Chebyshev polynomials of the first kind are defined by the formula: \( T_k(\cos \theta) = \cos k\theta \). We will prove the following result.

**Theorem 1.1.** Assume \( A_N \) and \( B_N \) are the real Hermitian overlapping matrices. Let \( T_k(x) \) denote the Chebyshev polynomials of the first kind. As \( N \to \infty \), the distribution of the centered linear statistics \( N^0(T_k, A_N) \) and \( N^0(T_l, B_N) \) converges to a two-variate Gaussian
distribution with the covariance equal to

\[
\begin{cases}
  \frac{k}{2} \left( \frac{\Delta}{\sqrt{ab}} \right)^k, & \text{if } k = l \geq 3, \\
  \frac{m_k - 1}{2} \left( \frac{\Delta}{\sqrt{ab}} \right)^2, & \text{if } k = l = 2, \\
  \frac{d}{4} \left( \frac{\Delta}{\sqrt{ab}} \right), & \text{if } k = l = 1, \\
  0, & \text{otherwise}.
\end{cases}
\]  

(1)

For the author, the motivation for this model came from the paper by Borodin [4] where a similar problem was considered and solved for matrices whose entries have the same first four moments as the Gaussian random variable. Formula (1) shows an interesting fact: the covariance structure of Chebyshev polynomials of degree higher than the second depends only on the first two moments of the off-diagonal matrix entries.

For the author, the motivation for this model came from the paper by Borodin [4] where a similar problem was considered and solved for matrices whose entries have the same first four moments as the Gaussian random variable. Formula (1) shows an interesting fact: the covariance structure of Chebyshev polynomials of degree higher than the second depends only on the first two moments of the off-diagonal matrix entries.

Figure 2. Overlapping non-symmetric matrices

1.2. Sample covariance matrices. Next, we study the singular values of non-symmetric overlapping matrices. This is equivalent to the study of eigenvalues of certain sample covariance matrices. Namely, let $A_N$ and $B_N$ be two submatrices of an infinite random matrix $X$ with independent identically distributed random entries. Let the entries of $X$ be real random variables with all moments finite and assume that $\mathbb{E}(X_{ij}) = 0$, $\mathbb{E}(X_{ij}^2) = 1$, and $\mathbb{E}(X_{ij}^4) = m_4$.

Suppose that $A_N$ has $a_1(N)$ rows and $a_2(N)$ columns, and that $B_N$ has $b_1(N)$ rows and $b_2(N)$ columns. In addition, suppose that $A_N$ and $B_N$ have $\Delta_1(N)$ rows and $\Delta_2(N)$ columns in common. (See Figure 1.2)

As before, we think about $A_N$ and $B_N$ as sequences of matrices of increasing size and we assume that there is a parameter $t_N$ that approaches infinity as $N \to \infty$, and that quantities
Let us define the normalized sample covariance matrices

\[ W_{AN} := \frac{1}{2\sqrt{a_1(N)a_2(N)}} \left[ A_N A_N^* - \left( a_1^{(N)} + a_2^{(N)} \right) I_{a_1(N)} \right], \]

where \( I_{a_1(N)} \) is the \( a_1(N) \)-by- \( a_1(N) \) identity matrix, and

\[ W_{BN} := \frac{1}{2\sqrt{b_1(N)b_2(N)}} \left[ B_N B_N^* - \left( b_1^{(N)} + b_2^{(N)} \right) I_{b_1(N)} \right]. \]

Again, the normalization is chosen in such a way that the empirical distribution of eigenvalues of \( W_{AN} \) and \( W_{BN} \) converges to a distribution supported on the interval \([-1, 1]\).

If \( f : \mathbb{R} \to \mathbb{R} \) is a test function, then we define the corresponding linear statistic for matrix \( W_{AN} \) as

\[ \mathcal{N}(f, W_{AN}) := \text{Tr} \left[ f(W_{AN}) \right] = \sum_i f(\lambda_i), \]

where \( \lambda_i \) are the eigenvalues of the matrix \( W_{AN} \). Hence, this is a linear statistic of the squares of matrix \( A_N \)'s singular values. We define linear statistics for matrix \( W_{BN} \) similarly. Let us also define the centered statistics,

\[ \mathcal{N}^o(f, W_{AN}) := \mathcal{N}(f, W_{AN}) - \mathbb{E} \mathcal{N}(f, W_{AN}), \quad \text{and} \]

\[ \mathcal{N}^o(g, W_{BN}) := \mathcal{N}(g, W_{BN}) - \mathbb{E} \mathcal{N}(g, W_{BN}). \]

Finally, let

\[ \gamma := \frac{\Delta_1 \Delta_2}{\sqrt{a_1a_2b_1b_2}}. \]

**Theorem 1.2.** Assume \( A_N \) and \( B_N \) are the real random overlapping matrices with the matrix entries that have the moments described above. Let \( T_k(x) \) denote the Chebyshev polynomials of the first kind. As \( N \to \infty \), the distribution of the centered linear statistics \( \mathcal{N}^o(T_k, W_{AN}) \) and \( \mathcal{N}^o(T_l, W_{BN}) \) converges to a two-variate Gaussian distribution with the covariance equal to

\[ \begin{align*}
\delta_{kl} k^2 \gamma, & \quad \text{if } k > 1, \\
\delta_{kl} \frac{(m_k-1)k}{4} \gamma, & \quad \text{if } k = 1.
\end{align*} \]

**1.3. CLT for continuously differentiable functions.** In order to extend our results to continuously differentiable functions, we have to restrict the scope to the models with matrix entries that satisfy the Poincaré inequality. The reason for this is that we will need a measure concentration inequality which is proven only for these models. Recall that a matrix entry \( X_{ij} \) satisfies the Poincaré inequality if there is a constant \( c > 0 \) such that for every continuously differentiable function \( f(x) \), we have

\[ \text{Var} \left( f(X_{ij}) \right) \leq c \mathbb{E} \left( |\nabla f(X_{ij})|^2 \right). \]
For example, the Poincare inequality holds for models with Gaussian entries or with complex entries uniformly distributed on the unit circle but not for the model with ±1 entries. In this section we consider only the case of Wigner matrices with real entries that satisfy the Poincare inequality. The results can be extended to the case of sample covariance matrices or to the complex case.

First, let us define the coefficients in the expansion of a function $f$ over Chebyshev polynomials:

$$
\hat{f}_k := \begin{cases} 
\frac{2}{\pi} \int_{-1}^{1} f(x) T_k(x) \frac{dx}{\sqrt{1-x^2}}, & \text{for } k \geq 1, \\
\frac{1}{\pi} \int_{-1}^{1} f(x) \frac{dx}{\sqrt{1-x^2}}, & \text{for } k = 1.
\end{cases}
$$

Let $F$ be the linear subspace of functions $f : \mathbb{R} \to \mathbb{R}$ which are differentiable with continuous derivative in the interval $I_{\delta} = [1 - \delta, 1 + \delta]$.

**Theorem 1.3.** Assume that $A_N$ and $B_N$ are the overlapping real Hermitian matrices with the moments of matrix entries described in Section 1.1. Assume that matrix entries satisfy the Poincare inequality. Then for every pair of functions $f$ and $g$ in $F$, the linear statistics $N_o(f, A_N)$ and $N_o(g, B_N)$ converge in distribution to the bivariate Gaussian random variable with the covariance matrix $V$:

$$
V_{11} = \frac{1}{2} \left( \frac{d_2}{2} (\hat{f}_1)^2 + (m_4 - 1) (\hat{f}_2)^2 + \sum_{k=3}^{m} k (\hat{f}_k)^2 \right),
$$

$$
V_{12} = \frac{1}{2} \left( \frac{d_2}{2} \hat{f}_1 \hat{g}_1 \gamma + (m_4 - 1) \hat{f}_2 \hat{g}_2 \gamma^2 + \sum_{k=3}^{m} k \hat{f}_k \hat{g}_k (\gamma)^k \right),
$$

$$
V_{22} = \frac{1}{2} \left( \frac{d_2}{2} (\hat{g}_1)^2 + (m_4 - 1) (\hat{g}_2)^2 + \sum_{k=3}^{m} k (\hat{g}_k)^2 \right),
$$

where $\gamma = \Delta / \sqrt{ab}$.

Let us put the results of this paper in a more general prospective. Linear statistics of sample covariance matrices have been first investigated in Jonsson [15], who established the joint CLT for the moments of eigenvalues. A recent contribution is by Anderson and Zeitouni [2], who extended the study of linear statistics to a very general class of matrices with independent entries of changing variance. They derived a formula for the covariance of linear eigenvalue statistics, proved a CLT theorem for continuously differentiable test functions, and noted a relation to the Chebyshev polynomials. For more restricted classes of matrices, namely, for Gaussian and unitarily invariant matrices, important results about linear statistics were established in Diaconis-Shahshahani [8], Costin-Lebowitz [6], Johansson [13], Soshnikov [20] and [21], and Diaconis-Evans [7].

The main novelty of our results is that they extend the investigation to the spectra of overlapping submatrices. In addition, they explain the central role of the Chebyshev polynomials of the first kind by relating these polynomials to the number of non-backtracking tailless paths on a finite graph. This extends previous results by Feldheim and Sodin in [9], which were concerned with the Chebyshev polynomials of the second type.
In addition, we have found that the covariance of Chebyshev polynomials of the degree higher than the second depends only on the matrix entry variance, not on their higher moments. This result holds both for Wigner and sample covariance matrices. In order to see the significance of this result, recall that from the “four moment theorem” by Tao and Vu [23], we can expect that the covariance matrix in the CLT for linear statistics depends only on the first four moments of the matrix entries. However, the fact that the covariance for high-degree Chebyshev polynomials depends actually only on the first two moments is surprising and it seems that it was not noted before in the literature.

The rest of the paper is organized as follows. In Section 2, we will show how the Chebyshev polynomials of the first type are related to the non-backtracking tailless paths. In Section 3, we will prove the CLT results for simpler models in which the entries are either \( \pm 1 \) or uniformly distributed on the unit circle (Theorems 3.1 and 3.2). In Section 4, we will prove Theorems 1.1 and 1.2. Section 5 is devoted to the proof of Theorem 1.3, and Section 6 concludes.

2. The Chebyshev T-polynomials and non-backtracking paths

The Chebyshev polynomials of the first and the second kind are defined by the formulas

\[
T_k (\cos \theta) = \cos (k \theta) \quad \text{and} \quad U_k (\cos \theta) = \frac{\sin (k \theta)}{\sin \theta},
\]

respectively. For \( k \geq 2 \), both the \( T \) and \( U \) polynomials satisfy the same recursion. For example, for \( U \) polynomials, it is

\[
U_k (x) = 2x U_{k-1} (x) - U_{k-2} (x).
\]

The initial conditions in this recursion are \( T_0 (x) = 1 \) and \( T_1 (x) = x \) for \( T \) polynomials and \( U_0 (x) = 1 \) and \( U_1 (x) = 2x \) for \( U \) polynomials.

The \( T \) and \( U \) polynomials can be related as follows:

\[
U_k (x) = 2 \left[ T_k (x) + T_{k-2} (x) + \ldots + T_1 (x) \right] + (\varepsilon - 1),
\]

where \( \varepsilon = 1 \) if \( k \) is odd and \( \varepsilon = 0 \) if \( k \) is even.

2.1. Wigner matrices. Let \( G \) be a \((d + 1)\)-regular graph with the vertex set \( V = \{1, \ldots, n\} \). We will say that the matrix \( A \) is a generalized adjacency matrix if it is Hermitian and if \( |A_{uv}| = 1 \) if \((u, v)\) is an edge of \( G \) and 0 otherwise.

A path \( \gamma \) is a sequence of vertices \((u_0, u_1, \ldots, u_k)\) which are adjacent in graph \( G \). The length of this path is \( k \). The path is called non-backtracking if \( u_{j+1} \neq u_j \). It is closed if \( u_k = u_0 \). A closed path is non-backtracking tailless if it is non-backtracking and \( u_{k-1} \neq u_1 \).

Theorem 2.1 and the following remark are essentially due to Feldheim and Sodin [9]. (See Lemma II.1.1 and Claim II.1.2 in their paper, where this result is proved for complete graphs.)

**Theorem 2.1** (Feldheim-Sodin). Suppose that \( A \) is a generalized adjacency matrix for a \((d + 1)\)-regular graph \( G \). Then for all \( k \geq 1 \),

\[
\sum_{u_0 = u, u_k = v} A_{u_0 u_1} A_{u_1 u_2} \ldots A_{u_{k-1} u_k} = [P_k (A)]_{uv}
\]

where the sum is over all non-backtracking length-\( k \) paths from \( u \) to \( v \), and \( P_k (x) \) is a polynomial defined for \( k = 1, 2 \) as \( P_1 (x) := x \), \( P_2 (x) = x^2 - (d + 1) \), and for \( k \geq 3 \) by the
recursion:

\[ P_k(x) = xP_{k-1}(x) - dP_{k-2}(x). \]  

(7)

Remark: \( P_k(x) \) can be expressed in terms of Chebyshev’s \( U \)-polynomials as follows:

\[ P_k(x) = d^{k/2}U_k \left( \frac{x}{2\sqrt{d}} \right) - d^{(k-2)/2}U_{k-2} \left( \frac{x}{2\sqrt{d}} \right). \]  

(8)

Theorem 2.2. Suppose that \( A \) is a generalized adjacency matrix for a \((d + 1)\)-regular graph \( G \). Then for all \( k \geq 1 \),

\[ \sum A_{u_0u_1}A_{u_1u_2} \cdots A_{u_{k-1}u_0} = \begin{cases} \text{Tr} \left[ 2d^{k/2}T_k \left( \frac{A}{2\sqrt{d}} \right) \right], & \text{if } k \text{ is odd,} \\ \text{Tr} \left[ 2d^{k/2}T_k \left( \frac{A}{2\sqrt{d}} \right) + (d - 1) I_n \right], & \text{if } k \text{ is even,} \end{cases} \]

where the sum on the left hand-side is over all closed non-backtracking tailless paths \((u_0, u_1, \ldots, u_{k-1}, u_0)\) of length \( k \).

Proof: Let

\[ Q_k(A, u_0) := \sum A_{u_0u_1}A_{u_1u_2} \cdots A_{u_{k-1}u_0}, \]

where the sum over all closed non-backtracking tailless paths of length \( k \) that start from \( u_0 \).

Also, define

\[ Q'_{k-2}(A, u_1, u_0) := \sum A_{u_1u_2} \cdots A_{u_{k-2}u_1}, \]

where the sum is over all closed non-backtracking tailless paths of length \( k - 2 \) that start from \( u_1 \), except that the paths that start with \((u_1, u_0)\) and those that end with \((u_0, u_1)\) are excluded. Then we can write:

\[ [P_k(A)]_{u_0u_0} = Q_k(A, u_0) + \sum_{(u_0, u_1)} Q'_{k-2}(A, u_1, u_0) + \sum_{(u_0, u_1, u_2)} Q'_{k-4}(A, u_2, u_1) + \cdots, \]

where the sums are over non-backtracking paths that start from \((u_0, \ldots, u_s)\) and return along the same path in the opposite direction.

For example, the second term corresponds to paths that start at \( u_0 \), go to \( u_1 \), then perform a non-backtracking tailless closed path of length \( k - 2 \) and then return to \( u_0 \).

Next, we sum this expression over all possible \( u_0 \) and count how many times a term corresponding to a particular non-backtracking tailless path of length \( k - 2s \), say \( \gamma = (v_0, v_1, \ldots, v_{k-2s-1}, v_0) \), enters the resulting expression. It is easy to see that such a term must be a part of the sum

\[ \sum_{(u_0, \ldots, u_s)} Q'_{k-2s}(A, u_s, u_{s-1}), \]

where \( u_0 \) is now arbitrary. In order for \( \gamma \) to be counted by a term indexed by \((u_0, \ldots, u_s)\) in this sum, the non-backtracking path \((u_0, \ldots, u_s)\) has to satisfy the following constraints: \( u_s = v_0, u_{s-1} \neq v_1, u_{s-1} \neq v_{k-2s-1} \). Therefore, \( \gamma \) will be counted \((d - 1) d^{s-1} \) times.
Hence, we can write:

\[ \text{Tr} [ P_k (A) ] = \sum_{v_0} Q_k (A, v_0) + (d - 1) \sum_{v_0} Q_{k-2} (A, v_0) + (d - 1) d \sum_{v_0} Q_{k-4} (A, v_0) + \ldots. \]

In addition, \( \sum_{v_0} Q_k (A, v_0) = 0 \) for \( k = 1 \) and \( k = 2 \). Then, formulas (5), (8), and (9) imply immediately that

\[ \sum_{v_0} Q_k (A, v_0) = \left\{ \begin{array}{ll}
\text{Tr} \left[ 2d^{k/2} T_k \left( \frac{A}{\sqrt{d}} \right) \right], & \text{if } k \text{ is odd}, \\
\text{Tr} \left[ 2d^{k/2} T_k \left( \frac{A}{\sqrt{d}} \right) + (d - 1) I_n \right], & \text{if } k \text{ is even}.
\end{array} \right. \]

\[ \square \]

2.2. Sample Covariance Matrices. A bipartite graph is a graph whose vertices belong to two sets, \( V \) and \( W \), such that the vertices in \( V \) are connected only to vertices in \( W \) and vice versa. A bipartite graph is \( (c + 1, d + 1) \)-regular if every vertex in \( V \) is connected to \( c + 1 \) vertices in \( W \) and every vertex in \( W \) is connected to \( d + 1 \) vertices in \( V \).

Let \( G \) be a \( (c + 1, d + 1) \)-regular graph with \( |V| = n \) and \( |W| = m \). Consider an \( n \times m \) matrix \( A \). We will identify row indices with elements of \( V \) and column indices with elements of \( W \). We say that an \( n \times m \) matrix \( A \) is a generalized adjacency matrix for a bipartite graph \( G \), if \( |A_{uv}| = 1 \) for \( (u, v) \in G \) and \( A_{uv} = 0 \) otherwise.

Let us define the following object

\[ R_k (A, v_0, v_k) = \sum A_{v_0 w_1} A_{v_1 w_1} \ldots A_{v_{k-1} w_k} A_{v_k w_k}, \]

where the summation is over all non-backtracking paths \((v_0, w_1, v_1, \ldots, w_k, v_k)\) of length \( 2k \) from \( v_0 \) to \( v_k \).

The following two results are essentially due to Feldheim and Sodin [2]. (However, their expressions for \( R_k \) in terms of Chebyshev polynomials are incorrect. Compare their Lemma IV.1.1 and Claim IV.1.2.)

**Theorem 2.3.** Suppose that the matrix \( A \) is a generalized adjacency matrix for a \((c + 1, d + 1)\)-regular bipartite graph \( G \). Then for all \( k \geq 1 \),

\[ R_k (A, v_0, v_k) = [R_k (AA^*)]_{v_0 v_k}, \]

where \( R_k (x) \) are polynomials, which for \( k = 1, 2 \) are defined as \( R_1 (x) = x - (c + 1) \) and \( R_2 (x) = x^2 - (2c + d + 1) x + (c + 1)c \), and for \( k \geq 3 \), by the following recursion:

\[ R_k (x) := (x - (c + d)) R_{k-1} (x) - cd R_{k-2} (x). \]

**Proof:** It is easy to check the statement for \( k = 1 \). Indeed, \( R_1 (A, v_0, v_1) = (AA^*)_{v_0 v_1} \) if \( v_0 \neq v_1 \), and \( = 0 = ((AA^*)_{v_0 v_0}) - (c + 1) \) if \( v_0 = v_1 \). Hence \( R_1 (A, v_0, v_1) = [AA^* - (c + 1) I]_{v_0 v_1} \).

For \( k \geq 2 \), we multiply \( R_{k-1} (A, v_0, v_{k-1}) \) by

\[ R_1 (A, v_{k-1}, v_k) = [AA^* - (c + 1) I]_{v_{k-1} v_k} = \left\{ \begin{array}{ll}
\sum w_k A_{v_{k-1} w_k} A_{v_k w_k}, & \text{if } v_k \neq v_{k-1}, \\
0, & \text{if } v_k = v_{k-1},
\end{array} \right. \]
and sum over $v_{k-1}$ and $w_k$. We use definition (10) for $R_{k-1} (A, v_0, v_{k-1})$. If $w_k \neq w_{k-1}$, we get an element of the defining sum for $R_k (A, v_0, v_k)$. If $w_k = w_{k-1}$ and $v_k \neq v_{k-2}$ then we get an element of $R_{k-1} (A, v_0, v_k)$ and every of these elements will be counted $d - 1$ times (once for each of possible $v_{k-1} \neq v_{k-2}, v_{k-1} \neq v_k$). If $w_k = w_{k-1}$ and $v_k = v_{k-2}$ then we get an element of the sum $R_{k-2} (A, v_0, v_k)$ and every of these elements will be counted $cd$ times if $k \geq 3$ and $(c + 1) d$ times if $k = 2$ (every element will be counted once for each of possible $v_{k-1} \neq v_k$ and $w_{k-1} \neq w_{k-2}$). For $k \geq 3$, this gives the relation

$$R_k (A, v_0, v_k) = \sum_{v_{k-1}} R_{k-1} (A, v_0, v_{k-1}) (AA^*)_{v_{k-1}v_k} - (c + d) R_{k-1} (A, v_0, v_k) - cdR_{k-2} (A, v_0, v_k).$$

It is also easy to check the formula for $R_k (A, v_0, v_k)$ in the case $k = 2$. Altogether, these relations show that $R_k (A, v_0, v_k)$ is a matrix element of the polynomial $R_k (AA^*)$. □

**Lemma 2.4.** Let

$$\bar{U}_k (x) := (cd)^{k/2} U_k \left( \frac{x - (c + d)}{2\sqrt{cd}} \right),$$

where $U_k$ are the Chebyshev polynomials of the second kind and $\bar{U}_k := 0$ for $k < 0$. Then, for $k \geq 1$

$$R_k (x) := \bar{U}_k (x) + (d - 1) \bar{U}_{k-1} (x) - d\bar{U}_{k-2} (x).$$

Proof of the lemma is by verifying that the formula satisfies the recursion and that it gives the correct initial terms of the recursion.

Now, let us define

$$\bar{T}_k (x) := (cd)^{k/2} T_k \left( \frac{x - (c + d)}{2\sqrt{cd}} \right),$$

where $T_k$ are the Chebyshev polynomials of the first kind.

**Theorem 2.5.** Suppose that the matrix $A$ is a generalized adjacency matrix for a $(c + 1, d + 1)$-regular bipartite graph $G$ with vertex set $V \cup W$. Assume $c \geq d$. Then for all $k \geq 1$,

$$\sum A_{v_0w_1A_{v_1w_2}A_{v_2w_2}} \cdots A_{v_{k-1}w_kA_{v_kw_k}} = \text{Tr} \left[ 2\bar{T}_k (AA^*) + \frac{(c - d) (-d)^k + cd - 1}{d + 1} I \right]$$

where the sum is over all closed non-backtracking tailless paths $(v_0, w_1, v_1, w_2, \ldots, v_{k-1}, w_k, v_0)$ of length $2k$ that start with a vertex in $V$.

**Proof:** Let

$$S_k (v_0) := \sum A_{v_0w_1A_{v_1w_1}A_{v_2w_2}A_{v_2w_2}} \cdots A_{v_{k-1}w_kA_{v_kw_k}},$$

where the sum over all closed non-backtracking tailless paths $(v_0, w_1, v_1, w_2, \ldots, v_{k-1}, w_k, v_0)$ of length $2k$ that start from $v_0 \in V$. Also, define $S_k' (v_1, w_1)$ as the sum over all closed non-backtracking tailless paths of length $2k$ that originate at $v_1 \in V$ then proceed along an edge different from $(v_1, w_1)$ and end along an edge different from $(w_1, v_1)$.
We define $\tilde{S}_k (w_1)$ and $\tilde{S}'_k (w_1, v_0)$ similarly except now the paths start with a vertex in $W$, for example,

$$\tilde{S}_k (w_1) := \sum A_{v_1 w_1} A_{v_1 w_2} A_{v_2 w_2} \cdots A_{v_{k-1} w_k} A_{v_k W} A_{v_k w_1},$$

Then, we can write the following identity:

$$R_k (A, v_0, v_0) = S_k (v_0) + \sum_{(v_0, v_1)} \tilde{S}_{k-1} (w_1, v_0) + \sum_{(v_0, v_1, v_2)} S'_{k-2} (w_1, v_1) + \sum_{(v_0, v_1, v_2, v_3)} \tilde{S}'_{k-3} (w_2, v_1) + \ldots,$$

where sums are over non-backtracking paths.

Let us sum this identity over $v_0$, and consider, for example, a specific path of length $2 (k - 1)$, namely $\gamma = (w_1, v_1, \ldots, w_{k-1}, v_{k-1}, w_1)$. In the sum $\sum_{v_0, w_1} \tilde{S}_{k-1} (w_1, v_0)$ this path will be counted $d - 1$ times (once for all possible $v_0$ not equal to $v_1$ and not equal to $v_{k-1}$). By a similar counting we find that the following identity holds:

$$\sum_{v_0} R_k (A, v_0, v_0) = \sum_{v_0} S_k (v_0) + (d - 1) \sum_{v_0} \tilde{S}_{k-1} (w_0) + (c - 1) c \sum_{v_0} S_{k-2} (v_0)$$

$$+ (d - 1) c d \sum_{v_0} \tilde{S}_{k-3} (w_0) + (c - 1) c d^2 \sum_{v_0} S_{k-4} (v_0) + \ldots$$

Next, we note that the shift transformation $(v_0, w_1, v_1, \ldots, w_k, v_0) \rightarrow (w_1, v_1, \ldots, w_k, v_0, w_1)$ defines a bijection of the non-backtracking tailless paths that start with a vertex in $V$ to non-backtracking tailless paths that start with a vertex in $W$. In addition, obviously

$$A_{v_0 w_1} A_{v_1 w_1} A_{v_1 w_2} A_{v_2 w_2} \cdots A_{v_{k-1} w_k} A_{v_k W} = A_{v_1 w_1} A_{v_1 w_2} A_{v_2 w_2} \cdots A_{v_{k-1} w_k} A_{v_k W} A_{v_k w_1}.$$

Hence, $\sum_{v_0} \tilde{S}_k (w_0) = \sum_{v_0} S_k (v_0)$ for all $k$, and we can write

$$\sum_{v_0} R_k (A, v_0, v_0) = \left[ \sum_{v_0} S_k (v_0) + c d \sum_{v_0} S_{k-2} (v_0) + (c d)^2 \sum_{v_0} S_{k-4} (v_0) + \ldots \right]$$

$$+ (d - 1) \left[ \sum_{v_0} S_{k-1} (v_0) + c d \sum_{v_0} S_{k-3} (v_0) + \ldots \right]$$

$$- d \left[ \sum_{v_0} S_{k-2} (v_0) + c d \sum_{v_0} S_{k-4} (v_0) + \ldots \right].$$

It remains to check that if we define the polynomial $S_k (x)$ by the formula

$$S_k (x) = 2 \tilde{T}_k (x) + \frac{x^{d-1}}{d + 1}$$

for $k \geq 1$ and $S_k (x) = 0$ for $k \leq 0$, and if $R_k (x)$ are defined as in Lemma 2.4, then the following identity is true:

$$R_k (x) = \left[ S_k (x) + c d S_{k-2} (x) + (c d)^2 S_{k-4} (x) + \ldots \right]$$

$$+ (d - 1) \left[ S_{k-1} (x) + c d S_{k-3} (x) + \ldots \right]$$

$$- d \left[ S_{k-2} (x) + c d S_{k-4} (x) + \ldots \right].$$

This can be verified by using the identity (5). □
3. Linear statistics of polynomials for simple models.

We will use the results of the previous section to prove Theorems 3.1 and 3.2 below.

3.1. Wigner matrices. In this section we consider the following simple model. Assume that
the diagonal entries are zero. We consider two cases for off-diagonal entries. Either the entries
are uniformly distributed on the unit circle or they take values ±1 with probability 1/2.

In this section the normalized matrices \( \tilde{A}_N \) will be defined as follows:
\[
\tilde{A}_N := \frac{1}{2\sqrt{(a(N) - 2)}} A_N, \quad \text{and} \quad \tilde{B}_N := \frac{1}{2\sqrt{(b(N) - 2)}} B_N.
\]

The choice of \( a(N) - 2 \) instead of \( a(N) \) and \( b(N) - 2 \) instead of \( b(N) \) is clearly not essential for
first-order asymptotics. However it makes some formulas shorter. Recall that if \( f : \mathbb{R} \rightarrow \mathbb{R} \)
is a test function, then we define its linear statistic for matrix \( A_N \) as
\[
\mathcal{N}(f, A_N) := \text{Tr} \left[ f \left( \tilde{A}_N \right) \right] = \sum f(\lambda_i),
\]
where \( \lambda_i \) are the eigenvalues of the matrix \( \tilde{A}_N \). The quantity \( \mathcal{N}(f, B_N) \) is defined similarly.
The centered statistics \( \mathcal{N}^0(T_k, A_N) \) and \( \mathcal{N}^0(T_l, B_N) \) are obtained by subtracting the corre-
spending expectation values.

**Theorem 3.1.** Let \( T_k(x) \) denote the Chebyshev polynomials of the first kind. As \( N \rightarrow \infty \),
the distribution of the centered linear statistics \( \mathcal{N}^0(T_k, A_N) \) and \( \mathcal{N}^0(T_l, B_N) \) converges to a
two-variate Gaussian distribution with the covariance equal to
\[
\begin{cases}
\delta_{kl} \frac{k}{2\beta^3} \left( \frac{a}{\sqrt{ab}} \right)^k, & \text{if } \min \{k,l\} \geq 3, \\
0, & \text{if } \min \{k,l\} \leq 2,
\end{cases}
\]
with \( \beta = 1 \) for the model with ±1 entries and \( \beta = 2 \) for the model with entries on the unit
circle.

**Proof of Theorem 3.1** We will only compute covariance. The proof that other moments
are determined by covariance follows from a similar combinatorial analysis. This analysis is
sketched below in the proof of the corresponding theorem for the sample covariance matrices.

Let \( A(v_0, v_1, \ldots, v_{k-1}, v_0) \) denote the product \( A_{v_0v_1} A_{v_1v_2} \cdots A_{v_{k-1}v_0} \), and let us estimate
the following object:
\[
\frac{1}{(t_N)^k} \left[ \sum \mathbb{E} A(v_0, \ldots, v_{k-1}, v_0) B(v_0', \ldots, v_{k-1}', v_0') - \sum \mathbb{E} A(v_0, \ldots, v_{k-1}, v_0) \sum \mathbb{E} B(v_0', \ldots, v_{k-1}', v_0') \right],
\]
where the sums are over all pairs of non-backtracking tailless ("NBT") cyclic paths of length
\( k \). Consider the sum
\[
\sum \mathbb{E} A(v_0, \ldots, v_{k-1}, v_0) B(v_0', \ldots, v_{k-1}', v_0'),
\]
and assign a type and a reduced type to each term in this sum. The type of a term is a graph
and a pair of paths on this graph. The graph is formed by the edges of the NBT closed paths
\( (v_0, \ldots, v_{k-1}, v_0) \) and \( (v_0', \ldots, v_{k-1}', v_0') \) and the paths are induced by these two paths. We
understand that the original labels of the vertices are removed in the type graph. In other words, the type is an equivalence class defined up to isomorphisms induced by arbitrary relabellings of the vertices. For example, any two disjoint cyclic NBT paths (in which no vertex except the original one is repeated twice) correspond to the same type. The reduced type is simply the type graph with the information about the paths removed.

Note that due to the expectation sign, the only types that bring a non-zero contribution to the sum are those in which every edge is traversed even number of times by the NBT paths.

In addition, if a term has a disconnected reduced type graph then its contribution will be cancelled by the contribution of a corresponding term from the product of the sums \( \sum \mathbb{E} A \left( v_0, \ldots, v_{k-1}, v_0 \right) \) and \( \sum \mathbb{E} B \left( v'_0, \ldots, v'_{k-1}, v'_0 \right) \). Hence we only need to consider terms with a connected reduced type graph.

Finally, note the crucial observation that since the paths are non-backtracking and tailless, hence every vertex in the reduced type graph has the degree greater or equal to two.

Consider the sum of all terms that have the reduced type graph equal to the cycle on \( k \) vertices. Since both paths are on this cycle, hence the vertices must have the labels from rows and columns that are common to both \( A \) and \( B \) matrices. Hence, the number of valid labelings of the reduced type graph is asymptotically close to \( \left( \Delta^{(N)} t_N \right)^k \), with the error \( O \left( k t_N^{k-1} \right) \). Next, note that each reduced type contains \( 2^k \) different types since every NBT path can start from any of \( k \) possible vertices in \( V \) and the path can have either the same or opposite orientation. In the case \( \beta = 1 \) both orientations contribute and in the case \( \beta = 2 \), only one orientation contribute. Hence, the total number of terms in this reduced type can be estimated as

\[
\frac{2}{\beta} k \left( \Delta^{(N)} \right)^k (t_N)^k + O \left( k t_N^{k-1} \right)
\]

Each term of this type brings a contribution of 1 to the sum \[14\]. After taking into account the normalization, we find that the contribution of this reduced type to the covariance is \( \frac{2}{\beta} k \left( \Delta^{(N)} \right)^k (t_N)^k + O \left( k t_N^{k-1} \right) \).

The next step is to show that the contribution of all other terms is negligible for all other types if \( N \) is large. This is straightforward. Indeed, if any edge is traversed by the NBT paths more than twice, then the total number of edges in the reduced type graph is \( < k \), hence the number of vertices is also less than \( k \) (using the fact that each vertex must have the degree of at least 2). The total number of paths that connect \( r \) vertices is smaller than \( r^r \). Hence the number of terms with this reduced type is of order not greater than \( O \left( k t_N^{k-1} \right) \) or less, which implies that these terms give a negligible contribution for large \( N \) provided that \( k^k \ll t_N \). This holds if \( k \ll \log N / (\log \log N) \), in particular for fixed \( k \).

Next, suppose that every edge is traversed exactly twice by the NBT paths, hence the number of edges is \( k \). Suppose, however, that the graph is not a cycle and therefore one of the vertices must have the degree of at least 3 (since we ruled out disconnected graphs and all vertices have the degree of at least 2). Since sum of degrees is twice the number of edges it
follows again that the number of vertices is \( k \) and the number of terms in this reduced type is bounded by \( O \left( k^k N^{-1} \right) \). Hence the contribution of these terms is negligible.

For \( k \neq l \), we find that contributions of all types are negligible.

Therefore, by using Theorem 2.2, we find that

\[
\lim_{N \to \infty} \text{Cov} \left( \mathcal{N}^\circ \left( T_k, A_N \right), \mathcal{N}^\circ \left( T_l, B_N \right) \right) = \begin{cases} 
\delta_{kl} k^2 \beta \left( \frac{\Delta}{\sqrt{ab}} \right)^k , & \text{if } \min\{k, l\} \geq 3, \\
0 , & \text{if } \min\{k, l\} \leq 2.
\end{cases}
\]  

(15)

The condition on \( k \) and \( l \) arises because there are no cycles of length 1 and 2 on the underlying graph. □

3.2. Sample covariance matrices. In this section we will assume that the entries of matrices \( A_N \) and \( B_N \) are either \( \pm 1 \) with equal probability, which we call \( \beta = 1 \) or the real case, or are uniformly distributed on the unit circle, which we call \( \beta = 2 \) or the complex case.

Let us define the normalized sample covariance matrices in the following form:

\[
W_{A_N} := \frac{1}{2} \frac{1}{\left( a_1^{(N)} - 1 \right) \left( a_2^{(N)} - 1 \right)} \left[ A_N A_N^* - \left( a_1^{(N)} + a_2^{(N)} - 2 \right) I_{a_1^{(N)}} \right],
\]

where \( I_{a_1^{(N)}} \) is the \( a_1^{(N)} \)-by-\( a_1^{(N)} \) identity matrix, and

\[
W_{B_N} := \frac{1}{2} \frac{1}{\left( b_1^{(N)} - 1 \right) \left( b_2^{(N)} - 1 \right)} \left[ B_N B_N^* - \left( b_1^{(N)} + b_2^{(N)} - 2 \right) I_{b_1^{(N)}} \right].
\]

The normalization is chosen in such a way that the empirical distribution of eigenvalues of \( W_{A_N} \) and \( W_{B_N} \) converges to a distribution supported on the interval \([-1, 1]\). The choice of \( a_i^{(N)} - 1 \) instead of \( a_i^{(N)} \) and \( b_i^{(N)} - 1 \) instead of \( b_i^{(N)} \) is not essential for asymptotics, and it makes some formulas shorter. Define

\[
\mathcal{N} \left( f, W_{A_N} \right) := \text{Tr} \left[ f \left( W_{A_N} \right) \right] = \sum f \left( \lambda_i \right),
\]

where \( \lambda_i \) are the eigenvalues of the matrix \( W_{A_N} \), and let

\[
\mathcal{N}^\circ \left( f, W_{A_N} \right) := \mathcal{N} \left( f, W_{A_N} \right) - \mathbb{E} \mathcal{N} \left( f, W_{A_N} \right).
\]

(16)

The linear statistics for the matrix \( W_{B_N} \) are defined similarly.

**Theorem 3.2.** Let \( T_k (x) \) denote the Chebyshev polynomials of the first kind. As \( N \to \infty \), the distribution of the centered linear statistics \( \mathcal{N}^\circ \left( T_k, W_{A_N} \right) \) and \( \mathcal{N}^\circ \left( T_l, W_{B_N} \right) \) converges to a two-variate Gaussian distribution with the covariance equal to

\[
\delta_{kl} k^2 \beta \gamma^k ,
\]

if \( \min\{k, l\} > 1 \) and 0 otherwise. Here \( \beta = 1 \) for the model with \( \pm 1 \) entries and \( \beta = 2 \) for the model with entries on the unit circle.
We will prove Theorem 3.2 in a slightly different form. First, recall the definition of the rescaled Chebyshev polynomials:

\[ T_k[c, d](x) = (cd)^{k/2} T_k \left( \frac{x - (c + d)}{2 \sqrt{cd}} \right), \]

where \( T_k \) are the Chebyshev polynomials of the first kind. (Here we explicitly indicate that \( T_k \) depends on the parameters \( c \) and \( d \).)

Next, recall that there is a parameter \( t_N \) that approaches infinity as \( N \to \infty \), and that quantities \( a_1(N)/t_N, a_2(N)/t_N, b_1(N)/t_N, b_2(N)/t_N, \Delta_1(N)/t_N, \) and \( \Delta_2(N)/t_N \) approach some positive limits \( a_1, a_2, b_1, b_2, \Delta_1, \) and \( \Delta_2 \), respectively.

Define the rescaled linear statistics:

\[
X_k(A_N) := (t_N)^{-k} \left[ {\mathcal{N}}_{\left( \tilde{T}_k[a_1(N) - 1, a_2(N) - 1], A_N A_N^* \right)} - {\mathbb{E}}{\mathcal{N}}_{\left( \tilde{T}_k[a_1(N) - 1, a_2(N) - 1], A_N A_N^* \right)} \right],
\]

\[
X_k(B_N) := (t_N)^{-k} \left[ {\mathcal{N}}_{\left( \tilde{T}_k[b_1(N) - 1, b_2(N) - 1], B_N B_N^* \right)} - {\mathbb{E}}{\mathcal{N}}_{\left( \tilde{T}_k[b_1(N) - 1, b_2(N) - 1], B_N B_N^* \right)} \right].
\]

Note that \( X_k(A_N) = \left( (a_1(N) - 1)(a_2(N) - 1) \right)^{k/2} N^\alpha(T_k, W_{A_N}) \) with \( N^\alpha(T_k, W_{A_N}) \) as defined in (16) and \( X_k(B_N) \) can be interpreted similarly.

We will show that the following result holds. It is easy to check that it is equivalent to the result in Theorem 3.2.

**Proposition 3.3.** As \( N \to \infty \), the distribution of the rescaled linear statistics \( X_k(A_N) \) and \( X_k(B_N) \) converges to a two-variate Gaussian distribution with the covariance equal to

\[
\delta_{kl} \frac{k^{\alpha}}{2^\beta} \begin{pmatrix}
(\Delta_1 \Delta_2)^k \\
(b_1 b_2)^k
\end{pmatrix},
\]

if \( \min \{ k, l \} > 1 \) and 0 otherwise.

**Proof of Proposition 3.3:** Note that matrix \( A_N \) is a generalized adjacency matrix for a complete bipartite graph \( G \). The vertex sets of \( G \) have \( a_1(N) \) and \( a_2(N) \) vertices, respectively. In particular, the graph is \( \left( a_1(N), a_2(N) \right) \)-regular. By Theorem 2.5, we see that

\[
2(t_N)^{-k} N\left( \tilde{T}_k[a_1(N) - 1, a_2(N) - 1], A_N A_N^* \right) = \frac{1}{(t_N)^k} \sum A(v_0, w_1, v_1, \ldots, w_k, v_0) - A^{(N)}_A,
\]

where

\[
r_A^{(N)} = \frac{1}{(t_N)^k} \frac{a_1(N)}{a_2(N)} \left[ \left( a_1(N) - a_2(N) \right) \left( -a_2(N) + 1 \right)^k + \left( a_1(N) - 1 \right) \left( a_2(N) - 1 \right) - 1 \right].
\]

In this formula, \( A(v_0, w_1, v_1, \ldots, w_k, v_0) \) denotes the product \( A_{v_0w_1}A_{v_1w_2}A_{v_2w_3} \ldots A_{v_{k-1}w_k}A_{v_kw_k} \) and the sum is over all non-backtracking tailless paths that start with a vertex in \( V \).

A similar formula holds for the linear statistics of the \( B_N \) matrix.

Since \( r_A^{(N)} \) and \( r_B^{(N)} \) are not random, we only need to understand the joint distribution of the normalized sums \( S_k(A_N) := (t_N)^{-k} \sum A(v_0, w_1, v_1, \ldots, w_k, v_0) \) and \( S_k(B_N) := (t_N)^{-k} \sum B(v_0, w_1, v_1, \ldots, w_k, v_0) \). In particular, we need to show that in the limit of large
Consider the case \( k = l \). (The case of \( k \neq l \) is similar and will be omitted.)

1. Covariance

We are interested in estimating the following object:

\[
\frac{1}{(t_N)^2k} \left[ \sum \mathbb{E}A(v_0, \ldots, w_k, v_0) B(v'_0, \ldots, w'_k, v'_0) - \sum \mathbb{E}A(v_0, \ldots, w_k, v_0) \sum \mathbb{E}B(v'_0, \ldots, w'_k, v'_0) \right],
\]

where the sum is over all pairs of non-backtracking tailless ("NBT") cyclic paths of length \( 2k \) each. Note that for \( k = 1 \), the number of such paths is zero. In the following we assume \( k > 1 \). Consider the sum

\[
\sum \mathbb{E}A(v_0, \ldots, w_k, v_0) B(v'_0, \ldots, w'_k, v'_0),
\]

and assign a type and a reduced type to each term in this sum. This is done as for Wigner matrices with a small modification. Namely, since the original graph is bipartite, the reduced type graphs are also bipartite and we will keep the information about the partition. In other words, while we remove the labels from the vertices, we will still keep information whether a given vertex corresponds to a row or a column of the matrix. (This information is not very important for this model. However it will be useful in a later model with more general matrix entries.)

First, note that due to the expectation sign, the only types that bring a non-zero contribution to the sum are those in which every edge is traversed even number of times by the NBT paths. In addition, if a term has a disconnected reduced type graph then its contribution will be cancelled by the contribution of a corresponding term from the product of the sums \( \sum \mathbb{E}A(v_0, \ldots, w_k, v_0) \) and \( \sum \mathbb{E}B(v'_0, \ldots, w'_k, v'_0) \). Hence we only need to consider terms with a connected reduced type graph.

Crucially, since the paths are non-backtracking and tailless, hence every vertex in the reduced type graph must have the degree greater or equal to two.

Now, consider the sum of all terms that have the reduced type graph equal to the cycle on \( 2k \) vertices. Since both paths are on this cycle, hence the vertices must have the labels from rows and columns that are common to both \( A \) and \( B \) matrices. Hence, the number of valid labelings of the reduced type graph is asymptotically close to \( (\Delta_1^{(N)} t_N)^k (\Delta_2^{(N)} t_N)^k \), with the error \( O\left(k t_N^{2k-1}\right) \). Next, note that each reduced type contains \( 2k \) different types since every NBT path can start from any of \( k \) possible vertices in \( V \) and the path can have either the same or opposite orientation. In the case \( \beta = 1 \) both orientations contribute and in the case \( \beta = 2 \), only one orientation contribute. Hence, the total number of terms belonging to this reduced type can be estimated as

\[
(2/\beta)k \left( \Delta_1^{(N)} \Delta_2^{(N)} \right)^k (t_N)^{2k} + O \left(k t_N^{2k-1}\right).
\]
Each term of this type brings a contribution of 1 to the sum. After taking into account the normalization, we find that the contribution of this reduced type to the covariance is 
\[
(2/\beta)k \left( \Delta_1^{(N)} \Delta_2^{(N)} \right)^k + O \left( k/t_N \right).
\]

The next step is to show that the contribution of all other terms is negligible for all other types if \( N \) is large. This is straightforward. Indeed, if any edge is traversed by the NBT paths more than twice, then the total number of edges in the reduced type graph is \(< 2k\), hence the number of vertices is also less than \( 2k \) (using the fact that each vertex must have the degree of at least 2). The total number of paths that connect \( r \) vertices is smaller than \( r^r \).

Hence the number of terms with this reduced type is of order not greater than \( O \left( k^2 k^2 t^2 - 1 \right) \) or less, which implies that these terms give a negligible contribution for large \( N \) provided that \( k^2 \ll t_N \). This holds if \( k \ll \log N / (2 \log \log N) \), in particular for fixed \( k \).

Next, suppose that each edge is traversed twice by the NBT paths but the reduced type graph is not a cycle. Then one of the vertices must have the degree of at least 3 (since we ruled out disconnected graphs and since all vertices must have the degree of at least 2.) Since the sum of degrees is twice the number of edges it follows again that the number of vertices is \(< 2k\) and the number of terms in this reduced type is bounded by \( O \left( k^2 k^2 t^2 - 1 \right) \). Hence the contribution of these terms is negligible. Therefore, we arrive to the conclusion that

\[
\lim_{N \to \infty} \text{Cov} \left( X_k (A_N), X_k (B_N) \right) = \lim_{N \to \infty} \text{Cov} \left( S_k (A_N), S_k (B_N) \right) = \frac{2k}{\beta} (\Delta_1 \Delta_2)^k.
\]

Similarly,

\[
\lim_{N \to \infty} \text{Cov} \left( X_k (A_N), X_k (A_N) \right) = \frac{2k}{\beta} (a_1 a_2)^k, \text{ and}
\]

\[
\lim_{N \to \infty} \text{Cov} \left( X_k (B_N), X_k (B_N) \right) = \frac{2k}{\beta} (b_1 b_2)^k.
\]

2. Higher moments

The argument for higher moments is similar. Consider an \( m \)-th joint moment,

\[
\mathbb{E} \left( [S_k (A_N)]^{m_a} [S_k (B_N)]^{m_b} \right),
\]

where \( m_a + m_b = m \). The type of a term in the expansion of this moment is given by a graph and \( m \) non-backtracking tailless paths which correspond to the factors in the product \( [S_k (A_N)]^{m_a} [S_k (B_N)]^{m_b} \). Non-negligible contributions come from the terms whose reduced type is the union of disjoint cycles of length \( 2k \). These cycles must be traversed by the NBT paths exactly twice, which implies, in particular that \( m \) must be even. This defines a matching on the set of NBT paths. Every pair in this matching contributes to the limit of the \( m \)-th moment a factor equal to the limit covariance of the terms corresponding to the NBT paths. Hence, the resulting sum coincides exactly with the Wick formula for the higher moments of the Gaussian distribution. (See Zee [24]). □
4. More General Matrices

4.1. Wigner. Proof of Theorem 1.1. Following Borodin [4], we can write the following expression for the limit of the covariance of \( \text{tr} \left( (A_N/\sqrt{\lambda N})^k \right) \) and \( \text{tr} \left( (B_N/\sqrt{\lambda N})^l \right) \) as \( N \to \infty \):

\[
d_2klC_{(k-1)/2}C_{(l-1)/2} \Delta a^{k-1}\frac{b}{2} \frac{l}{2} + (m_4 - 1) klC_{k/2}C_{l/2} \Delta^2 a^{k-1}b^{l-1}.
\]

\[+
\sum_{r=3}^{\infty} \frac{2kl}{r} \left( \sum_{s_i \geq 0} \prod C_{s_i} \right) \left( \sum_{l_i \geq 0} \prod C_{l_i} \right) \Delta^r a^{k-r}b^{l-r}.
\]

Here \( C_k := \binom{2k}{k} / (k + 1) \) are the Catalan numbers. Among other applications in combinatorics, they count the number of rooted planar trees with \( k \) edges.

For the convenience of the reader, let us explain the second term in this expression. Other terms can be obtained by a similar combinatorial analysis. Let us call “higher-order terms” all those terms in the expression for the covariance of \( \text{tr} \left( (A/\sqrt{\lambda N})^k \right) \) and \( \text{tr} \left( (B/\sqrt{\lambda N})^l \right) \) that involve matrix entry moments that are of the order higher than the second.

**Lemma 4.1.** The contribution of the higher-order terms to the covariance of \( \text{tr} \left( (A_N/\sqrt{\lambda N})^k \right) \) and \( \text{tr} \left( (B_N/\sqrt{\lambda N})^l \right) \) converges to

\[(m_4 - 1) klC_{k/2}C_{l/2} \Delta^2 a^{k-1}b^{l-1}.
\]

**Proof:** Recall that every term in the expression for the covariance of \( \text{tr} \left( (A_N/\sqrt{\lambda N})^k \right) \) and \( \text{tr} \left( (B_N/\sqrt{\lambda N})^l \right) \) is coded by two paths, \( \gamma_1 \) and \( \gamma_2 \), of the lengths \( k \) and \( l \), respectively. These paths trace the graphs \( G_1 \) and \( G_2 \). Let them have \( e_1 \) and \( e_2 \) edges, respectively, and let \( f \) of these edges be common to both of these graphs. Then the union of these graphs, \( G_1 \cup G_2 \), has \( e_1 + e_2 - f \) edges.

We can assume that \( G_1 \cup G_2 \) is connected, since otherwise the term that corresponds to \( (\gamma_1, \gamma_2) \) would give a zero contribution to the covariance. Hence, the number of vertices in \( G_1 \cup G_2 \) is \( e_1 + e_2 - f + \varepsilon \), where \( \varepsilon = 1 \) if \( G_1 \cup G_2 \) is a tree, \( \varepsilon = 0 \) if \( G_1 \cup G_2 \) contains exactly one cycle as a subgraph, and \( \varepsilon < 0 \) in all other cases.

Every edge in \( G_1 \cup G_2 \) must be traversed at least twice by the union of paths \( \gamma_1 \) and \( \gamma_2 \). Since we are interested in the contribution of higher-order terms, hence at least one edge must be traversed 3 times or more. Hence we have \( k + l - 2 (e_1 + e_2 - f) = \delta \), where \( \delta = 1 \) if one of the edges is traversed 3 times and all others - two times, \( \delta = 2 \) if one of the edges is traversed 4 times and all others - two times, or if two of the edges are traversed 3 times and all others - two times, and \( \delta > 2 \) in all other cases.

The contribution of the type \((G_1 \cup G_2, \gamma_1, \gamma_2)\) to the covariance is

\[< (c_Nt_N)|V|^\frac{1}{2} t_N^{-\frac{1}{2}}(k+l),\]
where \( c_N = \max \{ a(N), b(N) \} \) and \(|V|\) is number of vertices in \( G_1 \cup G_2 \). Hence, for large \( N \) it is bounded by

\[
c \exp \left( -\frac{1}{2} \log t_N \left[ k + l - 2 \left( e_1 + e_2 - f + \varepsilon \right) \right] \right) = c \exp \left( (2\varepsilon - \delta) \frac{\log t_N}{2} \right).
\]

Since \( \delta \geq 1 \) and \( \varepsilon \leq 1 \), this is non-negligible for large \( N \) only if \( \delta = 1 \) and \( \varepsilon = 1 \), or if \( \delta = 2 \) and \( \varepsilon = 1 \). However, if \( \varepsilon = 1 \), hence \( G_1 \cup G_2 \) is a tree, hence both \( G_1 \) and \( G_2 \) are trees, and therefore each of the closed paths \( \gamma_1 \) and \( \gamma_2 \) traverses each edge of the corresponding graph at least twice. Hence the case \( \delta = 1 \) and \( \varepsilon = 1 \) is impossible. The case \( \delta = 2 \) and \( \varepsilon = 1 \) correspond to the situation when \( G_1 \) and \( G_2 \) are two trees that have one edge in common. This common edge is traversed \( 4 \) times by paths \( \gamma_1 \) and \( \gamma_2 \), while all other edges are traversed two times each.

It is easy to see that this situation can happen if and only if \( k \) and \( l \) are odd and the contribution of each of these types converges to

\[
(m_4 - 1) \Delta^2 a^{\frac{k}{2} - 1} b^{\frac{l}{2} - 1}.
\]

This must be multiplied by \( klC_{k/2}C_{l/2} \), where \( C_{k/2}C_{l/2} \) is the number of pairs of trees, and \( kl \) correspond to the choice of the edges on these trees. The result is the limiting contribution stated by the lemma. □

In order to continue the proof of Theorem 1.1 note that from Theorem 3.1 we already know the value of the covariance of \( \text{tr} \left[ \left( A_N / \sqrt{t_N} \right)^k \right] \) and \( \text{tr} \left[ \left( B_N / \sqrt{t_N} \right)^l \right] \) for a particular model, in which \( d_2 := \mathbb{E} \left( X_{ii}^2 \right) = 0 \) and \( m_4 := \mathbb{E} \left( X_{ij}^4 \right) = 1 \). Hence we only need to find how the expressions

\[
\alpha_1 = \frac{k}{k-1} \left( \frac{l}{l-1} \right) \Delta a^{\frac{k}{2} - 1} b^{\frac{l}{2} - 1}, \quad \text{and}
\]
\[
\alpha_2 = \frac{k}{k-2} \left( \frac{l}{l-2} \right) \Delta^2 a^{\frac{k}{2} - 1} b^{\frac{l}{2} - 1}
\]

transform when we change the scaling and go from monomials \( x^k \) and \( x^l \) to the Chebyshev polynomials \( T_k(x) \) and \( T_l(x) \). First, note that these expressions can be written as contour integrals

\[
\alpha_1 = \frac{\Delta}{(2\pi i)^2} \int_{|z| < |w| = c_2} \left( z + \frac{a}{z} \right)^{k-1} \left( w + \frac{b}{w} \right)^{l-1} \frac{dz \; dw}{z \; w} \quad \text{and}
\]
\[
\alpha_2 = \frac{\Delta^2}{(2\pi i)^2} \int_{|z| < |w| = c_2} \left( z + \frac{a}{z} \right)^{k-2} \left( w + \frac{b}{w} \right)^{l-2} \frac{dz \; dw}{z \; w}.
\]
If we use the fact that \( \tilde{T}_{k,a} (x) := T_k (x/2\sqrt{a}) \), then we have
\[
\tilde{T}_{k,a} \left( z + \frac{a}{z} \right) = \frac{1}{2} \left[ \left( \frac{z}{\sqrt{a}} \right)^k + \left( \frac{\sqrt{a}}{z} \right)^k \right].
\]

Hence the covariance of linear statistics of \( \tilde{T}_{k,p,a} (x) \) and \( \tilde{T}_{k,q,b} (x) \) will include the additional term due to a linear combination of \( \alpha_1 \) terms:
\[
d_2 \frac{\Delta/4}{(2\pi i)^2} \int_{|z|=c_1} \left[ \left( \frac{z}{\sqrt{a}} \right)^k + \left( \frac{\sqrt{a}}{z} \right)^k \right] \frac{dz}{z (z + \frac{a}{z})} \times \int_{|w|=c_2} \left[ \left( \frac{w}{\sqrt{b}} \right)^l + \left( \frac{\sqrt{b}}{w} \right)^l \right] \frac{dw}{w (w + \frac{a}{w})}.
\]

This is different from zero if and only if \( k = l = 1 \), in which case we find the contribution of
\[
d_2 \frac{\Delta}{4 \sqrt{ab}}.
\]

Similarly, the contribution of the linear combination of the \( \alpha_2 \) terms is not zero for \( \tilde{T}_{k,p,a} (x) \) and \( \tilde{T}_{k,q,b} (x) \) if and only if \( k = l = 2 \), in which case it is
\[
m_4 - 1 \Delta^2 \frac{ab}{2}.
\]

Hence, the correction terms influence only the covariances of the first two Chebyshev polynomials (\( T_1 \) and \( T_2 \)). By using the result from Theorem 3.1, we easily find the correct values of covariances for the statistics of \( T_1 \) and \( T_2 \). \( \square \)

### 4.2. Sample Covariance

We proceed as in the previous section about Wigner matrices. Since we already know the covariance of the linear statistics for a particular model (from Theorem 3.2), hence we only need to estimate how a change in higher-order moments affects the covariance.

First, let us define the following functions of the parameters:
\[
R_k (a_1, a_2) = \left( \frac{a_1 + a_2}{2\sqrt{a_1 a_2}} \right)^k \sqrt{\frac{a_2}{a_1}} \times \sum_{s=1}^{k} \left( -1 \right)^s \frac{1}{a_1 + a_2}^s \binom{k}{s} \sum_{t=1}^{s} \binom{s}{t} a_1^t a_2^{s-t},
\]
where \( k \geq 1 \) and
\[
\binom{s}{t} := \frac{1}{s} \binom{s-1}{t-1} \binom{s}{t}.
\]

These coefficients are called Narayana numbers and they frequently occur in combinatorial problems. In particular, they count the number of rooted planar trees with \( n \) edges and \( k \) leaves, and the number of partitions of the set \( [n] = \{1, \ldots, n\} \) that have \( k \) blocks.
Proposition 4.2. Let $\Upsilon_{kl}^{(N)}$ be the additional term in the expression for the covariance of the statistics $N \to (x^k, W_A)$ and $N \to (x^l, W_B)$ which is due to the matrix entry moments of the order higher than the second. Then, as $N \to \infty$,
\[
\Upsilon_{kl}^{(N)} \to (m_4 - 1) \gamma R_k(a_1, a_2) R_l(b_1, b_2).
\]

Proof: Consider
\[
\mathbb{E} \left( \text{tr} \left( A_N A_N^* \right)^k \text{tr} \left( B_N B_N^* \right)^l \right).
\]
Let us expand this expression as a sum of products of matrix entries and take the expectation. We observe that the only reduced graph types that involve the entry moments higher than the second and that bring non-negligible contributions for large $N$ are given by two trees glued along an edge. One of these trees has $k$ edges and another one has $l$ edges. The paths have lengths $2k$ and $2l$, respectively, and they traverse every edge of the corresponding tree twice.

Now suppose that the tree with $k$ edges has $t_1$ row vertices and $k + 1 - t_1$ column vertices, and the tree with $l$ edges has $t_2$ row vertices and $k + 1 - t_2$ column vertices. (To avoid confusion, note that for simplicity we set the matrix size parameter $t_N$ equal to $N$ here, and we use letter $t$ for a different purpose.) Then the contribution of this reduced graph type is
\[
(m_4 - 1) \left( a_1^{(N)} N \right)^{t_1-1} \left( a_2^{(N)} N \right)^{k-t_1} \left( \Delta_1^{(N)} N \right) \left( b_1^{(N)} N \right)^{t_2-1} \left( b_2^{(N)} N \right)^{l-t_2} \left( \Delta_2^{(N)} N \right) + o(N^{-1}).
\]
Hence, for
\[
\text{Cov} \left( \text{tr} \left( \frac{A_N A_N^*}{2\sqrt{a_1 a_2 N}} \right)^k, \text{tr} \left( \frac{B_N B_N^*}{2\sqrt{b_1 b_2 N}} \right)^l \right), \tag{18}
\]
as $N \to \infty$, the contribution of this type will converge to
\[
(m_4 - 1) \frac{1}{2^{k+l}} \left( \frac{a_2}{a_1} \right)^{k-2t_1} \left( \frac{b_2}{b_1} \right)^{l-2t_2} \frac{\Delta_1}{a_1} \frac{\Delta_2}{b_1} = (m_4 - 1) \gamma \frac{1}{2^{k+l}} \left( \frac{a_2}{a_1} \right)^{k-2t_1+1/2} \left( \frac{b_2}{b_1} \right)^{l-2t_2+1/2}. \tag{19}
\]

In order to calculate the number of these reduced graph types, we need the following lemma.

Lemma 4.3. The number of the non-isomorphic bipartite planar rooted trees with $n$ edges that have $k$ vertices in one of the partitions equals the Narayana number $\left[ \begin{array}{c} n \\ k \end{array} \right]$.

Proof: We will prove the claim of the lemma by showing that the number of the bipartite planar rooted trees with $n$ edges and $k$ vertices in one of the partitions equals the number of non-crossing partitions of $[n]$ with $k$ blocks, which is the Narayana number $\left[ \begin{array}{c} n \\ k \end{array} \right]$.

Recall that a non-crossing partition of $[n] = \{1, \ldots, n\}$ is one for which no 4-tuple $a < b < c < d$ has $a$ and $c$ in one block and $b$ and $d$ in another. This implies that if the elements of $[n]$ are realized as points on the circle, and neighboring elements within each block are joined by line segments, then a non-crossing partition will appear as a system of non-overlapping polygons. The number of polygons equals to the number of blocks in the partition. Figure
Figure 3. An NC polygon diagram with superimposed bicolored plane tree. Polygon sides ↔ tree edges. (Graphics is courtesy Callan and Smiley.)

4.2 shows the bijection between such polygon systems and bipartite rooted planar trees. (This bijection was invented by Callan and Smiley in [5].) In this bijection the number of polygons corresponds to the number of vertices in one of the partitions of the tree. □

Hence, the number of the reduced types with contribution (19) is given by

\[
kl \left[ \begin{array}{c} k \\ t_1 \\ t_2 \end{array} \right].
\]

(The prefactor \( k_l \) corresponds to the choice of the tree edges that are glued together in the reduced type.) Therefore, the total contribution of higher entry moments to (18) converges to

\[
\frac{(m_4 - 1) \gamma k_l}{2^{k+l}} \sum_{t_1=1}^{k} \left[ \begin{array}{c} k \\ t_1 \end{array} \right] \left( \frac{a_2}{a_1} \right)^{k/2 - t_1 + 1/2} \sum_{t_2=1}^{l} \left[ \begin{array}{c} l \\ t_2 \end{array} \right] \left( \frac{b_2}{b_1} \right)^{l/2 - t_2 + 1/2}.
\]

Next, we note that by the binomial theorem

\[
\text{tr} \left( \frac{A_N A_N^*}{2 \sqrt{a_1 a_2 N}} - \frac{a_1 + a_2}{2 \sqrt{a_1 a_2}} I_N \right)^k = \sum_{s=1}^{k} (-1)^s \binom{k}{s} \text{tr} \left( \frac{A_N A_N^*}{2 \sqrt{a_1 a_2 N}} \right)^s \left( \frac{a_1 + a_2}{2 \sqrt{a_1 a_2}} \right)^{k-s} \left( \frac{a_1 + a_2}{2 \sqrt{a_1 a_2}} \right)^{k}.
\]

By using (20), we obtain the formula in the claim of the proposition. □

In fact, \( R_k (a_1, a_2) \) can be computed explicitly, and surprisingly, its value does not depend on \( a_1 \) or \( a_2 \).

Lemma 4.4.

\[
R_k (a_1, a_2) = \begin{cases} \frac{1}{2^k} \binom{k}{k/2}, & \text{if } k \text{ is odd}, \\ 0, & \text{if } k \text{ is even}. \end{cases}
\]
**Proof:** Recall that the Narayana polynomials are defined as

\[ N_n(x) = \sum_{k=1}^{n} \binom{n}{k} x^k. \]

Hence, if we take \( x = \frac{a_1}{a_2} \), then we have

\[
R_k(a_1, a_2) = \frac{1}{2^k} (1 + x)^k x^{-\frac{k+1}{2}}
\times \sum_{s=1}^{k} \left( \frac{-1}{1 + x} \right)^s \binom{k}{s} N_s(x).
\]

Next, we use the fact that \( N_s(x) \) are related to a particular case of the Jacobi polynomials. Namely,

\[ N_n(x) = \frac{x}{n} (x - 1)^{n-1} P^{(1,1)}_{n-1} \left( \frac{x + 1}{x - 1} \right). \]

(This fact was apparently first noted in Proposition 6 of [16].) By substituting this identity into the previous formula, we get

\[
R_k(a_1, a_2) = \frac{1}{2^k} \left( \frac{1 + x}{\sqrt{x}} \right)^k \sqrt{x} x^{1 - \frac{k+1}{2}}
\times \sum_{s=1}^{k} \left( \frac{-1}{1 + x} \right)^s \binom{k}{s} P^{(1,1)}_{s-1} \left( \frac{x + 1}{x - 1} \right).
\]

Next, we use the contour integral formula for the Jacobi polynomials:

\[
P^{(\alpha,\beta)}_n(t) = \frac{1}{2\pi i} \oint \frac{1}{(1 + t + \frac{1}{2} z)^{n+\alpha}} \frac{1}{1 + t - \frac{1}{2} z}^{n+\beta} z^{-n-1} dz,
\]

with the integration along a small circle around the zero. (See formula 4.4.1 in [22].) It follows that

\[
\sum_{s=1}^{k} (-t)^{-s} \binom{k}{s} P^{(1,1)}_{s-1}(t) = \frac{1}{2\pi i} \oint \sum_{s=1}^{k} (-tz)^{-s} \binom{k}{s} \left( 1 + tz + \frac{t^2 - 1}{4} z^2 \right)^s dz
= \frac{1}{2\pi i} \oint (-tz)^{-k} \left( 1 + \frac{t^2 - 1}{4} z^2 \right)^k dz,
\]

where we used the binomial theorem in the last step. This is zero for even \( k \). For odd \( k \) we calculate:

\[
\sum_{s=1}^{k} (-t)^{-s} \binom{k}{s} P^{(1,1)}_{s-1}(t) = -\frac{1}{t^k} \binom{k}{k-1} \left( \frac{t^2 - 1}{4} \right)^{\frac{k-1}{2}}.
\]

Next, we set \( t = (x + 1)/(x - 1) \) in (21). Since

\[
\left( \frac{1 + x}{2\sqrt{x}} \right)^k \frac{\sqrt{x}}{x - 1} = \left( \frac{t}{\sqrt{t^2 - 1}} \right)^k \frac{\sqrt{t^2 - 1}}{2},
\]

hence, for odd \( k \),

\[
R_k(a_1, a_2) = -2^{-k} \binom{k}{k-1}.
\]
Proof of Theorem 1.2. By using Proposition 4.2, Lemma 4.4 and a contour integral argument similar to the argument in the proof of Theorem 1.1, we find that the covariance of $N_o(T_k, W_{A_N})$ and $N_o(T_l, W_{B_N})$ is influenced by the fourth moment if and only if $k = l = 1$, in which case the change relative to the model with $\pm 1$ entries is given by

$$\frac{(m_4 - 1) \gamma}{4}.$$ 

5. Linear Statistics of Continuously Differentiable Functions

5.1. Preliminary remarks. Our goal here is to calculate the limit distribution for the centered linear statistics $N_o(f, A_N)$ and $N_o(g, B_N)$, when $f$ and $g$ are continuously differentiable functions. First, consider the case of arbitrary polynomial functions $f$ and $g$. Recall that the coefficients $\hat{f}_k$ and $\hat{g}_k$ are defined as follows:

$$\hat{f}_k := \begin{cases} \frac{2}{\pi} \int_{-1}^{1} f(x) T_k(x) \frac{dx}{\sqrt{1-x^2}}, & \text{for } k \geq 1, \\ \frac{1}{\pi} \int_{-1}^{1} f(x) \frac{dx}{\sqrt{1-x^2}}, & \text{for } k = 1, \end{cases} \quad (22)$$

and similarly for $\hat{g}_k$. By the orthogonality of Chebyshev polynomials, one can write:

$$f = \sum_{k=0}^{\infty} \hat{f}_k T_k(x) \quad \text{and} \quad g = \sum_{k=0}^{\infty} \hat{g}_k T_k(x),$$

and for polynomial $f$ and $g$, the summations in these series are over a finite number of terms.

A direct corollary of Theorem 1.1 is as follows.

Corollary 5.1. For the real overlapping Wigner matrices $A_N$ and $B_N$, and for polynomial functions $f$ and $g$, the random variables $N_o(f, A_N)$ and $N_o(g, B_N)$ converge in distribution to a two-variate Gaussian variable with the covariance

$$C(f, g) = \frac{1}{2} \left[ d_2 \hat{f}_1 \hat{g}_1 \gamma + (m_4 - 1) \hat{f}_2 \hat{g}_2 \gamma^2 + \sum_{k=3}^{\infty} k \hat{f}_k \hat{g}_k (\gamma)^k \right],$$

where $\gamma = \Delta/\sqrt{ab}$.

Now let us outline the plan of the proof of Theorem 1.3.

1. Take a sequence of polynomials $P_{f,m}$ and $P_{g,m}$ that approximate $f$ and $g$, respectively, in a suitable norm. Let $W_m$ be the two-variate Gaussian distribution which is the limit for the joint distributions of $N_o(P_{f,m}, A_N)$ and $N_o(P_{g,m}, B_N)$ as $N \to \infty$. Show that the sequence $W_m$ converges to a limit, a Gaussian distribution $\mathcal{W}$, as $m \to \infty$.

2. Prove that the joint distributions of pairs $N_o(f, A_N)$ and $N_o(g, B_N)$ form a tight family with respect to $N$. Let $\{Y_N\}$ denote this family and let $\mathcal{Y}$ be one of its limit points.

3. Show that a suitably defined distance between $\mathcal{Y}$ and $W_m$ converges to zero as $m \to \infty$. 
From (1) and (3), we can conclude that \( Y \) must coincide with \( W \). Since this is true for every limit point \( Y \), we will be able to infer that \( \mathcal{N}^\circ (f, A_N) \) and \( \mathcal{N}^\circ (g, B_N) \) converge to \( W \) as \( N \to \infty \).

Before proceeding to implementing this plan, let us derive some preliminary results.

First, we will need some additional facts about expansions in Chebyshev polynomials. Consider the change of variable \( x = \cos \theta \), where \( \theta \in [-\pi, \pi] \), and define \( F(\theta) = f(\cos \theta) \). If \( f \) is absolutely continuous on \([-1, 1]\), then \( F(\theta) \) is absolutely continuous on \([-\pi, \pi]\). The coefficients \( \hat{f}_n \) in the expansion of \( f \) in Chebyshev's polynomials correspond to the Fourier coefficients in the Fourier expansion of \( F(\theta) \):

\[
F(\theta) = \frac{1}{2} \sum_{n=0}^{\infty} \hat{f}_n \left( e^{i n \theta} + e^{-i n \theta} \right).
\]

First, we are going to show that if \( f \) is continuously differentiable, then

\[
\sum_{n=1}^{\infty} n |\hat{f}_n|^2 < \infty.
\]

This will show that the entries of the covariance matrix \( V \), defined in the statement of Theorem 1.3, are finite. In fact, this holds for a more general class of functions, namely, for the continuous embedding of the Sobolev class \( W^{1,p} \).

**Lemma 5.2.** If \( f' \in L^p([-1, 1], dx) \) with \( p > 1 \), then

\[
\sum_{n=1}^{\infty} n |\hat{f}_n|^2 \leq c_p \| f' \|_p^{1/2} < \infty.
\]

**Proof:** If \( f' \in L^p([-1, 1], dx) \) with \( p \geq 1 \), then

\[
\int |F'(\theta)|^p d\theta \ll \int_{-\pi}^\pi |f'(\cos \theta)\sin \theta|^p d\theta
\ll \int_{-1}^1 |f'(x)|^p (1 - x^2)^{p/2} \frac{dx}{\sqrt{1 - x^2}}
\ll \int_{-1}^1 |f'(x)|^p dx.
\]

Hence, \( F' \in L^p([-\pi, \pi], d\theta) \). Moreover, since the interval is finite, hence \( F' \in L^s \) if \( 1 \leq s \leq p \).

Recall that the Fourier coefficients of \( F' \) are \( \frac{i}{\pi} n \hat{f}_n \). Take an \( s \in (1, \min(2, p)) \) and define \( r := s / (2 - s) > 1 \). Then by the Hölder inequality,

\[
\sum_{n=1}^{\infty} n |\hat{f}_n|^2 \leq \left( \sum_{n=1}^{\infty} \frac{1}{n^r} \right)^{1/r} \left( \sum_{n=1}^{\infty} |n \hat{f}_n|^{2q} \right)^{1/q},
\]

where \( q = r / (r - 1) = \frac{1}{2} s / (s - 1) \). Since \( 2q > 1 \), hence the Hausdorff-Young inequality is applicable, and

\[
\left( \sum_{n=1}^{\infty} |n \hat{f}_n|^{2q} \right)^{1/(2q)} \leq \left( \frac{1}{2\pi} \int |F'(x)|^s x \right)^{1/s} \ll \| f' \|_s.
\]
It follows that
\[
\sum_{n=1}^{\infty} n \left| \hat{f}_n \right|^2 \leq c \left\| f' \right\|_s^{1/2} \leq \tilde{c} \left\| f' \right\|_p^{1/2} < \infty. \tag{23}
\]

Lemma 5.3. For sequences \( x := \{x_k\}_{k=1}^{\infty} \) and \( y = \{y_k\}_{k=1}^{\infty} \), define \( \langle x, y \rangle_K := \frac{d_2}{2} x_1 y_1 + (m_1 - 1) x_2 y_2 + \sum_{n=3}^{\infty} x_n y_n \) and \( \| x \|_K^2 := \langle x, x \rangle \). Then \( \| \cdot \|_K \) is a Hilbert norm induced by the scalar product \( \langle x, y \rangle_K \).

Proof is by verification that \( \langle c, d \rangle_K \) is a scalar product.

For functions \( f \in F = C^1([-1 - \delta, 1 + \delta]) \), we define \( \| f \|_K := \left\| \left\{ \hat{f}_i \right\}_{i=1}^{\infty} \right\|_K \), where \( \hat{f}_i \) is the coefficients of the expansion of \( f \) in Chebyshev polynomials. This is a seminorm on \( F \) (It is zero on the subspace spanned by constants.)

5.2. Proof of the theorem about linear statistics of continuously differentiable functions.

Proof of Theorem 1.3: First, let us approximate \( f' \) by polynomials \( \tilde{P}_{f,m}(x) \) of degree \( m \) in the uniform norm on the interval \( I_\delta = [-1 - \delta, 1 + \delta] \). Thus, assume that polynomials \( \tilde{P}_{f,m}(x) \) have the property that
\[
\max_{x \in I_\delta} \left| f'(x) - \tilde{P}_{f,m}(x) \right| \leq \varepsilon_m \to 0 \text{ as } m \to \infty. \tag{24}
\]
We define
\[
P_{f,m}(x) := f(-1 - \delta) + \int_{-1 - \delta}^{x} \tilde{P}_{m}(t) \, dt.
\]
Then, we have
\[
\left| f(x) - P_{f,m}(x) \right| \leq \int_{-1 - \delta}^{x} \left| f'(t) - \tilde{P}_{m}(t) \right| \, dt \leq c\varepsilon_m,
\]
where \( c \) is a constant. Hence
\[
\max_{x \in I_\delta} \left| f(x) - P_{f,m}(x) \right| \leq c\varepsilon_m \to 0 \text{ as } m \to \infty. \tag{25}
\]
From (24) and (25) it follows that \( \| f - P_{f,m} \|_{Lip} \to 0 \), where \( \| \cdot \|_{Lip} \) is the Lipschitz norm on the interval \( I_\delta \). (For differentiable functions, Lipschitz norm is defined by \( \| h \|_{Lip} := \sup_{x \in I_\delta} \left( |h(x)| + \sup_{x \in I_\delta} |h'(x)| \right) \). In addition, by (23) and (24) we have
\[
\| f - P_{f,m} \|_K < c\varepsilon_m \to 0 \text{ as } m \to \infty.
\]
In particular, by the triangle inequality, \( \| f \|_K - \| P_{f,m} \|_K \| \to 0 \) as \( m \to \infty \)(and also (since \( \| f \|_K + \| P_{f,m} \|_K \leq 3 \| f \|_K \) for sufficiently large \( m \)), \( \| f \|_K^2 - \| P_{f,m} \|_K^2 \| \to 0 \) as \( m \to \infty \).

We define \( P_{g,m}(x) \) similarly.
By Corollary 5.1 as $N \to \infty$, $\text{Tr} P_{f,m} \left( \tilde{A}_N \right)$ and $\text{Tr} P_{g,m} \left( \tilde{B}_N \right)$ converge in distribution to a bivariate Gaussian variable $\mathcal{W}_m$ with the covariance matrix $V_m$, where

$$(V_m)_{11} = \frac{1}{2} \|P_{f,m}\|_K^2$$

$$(V_m)_{12} = \frac{1}{2} \left( \frac{d_2}{2} (P_{f,m})_1 (P_{g,m})_1 \gamma + (m_4 - 1) (P_{f,m})_2 (P_{g,m})_2 \gamma^2 + \sum_{k=3}^{m} k (P_{f,m})_k (P_{g,m})_k \gamma^k \right)$$

$$(V_m)_{22} = \frac{1}{2} \|P_{g,m}\|_K^2.$$  

The diagonal entries of this matrix converge to the diagonal entries of the matrix $V$, defined in the statement of Theorem 1.3. The off-diagonal term $(V_m)_{12}$ can be written as

$$\left\langle P_{f,m} - f, P_{g,m}^{(\gamma)} \right\rangle_K + \left\langle f, P_{g,m}^{(\gamma)} - g^{(\gamma)} \right\rangle_K + \left\langle f, g^{(\gamma)} \right\rangle_K,$$

where

$$P_{g,m}^{(\gamma)} (x) = \sum_{k=1}^{m} (P_{g,m})_k \gamma^k T_k (x),$$

and

$$g^{(\gamma)} (x) = \sum_{k=1}^{m} g_k \gamma^k T_k (x).$$

The first two terms are small by the application of the Schwartz inequality for the scalar product $\left\langle \cdot, \cdot \right\rangle_K$, and the third term coincides with $V_{12}$. Hence we can conclude that $\|V_m - V\|$ converges to zero as $m \to \infty$. This implies that the Gaussian distributions $\mathcal{W}_m$ converge to the Gaussian distribution $W$ with the covariance matrix $V$. This finishes the first step of our proof.

In order to prove tightness for the distribution family of $\{ \mathcal{N}^\alpha (f, A_N), \mathcal{N}^\alpha (g, B_N) \}$ (with respect to $N$), we are going to prove that the norms of their covariance matrices are bounded. In fact, it is enough to prove that variances of each of $\mathcal{N}^\alpha (f, A_N)$ and $\mathcal{N}^\alpha (g, B_N)$ are bounded, since then the covariance will be bounded automatically.

Here, we rely heavily on the Poincare inequality property ("PI") of the matrix entries. The essential feature of the PI property is that it is well behaved with respect to taking the product of measures. By definition, the measure $\eta$ on $\mathbb{R}$ has the PI property, if for some $c_\eta > 0$ and all differentiable $f$,

$$\text{Var}_{\eta} (f) \leq c_\eta \int |f'(x)| \eta (dx).$$

Then, if $\eta_K = \otimes_{i=1}^{K} \eta_i$ with $\eta_i = \eta$ and if $h : \mathbb{R}^K \to \mathbb{R}$ is a differentiable function, then

$$\text{Var}_{\eta_K} (h) \leq c_\eta \int \|\nabla h (x)\| \eta (dx).$$

By approximation, this can be further extended to the case when $h$ is Lipschitz. In particular, if $h$ is a Lipschitz function on $\mathbb{R}$, then we have

$$\text{Var}_{\eta_K} (h) \leq c_\eta \|h\|_{\text{Lip}},$$

(26)
Fortunately, in this situation we can write
\[ f(x) = f_1(x) + f_2(x), \]
where the Lipschitz norm is taken over the interval \( I = [-1 - \delta, 1 + \delta] \). Outside of \( I \), we only know that it has a polynomial growth. Fortunately, in this situation we can write \( f \) as a sum of two functions: \( f = f_1 + f_2 \), with \( f_1 \) Lipschitz and bounded everywhere on \( \mathbb{R} \), \( \|f_1\|_{Lip} < \infty \), and \( f_2 \) vanishing on \( I_{\delta/2} : [-1 - 2, 1 + 2] \) and having a polynomial growth. Then (27) can be applied to bound
\[ \text{Var} \left( \text{Tr} f \left( A_N / \sqrt{d(N)} \right) \right) \leq C c_\eta \|f\|_{Lip}, \]

where \( C \) is an absolute constant and \( c_\eta \) depends only on the distribution of matrix entries.

A complication arises since under our assumptions, \( f \) is assumed Lipschitz only on the interval \( I = [-1 - \delta, 1 + \delta] \). Outside of \( I \), we only know that it has a polynomial growth. Fortunately, in this situation we can write \( f \) as a sum of two functions: \( f = f_1 + f_2 \), with \( f_1 \) Lipschitz and bounded everywhere on \( \mathbb{R} \), \( \|f_1\|_{Lip} < \infty \), and \( f_2 \) vanishing on \( I_{\delta/2} : [-1 - 2, 1 + 2] \) and having a polynomial growth. Then (27) can be applied to bound
\[ \text{Var} \left( \text{Tr} f_1 \left( A_N / \sqrt{d(N)} \right) \right) \].

In addition, from the results about the spectra of Wigner matrices, it is known that the probability for \( A_N / \sqrt{d(N)} \) to have an eigenvalue outside of \( I_{\delta/2} \) becomes exponentially small in \( N \), as \( N \) grows. This implies that
\[ \text{E} \left[ \text{Tr} f_2 \left( A_N / \sqrt{d(N)} \right) \right]^2 \to 0, \text{ as } N \to \infty. \]

Since for two random variables, \( \xi_1 \) and \( \xi_2 \), it is true that
\[ \sqrt{\text{Var} (\xi_1 + \xi_2)} \leq \sqrt{\text{Var} (\xi_1)} + \sqrt{\text{Var} (\xi_2)}, \]

we can conclude that
\[ \limsup_{N \to \infty} \text{Var} \left( \text{Tr} f \left( A_N / \sqrt{d(N)} \right) \right) \leq c \|f\|_{Lip}, \]

where the Lipschitz norm is taken over the interval \( I \).

A similar argument holds for the random variable \( \text{Tr} g \left( B_N / \sqrt{d(N)} \right) \), and therefore the norm of the covariance matrices of these two random variables is bounded. This shows that the distributions of the pairs \( \text{N}^o (f, A_N), \text{N}^o (g, B_N) \) form a tight family and concludes the second step of our proof.

Next, let \( \mathcal{Y} \) be a limit point for the distributions \( \mathcal{Y}_N \) of \( \{\text{N}^o (f, A_N), \text{N}^o (g, B_N)\} \), so that \( \mathcal{Y}_{N_k} \to \mathcal{Y} \) in distribution for a sequence of \( N_k \). We are going to estimate the difference between the characteristic functions of the distributions \( \mathcal{Y} \) and \( \mathcal{W}_m \).

For convenience, we will assume that all relevant random variables are realized on a single probability space so that convergence in distribution reflects convergence almost surely. In this realization, let \( Y_{N_k} \) and \( W_{m,N_k} \) denote random variables that have the same joint distribution as \( \{\text{N}^o (f, A_{N_k}), \text{N}^o (g, B_{N_k})\} \) and \( \{\text{N}^o (P_{f,m}, A_{N_k}), \text{N}^o (P_{g,m}, B_{N_k})\} \). The variables \( Y_{N_k} \) and \( W_{m,N_k} \) converge almost surely to random variables \( Y \) and \( W_m \), that have the
distributions $\mathcal{Y}$ and $W_m$, respectively. Then,
\[
|E e^{itY} - E e^{itW_m}| = \left|E \exp \left( it \lim_{N_k \to \infty} Y_{N_k} \right) - E \exp \left( it \lim_{N_k \to \infty} W_{m,N_k} \right) \right|
\leq \limsup_{N \to \infty} |E \exp (itY_{N_k}) - E \exp (itW_{m,N_k})|
= \limsup_{N \to \infty} |E \exp (it(Y_{N_k} - W_{m,N_k})) - 1|,
\]
where the inequality follows from Fatou’s lemma.

Next, note that by using (27), we have
\[
\limsup_{N_k \to \infty} \text{Var} \left[ \text{Tr} \left( A_N / \sqrt{4a(N)} \right) - \text{Tr} P_{f,m} \left( A_N / \sqrt{4a(N)} \right) \right] \leq c \| f - P_{f,m} \|_{\text{Lip}},
\]
which implies that for the first component of the vector $Y_{N_k} - W_{m,N_k}$ we have the following bound:
\[
\limsup_{N_k \to \infty} E \left[ N^o (f, A_{N_k}) - N^o (P_{f,m}, A_{N_k}) \right]^2 \leq c \| f - P_{f,m} \|_{\text{Lip}}.
\]
A similar expression can be written for the second component of $Y_{N_k} - W_{m,N_k}$.

Now, suppose $\xi_1$ and $\xi_2$ are two real random variables with finite variances. Assume $E \xi_1 = E \xi_2$. Then,
\[
|E \exp (i s (t_1 \xi_1 + t_2 \xi_2)) - 1| \leq 2s^2 \text{Var} (t_1 \xi_1 + t_2 \xi_2)
\leq 2s^2 \left( t_1 \sqrt{\text{Var} \xi_1} + t_2 \sqrt{\text{Var} \xi_2} \right)^2,
\]
where the first inequality is a consequence of inequality II.3.14 on p.278 in Shiryaev [19]. By applying this inequality to the two components of the vector $Y_{N_k} - W_{m,N_k}$, and by using (28) and its analogue for the function $g$, we find:
\[
\limsup_{N_k \to \infty} |E \exp (it(Y_{N_k} - W_{m,N_k})) - 1| \leq c \| t \|_2 \max \left\{ \| f - P_{f,m} \|_{\text{Lip}}, \| g - P_{g,m} \|_{\text{Lip}} \right\},
\]
which implies that
\[
|E e^{itY} - E e^{itW_m}| \leq c \| t \|_2 \max \left\{ \| f - P_{f,m} \|_{\text{Lip}}, \| g - P_{g,m} \|_{\text{Lip}} \right\}.
\]
By our choice, as $m \to \infty$, $P_{f,m}$ and $P_{g,m}$ converge to $f$ and $g$, respectively, in the Lipschitz norm. Hence, the random variables $W_m$ converge in distribution to $Y$. This concludes the third and final step of our proof. As explained before, these three steps imply that $\{N^o (f, A_N), N^o (g, B_N)\}$ converge in distribution to the Gaussian random variable $W$.

6. Conclusion

We computed the joint distribution of the eigenvalue statistics for two random matrix models. For both Wigner and sample covariance matrices, we found that the covariance of the Chebyshev polynomials of the first type has the diagonal structure. The variances are different from those found in Borodin’s paper for Gaussian matrices. This is in agreement with the
results by Anderson and Zeitouni in [2] and by Tao and Vu in [23], which suggest that the linear statistic covariances should depend on the first four moments of matrix entries. However, somewhat surprisingly, we find that for high-degree Chebyshev polynomials the covariance depend only on the variance of matrix entries.

REFERENCES

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. An Introduction to Random Matrices, volume 118 of Cambridge studies in advanced mathematics. Cambridge University Press, 2009.
[2] Greg W. Anderson and Ofer Zeitouni. A CLT for a band matrix model. Probability Theory and Related Fields, 134:283–338, 2006.
[3] Yuliy Baryshnikov. Gues and queues. Probability Theory and Related Fields, 119:256–274, 2001.
[4] Alexei Borodin. CLT for spectra of submatrices of wigener random matrices. arxiv:1010.0898, 2009.
[5] David Callan and Len Smiley. Noncrossing partitions under rotation and reflection. arxiv:math/0510447 2005.
[6] Ovidiu Costin and Joel L. Lebowitz. Gaussian fluctuations in random matrices. Physical Review Letters, 76:69–72, 1995.
[7] Persi Diaconis and Steven N. Evans. Linear functionals of eigenvalues of random matrices. Transactions of American Mathematical Society, 353(7):2615–2633, 2001.
[8] Persi Diaconis and Mehrdad Shahshahani. On eigenvalues of random matrices. Journal of Applied Probability, 31:49–62, 1994.
[9] Ohad N. Feldheim and Sasha Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. Geometric and Functional Analysis, 20:88–123, 2010. arxiv:0812.1961
[10] Peter J. Forrester and Taro Nagao. Determinantal correlations for classical projection processes. Journal of Statistical Mechanics: Theory and Experiment, 2011(08):P08011, 2011. arxiv:0801.0100
[11] Peter J. Forrester and Eric Nordenstam. The anti-symmetric GUE minor process. Moscow Mathematical Journal, 9:749–774, 2009.
[12] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. Electronic Communications in Probability, 5:119–136, 2000.
[13] Kurt Johansson. On fluctuation of eigenvalues of random Hermitian matrices. Duke Mathematical Journal, 91:151–204, 1998.
[14] Kurt Johansson and Eric Nordenstam. Eigenvalues of gue minors. Electronic Journal of Probability, 11:1342–1371, 2006.
[15] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. Journal of Multivariate Analysis, 12:1–38, 1982.
[16] Vladimir P. Kostov, Andrei Martinez-Finkelshtein, and Boris Z. Shapiro. Narayana numbers and Schur-Szego composition. Journal of Approximation Theory, 161:464–476, 2009. available at http://people.su.se/shapiro/Articles/EVCSS.pdf
[17] Anthony P. Metcalfe. Universality properties of Gelfand-Tsetlin patterns. Probability Theory and Related Fields, 155:303–346, 2013. arxiv:1105.1272
[18] Matthew Reed. PhD thesis, University of California, Davis, 2014.
[19] A. N. Shiryaev. Probability. Springer, second edition, 1996.
[20] Alexander Soshnikov. The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. Annals of Probability, 28:1353–1370, 2000.
[21] Alexander B. Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. Journal of Statistical Physics, 100:491–522, 2000.
[22] Gabor Szegö. Orthogonal Polynomials. American Mathematical Society, third edition, 1967.
[23] T. Tao and V. Vu. Random matrices: universality of the local eigenvalue statistics. *Acta Mathematica*, 206:127–204, 2011.

[24] A. Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 2003.