GLOBAL FLUCTUATIONS FOR LINEAR STATISTICS OF $\beta$-JACOBI ENSEMBLES

IOANA DUMITRIU AND ELLIOT PAQUETTE

Abstract. We study the global fluctuations for linear statistics of the form $\sum_{i=1}^{n} f(\lambda_i)$ as $n \to \infty$, for $C^1$ functions $f$, and $\lambda_1, \ldots, \lambda_n$ being the eigenvalues of a (general) $\beta$-Jacobi ensemble $[18, 29]$. The fluctuation from the mean ($\sum_{i=1}^{n} f(\lambda_i) - E \sum_{i=1}^{n} f(\lambda_i)$) is given asymptotically by a Gaussian process.

We compute the covariance matrix for the process and show that it is diagonalized by a shifted Chebyshev polynomial basis; in addition, we analyze the deviation from the predicted mean for polynomial test functions, and we obtain a law of large numbers.

1. Introduction

Global fluctuations for linear statistics, also known as central limit theorems, have been of interest to the random matrix community for almost as long as the limiting properties of empirical spectral distributions (also known sometimes as laws of large numbers). A variety of models and eigenvalue distributions have been studied from this point of view, starting with the classical Gaussian and Wishart matrices $[34, 36]$, generalizations thereof (Wigner and Wishart-like matrices $[3, 11, 21, 32, 37, 42, 43]$), tridiagonal models $[17, 39]$, different eigenvalue potentials $[25]$, $\beta$-ensembles $[19, 28]$, classical compact groups $[15, 44]$, banded matrices $[2, 21]$, permutations $[7]$ and so on. The methods of approach range from the classical method of moments $[2, 17]$, to free probability $[12, 21, 30, 35]$ and stochastic calculus $[11]$.

To put it more concretely, we are interested in the following problem. A linear statistic of an $n \times n$ matrix $A$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ is a functional of the form

$$F(A) := \sum_{i=1}^{n} f(\lambda_i),$$

where $f$ is a function (we sometimes refer to them as test functions) belonging to a certain class (which, depending on the ensemble to whom $A$ belongs, may be as restrictive as the class of polynomials, or as wide as $L^2$). The first issue at hand is to calculate the limit of $\frac{1}{n} F(A)$ as $n \to \infty$ (in case this exists), in other words, to find the limiting empirical spectral distribution for the eigenvalues of $A$ (also known as the law of large numbers). The second issue is to examine the fluctuation from the mean, e.g., study

$$X_{f,A} := F(A) - E F(A),$$

perhaps under a suitable scaling, and prove that $X_{f,A}$ converges in distribution, here to a centered Gaussian variable whose variance depends on $f$. The term “global” in “global fluctuations” refers to the fact that all eigenvalues contribute similarly to $F(A)$.

The Jacobi ensemble (also known as Double Wishart) is one of many on which such studies have been performed. They were introduced in connection with the MANOVA procedure of statistics for measuring the likelihood of a multivariate linear model $[4, 36]$, and found to be of interest in quantum conductance and log-gas theory $[6, 20]$. One can describe them through their eigenvalue distributions

$$d\mu_f(\lambda_1, \ldots, \lambda_n) := \frac{1}{Z} \prod_{i=1}^{n} \lambda_i^{[n_1-n+1]-1} (1-\lambda_i)^{[n_2-n+1]-1} \prod_{i<j} |\lambda_i - \lambda_j|^\beta,$$

where $Z = Z(n, n_1, n_2, \beta)$ is a normalization constant. In full generality, $\beta > 0$, while $n_1$ and $n_2$ need not be positive integers; in fact, the only constraints (which relate to the integrability of the measure) are that $n_1, n_2 \geq n - 1$.

Date: May 5, 2014.

Both authors acknowledge the support of the NSF through CAREER award DMS-0847661. The second author was additionally supported by NSF grant DMS-0806024.
In the case that \( \beta \in \{1, 2, 4\} \) and \( n_1, n_2 \in \mathbb{N} \), they admit full matrix models (as \( J = W_1^{1/2}(W_1 + W_2)^{-1}W_1^{1/2} \), where \( W_1, W_2 \) are Wishart matrices, hence the “double Wishart” name. For an extensive study of the \( \beta = 1 \) case, as well as a clear exposition of how these models arose, we refer to [36]; the other cases (\( \beta = 2, 4 \)) can be dealt with similarly.

Recently, it was shown that in these “classical” cases a different kind of model can be constructed, starting from random projections, rather than random Wishart matrices; or, equivalently, that “chopping off” an appropriate corner of a unitary Haar matrix will yield a matrix whose singular values, squared, are distributed from random projections, rather than random Wishart matrices; or, equivalently, that “chopping off” an appropriate corner of a unitary Haar matrix will yield a matrix whose singular values, squared, are distributed.

The greatest generality is achieved by the tridiagonal model [18,29], which covers any \( \beta > 0 \), and removes the condition that \( n_1, n_2 \in \mathbb{N} \). We give the model below (hereafter referred to as the Edelman-Sutton model, as it appears most clearly in their work [18]).

Given the matrix \( B_\beta \) defined as

\[
B_\beta = \begin{pmatrix}
c_n s_{n-1}' & c_{n-1} s_{n-2}' & \vdots & \vdots & -c_n s_{n-1}' \\
-s_{n-1} c_{n-1}' & -s_{n-2} c_{n-2}' & \vdots & \vdots & -s_{n-1} c_{n-1}' \\
& \ddots & \ddots & \ddots & \vdots \\
& & \ddots & \ddots & -s_2 c_1' \\
& & & -s_1 c_1' & c_1
\end{pmatrix},
\]

with the variables \( c_i, s_i, i = 1, \ldots, n \), and \( c_j', s_j', j = 1, \ldots, (n-1) \) obeying the distribution laws and relationships

\[
\{c_1, c_2, \ldots, c_n, c_1', c_2', \ldots, c_{n-1}'\} \text{ mutually independent,}
\]

\[
c_i \sim \text{Beta}(\frac{1}{2}(n_1 - n + i), \frac{1}{2}(n_2 - n + i)) \text{ and } c_j' \sim \text{Beta}(\frac{1}{2}j, \frac{1}{2}(n_1 + n_2 - 2n + 1 + j))
\]

\[
s_i = \sqrt{1 - c_i^2} \text{ and } s_j' = \sqrt{1 - c_j'^2}.
\]

the eigenvalues of \( A = B_\beta B_\beta^T \) are distributed according to \( \Pi \) (see [18]).

We are interested in the behavior of \( X_{f,A} \) as \( n \to \infty \) with \( (n_1 + n_2 - 2n) \) growing linearly in \( n \) and with \( \beta \) fixed. This is the only scaling regime in which the limiting spectral distribution is truly Jacobi.

If either \( n_1 \gg n \) or \( n_2 \gg n \), in the case when \( \beta = 1, 2, 4 \), the Wishart matrices in the full models have \( W_1 \approx \beta n_1 I_n \), respectively, \( W_2 \approx \beta n_2 I_n \). For example if \( n_2 \gg n \) and \( n_2 \gg n_1 \), this heuristics predicts that the Double Wishart model behaves like

\[
W_1(W_1 + W_2)^{-1} \approx W_1(W_1 + \beta n_2 I_2)^{-1} \approx W_1/(\beta n_2),
\]

so that appropriately rescaling, Wishart behavior should appear. These heuristics are studied rigorously in Jiang [23]. (The symmetric regime, \( n_1 \gg n \) and \( n_1 \gg n_2 \), predicts Wishart behavior with a huge shift in eigenvalues.)

Conversely, in the sublinear growth cases, i.e. where \( (n_1 + n_2 - 2n) \ll n \), the Jacobi ensemble takes on behavior that looks much more like the classical compact groups. This connection is explicit for \( \beta = 1, 4 \) and fixed values of \( n_1 - n \) and \( n_2 - n \) (see Proposition 3.1 of [14]). These heuristics predict the correct limiting spectral distributions as well. In the superlinear case, the limiting spectral distribution is a point mass (easily seen also from [3] which shows that the matrix \( B_\beta B_\beta^T \) is very close to a multiple of the identity), while in the sublinear case, the limiting spectral distribution is the arcsine law. These statements about the limiting spectral distributions are straightforward exercises following the approach of Trotter [49]. We sketch this approach in the proof of the following proposition.

**Proposition 1.1.** Let \( f \) be a continuous test function on \([0, 1]\).

1. If \( n_1 + n_2 - 2n = o(n) \), then
   \[
   \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i) \to_p \frac{1}{\pi} \int_0^1 \frac{f(x)}{\sqrt{x(1-x)}} \, dx.
   \]

2. If \( n_1/n \to p \) and \( n_2/n \to q \), then
   \[
   \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i) \to_p \int_0^1 f(x) \, d\mu(x),
   \]
where $\mu$ has density
\[
d\mu(x) := \frac{p+q}{2\pi x(1-x)} \sqrt{(x-\lambda_-)(x-\lambda_+)} 1_{[\lambda_-,\lambda_+]} dx,
\]
and
\[
\lambda_{\pm} := \left[ \sqrt{\frac{p}{p+q}(1-\frac{1}{p+q})} \pm \sqrt{\frac{1}{p+q}(1-\frac{p}{p+q})} \right]^2.
\]

(3) If $n_1 + n_2 - 2n = \omega(n)$ and if $(n_1 - n)/(n_1 + n_2 - 2n) \to \lambda$, then
\[
\frac{1}{n} \sum_{i=1}^{n} f(\lambda_i) \to f(\lambda).
\]

Proof. Regardless of the scales of $n_1 - n$ and $n_2 - n$, the limiting eigenvalue distribution can be understood by computing $A_\infty = B_\infty B^T_\infty$. (Note that on taking the $\beta$ parameter to infinity, the $\text{Beta}(\beta x, \beta y)$ variable in the matrix model converges in probability to $\frac{x}{x+y}$. Replacing the Beta variables by these limits in $B_\beta$ gives the matrix $B_\infty$.)

By applying Stirling’s approximation, it can be shown that there is a constant $C$ depending only on $\beta$ so that
\[
E \left| c_i - \sqrt{\frac{n_1 - n + i}{n_1 + n_2 - 2n + 2i}} \right|^2 \leq \frac{C}{i}.
\]

A similar bound holds for $c_1^i$ and for $c_is_i$. Applying all these bounds, it follows that
\[
E \|B_\beta B^T_\beta - B_\infty B^T_\infty\|_F^2 = O(\log n).
\]

From the fundamental realization of Trotter \[9\], any $o(n)$ bound on the expected-square Frobenius norm suffices to show that the ESDs of two matrix models are converging together as $n \to \infty$.

It is now elementary to check that the limiting spectral distribution for $B_\infty B^T_\infty$ is that which is stated in the theorem in the sublinear and superlinear cases. In the linear case, we compute the limiting distribution by way of the Jacobi differential recurrence formula, which we do in proving Theorem 5.1 (see \[15\]).

In our study of the linear scaling regime, we apply a wide array of methods, starting with the method of moments (which often boils down to path-counting), special functions (orthogonal polynomial) theory and generating functions, as well as one important result from the work of Anderson and Zeitouni \[2\] (more details in Section 4).

As mentioned in the introductory paragraph, the study of global fluctuations of linear statistics for random matrices spans a wide literature, and covers a broad spectrum of models. We will only mention here a few works that are either closely related in scope, in model, or those that have served as inspiration for our study.

The method of moments, introduced by Wigner himself \[51,52\] and used for proving central limit theorems for polynomials of Wishart matrices by Jonsson \[20\], has been employed with great success by Sinai and Soshnikov \[42\], Soshnikov \[43\], Pégé and Soshnikov \[38\], etc., to obtain both central limit theorems for traces of large powers of random matrices and universality results for the fluctuations of the extremal eigenvalues in the case of Wigner and Wishart matrices. The method of moments has also been used by Dumitriu and Edelman \[17\] to calculate the fluctuations in the case of $\beta$-Hermite and $\beta$-Laguerre ensembles (generalizations of the Gaussian and central Wishart ensembles for $\beta = 1, 2, 4$), in the case of polynomial functions. It is also one essential ingredient in the work of Anderson-Zeitouni \[2\] on band matrices.

It is worth mentioning that the method of moments is essentially equivalent in spirit (though not necessarily in form) to the Stieltjes transform methods used by Bai and Silverstein (e.g., \[3\]) to calculate central limit theorems for generalized Wishart matrices; for a good reference on the methodology involved, we recommend \[4\].

Another method for computing fluctuations of linear statistics involves a stochastic calculus approach introduced by Cabanal-Duvillard \[11\] to prove a central limit theorem for Wishart matrices in the case $\beta = 2$; stochastic calculus was also used by Guionnet \[21\] in computing fluctuations for a class of band matrices and sample covariance matrices, and by Guionnet and Zeitouni \[22\] to calculate large deviations for a wide class of random matrices.
Other approaches to calculating fluctuations for linear functionals for $\beta$-ensembles include the Capitaine and Casalis work [12], which, through free probability, obtains results for both Wishart and Jacobi (Double Wishart) matrices in the case $\beta = 2$. The later work of Kusakil, Mingo, and Speicher [30] builds on [12] and on results obtained by Mingo and Nica [35] to obtain fluctuations (second-order asymptotics) for random matrices (also in the case $\beta = 2$). Finally, Chatterjee [13] has introduced the Stein method to computing central limit theorems for a wide class of random matrices, for analytic potentials.

Specifically in the case of $\beta$-Jacobi ensembles, for an “extremal” class of $\beta$-Jacobi ensembles (when $n_1 = o(\sqrt{n_2})$ and $n = o(\sqrt{n_2})$), as mentioned before, Jiang [23] has established a series of important results, among which are the calculations of fluctuations, through approximation methods.

For all $\beta$-Jacobi ensembles of fixed parameters, Killip [28] proved that the fluctuations of macroscopic statistics obey a CLT; this result is similar to the one we obtain, but in the case that $f = \chi_I$ where $I$ is a (fixed, independent of $n$) finite union of intervals in $[0, 1]$ and under a different normalization. It is unclear how Killip’s result changes if the parameters of the ensemble scale with $n$, which is the regime studied here. In addition, while our method does not allow us to obtain any results for discontinuous functions, it seems that going in the opposite direction – using Killip’s results to obtain fluctuation theorems for smooth functions – would need microscopic statistics, i.e. where the lengths of the intervals shrink with $n$. As Killip notes, the microscopic regime is much more difficult and is not covered in [28].

Last but by no means least, we would like to mention that the most comprehensive results for linear functionals for $\beta$-ensembles in the case of Hermitian matrices, for a large class of (polynomial) potentials, and for a large class of functions $f$ (in its full generality, Johansson’s work is applicable to $H^{17/2}$ functions, where $H^n$ stands for the corresponding Sobolev space). The methods are analytical and make heavy use of potential theory. In addition to the fluctuations, Johansson was also able to obtain the deviation from the mean (second-order asymptotics), for the same class of functions.

Johansson’s results subsume the work [17] in the case of $\beta$-Hermite matrices (general $\beta$, fixed potential $V(x) = x^2$), and have served as a “moral” (albeit not technical) inspiration to us in our quest.

1.1. Our results. Our purpose in this paper is to calculate the global fluctuations for $\beta$-Jacobi ensembles, for as large a class of functions $f$ as possible. By using concentration properties of the Jacobi ensemble and making use of a theorem by Anderson and Zeitouni [2], cited below, we were able to obtain the fluctuations for all $\beta$ in the case of $C^1$ test functions on $[0, 1]$. We only obtain the deviation from the mean for polynomial test functions, and conjecture the deviation should extend to a larger class of functions.

Our asymptotic analysis will occur in the proportional scaling regime, and so we will make the following assumptions on the growth of $n_1$ and $n_2$.

Assumption 1.2. Let $n_1 = pn$ and $n_2 = qn$ for some fixed $p, q \geq 1$ having $p + q > 2$.

Chebyshev polynomials are an essential ingredient to our proof, both for their analytic properties and their combinatorial ones. We define the shifted Chebyshev polynomials of the first kind, $\Gamma$, by

$$
\Gamma_n(x) = 2T_n\left(\frac{2x - \lambda_+ - \lambda_-}{\lambda_+ - \lambda_-}\right)
$$

where $T_n$ are the standard Chebyshev polynomials of the first kind, satisfying $T_n(\cos \theta) = \cos n\theta$. By making a change of variables, it immediately follows that $\{\Gamma_n(x)\}^\infty_{n=0}$ are a complete orthonormal system for $L^2(\Omega)$, the weighted $L^2$ space induced by the inner product

$$
(f, g) = \frac{1}{2\pi} \int_{\Lambda_+} f(x)g(x) \frac{1}{\sqrt{(\lambda_+ - x)(x - \lambda_-)}} dx.
$$

Using this inner product, we define the Chebyshev coefficients

$$
\hat{f}(n) = (f, \Gamma_n) = \frac{1}{2\pi} \int_{\Lambda_+} f(x)\Gamma_n(x) \frac{1}{\sqrt{(\lambda_+ - x)(x - \lambda_-)}} dx.
$$

Our main result is given below.
Theorem 1.3. Let $A$ be an $n \times n$ $\beta$-Jacobi matrix, with $n_1, n_2$ satisfying Assumption 1.2. For any fixed $k \in \mathbb{N}$, the $k$-tuple $(X_{1,A}, \ldots, X_{k,A})$ converges in distribution to the $k$-tuple of independent centered normal variables $(Y_1, \ldots, Y_k)$, where $Y_i$ has variance $\frac{2}{\pi}$. Further, for any $f$ continuously differentiable on $[0,1]$, the variable $X_{f,A}$ converges in distribution to a centered normal variable $Y_f$, with variance given by

$$
\sigma_f^2 := \frac{2}{\pi} \sum_{n=1}^{\infty} n |\hat{f}(n)|^2,
$$

where $\hat{f}(n)$ is the $n^{th}$ Chebyshev coefficient, defined as in (7).

Remark 1.4. In analogy with Fourier series on the unit circle, it is alluring to consider the condition above for $f$ as requiring one half a derivative, in the $L^2$ sense; we would expect for

$$
\tau_f^2 = \sum_{n=1}^{\infty} n^2 |\hat{f}(n)|^2
$$

to behave like the square-$L^2$ norm of $f'$, and this can be easily established. Precisely,

$$
\tau_f^2 = \frac{1}{2\pi} \int_{\lambda_-}^{\lambda_+} |f'(x)|^2 \sqrt{(\lambda_+ - x)(x - \lambda_-)} \, dx,
$$

where the proof follows from the identity $T_n'(x) = n U_{n-1}(x)$, with $U$ the Chebyshev polynomial of the second kind, and the orthonormality of $\{U_n\}$ with the weight $\sqrt{1-x^2}$. Since $\tau_f^2 \geq \sigma_f^2$, given the $C^1$ condition for $f$ on $[\lambda_-, \lambda_+]$, the variance in the Theorem 1.3 is finite.

Remark 1.5. Note that the case when $p = q = 1$ is not covered. This is the case when neither one of the exponents of the ensemble grows to $\infty$; the method of proof collapses since one of the main ingredients, the ability to get uniform tail bounds for entries of the matrix is no longer true at the “bottom right” corner of the matrix, and as such the errors can no longer be accurately estimated by the same means. We present the results of some numerical simulations for this case in Section 6. We also note that the theorem is proven by Johansson in the $\beta = 2$ case by methods of orthogonal polynomial theory [24].

Our second result concerns the deviation from the mean, and is restricted to polynomial functions.

Theorem 1.6. For any polynomial $\phi$,

$$
\mathbb{E} \text{tr}(\phi(A)) = n \int_{\lambda_-}^{\lambda_+} \phi(x) d\mu(x) + \left(\frac{2}{\pi} - 1\right) \int_{\lambda_-}^{\lambda_+} \phi(x) d\nu(x) + o\left(\frac{1}{n}\right),
$$

where $\mu$ is as defined in Theorem 1.1 and $\nu$ is the signed measure with density

$$
d\nu := \frac{1}{2\pi} \delta_{\lambda_-} + \frac{1}{\pi} \delta_{\lambda_+} - \frac{1}{2\pi \sqrt{-(x-\lambda_+)(x-\lambda_-)}} 1_{(\lambda_-, \lambda_+)} \, dx.
$$

To structure of the paper follows the method of proof, which takes the following steps:
Step 1. Prove a “central limit theorem” for polynomials;
Step 2. Find the class of polynomials which diagonalizes the covariance matrix for the resulting Gaussian process;
Step 3. Use concentration techniques to show that $C^1[0,1]$ linear statistics can be approximated by polynomial test functions in such a way that the variance of the difference of the two is small for all $n$.

Step 4. Prove that the approximation works asymptotically.

The rest of the paper is structured as follows: after a reparameterization of the model (Section 1.2), Section 2 covers Step 1 in the above “recipe”: show that the fluctuations are Gaussian when the test functions are the monomials. The proof extends the mechanism that was employed in [17] for the $\beta$-Hermite and $\beta$-Laguerre ensembles. In Section 3 we show that the limiting covariance is diagonalized in shifted Chebyshev basis; the method employed is original and has to do with the generating function of the covariance matrix. Section 4 contains the proof that the matrix model satisfies the necessary conditions to apply the Anderson-Zeitouni
Theorem. Section [3] contains the proof of Theorem 1.6 (calculating the deviation from the mean for analytic functions). Section [4] contains experimental results for the case that \( p = q = 1 \). Finally, we have included three Appendices. Appendix [A] which is the longest of the three, contains the symmetric function theory results necessary for the calculation of the deviation (Section 5); more explicitly, it contains the proof that the series expansion of the functional \( F(A) \) for monomial \( f \) has a “palindromic” quality (the mechanism here is similar to the one employed in [17]). Appendix [B] shows the existence of a Poincaré inequality for Beta variables that is stronger than what can be proven using general log-concave theory. Finally Appendix [C] shows a theorem of independent interest, which we proved in the course of an unsuccessful attempt to obtain our main result by a different approximation method: that “square root of beta” variables can be coupled to Gaussian variables in such a way as to have small variance.

1.2. Reparameterization. While the parameters given naturally arise in the full matrix model (which exists only for \( \beta = 1, 2, 4 \), e.g., as the size ratios of the two Wishart matrices involved, we choose to work with a slightly different set of parameters for the purposes of this problem. Define parameters \( a \) and \( b \) by

\[
a := \frac{1}{p + q} \quad \text{and} \quad b := \frac{p}{p + q}.
\]

As we shall see, \( a \) and \( b \) allow us to express the results in a “cleaner”, perhaps more natural form. They expose symmetries of the asymptotics, which are invariant under the involution \( a \mapsto 1 - b \), \( b \mapsto 1 - a \).

For the regime of consideration of Theorem 1.3 the parameters \( a \) and \( b \) take on values in the triangle \( 0 < a < \frac{1}{2} \), and \( a < b < 1 - a \). The limiting spectral distribution will have support given by

\[
\lambda_{\pm} = \left[ \sqrt{b(1 - a)} \pm \sqrt{a(1 - b)} \right]^2.
\]

The reciprocal expression \( \frac{2}{\beta} \) appears frequently, with some terms having polynomial dependence upon it. Thus in the proofs we have used \( \alpha \) in place of \( \frac{n}{\beta} \). The Jacobi ensemble density, with these parameters, is expressed as

\[
d\mu_J(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z} \prod_i \lambda_i^{n \left[ \frac{b}{a} - 1 \right] - 1} (1 - \lambda_i)^{n \left[ \frac{1 - b}{a} - 1 \right] + 1} \prod_{i < j} |\lambda_i - \lambda_j|^{\frac{2}{a}}.
\]

The tridiagonal matrix model with these parameters is given \( A = B_\beta B_\beta^t \) where

\[
B_\beta = \begin{pmatrix}
c_n s_{n-1} & c_{n-1} s_{n-2} & \cdots & 0 \\
-s_{n-1} c'_n & -s_{n-2} c'_n & \cdots & c_1 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & -s_{n-1} c'_1 & c_1
\end{pmatrix},
\]

\( \{c_1, c_2, \ldots, c_n, c'_1, c'_2, \ldots, c'_n\} \) mutually independent,

\[
c_i \sim \sqrt{\text{Beta}(\frac{ab}{\alpha a + \alpha^{-1}(i - n)}, \frac{n(1-b)}{\alpha a} + \alpha^{-1}(i - n))} \quad \text{and} \quad c'_i \sim \sqrt{\text{Beta}(\alpha^{-1}i, \frac{n}{\alpha a} + \alpha^{-1}(i - 2n + 1))},
\]

\[
s_i = \sqrt{1 - c_i^2} \quad \text{and} \quad s'_i = \sqrt{1 - c'_i^2}.
\]

2. Polynomial Fluctuations

2.1. Traces of Powers and Path Counting. When the linear statistic \( f \) is a polynomial, it can be computed explicitly using powers of the matrix model. By linearity, this reduces to the study of monomials \( \text{tr}(A^k) \), and by the tridiagonality of \( A \), there is a simple combinatorial expansion for this trace. In particular, these traces can be expressed in terms of certain lattice paths. In this section we will study these lattice paths and develop their combinatorial properties. We will use these combinatorial properties to compute the covariance of the limiting Gaussian process for polynomial test functions. Their properties are not needed for the proof that the limiting fluctuations are Gaussian.
Definition 2.1. An alternating bridge is a lattice path from \((0,0)\) to \((2k,0)\) using only the steps \((1,1)\), \((1,0)\), and \((1,-1)\) none of whose odd steps travel up and none of whose even steps travel down. Let \(A_{2k}\) denote the collection of all such lattice paths. Likewise, let \(L_k\) denote the collection of all lattice paths of length \(k\) without the alternating property.

Remark 2.2. These paths bear some similarity to the alternating Motzkin Paths which have been used to study the Laguerre Ensemble [16]. These paths differ in that Motzkin paths are restricted to stay above the \(x\)-axis, while these are allowed to go above and below the axis.

For a lattice path \(\bar{w}\) starting at \((0,k)\) with sequence of vertical coordinates \(\{w_0 = k, w_1, w_2, \ldots\}\) and an \(n \times n\) matrix \(M\), define \(M_{\bar{w}}\) to be the product

\[
M_{\bar{w}} = \prod_{i=0}^{n} M_{w_{2i},w_{2i+1}}^T M_{w_{2i+1},w_{2i+2}}^T = \prod_{i=0}^{n} M_{w_{2i+1},w_{2i+2}}^T M_{w_{2i+2},w_{2i+3}}^T,
\]

provided that all \(n \leq w_i \leq 1\). If the lattice path \(\bar{w}\) walks off the edge of the matrix, in the sense that either some \(w_i > n\) or \(w_i < 1\), then define \(M_{\bar{w}} = 0\).

Example 2.3. A lattice path \(\bar{w}\) and its associated product \(M_{\bar{w}}\).

Expanding the trace,

\[
\text{tr } A^k = \sum_{i=1}^{n} \left[ (B_{\beta} B_{\beta}^T)^k \right]_{i,i}.
\]

The diagonal entries \([ (B_{\beta} B_{\beta}^T)^k ]_{i,i}\) can be written in terms of alternating bridges, since for all \(1 \leq i \leq n\),

\[
\left[ (B_{\beta} B_{\beta}^T)^k \right]_{i,i} = \sum_{\bar{w} \in A_{2k}} (B_{\beta})_{\bar{w}+i},
\]

where \(\bar{w} + i\) is the lattice path \(\bar{w}\) shifted up by \(i\). For convenience, define \(\bar{A}_{2k,n}\) to be all alternating bridges that are shifted up to start at coordinates between 1 and \(n\); we will refer to these lattice paths as tridiagonal trace paths. In terms of these paths, we can write the trace of a power of a matrix as

\[
\text{tr } A^k = \sum_{\bar{w} \in \bar{A}_{2k,n}} A_{\bar{w}}.
\]

When \(n\) is large and \(k\) is fixed, each \(A_{\bar{w}}\) is approximated by a substantially simpler quantity: every entry in a \(2k \times 2k\) principal submatrix on the diagonal of \(A\) is strongly approximated by a deterministic tridiagonal band matrix (c.f. Lemmas 2.19 and 2.19). Thus, endow an alternating bridge with a weight by giving each horizontal edge weight \(x\) and each inclined edge weight \(y\). Define the weight of the bridge to be the product of the weights of its edges, and define \(p_k(x,y)\) to be the sum of all the weights over all the paths in \(A_{2k}\). If we let \(h(\bar{w})\) denote the number of horizontal steps taken by path \(\bar{w}\), then

\[
p_k(x,y) = \sum_{\bar{w} \in A_{2k}} x^{h(\bar{w})} y^{2k-h(\bar{w})}.
\]

We are interested in finding the exponential generating function for these \(p_k\), i.e. we will compute

\[
\mathcal{P}(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} p_k(x,y),
\]

and show that

\[
\mathcal{P}(t) = e^{t(x+y^2)} I_0(2xyt),
\]

(10)
where $I_0$ is the modified Bessel function of the first kind.

These polynomials exhibit some nice combinatorial properties. Suppose that a path $\bar{w} \in \mathcal{A}_{2k}$ has $i$ up-steps. Because the path returns to 0, it must also have $i$ down-steps. Down-steps must be placed in odd positions, and up-steps must be placed in even positions; as a result, the placement of the up-steps is independent from the placement of the down-steps. Thus, there are exactly $\binom{k}{i} \binom{k}{k-i}$ paths in $\mathcal{A}_{2k}$ having $2i$ inclined steps. Note, this argument also shows that the number of inclined steps must be even. Consequently, the number of horizontal steps is even as well, and we have shown

$$p_k(x, y) = \sum_{l=0}^{k} \binom{k}{l}^2 x^{2l} y^{2(k-l)} = y^{2k} F_1(-k, -k, 1; (x/y)^2).$$

For definitions and properties of the hypergeometric function $F_1$ see [1, page 556]. As a consequence, we are able to compute the size of $\mathcal{A}_{2k}$ by simply evaluating this polynomial at $x = y = 1$,

$$|\mathcal{A}_{2k}| = \sum_{i=0}^{k} \binom{k}{i}^2 = \binom{2k}{k}.$$

While the alternating structure naturally lends itself to describing traces of $A$, there is another way to view $\mathcal{A}_{2k}$ which lends itself better to computing $\mathcal{S}(t)$. If $\bar{w} = w_1 w_2 \cdots w_{2k-1} w_{2k}$, for steps $w_i$ then the concatenation of the steps $w_{2i-1} w_{2i}$ is one of $(2, 1)$, $(2, -1)$ or $(2, 0)$. Moreover, if it is either of the first two, then by the alternating structure, $w_{2i-1} w_{2i}$ must have been $(1, 0)(1, 1)$ or $(1, -1)(1, 0)$ respectively. If it was a horizontal step, then there are two possibilities, either $(1, 0)(1, 0)$ or $(1, -1)(1, 1)$.

**Definition 2.4.** By concatenating pairs of steps, alternating bridges $\bar{w}$ are in bijective correspondence with lattice paths in $\mathcal{L}_k$ whose horizontal steps are 2-colored. Let those horizontal steps corresponding to $(1, 0)(1, 0)$ be colored red, and let those horizontal steps corresponding to $(1, -1)(1, 1)$ be colored blue.

**Example 2.5.** Two alternating bridges with the overlaid $\mathcal{L}_k$ path.

**Lemma 2.6.** Let $\bar{w} \in \mathcal{A}_{2k,n}$ be given, and define $S_{\bar{w}}^+(m)$ to be the number of times $\bar{w}$ walks from height $m$ to height $m + 1$ or back, and let $S_{\bar{w}}^-(m)$ be the number of times that $\bar{w}$ walks horizontally at height $m$. Both $S_{\bar{w}}^+(m)$ and $S_{\bar{w}}^-(m)$ are even.

**Proof.** Let $\bar{u}$ be the colored lattice path from $\mathcal{L}_k$ that corresponds to $\bar{w}$. Let $v$ be the number of steps that $\bar{u}$ makes between height $m$ and height $m + 1$ and back. Because $\bar{u}$ returns to its starting height, $v$ is even. Let $R$ be the number of red horizontal steps (i.e. those resulting from a $(1, 0)(1, 0)$ pattern) that $\bar{u}$ makes at height $m$, and let $B$ be the number of blue horizontal steps (those resulting from a $(1, -1)(1, 1)$ pattern) that $\bar{u}$ makes at height $m + 1$. Because $S_{\bar{w}}^+(m) = v + 2R$ and $S_{\bar{w}}^-(m) = v + 2B$, both are always even. □

The correspondence between colored $\mathcal{L}_k$ and $\mathcal{A}_{2k}$ allows the polynomials $p_k(x, y)$ to be represented in a third way. We will define the weight of an uncolored path $p \in \mathcal{L}_k$ to equal the sum of the weights over all alternating bridges $\bar{w}$ to which its colorings correspond. Suppose that an alternating bridge $\bar{w}$ is in correspondence with a colored path $p$, one with $r$ red edges and $b$ blue edges. Recall that $h(p)$ is the number of horizontal steps the path takes, and therefore the weight of $\bar{w}$ is $(xy)^{h(p)} x^{2r} y^{2b}$. There are $h(p)$ ways of placing the $r$ red edges on the path (after which the placement of the $b$ blue edges is determined). As the possible colorings of a fixed path $p$ are in bijective correspondence with $\{1, 0\}^{h(p)}$, it follows that the sum of
the weights corresponding to all different colorings of a given path \( p \) is \((xy)^{k-h(p)}(x^2 + y^2)^{h(p)}\). In conclusion, \( p_k(x, y) \) can be written as

\[
p_k(x, y) = \sum_{p \in \mathcal{L}_k} (xy)^{k-h(p)}(x^2 + y^2)^{h(p)}.
\]

The subset of the lattice paths \( \mathcal{L}_k \) that fixes a given horizontal edge is in bijective correspondence with \( \mathcal{L}_{k-1} \), simply by removing the given edge. By inclusion-exclusion, it follows immediately that the lattice paths in \( \mathcal{L}_k \) that have no horizontal steps are counted by

\[
|\mathcal{L}_k| - \binom{k}{1} |\mathcal{L}_{k-1}| + \binom{k}{2} |\mathcal{L}_{k-2}| - \binom{k}{3} |\mathcal{L}_{k-3}| + \cdots = \begin{cases} \binom{k}{k} & k \text{ even} \\ \binom{k}{0} & k \text{ odd} \end{cases}
\]

The correspondence between \( \mathcal{L}_k \) with a fixed horizontal edge and \( \mathcal{L}_{k-1} \) decreases the statistic \( h(p) \) by exactly 1, and so this inclusion-exclusion formula carries over to \( p_k \) as

\[
p_k(x, y) - \binom{k}{1} (x^2 + y^2)p_{k-1}(x, y) + \binom{k}{2} (x^2 + y^2)^2 p_{k-2}(x, y) + \cdots = \begin{cases} (xy)^k \binom{k}{k} & k \text{ even} \\ 0 & k \text{ odd} \end{cases}.
\]

This recurrence can be recast in terms of the exponential generating function \( \mathcal{P}(t) \) to read

\[
\mathcal{P}(t)e^{-t(x^2 + y^2)} = 1 + \frac{x^2y^2t^2}{2!} + \frac{x^4y^4t^4}{4!} + \frac{x^6y^6t^6}{6!} + \cdots = I_0(2xyt).
\]

Thus, we have shown \( \ref{eq:laplace-transform} \),

\[
\mathcal{P}(t) = e^{t(x^2 + y^2)}I_0(2xyt).
\]

Working with this function proves to be somewhat complicated, and it will be convenient to instead use the Laplace transform of \( \mathcal{P}(t) \). Let \( \mathcal{L}_t[f(t)](\omega) \) denote the Laplace transform in the variable \( t \)

\[
\mathcal{L}_t[f(t)](\omega) = \int_0^\infty e^{-\omega t}f(t)dt.
\]

When applicable, \( \mathcal{L}_{s,t} \) will denote the Laplace transform in both variables. The calculation of the Laplace transform of \( \mathcal{P}(t) \) is simplified greatly by some elementary properties of the Laplace transform and the known Laplace transforms of modified Bessel functions. All of these properties are available for reference in \( \ref{bib:abramowitz} \) Chapter 29; properties of the modified Bessel functions are available in \( \ref{bib:abramowitz} \) Chapter 9. The Laplace transform of the modified Bessel functions \( I_n \) is given by

\[
\mathcal{L}_t[I_n(\omega t)](\omega) = \frac{e^{\omega t}}{(\omega + \sqrt{\omega^2 - c^2})^n} \omega^n \left( \frac{1}{\sqrt{\omega^2 - c^2}} \right), \quad \omega > c.
\]

If for some real value of \( \omega_0 \), the Laplace transform is finite, then for any \( \omega \) in the half plane \( \Re \omega > \omega_0 \), the Laplace transform is finite. Further, the transform satisfies the following identities

\[
\mathcal{L}_t[e^{kt}f(t)](\omega) = \mathcal{L}_t[f(t)](\omega - k),
\]

\[
\mathcal{L}_t[tf(t)](\omega) = -\frac{d}{d\omega}\mathcal{L}_t[f(t)](\omega).
\]

We will show that \textit{a priori}, the Laplace transform of \( \mathcal{P}(t) \) is finite in the half plane \( \Re \omega > (x + y)^2 \). This follows as \( I_n(2xyt) \) satisfies the simple estimate

\[
0 \leq I_n(2xyt) \leq e^{2xyt},
\]

for \( t > 0, 2xy > 0 \), and thus

\[
0 \leq \mathcal{P}(t) \leq e^{t(x+y)^2}.
\]

Identity \( \ref{eq:laplace-transform} \) makes computing the Laplace transform of \( \mathcal{P}(t) \) a simple substitution into \( \ref{eq:laplace-transform} \), as

\[
\mathcal{L}_t[\mathcal{P}(t)](\omega) = \mathcal{L}_t[e^{t(x^2 + y^2)}I_0(2xyt)](\omega) = \mathcal{L}_t[I_0(2xyt)](\omega - x^2 - y^2) = \frac{1}{\sqrt{(\omega - x^2 - y^2)^2 - 4x^2y^2}}.
\]

Using \( \ref{eq:laplace-transform} \), it is possible to compute the Laplace transform of \( \partial_x \mathcal{P}(t) \), which arises later.
Lemma 2.7.
\[ L_t[\partial_x \mathcal{P}(t)](\omega) = \frac{2x(\omega + y^2 - x^2)}{\left((\omega - x^2 - y^2)^2 - 4x^2y^2\right)^{3/2}}, \quad \omega > (x + y)^2. \]

Proof. This is a straightforward application of (13), (14) and the identity \( I_0(t)' = I_1(t) \).
\[ L_t[\partial_x \mathcal{P}(t)](\omega) = L_t[2xte^{tx^2+y^2}I_0(2xyt) + 2yte^{tx^2+y^2}I_1(2xyt)](\omega) = -\partial_\omega L_t[2xte^{tx^2+y^2}I_0(2xyt) + 2yte^{tx^2+y^2}I_1(2xyt)](\omega) = -\partial_\omega \left[ \frac{2x}{\sqrt{\omega^2 - 4x^2y^2}} + \frac{2y}{\sqrt{\omega^2 - 4x^2y^2} \omega + \sqrt{\omega^2 - 4x^2y^2}} \right], \]
where \( \tilde{\omega} \) is \( \omega - x^2 - y^2 \). Thus
\[ L_t[\partial_x \mathcal{P}(t)](\omega) = \frac{2x(\tilde{\omega} + 2y^2)}{(\tilde{\omega}^2 - 4x^2y^2)^{3/2}} = \frac{2x(\omega + y^2 - x^2)}{\left((\omega - x^2 - y^2)^2 - 4x^2y^2\right)^{3/2}}, \quad \omega > (x + y)^2. \]

Remark 2.8. In a manner of speaking, we have circuitously arrived at the regular generating function for \( p_k(x, y) \), since it is possible to deduce the generating function from the exponential generating function by way of the Laplace transform, as follows. Let \( \mathcal{P}^R(t) \) denote the generating function,
\[ \mathcal{P}^R(t) = \sum_{k=0}^{\infty} t^k p_k(x, y). \]
The effect of taking the Laplace transform on an exponential generating function can be understood using the Gamma function.
\[ L_t[\mathcal{P}(t)](\omega) = \int_0^\infty e^{-\omega t} \mathcal{P}(t) dt = \int_0^\infty e^{-\omega t} \sum_{k=0}^{\infty} \frac{t^k}{k!} p_k(x, y) dt. \]
The order of summation and integration can be interchanged because \( \frac{t^k}{k!} p_k(x, y) \) is always positive for \( t > 0, x, y \in \mathbb{R} \),
\[ L_t[\mathcal{P}(t)](\omega) = \sum_{k=0}^{\infty} \int_0^\infty e^{-\omega t} \frac{t^k}{k!} p_k(x, y) dt. \]
Make the change of variables \( s = \omega t \), so that
\[ L_t[\mathcal{P}(t)](\omega) = \sum_{k=0}^{\infty} \int_0^\infty e^{-s} \frac{s^k}{\omega^{k+1} k!} p_k(x, y) ds = \sum_{k=0}^{\infty} \omega^{-k-1} p_k(x, y) ds = \mathcal{P}^R(\omega^{-1}) \omega^{-1}. \]
Thus, putting everything together,
\[ \mathcal{P}^R(\omega) = \frac{1}{\sqrt{(1 - \omega(x^2 + y^2))^2 - 4\omega^2 x^2 y^2}}. \]
2.2. Asymptotic Normality of Fluctuations. We show in this subsection that polynomial test functions asymptotically have jointly normal fluctuations. This is the first component of Theorem 1.3 and we summarize the precise claim in the following proposition.

**Proposition 2.9.** Let $A$ be an $n \times n$ $\beta$-Jacobi matrix, with parameters as described in Section 1.2. For any fixed $k \in \mathbb{N}$, the $k$-tuple $(X_{x^1, A}, X_{x^2, A}, \ldots, X_{x^n, A})$ converges in distribution to a centered multivariate normal random variable.

The method of proof will be the computation of the moments. Recall that a multivariate normal variable has mixed moments characterized by the Wick formula, which we will state precisely.

**Proposition 2.10.** A centered random vector $(Z_1, Z_2, \ldots, Z_k)$ is a multivariate normal if and only if for each word $m \in [k]^l$, the mixed moments satisfy

$$\mathbb{E} \prod_{i \in m} Z_i = \begin{cases} 0 & \text{if } l \text{ is odd}, \\ \sum_G \prod_{\{a,b\} \in \mathcal{E}(G)} \mathbb{E}Z_{m_a}Z_{m_b} & \text{if } l \text{ is even}, \end{cases}$$

where the sum is over all graphs $G$ that are perfect matchings on the vertices $[k]$, and where $\mathcal{E}(G)$ is the edge set of this graph.

To prove Proposition 2.10 it suffices to show that all the mixed moments asymptotically obey the Wick formula. Thus, our first goal is to show that the moments have the correct form.

**Proposition 2.11.** For a fixed word $m \in [k]^l$,

$$\mathbb{E} \prod_{i \in m} X_{x^i, A} = \begin{cases} O(n^{-1/2}) & \text{if } l \text{ is odd}, \\ \sum_G \prod_{\{a,b\} \in \mathcal{E}(G)} \mathbb{E}X_{x^{m_a}, A}X_{x^{m_b}, A} + O(n^{-1/2}) & \text{if } l \text{ is even}, \end{cases}$$

where the sum is over all graphs $G$ that are perfect matchings on the vertices $[k]$, and where $\mathcal{E}(G)$ is the edge set of this graph.

This nearly proves Proposition 2.9 but it remains to show that the covariances have a limit. We will delay this proof as we will identify the limiting covariance explicitly, and we begin in the direction of proving Proposition 2.11. In the sequel, fix some word $m \in [k]^l$. We will write the mixed moment indicated by $m$ in a way that exposes its asymptotically relevant terms. The first step is to write the mixed moment in terms of tridiagonal trace paths.

$$\mathbb{E} \prod_{u \in m} X_{x^u, A} = \mathbb{E} \prod_{u \in m} \left[ \sum_{\tilde{w} \in \tilde{\mathcal{A}}_{2m_1,n}} A_{\tilde{w}} - \mathbb{E}A_{\tilde{w}} \right]$$

(15)

where the sum is over all tridiagonal trace paths $(\tilde{w}_1, \ldots, \tilde{w}_l) \in \tilde{\mathcal{A}}_{2m_1,n} \times \tilde{\mathcal{A}}_{2m_2,n} \times \cdots \times \tilde{\mathcal{A}}_{2m_l,n}$.

Each nonzero random variable $A_{\tilde{w}}$ is a product of terms of matrix entries. More specifically, by Lemma 2.6 trace paths visit each matrix entry an even number of times, and so $A_{\tilde{w}}$ is a polynomial in the random variables $\{c_i^2\}$ and $\{(c_i')^2\}$. Thus for each tridiagonal trace path $\tilde{w}_i$ for which $A_{\tilde{w}_i} \neq 0$, it is possible to define random variables $q_{j_i}^{\tilde{w}_i}$ with $1 \leq j_i \leq 2n_i - 1$ so that

1. $q_{j_i}^{\tilde{w}_i}$ is a polynomial in $c_j^2$ for $1 \leq j \leq n$;
2. $q_{j_i + n_i}^{\tilde{w}_i}$ is a polynomial in $(c'_j)^2$ for $1 \leq j \leq n - 1$;
3. $A_{\tilde{w}_i} = \prod_{j=1}^{2n_i-1} q_{j_i}^{\tilde{w}_i}$;
4. The smallest nonzero coefficient of each $q_{j_i}^{\tilde{w}_i}$ is 1.

We will write $q_{j_i}^{\tilde{w}_i}(x)$ for the corresponding polynomial in $x$, while when no argument is provided, we mean the random variable defined above. This decomposition breaks a random variable $A_{\tilde{w}_i}$ into a product of independent random variables. Further, each polynomial has the form $q_{j_i}^{\tilde{w}_i}(x) = x^{a_{i,j}}(1 - x)^{b_{i,j}}$ for some non-negative integer powers. Note, however, that most of these polynomials are identically 1.
Proof. In the current parameterization, we recall that variables with parameters so that Lemma 2.12.

The primary tool in this proof is the Poincaré inequality for Beta random variables. From Lemma B.1, a hypothesis on the parameters of the matrix, (16) nonempty

In this last step we omit the empty set precisely because it is the term canceled by $EA_{\bar{w}_i}$.

Note that in (16) we require a product of $l$ of these terms. Thus, by applying the (16) multiple times, we can write

$$
\prod_{i=1}^{l} [A_{\bar{w}_i} - EA_{\bar{w}_i}] = \sum_{S \subseteq [2n-1]} \prod_{i \in S} (q_{j_i}^{\bar{w}_i} - EQ_{j_i}^{\bar{w}_i}) \prod_{j \not\in S} EQ_{j_i}^{\bar{w}_i},
$$

where it is important to note that the sum is over nonempty subsets of $[2n-1]$.

In expectation, we will see that each difference term $q_{j_i}^{\bar{w}_i} - EQ_{j_i}^{\bar{w}_i}$ that appears in the product contributes a factor of $n^{-1/2}$, and thus that the magnitude of (17) is at most $O(n^{-1/2})$. To show this, we require the ability to estimate moments of the terms that appear in the right hand side of (17). This is expressed in the following lemma.

**Lemma 2.12.** Fix a polynomial $q(x) = x^{a_1} (1-x)^{a_2}$, and fix an $n \in \mathbb{N}$. There is a constant $C = C(m, a_1, a_2)$ so that

$$
\max_{1 \leq i \leq n} E \left| q(c_i^2) - EQ(c_i^2) \right|^m \leq Cn^{-m/2}, \quad \text{and} \quad 
\max_{1 \leq i \leq n-1} E \left| q((c_i')^2) - EQ((c_i')^2) \right|^m \leq Cn^{-m/2}.
$$

**Proof.** In the current parameterization, we recall that $c_i^2$ and $(c_i')^2$ are mutually independent Beta random variables with parameters

- $c_i^2 \sim \text{Beta}(\frac{m}{a_0} + \alpha^{-1}(i - n), \frac{n(1-n)}{a_0} + \alpha^{-1}(i - n))$, and
- $(c_i')^2 \sim \text{Beta}(\alpha^{-1} i, \frac{n}{a_0} + \alpha^{-1}(i - 2n + 1))$.

The primary tool in this proof is the Poincaré inequality for Beta random variables. From Lemma B.1 a Beta variable $X \sim \text{Beta}(p_1, p_2)$ satisfies a Poincaré inequality

$$
\text{Var} (f(X)) \leq \frac{1}{4(p_1 + p_2)} \mathbb{E} \left| f'(X) \right|^2,
$$

for any Lipschitz function $f$ on $[0,1]$. Let $\mathcal{M}$ denote the collection of all Beta variables appearing in the matrix model. We note that for all these variables, the sum of their parameters is at least $\frac{n}{a} \left( \frac{1}{a} - 2 \right)$. By hypothesis on the parameters of the matrix, $a < 1/2$, and thus there is a constant $C$ so that

$$
\max_{X \in \mathcal{M}} \sup_{\|f\|_{Lip} < \infty} \left[ \frac{\text{Var} f(X)}{\mathbb{E} \left| f'(X) \right|^2} \right] \leq C \frac{1}{n}.
$$

Further, by applying each of these inequalities to $q(X)$ for any $X \in \mathcal{M}$, we see that for any Lipschitz $f$,

$$
\text{Var} f(q(X)) \leq \frac{C}{n} \mathbb{E} \left| f'(q(X)) q'(X) \right|^2.
$$

Note that $|q'(x)| \leq (a_1 + a_2)$ on $[0,1]$, and thus

$$
\text{Var} f(q(X)) \leq \frac{C(a_1 + a_2)^2}{n} \mathbb{E} \left| f'(q(X)) \right|^2,
$$
for all Lipschitz functions on the interval and any $X \in \mathcal{M}$. It is well known that a Poincaré inequality implies exponential integrability (see [10]). Precisely,

$$\mathbb{E} \exp \left[ \frac{(g(X) - \mathbb{E}g(X))}{\sqrt{12(a_1 + a_2)/C}} \right] \leq 2,$$

for every $X \in \mathcal{M}$. By expanding the exponential in its series, the claim follows. \hfill \square

As a consequence of Lemma 2.12 it is possible to estimate the contribution of any product of terms as in (17).

**Lemma 2.13.** There is a constant $C = C(l, \max_{1 \leq i \leq l} m_i)$ so that for any $l$-tuple $(\tilde{w}_1, \ldots, \tilde{w}_l) \in \tilde{A}_{2m_1,n} \times \tilde{A}_{2m_2,n} \times \cdots \times \tilde{A}_{2m_l,n}$,

$$\left| \mathbb{E} \prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}] \right| \leq Cn^{-l/2}.$$  

Furthermore, the dominant contribution is given by

$$D_{(\tilde{w}_i)} := \sum_{s_1, \ldots, s_l} \prod_{i=1}^l \left( q_{s_i}^{\tilde{w}_i} - \mathbb{E}q_{s_i}^{\tilde{w}_i} \right) \prod_{j \not\in s_i} \mathbb{E}q_j^{\tilde{w}_i},$$

with the sum over all $l$-tuples $(s_1, \ldots, s_l) \in [l]^{2n-1}$, and

$$\left| \mathbb{E} \prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}] - \mathbb{E}D_{(\tilde{w}_i)} \right| \leq Cn^{-(l+1)/2}.$$  

**Proof.** We recall (17):

$$\prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}] = \sum_{S_1, \ldots, S_l} \prod_{i=1}^l \prod_{j \in S_i} \left( q_j^{\tilde{w}_i} - \mathbb{E}q_j^{\tilde{w}_i} \right) \prod_{j \not\in S_i} \mathbb{E}q_j^{\tilde{w}_i},$$

where the sum is over nonempty subsets $S_i \subset [2n-1]$. Taking expectations, most of these of summands will be 0. This is because for each word $\tilde{w}_i$, there are at most $4m_i$ nontrivial polynomials $q_j^{\tilde{w}_i}$, where $\tilde{w}_i \in \tilde{A}_{2m_i,n}$. Thus, there are at most $2^{4m_1}2^{4m_2}\cdots2^{4m_l}$ nonzero summands of the form

(18)

$$P_{S_1, \ldots, S_l} := \mathbb{E} \prod_{i=1}^l \prod_{j \in S_i} \left( q_j^{\tilde{w}_i} - \mathbb{E}q_j^{\tilde{w}_i} \right) \prod_{j \not\in S_i} \mathbb{E}q_j^{\tilde{w}_i},$$

and thus it suffices to show the desired bound for an arbitrary term such as this. From each $S_i$, pick an arbitrary $j_i$. Each $q_j^{\tilde{w}_i}$ is a random variable supported on $[0,1]$, and thus both $|q_j^{\tilde{w}_i} - \mathbb{E}q_j^{\tilde{w}_i}| \leq 1$ and $|\mathbb{E}q_j^{\tilde{w}_i}| \leq 1$. Therefore, the term in (18) can be bounded by

$$|P_{S_1, \ldots, S_l}| \leq \mathbb{E} \left| \prod_{i=1}^l (q_{j_i}^{\tilde{w}_i} - \mathbb{E}q_{j_i}^{\tilde{w}_i}) \right| \leq \frac{1}{l} \sum_{i=1}^l \mathbb{E} \left| q_{j_i}^{\tilde{w}_i} - \mathbb{E}q_{j_i}^{\tilde{w}_i} \right|,$$

where we have applied the arithmetic-geometric mean inequality. By applying Lemma 2.12, we conclude that there is a constant $C$ that depends only on $\max_{1 \leq i \leq l} m_i$ and $l$ so that

$$|P_{S_1, \ldots, S_l}| \leq Cn^{-l/2}.$$  

Summing over all possible nonzero summands, the first conclusion follows. Note that the same argument shows that if $\sigma := |S_1| + |S_2| + \cdots + |S_l| > l$, then the same argument (with the same constant no less) shows

$$|P_{S_1, \ldots, S_l}| \leq Cn^{-\sigma/2},$$

from which the second conclusion follows. \hfill \square

Having established these bounds, we introduce the notion of a dependency graph.

**Definition 2.14.** For any tuple of tridiagonal trace paths $(\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_l)$, define the dependency graph $G$ to be a graph with vertex set $[l]$ and $i \not\sim j$ if and only if $A_{\tilde{w}_i}$ and $A_{\tilde{w}_j}$ are functions of mutually independent random variables.
The family of vector variables
\[ \Xi := \{(A_{\bar{w}_i})_{i \in S}\}_{S}, \]
where \( S \) ranges over all connected components of \( G \), is a mutually independent family of random variables. The importance of these connected components is that there are very few \( l \)-tuples of tridiagonal trace paths that have few connected components in their dependency graph. Moreover, it is possible to estimate exactly how many trace paths have such dependency graphs. This motivates the following definition.

**Definition 2.15.** For any \( \chi \in \{1, 2, \ldots, \lfloor l/2 \rfloor \} \), let \( B_\chi \) be the collection of all \( l \)-tuples in \( \tilde{A}_{2m_1,n} \times \tilde{A}_{2m_2,n} \times \cdots \times \tilde{A}_{2m_\chi,n} \), whose dependency graphs have \( \chi \) connected components and no isolated vertices. For any such word tuple of words, let \( \mathcal{E} = \mathcal{E}(\tilde{w}_1, \ldots, \tilde{w}_l) \) denote the edge set of the dependency graph.

When \( l \) is even, \( B_{l/2} \) is the collection of all \( l \)-tuples of trace paths whose dependency graphs are perfect matchings. With this definition, we can count the number of \( l \)-tuples of trace paths having a particular number of connected components.

**Lemma 2.16.** For any \( \chi \in \mathbb{N} \), there is a constant \( C = C(\chi, \max_{1 \leq i \leq l} m_i) \) so that \( |B_\chi| \leq Cn^\chi \).

**Proof.** This ultimately stems from the observation that there are only finitely many entries in the matrix that depend on a given entry. Thus, once any arbitrary trace path in a connected component has been chosen, the remainder of the trace paths must start nearby. Formally, we begin by bounding the number of ways to construct a connected component on \( s \) vertices.

Without loss of generality, suppose these \( s \)-tuples are chosen from \( \tilde{A}_{2m_1,n} \times \tilde{A}_{2m_2,n} \times \cdots \times \tilde{A}_{2m_\chi,n} \). As we would like choices having a connected dependency graph, we overcount by first choosing a desired spanning tree and then filling out the graph. As there are only \( s^{\chi-2} \) such spanning trees, we lose at most a constant factor.

Let \( M = \max_{1 \leq i \leq l} m_i \), and choose the first trace path in the tuple arbitrarily; there are \( |\tilde{A}_{2m_\chi,n}| \) possible choices for this path. Traversing the vertices of the tree in a depth first search, each vertex traversed must depend on the previously chosen path \( \bar{w}_{prev} \in \tilde{A}_{2m_{prev},n} \). This forces the choice of \( \bar{w}_{new} \in \tilde{A}_{2m_{new},n} \) to have that \( A_{\bar{w}_{new}} \) depends on \( A_{\bar{w}_{prev}} \), and thus the starting point of \( \bar{w}_{new} \) must be no more than \( m_{new} + m_{prev} \) steps from the starting point of the previous. Thus there are at most \( 4M |A_{2M}| \) ways to choose the new path. This bound holds for every vertex explored in the depth first search, and we arrive at the bound that there are at most \( |A_{2M}|^s \cdot n \) ways to choose trace paths having dependency graph spanned by a given tree.

Summing over all possible partitions of \( l \) with \( \chi \) parts, i.e. all multisets of naturals \( \{s_i\} \) so that \( s_1 + s_2 + \cdots + s_\chi = l \), and choosing components of these sizes for each, we arrive at the bound that there is a constant \( C \) so that \( |B_\chi| \leq Cn^\chi \).

It is now possible to identify the asymptotically relevant portions of an arbitrary mixed moment, and hence prove Proposition 2.11.

**Proof of Proposition 2.11** In terms of the notation \( B_\chi \), we recall (15) and rewrite it as
\[ \mathbb{E} \prod_{u \in m} X_{u^\chi,A} = \sum_{\tilde{w}_1, \ldots, \tilde{w}_l} \mathbb{E} \prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}] \]
\[ = \sum_{\chi=1}^{\lfloor l/2 \rfloor} \sum_{\{\tilde{w}_1, \ldots, \tilde{w}_l\} \in B_\chi} \mathbb{E} \prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}], \]
noting that this sum contains no \( l \)-tuples of words with isolated vertices in their dependency graphs, as these vanish identically on taking expectations. By Lemma 2.13 there is a constant \( C_1 \) sufficiently large that
\[ \left| \mathbb{E} \prod_{i=1}^l [A_{\tilde{w}_i} - \mathbb{E}A_{\tilde{w}_i}] \right| \leq C_1 n^{-l/2}, \]

14
for every word in the sum. Also, by Lemma 2.16 there is a constant $C_2$ sufficiently large that for all $1 \leq \chi \leq \lfloor l/2 \rfloor$, $|B_\chi| \leq C_2 n^\chi$. It is immediate that if $l$ is odd, then by (19),

$$\left| \mathbb{E} \prod_{u \in m} X_{x^{u,A}} \right| \leq \sum_{\chi=1}^{\lfloor l/2 \rfloor} C_1 n^{-\chi/2} C_2 n^\chi = O(n^{-1/2}).$$

If $l$ is even, however, then applying the same bound to terms for which $\chi < l/2$,

$$\mathbb{E} \prod_{u \in m} X_{x^{u,A}} = \sum_{(\bar{w}_1, \ldots, \bar{w}_l) \in B_{l/2}} \mathbb{E} \prod_{r=1}^l [A_{\bar{w}_r} - \mathbb{E} A_{\bar{w}_r}] + O(n^{-1/2})$$

(20)

$$= \sum_{(\bar{w}_1) \in B_{l/2}} \prod_{(a,b) \in E} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}] + O(n^{-1/2}).$$

It only remains to show that the Wick word has the same form, i.e. it should be shown that

$$W := \sum_{G} \prod_{(a,b) \in E(G)} \mathbb{E} [X_{x^{a,b,A}} X_{x^{a,b,A}}],$$

where $G$ ranges over all perfect matchings of $[l]$, has the same asymptotically relevant terms as (20). We recall (13), due to which we may rewrite

$$W = \sum_{G} \prod_{(a,b) \in E(G)} \sum_{A_{2m(a,n), A_{2m(n,b)}}} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}],$$

where the inner sum may be taken over all pairs of $l$-tuples. For a fixed perfect matching $G$, every possible tuple $(\bar{w}_1, \ldots, \bar{w}_l)$ is represented exactly once. After commuting the inner sum and the product, we may write

$$W = \sum_{(\bar{w}_1, \ldots, \bar{w}_l) \in B_{l/2}} \prod_{G} \sum_{(a,b) \in E(G)} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}].$$

As before, we may ignore $l$-tuples whose dependency graphs have an isolated vertex, and thus we write

$$W = \sum_{\chi=1}^{l/2} \sum_{(\bar{w}_1) \in B_\chi} \sum_{G} \prod_{(a,b) \in E(G)} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}].$$

We will bound the contribution of terms having $\chi < l/2$, and we note that there is a constant $C_3$ so that for any pairing $G$ and any tuple of paths $(\bar{w}_i)_i$,

$$\left| \prod_{(a,b) \in E(G)} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}] \right| \leq C_3 n^{-\chi/2},$$

which follows from applying Lemma 2.13. Writing $C_4 = (2l)! / 2^l l!$ for the number of perfect matchings on $[l]$, we have

$$\sum_{\chi=1}^{l/2-1} \sum_{(\bar{w}_1) \in B_\chi} \sum_{G} \prod_{(a,b) \in E(G)} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}] \leq \sum_{\chi=1}^{l/2-1} C_2 n^\chi \cdot C_4 \cdot C_3 n^{-\chi/2} = O(n^{-1/2}).$$

For each tuple of words $(\bar{w}_i)_i \in B_{l/2}$, there is exactly one choice of pairing $G$ so that so that the product is nonzero, and thus

$$W = \sum_{(\bar{w}_i)_i \in B_{l/2}} \prod_{(a,b) \in E} \mathbb{E} [A_{\bar{w}_a} - \mathbb{E} A_{\bar{w}_a}] [A_{\bar{w}_b} - \mathbb{E} A_{\bar{w}_b}] + O(n^{-1/2}),$$

which completes the proof on comparison with (20). □
2.3. Computing the Covariance. We now turn to showing that all possible the pairwise covariances \( \text{Cov}(X_{x^k, A}, X_{x^l, A}) \) have limits and produce an expression for that limiting covariance. We will use \( C_{k,l} \) to denote the covariance we eventually show to be the limit. These covariances can be described in terms of the polynomials \( p_k(x, y) \) introduced in Section 2.1. The exact form of the covariance is given by an integral against a parameter \( \sigma \). In terms of \( \sigma \), define the expressions
\[
\begin{align*}
x &= \sqrt{(b + \sigma)(1 - a + \sigma)} / (1 + 2\sigma), \quad \text{and } y &= \sqrt{(1 - b + \sigma)(a + \sigma)} / (1 + 2\sigma).
\end{align*}
\]
The matrix \( C_{k,l} \) for \( k, l \geq 1 \) can now be defined by
\[
C_{k,l} := \alpha \int_{-a}^{a} \frac{1}{1 + 2\sigma} \left[ (\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m) (1 - x^2 - y^2) - (\partial_x p_k \partial_y p_m + \partial_y p_k \partial_x p_m) (2xy) \right] d\sigma.
\]

Remark 2.17. In this form, the integrand is separated into positive and negative parts. We can check that \( x^2 + y^2 < 1 \) for all \( -a \leq \sigma \leq 0 \). Furthermore, because \( p_k \) have all positive coefficients, \( x \) is nonnegative, and \( y \) is nonnegative, it follows that
\[
(\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m) (1 - x^2 - y^2) \geq 0,
\]
for all \( -a \leq \sigma \leq 0 \). To check that \( x^2 + y^2 < 1 \), we clear the denominator and expand the terms to show that this is equivalent to
\[
b(1 - a) + (1 - b)a < 1 + 2\sigma + 2a^2.
\]
The quadratic on the right is increasing for \(-1/2 < \sigma < 0\), and thus to show the inequality, it suffices to show that
\[
b(1 - a) + (1 - b)a = a + b(1 - 2a) < 1 - 2a + 2a^2.
\]
Using that \( 1 - 2a > 0 \) and \( b < 1 - a \), the inequality follows.

Our primary purpose in this section is to prove the following Proposition.

Proposition 2.18. For each fixed \( k, l \in \mathbb{N}, \) as \( n \to \infty, \)
\[
\text{Cov}(X_{x^k, A}, X_{x^l, A}) = \mathbb{E} \left[ X_{x^k, A}X_{x^l, A} \right] = C_{k,l} + O(n^{-1/2}).
\]

Note that combining this Proposition with Proposition 2.11 we have proven Proposition 2.9. We turn immediately towards proving Proposition 2.18. We recall that by (15), we have
\[
\mathbb{E} \left[ X_{x^k, A}X_{x^l, A} \right] = \sum_{\tilde{w}_k, \tilde{w}_l} \mathbb{E} \left[ A_{\tilde{w}_k} - \mathbb{E} A_{\tilde{w}_k} \right] \mathbb{E} A_{\tilde{w}_l} - \mathbb{E} A_{\tilde{w}_l}.
\]

By Lemma 2.16 there is a constant \( K_\chi \) so that there are at most \( K_\chi \cdot n \) such words. Applying the second portion of Lemma 2.13 we have that there is a constant \( K_{k,l} \) so that
\[
\left| \mathbb{E} \left[ X_{x^k, A}X_{x^l, A} \right] - \sum_{\tilde{w}_k, \tilde{w}_l} \mathbb{E} \left[ A_{\tilde{w}_k} - \mathbb{E} A_{\tilde{w}_k} \right] \mathbb{E} A_{\tilde{w}_l} - \mathbb{E} A_{\tilde{w}_l} \right| \leq K_\chi n \cdot K_{k,l} \cdot n^{-3/2},
\]
where we recall that \( D_{(\tilde{w}_k, \tilde{w}_l)} \) is given by
\[
D_{(\tilde{w}_k, \tilde{w}_l)} = \sum_{s_k, s_l} \prod_{i \in \{k, l\}} \left[ (q_{s_i} - \mathbb{E} q_{s_i}) \prod_{j \neq s_i} \mathbb{E} q_{s_j} \right],
\]
with the sum over all choices of \( s_k, s_l \in [2n - 1] \). Thus, it suffices to analyze the quantity \( \sum_{\tilde{w}_k, \tilde{w}_l} \mathbb{E} D_{(\tilde{w}_k, \tilde{w}_l)} \) and show it has the desired limit. Note that by the construction of \( q_{s_i} \), each of \( q_{s_k} \) and \( q_{s_l} \) are independent if \( s_k \neq s_l \), and thus we have
\[
\mathbb{E} \left[ X_{x^k, A}X_{x^l, A} \right] = \sum_{\tilde{w}_k, \tilde{w}_l} \sum_{t=1}^{2n-1} \mathbb{E} \left[ q_{t} \tilde{w}_k - \mathbb{E} q_{t} \tilde{w}_k \right] \left[ \tilde{w}_i - \mathbb{E} \tilde{w}_i \right] \prod_{j \neq t} \mathbb{E} q_{j} \tilde{w}_k \mathbb{E} q_{j} \tilde{w}_l + O(n^{-1/2}).
\]

We define \( r_t \) so that
\[
r_t := \sum_{\tilde{w}_k, \tilde{w}_l} \mathbb{E} \left[ q_{t} \tilde{w}_k - \mathbb{E} q_{t} \tilde{w}_k \right] \left[ \tilde{w}_i - \mathbb{E} \tilde{w}_i \right] \prod_{j \neq t} \mathbb{E} q_{j} \tilde{w}_k \mathbb{E} q_{j} \tilde{w}_l ,
\]
and note that by commuting sums in the previous equation, we have

\begin{equation}
E[X_{x^k,A}X_{x^l,A}] = \sum_{t=1}^{2n-1} r_t + O(n^{-1/2}).
\end{equation}

Let \(\{z_i\}_{i=1}^{2n-1}\) be the enumeration of all the Beta variables in \(\mathcal{M}\), where \(z_i = (c_i)^2\) for \(1 \leq i \leq n\) and \(z_i = (c_{i-n})^2\) when \(n+1 \leq i \leq 2n-1\). This makes each \(q_j^{\bar{w}_i}\) a polynomial in \(z_j\). The first step in the analysis amounts to using Taylor approximation to pull the expectations inside the \(q_j^{\bar{w}_i}\) polynomials.

**Lemma 2.19.** There is a constant \(K = K(k,l)\) so that for all \(1 \leq t \leq 2n-1\),

\[|r_t| \leq Kn^{-1}.\]

Moreover, it is possible to identify the dominant contribution \(r_t^D\), which is given by

\[r_t^D := \text{Var}(z_t) \sum_{\bar{w}_i, \bar{w}_j} \left[ q_t^{\bar{w}_i}(\mathbb{E}z_i) q_t^{\bar{w}_j}(\mathbb{E}z_j) \prod_{j \neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j) q_j^{\bar{w}_l}(\mathbb{E}z_l) \right],\]

and which has

\[|r_t - r_t^D| \leq Kn^{-3/2}.\]

**Proof.** The first claim follows from Lemma 2.12 and from the fact that the number of trace paths that depend on \(z_t\) is bounded by some \(K = K(k,l)\). The second claim will follow from Taylor approximation. For any polynomial \(q_j^{\bar{w}_k}\) or \(q_j^{\bar{w}_l}\), it is possible to bound the maximums of the derivatives over \([0,1]\) in terms of \(k\) and \(l\). Each polynomial has the form \(q_j^{\bar{w}_i}(x) = x^{a_1} - x^{a_2}\), and hence its first and second derivatives can be bounded by \(a_1 \cdot a_2\) and \((a_1^2 + a_2^2)^{1/2}\). These parameters \(a_1\) and \(a_2\) can in turn be bounded by \(i\), to yield

\[\max_{x \in [0,1]} |q_j^{\bar{w}_i}''| \leq 4i \quad \text{and} \quad \max_{x \in [0,1]} |q_j^{\bar{w}_i}'''| \leq (4i)^2,\]

for either \(i \in \{k,l\}\). These imply that the 0th order approximation has error

\[|q_j^{\bar{w}_i}(z_j) - q_j^{\bar{w}_i}(\mathbb{E}z_j)| \leq 4i |z_j - \mathbb{E}z_j|,\]

and the 1st order approximation has error

\[|q_j^{\bar{w}_i}(z_j) - q_j^{\bar{w}_i}(\mathbb{E}z_j) (z_j - \mathbb{E}z_j)| \leq 8i^2 |z_j - \mathbb{E}z_j|^2.\]

We recall the definition of \(r_t\), which was given by

\[r_t = \sum_{\bar{w}_i, \bar{w}_j} \left[ q_t^{\bar{w}_i} - q_t^{\bar{w}_j} \right] \prod_{j \neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j) q_j^{\bar{w}_l}(\mathbb{E}z_l).\]

Using 1st order approximation for term (i), we bound

\begin{equation}
D_{(i)} := \left| E \left[ q_t^{\bar{w}_i} - q_t^{\bar{w}_j} \right] \prod_{j \neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j) q_j^{\bar{w}_l}(\mathbb{E}z_l) - E \left[ q_t^{\bar{w}_i}(\mathbb{E}z_j) q_t^{\bar{w}_j}(\mathbb{E}z_l) (z_t - \mathbb{E}z_t)^2 \right] \right| \leq K_1 n^{-3/2},
\end{equation}

with the constant implicitly depending on \(k, l\), and the constants assured by Lemma 2.12. Using the 0th order approximation for term (ii), we will bound the difference between (ii) and its approximation. This will be done by replacing each \(z_j\) by \(\mathbb{E}z_j\) one term at a time. As there are at most \(2k + 2l\) non-constant polynomials \(q_j^{\bar{w}_i}\) and \(q_j^{\bar{w}_l}\), this reduces bounding (ii) to bounding, for any fixed \(u\),

\[\Delta_u := \left| q_u^{\bar{w}_i}(\mathbb{E}z_u) q_u^{\bar{w}_i}(\mathbb{E}z_u) - q_u^{\bar{w}_j}(z_u) q_u^{\bar{w}_j}(z_u) \prod_{j < u \land j \neq t} q_j^{\bar{w}_k}(\mathbb{E}z_j) q_j^{\bar{w}_l}(\mathbb{E}z_l) \prod_{j > u \land j \neq t} q_j^{\bar{w}_k}(z_j) q_j^{\bar{w}_l}(z_j) \right|.
\]

Recalling that all \(q_j^{\bar{w}_i}\) are almost surely less than 1, this can be bounded by

\[|\Delta_u| \leq \left| q_u^{\bar{w}_i}(\mathbb{E}z_u) q_u^{\bar{w}_i}(\mathbb{E}z_u) - q_u^{\bar{w}_j}(z_u) q_u^{\bar{w}_j}(z_u) \right| \leq (4k + 4l) |z_u - \mathbb{E}z_u| \leq K_2 n^{-1/2}.\]
These bounds applied to the difference of (ii) and its approximation show

$$D_{(ii)} := \left| \prod_{j \neq t} \mathbb{E} q_j^{\tilde{w}_k}(z_j) q_j^{\tilde{y}_i}(z_j) - \prod_{j \neq t} q_j^{\tilde{w}_k}(E z_j) q_j^{\tilde{y}_i}(E z_j) \right| \leq \sum_{1 \leq u \leq 2n - 1} \Delta_u \leq (2k + 2l) \cdot K_2 n^{-1/2}.$$  

By combining Lemma 2.12 with Cauchy-Schwarz, one has that (i) is at most $K_3 n^{-1}$. Therefore, we can combine both of (26) and (27) to show

$$|r_t - r_t^D| \leq \sum_{\tilde{w}_k, \tilde{w}_i} |(i)| |D_{(ii)}| + |D_{(i)}| \left| \prod_{j \neq t} q_j^{\tilde{w}_k}(E z_j) q_j^{\tilde{y}_i}(E z_j) \right| \leq \sum_{\tilde{w}_k, \tilde{w}_i} K_3 n^{-1} \cdot (2k + 2l) \cdot K_2 n^{-1/2} + K_1 n^{-3/2} \cdot 1.$$

As the sum is only over paths that depend upon $t$, the proof is complete.

All the expectations in $r_t^D$ are approximately equal to one of two values, $\mathbb{E} z_t$ and $\mathbb{E} z_{t+n}$ (or $t - n$ in the case $t > n$), on account of the trace paths being forced to overlap. Thus, this can be expressed in terms of the polynomials $p_k(x, y)$ for values of $t$ for which the trace paths are sufficiently far from the matrix edge. The values of $x$ and $y$ are given in terms of the expectations of matrix entries. Put $s(t) = t$ if $1 \leq t \leq n$, and put $s(t) = t - n$ if $n + 1 \leq t \leq 2n - 1$. The values of $x$ and $y$ are given by

$$x(t) := \sqrt{\mathbb{E} (c_k^2(1 - (c_k')^2))} \quad \text{and} \quad y(t) := \sqrt{\mathbb{E} [((c_k')^2(1 - (c_k')^2))]}.$$  

Note that these are not exactly the expressions for $x$ and $y$ given in (22), but we will show that these two quantities are strongly related. In what follows, we unequivocally mean the $x$ and $y$ given in (28).

**Lemma 2.20.** Define $\xi_t^D$ to be

$$\xi_t := r_t^D - \frac{1}{4} \left\{ \frac{\text{Var}(c_k^2)}{\text{Var}(c_k')} \left[ \left( \frac{\partial_x p_k(x, y)}{\text{E} c_k^2} - \frac{\partial_y p_k(x, y)}{\text{E} c_k'} \right) \left( \frac{\partial_x p_m(x, y)}{\text{E} c_k^2} - \frac{\partial_y p_m(x, y)}{\text{E} c_k'} \right) \right] \right\} \left( 1 \leq t \leq n \right) \quad \text{and} \quad \xi_t^D \leq K n^{-2}.$$

There is a constant $K = K(k, l)$ so that for all $k + l \leq t \leq n - k - l$ and $n + k + l \leq t \leq 2n - k - l - 1$,

$$|\xi_t^D| \leq K n^{-2}.$$

**Proof.** We show the proof for $1 \leq t \leq n$. The proof for $t > n$ is identical. We recall that $r_t^D$ is given by

$$r_t^D = \text{Var}(z_t) \sum_{\tilde{w}_k, \tilde{w}_i} \left[ q_t^{\tilde{w}_k}(E z_t) \right] \left[ q_t^{\tilde{y}_i}(E z_t) \right] \left[ \prod_{j \neq t} q_j^{\tilde{w}_k}(E z_j) q_j^{\tilde{y}_i}(E z_j) \right].$$

This splits nicely as $r_t^D = \text{Var}(z_t) M_t(\tilde{w}_k) M_t(\tilde{w}_i)$, where we define

$$M_t(\tilde{w}_i) := \sum_{\tilde{w}_i} \left[ q_t^{\tilde{w}_i}(E z_t) \right] \left[ \prod_{j \neq t} q_j^{\tilde{w}_i}(E z_j) \right].$$

This $M_t(\tilde{w}_i)$ is essentially computable from just two expectations, $\mathbb{E} z_t$ and $\mathbb{E} z_{t+n}$. Letting

$$M_t(\tilde{w}_i) := \sum_{\tilde{w}_i} \left[ q_t^{\tilde{w}_i}(E z_t) \right] \left[ \prod_{1 \leq j \leq n} q_j^{\tilde{w}_i}(E z_j) q_j^{\tilde{w}_i}(E z_{t+n}) \right],$$

we show that $|M_t(\tilde{w}_i) - M_t^D(\tilde{w}_i)|$ is $O(n^{-1})$. We will require the formulae for $\mathbb{E} z_t = \text{E} c_k^2$ and $\mathbb{E} z_{t+n} = \text{E} (c_k')^2$, and so we recall the precise distributions of these entries,

$$c_k^2 \sim \text{Beta}(\frac{n}{a_k} + \alpha^2(1 - n), \frac{n(1 - b)}{a_k} + \alpha^2(1 - n)) \quad \text{and} \quad (c_k')^2 \sim \text{Beta}(\alpha^2(1 - i - n), \frac{a_k}{\alpha} + \alpha^2(i - n)).$$

Their expectations are given by

$$\mathbb{E} z_t = \text{E} c_k^2 = \frac{a_k}{\alpha} + \alpha^2(t - n) = \frac{b - a + 2a \frac{t}{n}}{1 - 2a + 2a \frac{t}{n}}$$

and

$$\mathbb{E} z_{t+n} = \text{E} (c_k')^2 = \frac{\alpha^2}{1 - 2a + 2a \frac{t}{n}}.$$
Each of these expectations, as a function of \( t \), is uniformly Lipschitz continuous over \( 0 \leq t \leq n \) with constant \( K_1 \cdot n^{-1} \) for some \( K_1 \), depending only on the ensemble parameters. By the same method used in the proof of Lemma \( 2.19 \), it is straightforward to show that there is a constant \( K_2 = K_2(k, l) \) so that

\[
| M_t(\bar{w}_i) - M_t^D(\bar{w}_i) | \leq K_2 n^{-1}.
\]

We recall the notation of Lemma \( 2.20 \) where we defined \( S^{\bar{w}_i}(t) \) to be the number of horizontal steps of \( \bar{w}_i \) from level \( i \) to \( i \) and \( S^{\bar{w}_i}(t) \) to be the number of steps of \( \bar{w}_i \) from level \( i \) to \( i+1 \) or vice versa. The polynomial \( q^{\bar{w}_i}_t \) may be identified precisely in terms of these counts. Recalling the matrix model (3), the variables \( c_t \) and \( s_t \) appear only in the \( t \)th row from the bottom of the matrix. It follows that

\[
q^{\bar{w}_i}_t(c_t^i) = c_t^i S^{\bar{w}_i}(t) \left( 1 - c_t^i S^{\bar{w}_i}(t) / 2 \right),
\]

and thus, differentiating,

\[
q^{\bar{w}_i}_t(z) = \frac{S^{\bar{w}_i}(t) z S^{\bar{w}_i}(t) / 2 - 1}{2} \left( 1 - \frac{z}{2} \right)
- \frac{S^{\bar{w}_i}(t) z S^{\bar{w}_i}(t) / 2 \left( 1 - \frac{z}{2} \right)}{2}
\]

\[
= \frac{S^{\bar{w}_i}(t) q^{\bar{w}_i}_t(z)}{2z} - \frac{S^{\bar{w}_i}(t) q^{\bar{w}_i}_t(z)}{2(1 - z)}.
\]

We now relate \( M^D_t(\bar{w}_i) \) to expressions containing \( p_t(x, y) \). The essential realization is that

\[
\sum_{\bar{w}_i} S^{\bar{w}_i}(t) \left[ \prod_{1 \leq j \leq n} q^{\bar{w}_i}_t(E_{z_j}) q^{\bar{w}_i}_{j+n}(E_{z_{j+n}}) \right] = \sum_{\bar{w}_i \in A_{2t}} h(\bar{w}) x h(\bar{w}) y^{2l - h(\bar{w})} = x \partial_x p_t(x, y),
\]

where \( h(\bar{w}_i) \) is the number of horizontal steps \( \bar{w}_i \) makes, and \( x \) and \( y \) are defined earlier. This is a direct consequence of the bijection between paths \( \bar{w}_1 \in A_{2t, n} \) that have a single marked horizontal edge at level \( t \) and paths \( \bar{w}_2 \in A_{2t} \) having a single marked horizontal edge. This is given by the map that simply vertically shifts \( \bar{w}_1 \) to start at \( 0 \); note that this is invertible on account of the mark being forced to lie at level \( t \). For \( n - k - l \geq t \geq k + l \), every summand on the left hand side of (32) is exactly the summand given on the right when identifying paths via this bijection (note that for \( t \) too close to the matrix edge, some of the paths on the left hand side will be \( 0 \), destroying the identity). Similar reasoning shows

\[
\sum_{\bar{w}_i} S^{\bar{w}_i}_i(t) \left[ \prod_{1 \leq j \leq n} q^{\bar{w}_i}_j(E_{z_j}) q^{\bar{w}_i}_{j+n}(E_{z_{j+n}}) \right] = y \partial_y p_t(x, y).
\]

By combining (31), (32), and (33), it follows that

\[
M^D_t(\bar{w}_i) = \frac{x \partial_x p_t(x, y)}{2E_{z_t}} - \frac{y \partial_y p_t(x, y)}{2E[1 - z_t]}.
\]

The conclusion of the lemma follows more or less immediately. By Lemma \( 2.12 \), the variance of \( z_t \) can be controlled by \( K_3 n^{-1} \), with \( K_3 \) depending only on the matrix parameters. The moduli of \( M_t(\bar{w}_i) \) and \( M_t(\bar{w}_i)^D \) can be controlled by some \( K_4 = K_4(k, l) \), and so

\[
|\xi_t| = |\text{Var}(z_t) M_t(\bar{w}_k) M_t(\bar{w}_l) - \text{Var}(z_t) M_t^D(\bar{w}_k) M_t^D(\bar{w}_l)| \leq K_3 n^{-1} \cdot 2K_4 \cdot K_2 n^{-1},
\]

completing the proof.

On account of the variance being of the order of \( n^{-1} \), summing these expressions takes the form of a Riemann sum. We thus conclude the proof of the limiting covariance formula by showing that this Riemann sum converges to the integral given by \( C_{k, t} \).

**Proof of Proposition 2.13** By Lemma 2.19 and (25),

\[
\text{EX}_{x^t, A} X_{x^t, A} = \left[ \sum_{t=1}^{2n-1} r_t^D \right] + O(n^{-1/2}).
\]

For \( 1 \leq t \leq n \), Lemma 2.20 shows that

\[
\sum_{t=1}^{n} r_t^D = \sum_{t=k+l}^{n-k-l} \text{Var}(c_t^2) \left( \frac{x \partial_x p_k(x, y)}{E c_t^2} - \frac{y \partial_y p_k(x, y)}{1 - E c_t^2} \right) \left( \frac{x \partial_x p_m(x, y)}{E c_t^2} - \frac{y \partial_y p_m(x, y)}{1 - E c_t^2} \right) + O(n^{-1/2}).
\]
We will show that the variance of these Beta variables is of order \( n^{-1} \). To concisely describe the integrand that results in the limit, put \( \tau \) to be the variable over which the integral is taken, and define \( e(\tau) \) and \( e'(\tau) \) as

\[
e(\tau) := \frac{b-a+2a\tau}{1-2a+2a\tau} \quad \text{and} \quad e'(\tau) := \frac{a\tau}{1-2a+2a\tau} \quad 0 \leq \tau \leq 1,
\]

so that for \( \tau = t/n \), \( e(\tau) = E_t^2 \) and \( e'(\tau) = E(t)^2 \) (see (29)). We will reuse the notation \( x \) and \( y \) by putting

\[
x(\tau) := \sqrt{e(\tau)(1-e'(\tau))} \quad \text{and} \quad y(\tau) := \sqrt{e'(\tau)(1-e(\tau))}.
\]

This definition is now consistent with (22), after making a change of variables. We recall the variances of these Beta variables,

\[
\text{Var}(e_t^2) = \frac{\alpha}{n} \frac{(\frac{b}{n} + \frac{1}{n} - 1) (\frac{1}{n} + \frac{1}{n} - 1)}{\left(\frac{b}{n} + 2\frac{1}{n} - 2\right)^2} \left(\frac{1}{n} + \frac{2\frac{1}{n} - 2 + \frac{\alpha}{n}}{n}\right) = \frac{\alpha a}{n} \frac{b-a+a\frac{\alpha}{n}}{1-2a+2a\frac{\alpha}{n} + O(n^{-2})}
\]

\[
\text{Var}(e_t') = \frac{\alpha}{n} \frac{(\frac{b}{n} + \frac{1}{n} - 1) (\frac{1}{n} + \frac{1}{n} - 1)}{\left(\frac{b}{n} + 2\frac{1}{n} - 2\right)^2} \left(\frac{1}{n} + \frac{2\frac{1}{n} - 2 + \frac{\alpha}{n}}{n}\right) = \frac{\alpha a}{n} \frac{b-a+a\frac{\alpha}{n}}{1-2a+2a\frac{\alpha}{n} + O(n^{-2})},
\]

where we may choose the constants in the error terms to depend only on the ensemble parameters (and not \( t \)). By virtue of the \( \frac{\alpha}{n} \) factor, the sum \( \sum_{t=1}^{n} r_t^D \) takes the form of a Riemann sum. The integrand, exposed on the right hand side of (34), is Lipschitz continuous in \( \tau/n \), and thus the convergence of the Riemann sum to the integral occurs with rate \( O(n^{-1}) \). This shows

\[
\sum_{t=1}^{n} r_t^D = \frac{\alpha a}{4} \int_0^1 \frac{(b-a+2a\tau)(b-a+a\tau)}{(1-2a+2a\tau)^3} \left(\frac{x\partial_x p_k}{e(\tau)} - \frac{y\partial_y p_k}{1-e(\tau)}\right) \left(\frac{x\partial_x p_m}{e(\tau)} - \frac{y\partial_y p_m}{1-e(\tau)}\right) d\tau + O(n^{-1/2}).
\]

Applying the same reasoning to \( n + 1 \leq t \leq 2n - 1 \), it follows that

\[
\sum_{t=n+1}^{2n-1} r_t^D = \frac{\alpha a}{4} \int_0^1 \frac{(b-a+2a\tau)(b-a+a\tau)}{(1-2a+2a\tau)^3} \left(\frac{y\partial_y p_k}{e(\tau)} - \frac{x\partial_x p_k}{1-e(\tau)}\right) \left(\frac{y\partial_y p_m}{e(\tau)} - \frac{x\partial_x p_m}{1-e(\tau)}\right) d\tau + O(n^{-1/2}).
\]

The sum of these two integrals \([38]\) and \([39]\) and the associated error bounds show that the limiting covariance exists, and their sum provides an expression for the limit. The remainder of the proof will show that this expression can be alternately expressed in the form given by \( C_{k,l} \) (defined in (29)). The primary difference is a change of variables. \( \Delta = a(\tau - 1) \). The integrals become

\[
\sum_{t=1}^{n} r_t^D = \frac{\alpha a}{4} \int_{-a}^{0} \frac{e(\sigma)(1-e(\sigma))}{(1+2\sigma)} \left(\frac{x\partial_x p_k}{e(\sigma)} - \frac{y\partial_y p_k}{1-e(\sigma)}\right) \left(\frac{x\partial_x p_m}{e(\sigma)} - \frac{y\partial_y p_m}{1-e(\sigma)}\right) d\sigma + O(n^{-1/2})
\]

\[
\sum_{t=n+1}^{2n-1} r_t^D = \frac{\alpha a}{4} \int_{0}^{a} \frac{e'(\sigma)(1-e'(\sigma))}{(1+2\sigma)} \left(\frac{y\partial_y p_k}{e'(\sigma)} - \frac{x\partial_x p_k}{1-e'(\sigma)}\right) \left(\frac{y\partial_y p_m}{e'(\sigma)} - \frac{x\partial_x p_m}{1-e'(\sigma)}\right) d\sigma + O(n^{-1/2}).
\]

The sum of these integrals can be shown to equal \( C_{k,l} \) by checking the coefficients in front of the terms \( \partial_x p_k \partial_x p_m, \partial_y p_k \partial_y p_m, \partial_x p_k \partial_y p_m, \partial_y p_k \partial_x p_m \). The coefficient on \( \partial_x p_k \partial_x p_m \) in the sum of the integrands \([40]\) and \([41]\) is given by

\[
\frac{e(\sigma)(1-e(\sigma))}{(1+2\sigma)} \frac{x^2}{e(\sigma)^2} + \frac{e'(\sigma)(1-e'(\sigma))}{(1+2\sigma)} \frac{x^2}{1-e'(\sigma)^2} = \frac{(1-e(\sigma))(1-e'(\sigma)) + e(\sigma)e'(\sigma)}{1+2\sigma} = 1 - \frac{x^2 - y^2}{1+2\sigma}.
\]

Similar manipulations show that the coefficients on each of the other terms agree with the coefficients in the integrand of \( C_{k,l} \), completing the proof.

\[\square\]

3. Diagonalizing the Covariance Matrix

We proceed by showing that the covariances are diagonalized by the appropriate Chebyshev polynomial basis. This will be done by verifying that certain generating functions agree. We would like to show that the infinite covariance matrix can be decomposed as

\[C = LAL^T,\]

where \( L \) is the diagonal matrix of eigenvalues of \( C \) and \( A \) is the matrix of Chebyshev polynomials of the first kind. The proof involves constructing a basis of orthogonal polynomials that diagonalize \( C \) and then expressing the covariance matrix in terms of this basis. The approach is similar to the Gram-Schmidt process for constructing orthonormal polynomials.
for the diagonal matrix Λ = diag(0,1,2,3,4,…), and some lower triangular matrix L. The $L_{n,k}$ entry of this matrix is the coefficient of the $k^{th}$ Chebyshev polynomial $Γ_k(x)$ in the expansion of $x^n$. Define the exponential covariance generating function $G(s,t)$ as

\[ G(s,t) = \sum_{k,l>0} \frac{s^k t^l}{k! l!} C_{k,l}, \]

and define the exponential generating function of $LL^t$ analogously,

\[ T(s,t) = \sum_{k,l>0} \frac{s^k t^l}{k! l!} [LL^t]_{k,l}. \]

We will show that these generating functions are equal by computing their bivariate Laplace transforms and showing they are the same, from which it follows that $C = LL^t$.

3.0.1. Computing $L_s[t][T]$. The coefficients $L_{n,k}$ can be computed by a recursive formula, but they have a useful Fourier-like expansion. Define $θ$ in terms of $x$ so that

\[ \cos(θ) = \frac{2x - λ_+ - λ_-}{λ_+ - λ_-}, \]

from which it follows that

\[ 2 \cos(nθ) = 2T_n(\cos θ) = 2T_n \left( \frac{2x - λ_+ - λ_-}{λ_+ - λ_-} \right). \]

Expand $\frac{1}{2} e^{tx}$ as a series in $t$,

\[ \frac{1}{2} e^{tx} = \frac{1}{2} \sum_{n=0}^{∞} \frac{t^n}{n!} x^n = \frac{1}{2} \sum_{n=0}^{∞} \frac{t^n}{n!} \left( \frac{λ_+ + λ_-}{2} + \frac{λ_+ - λ_-}{2} \cos θ \right)^n = \frac{1}{2} \sum_{n=0}^{∞} \sum_{k=0}^{∞} L_{n,k} \frac{t^n}{n!} 2 \cos kθ = \sum_{n=0}^{∞} \sum_{k=0}^{∞} L_{n,k} \frac{t^n}{n!} \cos kθ, \]

where we have used the definition of $L_{n,k}$ as the coefficient of the $k^{th}$ Chebyshev polynomial in the expansion of $x^n$.

The Fourier interpretation allows for the matrix multiplication $LL^t$ to be carried out by an integral. Consider the kernel $K_N(θ, φ)$, which will formally play the role of $Λ$, given by

\[ K_N(θ, φ) = \sum_{k=1}^{N} k \cos kθ \cdot \cos kφ. \]

This allows for $T$ to be given by

\[ T(s,t) = \lim_{N→∞} \frac{1}{4π^2} \int_{0}^{π} \int_{0}^{π} e^{t \left( \frac{λ_+ + λ_-}{2} + \frac{λ_+ - λ_-}{2} \cos θ \right)} K_N(θ, φ) e^{s \left( \frac{λ_+ + λ_-}{2} + \frac{λ_+ - λ_-}{2} \cos θ \right)} dφ dθ, \]

as the coefficient on $t^k s^l$ would be

\[ \lim_{N→∞} \frac{1}{π^2} \int_{0}^{π} \int_{0}^{π} \left( \sum_{j=0}^{k} L_{k,j} \frac{t^k}{k!} \cos jθ \right) \left( \sum_{j=0}^{N} \sum_{k=1}^{l} k \cos kθ \cdot \cos kφ \right) \left( \sum_{j=0}^{∞} \sum_{l=0}^{∞} \frac{s^l}{l!} \cos jφ \right) dφ dθ, \]

which by the orthogonality of $\{ \cos jθ \}_{j=0}^{∞}$ on $[0, π]$, is exactly $(LL^t)_{k,l}$ when $N > \max(k,l)$. Further, these integrals can be evaluated, as the expression $e^z \cos θ$ has an expansion in terms of Bessel functions. Namely,

\[ e^z \cos θ = I_0(z) + 2 \sum_{k=1}^{∞} I_k(z) \cos kθ, \]
This defines the Fourier coefficients of $e^z \cos \theta$, from which it follows that $\mathcal{F}(s,t)$ can be rewritten as

$$
\mathcal{F}(s,t) = e^{c(t+s)} \lim_{N \to \infty} \frac{1}{4\pi^2} \int_0^\pi \int_0^\pi e^{rs \cos \phi} K_N(\theta, \phi) e^{r \sin \phi} \cos \phi \, d\theta d\phi
$$

$$= e^{c(t+s)} \sum_{k=1}^\infty kI_k(rt)I_k(rs),
$$

where $c = \left( \frac{\lambda_+ + \lambda_-}{2} \right)$ and $r = \left( \frac{\lambda_+ - \lambda_-}{2} \right)$.

Again, we will require the Laplace transform of this generating function. Each summand $kI_k(rt)I_k(rs)$ is positive for $s, t > 0$, and so commuting the sum and the Laplace transform is justified.

$$
\mathcal{L}_{s,t}[\mathcal{F}(s,t)](\eta, \omega) = \sum_{k=1}^\infty kL_{s,t} \left[ e^{c(t+s)}I_k(rt)I_k(rs) \right] (\eta, \omega)
$$

$$= \sum_{k=1}^\infty k \frac{1}{\sqrt{\eta^2 - rt - r^2}} \frac{1}{\sqrt{\eta^2 - r^2}} \frac{1}{\sqrt{\eta^2 - r^2}} e^{\frac{1}{2} \left( \frac{\eta^2}{\eta^2 - r^2} \right)},
$$

where $\tilde{\omega} = \omega - c$. This has the form for the series expansion of $(1 - x)^{-2}$. After simplifying, this expression is

$$
\mathcal{L}_{s,t}[\mathcal{F}(s,t)](\eta, \omega) = \frac{r^2}{\sqrt{\eta^2 - r^2} \sqrt{\eta^2 + r^2} \left[ \sqrt{(\tilde{\omega} + r)(\eta - r)} + \sqrt{(\tilde{\omega} - r)(\eta + r)} \right]^2}.
$$

3.0.2. Computing $\mathcal{L}_{s,t}[^C]$. We will now turn to computing the Laplace transform of $^C$. The integrand of $C_{k,l}$ is not positive, but it can be split into two integrals whose integrands are positive (see Remark 2.17)

$$
C_{k,l} = L_{k,l} - R_{k,l},
$$

with

$$
L_{k,l} = \frac{\alpha}{4} \int_{-a}^0 \frac{1}{1 + 2\sigma} \left[ (\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m) \right] \, d\sigma,
$$

and

$$
R_{k,l} = \frac{\alpha}{4} \int_{-a}^0 \frac{1}{1 + 2\sigma} \left[ (\partial_x p_k \partial_x p_m + \partial_y p_k \partial_y p_m) (x^2 + y^2) + (\partial_x p_k \partial_y p_m + \partial_y p_k \partial_x p_m) (2xy) \right] \, d\sigma.
$$

As $p_i(x,y)$ has all positive coefficients, and both $x$ and $y$ are positive on the domain of integration, each of these integrands is positive. Defining generating functions for each array,

$$
\mathcal{L}(s,t) = \sum_{k,l>0} \frac{s^k t^l}{k! l!} L_{k,l} \text{ and } \mathcal{R}(s,t) = \sum_{k,l>0} \frac{s^k t^l}{k! l!} R_{k,l},
$$

we can write

$$
\mathcal{L}(s,t) = \frac{\alpha}{4} \int_{-a}^0 \frac{1}{1 + 2\sigma} \left[ (\partial_x \mathcal{P}(s) \partial_x \mathcal{P}(t) + \partial_y \mathcal{P}(s) \partial_y \mathcal{P}(t)) \right] \, d\sigma, \quad \text{and}
$$

$$
\mathcal{R}(s,t) = \frac{\alpha}{4} \int_{-a}^0 \frac{1}{1 + 2\sigma} \left[ (\partial_x \mathcal{P}(s) \partial_x \mathcal{P}(t) + \partial_y \mathcal{P}(s) \partial_y \mathcal{P}(t)) (x^2 + y^2) + (\partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t) + \partial_y \mathcal{P}(s) \partial_x \mathcal{P}(t)) (2xy) \right] \, d\sigma,
$$

where we have commuted sum and integral by the positivity of the integrands. Recall that $\mathcal{P}(t) = \mathcal{P}(x, y, t)$ is the exponential generating function for the polynomials $p_k(x, y)$, and from (10), it is jointly analytic in all variables. As $-a > -1/2$, it follows that the integrands are continuous for all $-a \leq \sigma \leq 0$, and all $s, t$. In particular, each of $\mathcal{L}$ and $\mathcal{R}$ is finite for all $s, t$, and it follows that we can write $^C(s,t)$ as the sum of these two functions, so

$$
^C(s,t) = \mathcal{L}(s,t) - \mathcal{R}(s,t).
$$
The joint Laplace transforms in $s$ and $t$ will be computed for both of these expressions. This makes heavy use of Lemma 2.7. Additionally, it requires that the order of integration be switched, which requires an argument. We prove a simplified statement, by whose method it is easily seen that these integrals can be exchanged.

**Lemma 3.1.** Suppose that $\omega > \lambda_+$ and that $\eta > \lambda_+$, then

\[
\int_{s,t>0} \int_{-\alpha}^{0} e^{-\omega t - \eta s} \frac{\partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t)}{1 + 2\sigma} d\sigma ds dt = \int_{-\alpha}^{0} \int_{s,t>0} \frac{e^{-\omega t - \eta s}}{1 + 2\sigma} \partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t) d\sigma ds dt,
\]

and each is finite.

**Proof.** We begin by maximizing $x + y$ over $\sigma \in [-a, 0]$, where it is seen that the maximum is attained at $\sigma = 0$, at which point,

\[(x(0) + y(0))^2 = \left(\sqrt{b(1-a)} + \sqrt{a(1-b)}\right)^2 = \lambda_+.
\]

Thus, it follows that $(x + y)^2 < \lambda_+ < \omega$ for all $-a \leq \sigma \leq 0$. Recall that $\mathcal{P}(t)$ is given by $e^{t(s^2+y^2)} I_0(2xtyt)$, and thus

\[
\partial_x \mathcal{P}(t) = 2xte^{t(s^2+y^2)} I_0(2xtyt) + 2yte^{t(s^2+y^2)} I_1(2xtyt).
\]

Using that $0 \leq I_n(2xtyt) \leq e^{2xyt}$ for all $n$, it follows that

\[0 \leq \partial_x \mathcal{P}(t) \leq 2(x + y)e^{t(x+y)^2},
\]

for $x, y \geq 0$. It follows that there is a constant $C$ so that for all $-a \leq \sigma \leq 0$,

\[0 \leq \frac{e^{-\omega t - \eta s}}{1 + 2\sigma} \partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t) < C(x + y)^2 e^{-(\omega - (x+y)^2)t - (\eta - (x+y)^2)s}.
\]

Using the bound on $x + y$ derived above,

\[0 \leq \frac{e^{-\omega t - \eta s}}{1 + 2\sigma} \partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t) < C\lambda_+^2 e^{-(\omega - \lambda_+ t - (\eta - \lambda_+) s)}.
\]

Thus, provided that $\omega > \lambda_+$ and $\eta > \lambda_+$, the order of integration may be reversed by Fubini. $\Box$

We can now compute the bivariate Laplace transform of $\mathcal{G}(s,t)$.

**Lemma 3.2.**

\[L_{s,t}[\mathcal{G}(s,t)](\eta, \omega) = \frac{\alpha}{8} \int_{1-2\alpha}^{1} \frac{n_1 - n_2 \rho}{(p_1 - p_2 \rho)^2 (r_1 - r_2 \rho)^2} d\rho,
\]

where these parameters are given by

\[
n_1 = -\omega \eta [(1 - \lambda_- - \lambda_+)^2] - (\omega + \eta) \lambda_- \lambda_+ (1 - \lambda_- - \lambda_+) - (1 - \lambda_- - \lambda_+ + 2\lambda_- \lambda_+) \lambda_- \lambda_+,
\]

\[
n_2 = -\omega \eta [1 - \lambda_- - \lambda_+ + 2\lambda_- \lambda_+] + (\omega + \eta - 1) \lambda_- \lambda_+, \quad p_1 = \omega (1 - \lambda_- - \lambda_+) + \lambda_- \lambda_+,
\]

\[
p_2 = \omega - \omega^2, \quad r_1 = \eta (1 - \lambda_- - \lambda_+) + \lambda_- \lambda_+, \quad r_2 = \eta - \eta^2.
\]

**Proof.** We start by commuting the integration in $\sigma$ and the Laplace transform in $[13]$. To evaluate these Laplace transforms, we recall Lemma 2.7 where the Laplace transform $L_{s,t}[\partial_x \mathcal{P}(t)]$ was computed to be

\[L_{s,t}[\partial_x \mathcal{P}(t)](\omega) = \frac{2x(\omega + y^2 - x^2)}{(\omega - x^2 - y^2)^2 - 4x^2 y^2} \frac{1}{2}, \quad \omega > (x + y)^2.
\]

The quantity $x^2 - y^2$ simplifies to

\[x^2 - y^2 = \frac{b - a}{1 + 2\sigma}.
\]
The Laplace transform of \( \partial_x \mathcal{P}(t) \) can be rewritten as
\[
L_t[\partial_x \mathcal{P}(t)](\omega) = \frac{2x(1+2\sigma^2)(1+\sigma)(b-a)}{(\omega^2-\omega)(1+2\sigma^2) - (1-2a)(1-2b)\omega + (b-a)^2} \quad \omega > \lambda_+,
\]
for \( \sigma \in [-a,0] \). By symmetry, the Laplace transform \( \partial_y \mathcal{P}(t) \) is
\[
L_t[\partial_y \mathcal{P}(t)](\omega) = \frac{2y(1+2\sigma^2)(1+\sigma)(b-a)}{(\omega^2-\omega)(1+2\sigma^2) - (1-2a)(1-2b)\omega + (b-a)^2} \quad \omega > \lambda_+.
\]
Define \( \Delta(\omega) \) to be
\[
\Delta(\omega) = (\omega^2-\omega)(1+2\sigma^2) - (1-2a)(1-2b)\omega + (b-a)^2,
\]
and define \( \rho = (1+2\sigma^2) \). We will now split the computation of \( \mathcal{C}(s,t) \) into two pieces for simplicity’s sake.

The first piece is
\[
L_{s,t} [\partial_x \mathcal{P}(s) \partial_x \mathcal{P}(t) + \partial_y \mathcal{P}(s) \partial_y \mathcal{P}(t)](\eta,\omega)
= 4\rho^2 \frac{(x^2+y^2)(\omega \eta + (b-a)^2) - (b-a)^2(\omega + \eta)}{\sqrt{\Delta(\omega)} \Delta(\eta)^{3/2}}.
\]

The second piece is
\[
L_{s,t} [\partial_x \mathcal{P}(s) \partial_y \mathcal{P}(t) + \partial_y \mathcal{P}(s) \partial_x \mathcal{P}(t)](\eta,\omega)
= 8xy\rho^2 \frac{[(x^2+y^2)(1-x^2-y^2) - 4x^2y^2)\omega \eta + ((x^2+y^2)(1-x^2-y^2) + 4x^2y^2)(b-a)^2]}{\sqrt{\Delta(\omega)} \Delta(\eta)^{3/2}}.
\]

Combining these two pieces,
\[
L_{s,t}[\mathcal{C}](\eta,\omega) = \alpha \int_{-a}^{b} \rho^2 \left[ ((x^2+y^2)(1-x^2-y^2) - 4x^2y^2)\omega \eta + ((x^2+y^2)(1-x^2-y^2) + 4x^2y^2)(b-a)^2 \right] \frac{1}{\sqrt{\Delta(\omega)} \Delta(\eta)^{3/2}} d\sigma.
\]

We simplify some of these expressions,
\[
(x^2+y^2)(1-x^2-y^2) - 4x^2y^2 = \frac{1}{2} \rho^2 [-4a(1-a) + 1 - 4b(1-b) + 1] + \frac{1}{2} \rho^{-1} [1 - 2b(1-b) - 2a(1-a)],
\]
\[
(x^2+y^2)(1-x^2-y^2) + 4x^2y^2 = \frac{1}{2} \rho^{-1} [\rho + 1 - 2b(1-b) + 2a(1-a)],
\]
\[
1-x^2-y^2 = \frac{1}{2} \rho^{-1} [\rho + 1 - 2a(1-2b)].
\]

After changing the integration to be over \( \rho \), we produce the desired formula. \( \square \)

We will explicitly evaluate the integral in Lemma 3.2 to conclude that

Lemma 3.3.
\[
L_{s,t}[\mathcal{C}(s,t)] = \alpha \frac{d^2}{\sqrt{\omega+d}(\eta-d) + \sqrt{\omega-d}(\eta+d)} \left[ \sqrt{\omega^2-d^2} \sqrt{\eta-d^2} \right],
\]
where \( d = \left( \frac{\lambda_+ + \lambda_-}{2} \right), \omega = \omega - \left( \frac{\lambda_+ + \lambda_-}{2} \right) \) and \( \eta = \eta - \left( \frac{\lambda_+ + \lambda_-}{2} \right) \).

By comparing with the expression for \( L_{s,t}[\mathcal{F}(s,t)] \) derived in (42), this lemma completes the proof of the diagonalization of the covariances.

24
Proof. Differentiating both sides, it can be shown that
\[
\int \frac{n_1 - n_2 \rho}{(p_1 - p_2) \frac{3}{2} (r_1 - r_2)^{\frac{3}{2}}} \, d\rho = \frac{-2}{(p_1 r_2 - p_2 r_1)^2} \frac{2n_2 p_1 r_1 - n_1 (p_1 r_2 + p_2 r_1) + \rho (2n_1 p_2 r_2 - n_2 (p_1 r_2 + p_2 r_1))}{(p_1 - p_2) \frac{5}{2} (r_1 - r_2)^{\frac{5}{2}}}.
\]
The indefinite integral can be greatly simplified, plugging in some of the \(n, p, r\) terms.
\[
\int \frac{n_1 - n_2 \rho}{(p_1 - p_2) \frac{3}{2} (r_1 - r_2)^{\frac{3}{2}}} \, d\rho = 2\frac{(1 - \lambda_--\lambda_+)(\eta + \omega) + 2\lambda_--\lambda_+ - 2\rho(\omega + \eta - 2\omega\eta)}{(\eta - \omega)^2 (p_1 - p_2) \frac{7}{2} (r_1 - r_2)^{\frac{7}{2}}}.
\]
The antiderivative will now be evaluated at both endpoints. At \(\rho = 1\), it becomes
\[
2\frac{2\omega\eta - (\omega + \eta)(\lambda_--\lambda_+) + 2\lambda_--\lambda_+}{(\eta - \omega)^2 \sqrt{(\omega - \lambda_-)(\omega - \lambda_+)\sqrt{(\eta - \lambda_-)(\eta - \lambda_+)}}}.
\]
To evaluate at \(\rho = (1 - 2a)^2\), it is helpful to work with \(a\) and \(b\) instead of \(\lambda_{\pm}\). Using the formulae
\[
\lambda_-\lambda_+ = (b-a)^2 \text{ and } \lambda_- + \lambda_+ = 2(a + b - 2ab),
\]
the antiderivative evaluated at \(\rho = (1 - 2a)^2\) is simply
\[
\frac{4}{(\eta - \omega)^2}.
\]
At last we can give a single expression for the Laplace transform of the covariance function:
\[
\mathbf{L}_{s,t}[\mathcal{E}(s,t)] = \alpha \left[ \sqrt{(\omega - \lambda_-)(\eta - \lambda_+)} - \sqrt{(\omega - \lambda_+)(\eta - \lambda_-)} \right]^2.
\]
Recall that \(r = \left( \frac{\lambda_++\lambda_-}{2} \right), \tilde{\omega} = \omega - \left( \frac{\lambda_++\lambda_-}{2} \right)\) and \(\tilde{\eta} = \eta - \left( \frac{\lambda_++\lambda_-}{2} \right)\). We rewrite this expression in terms of these modified parameters to get
\[
\mathbf{L}_{s,t}[\mathcal{E}(s,t)] = \alpha \frac{d^2}{\sqrt{(\tilde{\omega} + d)(\tilde{\eta} - d) + \sqrt{(\tilde{\omega} - d)(\tilde{\eta} + d)}}^2} \sqrt{\tilde{\omega}^2 - d^2} \sqrt{\tilde{\eta} - d^2} = \alpha \mathbf{L}_{s,t}[\mathcal{E}(s,t)].
\]

4. Extension to Continuously Differentiable Test Functions

We learned the idea for the extending the CLT from the appendix of Anderson-Zeitouni [2]. Roughly speaking, one would like to extend a CLT for polynomial test functions to a CLT for a larger class of functions, the hope being to invoke the density of the polynomials. However, it needs to be assured that error-in-approximation produces small error in the fluctuations when evaluated on the empirical process. The property of a matrix ensemble that allows one to execute this is a type of global concentration of eigenvalues. See also Proposition 11.6 in [2] and Lemma 1 of [41] for related approaches.

**Proposition 4.1.** Let \(\{A_n\}\) be an ensemble of matrices with compact spectral support \(S\), and let \(V : C^1(S) \rightarrow \mathbb{R}\) be a positive semidefinite quadratic form for which there is constant \(C_1\) so that \(V(f) \leq C_1^2 \|f\|^2_{L^2}\) for all \(f \in C^1(S)\). Suppose that \(\{A_N\}\) satisfies a polynomial-type CLT, i.e., for all polynomials \(g\),
\[
\text{tr } g(A_n) - \mathbb{E} \text{tr } g(A_n) \Rightarrow N(0, V(g))
\]
and additionally \(\text{Var } \text{tr } g(A_n) \rightarrow V(g)\). If the ensemble satisfies a Poincaré type concentration inequality, i.e.,
\[
\text{Var } (\text{tr } f(A_n)) \leq C_2^2 \|f\|^2_{L^2}.
\]
for some constant \(C_2\) independent of \(n\) and any Lipschitz \(f\) on \(S\), then the polynomial CLT extends to all \(C^1\) functions \(f : S \rightarrow \mathbb{R}\), as
\[
\text{tr } f(A_n) - \mathbb{E} \text{tr } f(A_n) \Rightarrow N(0, V(f)).
\]
**Proof.** We recall the quadratic Wasserstein metric

\[ W_2(\mu, \nu)^2 = \inf \mathbb{E} (X - Y)^2, \]

with the infimum over all couplings \((X, Y)\) with marginals \(\mu\) and \(\nu\) respectively. For a random variable \(X\), we let \(\mathcal{L}X\) denote its law. It is well known that \(W_2(\mathcal{L}X_n, \mathcal{L}X) \to 0\) if and only if \(X_n \Rightarrow X\) and \(\mathbb{E}X_n^2 \to \mathbb{E}X^2\) (see Theorem 7.12 of [50]). For any \(f \in C^1(S)\), let \(Z_f\) denote a centered normal random variable with variance \(V(f)\). Thus for any polynomial \(g\), \(W_2(\mathcal{L}(g(A_n) - \mathbb{E} g(A_n)), \mathcal{L}Z_g) \to 0\).

Let \(f\) be any \(C^1(S)\) function. By Weierstrass approximation of the derivative of \(f\), there is a sequence of polynomials \(p_k\) so that \(\|f - p_k\|_{\text{Lip}} \to 0\) as \(k \to \infty\). It follows that \(V(p_k) \to V(f)\) from its continuity with respect to the Lipschitz seminorm, and hence that \(W_2(\mathcal{L}Z_p, \mathcal{L}Z_f) \to 0\) as \(k \to \infty\). For any \(k\) we can bound,

\[
W_2(\mathcal{L}(\text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ f(A_n)), \mathcal{L}Z_f) \leq \frac{W_2(\mathcal{L}(\text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ f(A_n)), \mathcal{L}(\text{tr}\ p_k(A_n) - \mathbb{E} \text{tr}\ p_k(A_n)))}{W_2(\mathcal{L}(\text{tr}\ p_k(A_n) - \mathbb{E} \text{tr}\ p_k(A_n)), \mathcal{L}Z_{p_k})}.
\]

By the concentration inequality, it is possible to bound

\[
\mathbb{E} |\text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ p_k(A_n) - \mathbb{E} \text{tr}\ p_k(A_n)|^2 \leq C_1^2 \|f\|_{\text{Lip}}^2,
\]

from which it follows that \(W_2(\mathcal{L}(\text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ f(A_n)), \mathcal{L}(\text{tr}\ p_k(A_n) - \mathbb{E} \text{tr}\ p_k(A_n))) \leq C_1 \|f\|_{\text{Lip}}\) by the definition of the Wasserstein metric as the infimum over couplings. Likewise

\[
W_2(\mathcal{L}Z_p, \mathcal{L}Z_f) = \sqrt{V(p_k) - \mathbb{E} (V(f))} \leq \sqrt{V(p_k - f)} \leq C_2 \|f\|_{\text{Lip}}.
\]

Therefore, from the polynomial CLT,

\[
\limsup_{n \to \infty} W_2(\mathcal{L}(\text{tr}\ f(A_n) - \mathbb{E} \text{tr}\ f(A_n)), \mathcal{L}Z_f) \leq (C_1 + C_2) \|f\|_{\text{Lip}}.
\]

Taking \(k \to \infty\) completes the proof. \(\square\)

Note that the moment-method proof used for the polynomial CLT implies \(\text{Var} \text{tr}(g(A_n)) \to V(g)\), and that the bound of \(V(f) \leq C \|f\|_{\text{Lip}}\) follows from Remark 4.2. To show that linear statistics of the Jacobi ensemble satisfy a Poincaré inequality, we will work directly with the joint eigenvalue density function. Recall [8], which stated

\[
d\mu_f(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z} \prod_i \lambda_i^\left(\frac{k}{2} - 1\right)^n \left(1 - \lambda_i^2\right)^{-\frac{k}{2} - 1} \prod_{i<j} \left|\lambda_i - \lambda_j\right|^2 d\mu_f(\lambda_1, \ldots, \lambda_n).
\]

We first show that the Jacobi ensemble satisfies a log-Sobolev inequality, which is strictly stronger than the Poincaré inequality. Define the entropy of a non-negative measurable function \(f\) with respect to a probability measure \(\mu\) by

\[
\text{Ent}_\mu(f) := \int f \log f d\mu - \left(\int f d\mu\right) \left(\int \log f d\mu\right),
\]

if \(f \log(1 + f)d\mu < \infty\) and +\(\infty\) otherwise. Our tool in this direction is a consequence of the well-known Bakry-Emery condition, the content of which is contained in the following proposition (see Proposition 3.1 of [2]).

**Proposition 4.2.** Suppose that \(d\mu = e^{-U}dx\) is supported on a convex set \(\Omega\). If there is a \(c > 0\) so that for all \(x \in \text{int}(\Omega)\), \(\text{Hess} U(x) \geq c I_d\), where \(I_d\) is the identity matrix and \(\geq\) is the partial ordering on positive semidefinite matrices, then for all smooth functions \(f\) on \(\mathbb{R}^n\),

\[
\text{Ent}_\mu(f^2) \leq \frac{2}{c} \int |\nabla f|^2 d\mu.
\]

To prove the log-Sobolev inequality with the appropriate constant, we need only check that the condition of Proposition 4.2 is satisfied. This we do in showing the following lemma.
Lemma 4.3. The Jacobi ensemble satisfies a log-Sobolev inequality

\[ \text{Ent}_{\mu_0}(f^2) \leq \frac{2}{c} \int |\nabla f|^2 \, d\mu, \]

with \( c = 4\alpha \min \{ \frac{b}{a} - 1, \frac{1-b}{a} - 1 \} \).

Proof. We will employ Proposition 4.2 and thus we begin by computing the Hessian of the logarithm of the density. Let \( p := \frac{\alpha}{\alpha - 1} + \frac{1}{\alpha - 1} \), and let \( q := \frac{\beta}{\alpha - 1} + \frac{1}{\alpha - 1} \). The first derivative is given by

\[ \frac{d}{d\lambda_i} (\log(d\mu_j)) = \frac{p}{\lambda_i} - \frac{q}{1 - \lambda_i} + \frac{1}{\alpha} \sum_{j \neq i} \frac{1}{|\lambda_i - \lambda_j|}. \]

The second derivative is thus

\[ \frac{d^2}{d\lambda_i^2} (\log(d\mu_j)) = -\frac{p}{\lambda_i^2} - \frac{q}{(1 - \lambda_i)^2} - \frac{1}{\alpha} \sum_{j \neq i} \frac{1}{(\lambda_i - \lambda_j)^2}. \]

The mixed partials are just

\[ \frac{d}{d\lambda_j} \frac{d}{d\lambda_i} (\log(d\mu_j)) = -\frac{1}{\alpha} \frac{1}{(\lambda_i - \lambda_j)^2}. \]

By the method of Gershgorin discs we conclude that the smallest eigenvalue of \( \text{Hess}(-\log d\mu_j) \) is at least

\[ \min_{1 \leq i \leq n} \left[ \frac{p}{\lambda_i^2} + \frac{q}{(1 - \lambda_i)^2} \right] \geq 4 \min\{p, q\} \geq \frac{4n}{\alpha} \min \{ \frac{b}{a} - 1, \frac{1-b}{a} - 1 \}. \]

It is now a simpler manner to show the needed concentration inequality and prove Theorem 1.3.

Proof of Theorem 1.3. From Proposition 2.9 and Proposition 4.1, it suffices to demonstrate a constant \( C \) so that \( \text{Var} \, \text{tr} \, f \leq C \|f\|_{L^p}^2 \), with the Lipschitz norm on \([0, 1]\), for all Lipschitz \( f \). This is turn follows from the somewhat sharper inequality that

\[ \text{Var} \, \text{tr} \, f \leq C \int |\partial_{\lambda_i}(f(\lambda_i))|^2 \, d\mu_j(\lambda_1, \ldots, \lambda_n) = \frac{C}{n} \int |\nabla \text{tr} \, f|^2 \, d\mu_j(\lambda_1, \ldots, \lambda_n), \]

where in the last step we have used the symmetry of the linear statistic. It is a standard fact that the log-Sobolev inequality implies the Poincaré inequality with half the constant (see [31] Chapter 5). Thus by Lemma 4.3 we have that for all smooth functions \( f \),

\[ \text{Var} \, \text{tr} \, f \leq \frac{4n}{\alpha} \min \{ \frac{b}{a} - 1, \frac{1-b}{a} - 1 \} \int |\nabla \text{tr} \, f|^2 \, d\mu_j(\lambda_1, \ldots, \lambda_n). \]

Extension to Lipschitz functions follows from the density of smooth functions in \( L^2 \), and the proof is complete.

\[ \Box \]

5. Computing the Expectation

In this section, we will prove Theorem 1.6. To establish the theorem for polynomial linear statistics \( \phi \), a proof will be given that follows a similar tract to the analogous statement proven for the Laguerre and Hermite ensembles in [17]. The key to this method of proof is establishing a certain palindromy. Recall that a polynomial \( p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 \) is palindromic in \( z \) if \( a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n \), or equivalently that \( p(z) = z^n p(z^{-1}) \).

Theorem 5.1. The scaled moment \( \frac{1}{n} \text{tr}(A^k) \) has a series expansion

\[ \frac{1}{n} \text{tr}(A^k) = \sum_{j=0}^{\infty} \eta_k(j, \alpha) n^{-j} \]

whose coefficients \( \eta_k(j, \alpha) \) are palindromic polynomials in \( (-\alpha) \) of degree \( j \).
While the proof of this palindromy works for all of these coefficients \( \eta \) simultaneously, only the palindromy of \( \eta_k(0, \alpha) \) and \( \eta_k(1, \alpha) \) are required for Theorem 1.6. Especially, palindromy forces \( \eta_k(0, \alpha) \) to have no \( \alpha \) dependence, and it forces \( \eta_k(1, \alpha) \) to be a multiple of \( 1 - \alpha \). As will be seen, this allows the \( \alpha = 0 \) case to be used to study the arbitrary \( \alpha \) case. As the proof of Theorem 5.1 requires symmetric function theory, we delay the proof to Appendix A to allow a brief introduction to the relevant symmetric function theory.

**Proof of Theorem 1.6 for polynomial \( \phi \).** Formally, let \( \hat{m}(x) \) be the moment generating function for the ensemble, and expand each moment asymptotically around \( n = \infty \), i.e.

\[
\hat{m}(x) = \frac{1}{n} \sum_{k=0}^{\infty} \mathbb{E} \left( \frac{\text{tr}(A^k)}{x^k} \right) = \sum_{k=0}^{\infty} x^{-k} \sum_{j=0}^{\infty} \eta_k(j, \alpha)n^{-j},
\]

then one has, to order \( \frac{1}{n} \),

\[
\hat{m}(x) = \sum_{k=0}^{\infty} x^{-k} \left( \eta_k(0, \alpha) + \eta_k(1, \alpha) \frac{1}{n} \right) + O(n^{-2}).
\]

The \( \alpha \)-dependence of either of these terms is completely determined by Theorem 5.1 as \( \eta_k(0, \alpha) \) can have no \( \alpha \) dependence, and \( \eta_k(1, \alpha) \) is a multiple of \( (1 - \alpha) \). Define \( m_0(x) \) and \( m_1(x) \) so that

\[
\hat{m}(x) \bigg|_{\alpha=0} = m_0(x) + \frac{1}{n} m_1(x) + O(n^{-2}).
\]

In this notation, the palindromy shows that

\[
\hat{m}(x) = m_0(x) + (1 - \alpha) \frac{1}{n} m_1(x) + O(n^{-2}).
\]

Further, the \( \alpha = 0 \) case, for fixed \( n \), is relatively simple. As observed by Sutton [46], the Jacobi matrix model tends to a deterministic one as \( \alpha \to 0 \); precisely, it has eigenvalues that are the roots of \( J_n^{r,s} \), the Jacobi polynomial of degree \( n \) and parameters

\[
r = n (\frac{b}{a} - 1), \quad s = n (\frac{1-b}{a} - 1).
\]

Suppose that the roots of \( J_n^{r,s} \) are given by \( \{\lambda_i\}_{i=1}^n \). Then for \( \alpha = 0 \), the moment generating function takes on the form

\[
\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{\infty} \sum_{k=0}^n \lambda_i^k x^k = \frac{1}{n} \sum_{i=1}^n \frac{1}{x - \lambda_i} = \frac{1}{n} \left( \ln J_n^{r,s}(x) \right)' = \frac{1}{n} J_n^{r,s}(x).
\]

Using the differential recurrence for Jacobi polynomials, it follows that \( \hat{m}(x) \) satisfies a formal power series equation

\[
(45) \quad \hat{m}^2 + \frac{r+1}{n} x \frac{J_n^{r+2}(x)}{J_n^{r}(x)} \hat{m} + \frac{1}{n} \frac{J_n^{r+1}(x)}{x} \hat{m}' = 0.
\]

It follows that the constant-order term \( m_0 \) satisfies

\[
am_0^2 + \frac{b-a-(1-2a)x}{x(1-x)} m_0 + \frac{1-a}{x(1-x)} = 0.
\]

This leads to an explicit form for \( m_0 \),

\[
m_0 = \frac{(a-b) + (1-2a)x - \sqrt{(b-a-(1-2a)x)^2 - 4a(1-a)x(1-x)}}{2ax(1-x)}
\]

\[
= \frac{(a-b) + (1-2a)x - \sqrt{(x - \lambda_-)(x - \lambda_+)}}{2ax(1-x)},
\]

where

\[
\lambda_{\pm} = \left[ \sqrt{b(1-a)} \pm \sqrt{a(1-b)} \right]^2.
\]

Note that \( \lambda_{\pm} \) are always real, and that they are always on \([0,1]\). They are 0 and 1 exactly when \( a = b \) and when \( a = 1 - b \), respectively. Taking an inverse Stieltjes transform gives absolutely continuous part

\[
d\mu(x) = \frac{\sqrt{-(x - \lambda_-)(x - \lambda_+)} \, x}{2\pi ax(1-x)} 1[\lambda_- \lambda_+].
\]
This integrates to 1, as it can be shown that
\[
\int_{\lambda}^{\lambda+} \frac{\sqrt{-x - \lambda_+}(x - \lambda)}{x(1 - x)} = \pi \left[1 - \sqrt{\lambda_+ + \sqrt{(1 - \lambda)(1 - \lambda)}}\right] = 2\pi a.
\]
Note that this implies that the distribution has no discrete part.

In the same fashion, one can also derive an explicit form for \(m_1\). Pulling out the \(\frac{1}{\pi}\) terms from (45), one is left with
\[
2a m_0 + \frac{(b - a) - (1 - 2a)x}{x(1 - x)} m_1 + \frac{1 - 2x}{x(1 - x)} a m_0 + \frac{a}{x(1 - x)} + a m' = 0.
\]
Solving for \(m_1\),
\[
m_1 = \frac{-x + \frac{1}{2}(\lambda_+ + \lambda) \pm \sqrt{(x - \lambda_+)(x - \lambda_-)}}{2(x - \lambda_+)(x - \lambda_-)}.
\]
To recover the density, one again applies the inverse Stieltjes transform. When \(x\) is neither \(\lambda_+\) nor \(\lambda_-\), the limit \(\lim_{\epsilon \to 0} m_1(x + i\epsilon)\) exists, and
\[
\lim_{\epsilon \to 0} m_1(x + i\epsilon) = -\frac{1}{2\pi \sqrt{-(x - \lambda_+)(x - \lambda_-)}} \mathbf{1}_{(\lambda_-, \lambda_+)}(x).
\]
Computing the inverse Stieltjes transform at either of the poles, it is seen that there are point masses, so that the entire signed measure is
\[
\nu(x) = \frac{1}{4} \delta_{\lambda_+}(x) + \frac{1}{4} \delta_{\lambda_-}(x) - \frac{1}{2\pi \sqrt{-(x - \lambda_+)(x - \lambda_-)}} \mathbf{1}_{(\lambda_-, \lambda_+)}(x).
\]

6. NUMERICS FOR THE EXTREMAL CASE

In this section, we investigate the choice \(p = q = 1\), which was not covered by Theorem 1.3. The method of proof breaks down in this extreme case, and so we have run a numerical simulation to help conjecture if the theorem extends.

![Experimental data for different values of \(\beta\), with \(n = 5000\) and \(50000\) samples of each. All experiments were run in Matlab R2010B, using the Edelman-Sutton matrix model.](image-url)
In the alternate parameterization we have that \( a = \frac{1}{2} \) and \( b = \frac{1}{2} \). The density of the Jacobi ensemble becomes

\[
d\mu_J(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z} \prod_i |\lambda_i^{\beta} - 1| (1 - \lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|^\beta.
\]

Note that the constraining potential no longer carries any dependence on \( n \). However, because the particles are forced to lie on \([0, 1]\) (physically speaking, they are trapped in an infinite potential well), it is likely that we have some limiting behavior. For polynomial test functions and \( \beta = 2 \), this case is covered by a theorem of Johansson (see Theorem 3.1 of [24]). However, the method of proof used here breaks down in the case \( a < \frac{1}{2} \), as it requires the entries of the sparse matrix model to have uniform variance estimates on the order of \( n^{-1} \). When \( a = \frac{1}{2} \), the matrix model entries are

\[
c_i \sim \sqrt{\text{Beta}(\frac{i}{2}, \frac{\beta}{2})} \quad \text{and} \quad c'_i \sim \sqrt{\text{Beta}(\frac{i}{2}, \frac{\beta}{2}(i+1))}.
\]

The variances of entries \( c_i \) and \( s_i \) are on the order of \( i^{-1} \), for which reason many of the arguments in later sections are no longer valid. To see how different the \( a = b = \frac{1}{2} \) case is from the \( a < \frac{1}{2} \) case, consider taking \( f(x) = x \). It is easily seen that

\[
X_{x^A, A} \to \sum_{i=1}^{\infty} (c_i^2 - \frac{1}{\beta})(1 - (c'_i)^2 - (c'_{i-1})^2),
\]

with the convergence in \( L^2 \). Note that while a normal limit is expected if the summands are becoming infinitesimal (and this is what happens when \( a < \frac{1}{2} \)), the normal limit here must follow from something else; in particular, the staircase dependency structure of the variable \( s \) can not be ignored. We invite the reader to check that the variable is symmetric and to note how much cancellation occurs in computing the second and fourth moments (they are \( 1/(8\beta) \) and \( 3/(64\beta^2) \) respectively). Again, the fact that this variable is normally distributed follows from the mentioned theorem of Johansson.

Appendices

A. Symmetric Functions

To find the asymptotic distribution of the traces, we will appeal to Kadell’s integral formula [27]. This formula makes use of Jack functions, and so we will provide a skeletal introduction to the relevant portions of symmetric function theory. A more expansive treatment is available in Macdonald’s book [33], whose notation we will follow.

By a partition \( \lambda \), we mean a non-increasing sequence of positive integers. The notation \( \lambda \vdash n \), read ‘\( \lambda \) partitions \( n \),’ means that the sum of the parts of \( \lambda \) equal \( n \). There is an important pictorial representation of a partition called a Young diagram. The diagram representation of a partition \( (\lambda_1, \ldots, \lambda_n) \) is drawn by placing \( \lambda_1 \) boxes horizontally in a row, placing \( \lambda_2 \) boxes horizontally below that, continuing through \( n \) and left justifying each row. Having drawn a diagram representation, we can easily define the conjugate \( \lambda' \) to be that partition represented by reflecting the diagram across the vertical axis and rotating counterclockwise by a quarter turn.

Example A.1. The partition \( \lambda = (5, 4, 1) \) is to the left, and its conjugate \( \lambda' = (3, 3, 2, 2, 1) \) is to the right.

\[\text{Example A.1. This is also called the transpose.}\]
Many formulas in symmetric function theory have sums or products computed from statistics of the diagram representation. For our purposes, we will need the arm length \( a \), arm co-length \( a' \), leg length \( l \), and leg co-length \( l' \) of a box \( s \). The statistics \( a(s) \) and \( a'(s) \) are the number of boxes to the right and to the left of box \( s \), respectively. Likewise, the statistics \( l(s) \) and \( l'(s) \) are the number of boxes below and above box \( s \).

Example A.2.

This is \( \lambda = (6, 5, 5) \).

\[
\begin{array}{cccccccccc}
\hline
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
& & & & & & & & & & \\
\hline
\end{array}
\]

\( a(s) = 1 \)

\( a'(s) = 3 \)

\( l(s) = 0 \)

\( l'(s) = 2 \)

The ring of symmetric functions \( \Lambda \) are all those formal power series with complex coefficients in the indeterminates \( \{x_1, x_2, \ldots\} \), that are symmetric under permutation of the indices. In this application, the symmetric functions will be evaluated at some point \( y = (y_1, y_2, \ldots, y_n) \in \mathbb{C}^n \), where it is understood that \( f(y) = f(y_1, y_2, \ldots, y_n, 0, 0, \ldots) \). In this way, symmetric functions specialize to symmetric polynomials.

The symmetric functions of interest here are the power sums, as they describe traces. For an integer \( k \), define \( p_k \) by

\[
p_k = x_1^k + x_2^k + x_3^k + \cdots,
\]

and for a partition \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \), define \( p_\lambda \) by

\[
p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_n}.
\]

These are called the **power sum symmetric functions**, and \( \{p_\lambda\}_\lambda \) are a basis for \( \Lambda \). Note that the trace of a power of a matrix \( \text{tr} A^k \) can alternately be expressed as \( p_k \) evaluated at the eigenvalues of \( A \).

The second basis we require are the **Jack symmetric functions** \( P_\alpha^\lambda \). For those interested, there is a concise introduction available in Stanley’s paper [45]. By virtue of being a basis, it is possible to write \( p_k \) as a finite linear combination of \( \{P_\alpha^\lambda\}_{\lambda \vdash k} \).

There are multiple normalizations for the Jack functions in the literature. In citing some theorems, we will require a second normalization, \( J_\lambda^\alpha \). The two are related, as \( J_\lambda^\alpha = c(\lambda, \alpha) P_\lambda^\alpha \), where

\[
(47) \quad c(\lambda, \alpha) = \prod_{s \in \lambda} (\alpha a(s) + l(s) + 1),
\]

using the arm length \( a(s) \) and leg length \( l(s) \).

One final tool we will use is the Macdonald automorphism \( \omega_\alpha \). It is defined in terms of the symmetric power functions by \( \omega_\alpha p_k = \alpha p_k \); it is extended to each \( p_\lambda \) as a multiplicative homomorphism; and at last it is extended to all \( \Lambda \) as a \( \mathbb{C} \)-linear transformation. This automorphism acts on the Jack functions in a nice way as well, as by a formula of Stanley [45],

\[
(48) \quad \omega_{-1/\alpha}^{-1} J_\lambda^\alpha = (-\alpha)^{|\lambda|} J_\lambda^{\bar{\alpha}}.
\]

More often in the literature on Jack functions, these coefficients are defined to be from \( \mathbb{Q}(\alpha) \), but the distinction here is immaterial.
A.1. Kadell’s Integral. Kadell’s integral (see [27]) is a generalization of Selberg’s integral [10], which states the following
\[
\int _{[0,1]^{n}} \prod _{i<j} |x_{i} - x_{j}|^{2/\alpha} \prod _{i=1}^{n} x_{i}^{r-1}(1-x_{i})^{s-1} \, dx = \prod _{i=1}^{n} \frac{\Gamma(1 + \frac{i+1}{\alpha})\Gamma(r + \frac{i-1}{\alpha})\Gamma(s + \frac{i-1}{\alpha})}{\Gamma(1 + \frac{i}{\alpha})\Gamma(r + s + \frac{n+1-2i}{\alpha})}.
\]

It was generalized to include the Jack function \( P_{\lambda}^{1/\alpha}(x) \) in the integrand. Letting \( W(n, \alpha, r, s) \) be the integrand of Selberg’s integral, Kadell’s integral is
\[
(49) \quad \int _{[0,1]^{n}} P_{\lambda}^{1/\alpha}(x)W(n, \alpha, r, s) \, dx = n! v_{\lambda}^{n} \prod _{i=1}^{n} \frac{\Gamma(\lambda_{i} + r + \frac{n-1}{\alpha})\Gamma(s + \frac{n-1}{\alpha})}{\Gamma(\lambda_{i} + r + s + \frac{2n-i-1}{\alpha})},
\]
where the term \( v_{\lambda}^{n} \) is defined as
\[
(50) \quad v_{\lambda}^{n} = \prod _{i<j} \frac{\Gamma(\lambda_{i} - \lambda_{j} + \frac{i-1}{\alpha})}{\Gamma(\lambda_{i} - \lambda_{j} + \frac{2i-1}{\alpha})}.
\]

Our goal is to show that
\[
\int _{[0,1]^{n}} \frac{P_{\lambda}^{1/\alpha}(x)}{P_{\lambda}^{1/\alpha}(I_{n})}W(n, \alpha, r, s) \, dx,
\]
where \( I_{n} = (1, 1, \ldots, 1) \) has \( n \) 1’s, has a quasi-palindromic property. The constant \( P_{\lambda}^{1/\alpha}(I_{n}) \) is computable in terms of diagram statistics. From formula VI.10.20 of [33],
\[
(51) \quad P_{\lambda}^{1/\alpha}(I_{n}) = \prod _{s \in \lambda} \left( \frac{n + \alpha a'(s) - l'(s)}{\alpha a(s) + l(s) + 1} \right) = \frac{1}{c(\lambda, \alpha)} \prod _{s \in \lambda} (n + \alpha a'(s) - l'(s))
\]
where \( c(\lambda, \alpha) \) is the constant that relates \( J_{\alpha}^{\lambda} \) and \( P_{\lambda}^{\alpha} \) (see [47]). To compare the two, we will convert Kadell’s expression using \( \Gamma \) functions into a Young diagram formula.

Recall that a quotient of \( \Gamma \) functions, also known as the Pochhammer symbol \((x)_{k}\), may be expressed alternately as
\[
\frac{\Gamma(x + k)}{\Gamma(x)} = (x)_{k} = (x)(x + 1) \cdots (x + k - 1),
\]
when \( k \) is a natural number. Define the generalized Pochhammer symbol \((t)_{\mu}\) (also known as the shifted factorial) to be
\[
(52) \quad (t)_{\mu} = \prod _{s \in \mu} (t + a'(s) - \frac{1}{\alpha} l'(s)).
\]

In terms of these expressions, (51) can be rewritten as
\[
(53) \quad P_{\lambda}^{1/\alpha}(I_{n}) = \frac{\left( \frac{\alpha}{\lambda} \right)_{\lambda}^{\lambda} |\lambda|}{c(\lambda, \alpha)}.
\]

We will need a closely related quantity to \( c(\lambda, \alpha) \), so define \( c'(\lambda, \alpha) \) to be
\[
\prod _{s \in \lambda} (\alpha a(s) + l(s) + \alpha).
\]

Both \( c(\lambda, \alpha) \) and \( c'(\lambda, \alpha) \) can be expressed as products of \( \Gamma \) terms, which will need to rewrite Kadell’s integral. Write out the terms in \( \alpha^{-|\lambda|} c'(\lambda, \alpha) \) by going from right to left along the first row of the diagram of \( \lambda \). There are \( \lambda_{1} - \lambda_{2} \) terms that have \( l(s) = 0 \):
\[
\left( \frac{n}{\alpha} + 1 + 0 \right) \left( \frac{n}{\alpha} + 1 + 1 \right) \cdots \left( \frac{n}{\alpha} + 1 + \lambda_{1} - \lambda_{2} - 1 \right) = \frac{\Gamma(\lambda_{1} - \lambda_{2} + 1)}{\Gamma(1)}.
\]

There are then \( \lambda_{2} - \lambda_{3} \) terms that have \( l(s) = 1 \):
\[
\left( \frac{1}{\alpha} + 1 + \lambda_{1} - \lambda_{2} \right) \left( \frac{1}{\alpha} + 1 + \lambda_{1} - \lambda_{2} + 1 \right) \cdots \left( \frac{1}{\alpha} + 1 + \lambda_{1} - \lambda_{3} - 1 \right) = \frac{\Gamma(\lambda_{1} - \lambda_{3} + 1 + \frac{1}{\alpha})}{\Gamma(\lambda_{1} - \lambda_{2} + 1 + \frac{1}{\alpha})}.
\]
This pattern continues until at last there are \( \lambda_n \) terms that have \( l(s) = n - 1 \):

\[
\left( \frac{n-1}{\alpha} + 1 + \lambda_1 - \lambda_n \right) \left( \frac{n-1}{\alpha} + 1 + \lambda_1 - \lambda_n + 1 \right) \cdots \left( \frac{n-1}{\alpha} + \lambda_1 \right) = \frac{\Gamma(\lambda_1 + 1 + \frac{n-1}{\alpha})}{\Gamma(\lambda_1 - \lambda_n + 1 + \frac{n-1}{\alpha})}.
\]

Writing out all the terms in the first row gives

\[
\frac{\Gamma(\lambda_1 - \lambda_2 + 1)}{\Gamma(1)} \frac{\Gamma(\lambda_1 - \lambda_3 + 1 + \frac{1}{\alpha})}{\Gamma(\lambda_1 - \lambda_2 + 1 + \frac{1}{\alpha})} \frac{\Gamma(\lambda_1 - \lambda_4 + 1 + \frac{2}{\alpha})}{\Gamma(\lambda_1 - \lambda_3 + 1 + \frac{2}{\alpha})} \cdots \frac{\Gamma(\lambda_1 + 1 + \frac{n-1}{\alpha})}{\Gamma(\lambda_1 - \lambda_n + 1 + \frac{n-1}{\alpha})}.
\]

Inducting over the rows, it follows that \( c'(\lambda, \alpha) \) can be written as

\[
(54) \quad c'(\lambda, \alpha) = (\alpha)^{|\lambda|} \prod_{i<j} \frac{\Gamma(\lambda_i - \lambda_j + 1 + \frac{i-j}{\alpha})}{\Gamma(\lambda_i - \lambda_j + 1 + \frac{i-j}{\alpha})} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + 1 + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + 1 + \frac{n-i}{\alpha})}
\]

If one does the same expansion along the first row for \( c(\lambda, \alpha) \), one gets

\[
\frac{\Gamma(\lambda_1 - \lambda_2 + \frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})} \frac{\Gamma(\lambda_1 - \lambda_3 + \frac{2}{\alpha})}{\Gamma(\lambda_1 - \lambda_2 + \frac{2}{\alpha})} \frac{\Gamma(\lambda_1 - \lambda_4 + \frac{3}{\alpha})}{\Gamma(\lambda_1 - \lambda_3 + \frac{3}{\alpha})} \cdots \frac{\Gamma(\lambda_1 + \frac{1}{\alpha})}{\Gamma(\lambda_1 - \lambda_n + \frac{1}{\alpha})}.
\]

Repeating the analogous procedure for the rest of the rows, we eventually conclude

\[
(55) \quad c(\lambda, \alpha) = (\alpha)^{|\lambda|} \prod_{i<j} \frac{\Gamma(\lambda_i - \lambda_j + \frac{i-j}{\alpha})}{\Gamma(\lambda_i - \lambda_j + \frac{i-j}{\alpha})} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + \frac{n-i+1}{\alpha})}{\Gamma(\lambda_i + \frac{n-i+1}{\alpha})}.
\]

Equations (54) and (55) allow (53) to be rewritten as

\[
(56) \quad v_\lambda^n = \frac{(\alpha)^{|\lambda|}}{\Gamma(\lambda, \alpha)} \prod_{i=1}^{n} \Gamma(\lambda_i + \frac{n-i}{\alpha}) \Gamma(\frac{n}{\alpha}).
\]

We can repeat the same procedure as used for \( c \) and \( c' \) to show that \( (t)_\lambda \) can be computed by

\[
(57) \quad (t)_\lambda = \prod_{i=1}^{n} \frac{\Gamma(t - \frac{i-1}{\alpha} + \lambda_i)}{\Gamma(t - \frac{i-1}{\alpha})}.
\]

This allows the expression in (50) for \( v_\lambda^n \) to be replaced by

\[
(58) \quad v_\lambda^n = \frac{(\alpha)^{|\lambda|}}{\Gamma(\lambda, \alpha)} \prod_{i=1}^{n} \Gamma(\lambda_i + \frac{n}{\alpha} - \frac{i-1}{\alpha}) \frac{\Gamma(\alpha)}{\Gamma(\alpha - \frac{i-1}{\alpha})} = \frac{(\alpha)^{|\lambda|}}{\Gamma(\lambda, \alpha)} \prod_{i=1}^{n} \Gamma(\lambda_i + \frac{i}{\alpha}).
\]

Combine this expression for \( v_\lambda^n \) with Kadell’s integral formula (19) and the simplified expression (53) for \( P_{1/\alpha}(I_n) \) to get

\[
\int_{[0,1]^n} P_{1/\alpha}(x) W(n, \alpha, r, s) dx = \frac{c(\lambda, \alpha)}{(\frac{n}{\alpha})_\lambda (\alpha)^{|\lambda|}} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + \frac{n-i}{\alpha})}{\Gamma(\frac{n}{\alpha} - \frac{i-1}{\alpha})} \prod_{i=1}^{n} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i}{\alpha})}
\]

\[
= \prod_{i=1}^{n} \frac{\Gamma(1 + \frac{i}{\alpha})}{\Gamma(1 + \frac{i}{\alpha})} \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha})}{\Gamma(\lambda_i + r + s + \frac{2n-i}{\alpha})}.
\]

Let \( \mu_f \) be the \((\frac{1}{\alpha}, r, s)\)-Jacobi ensemble measure on \([0,1]^n\). This has density function proportional to \( W(n, \alpha, r, s) \), but it is appropriately renormalized to be a probability measure. This normalization is given by Selberg’s integral.
The integral expression above can be rewritten as

\[
\int_{[0,1]^n} \frac{P^{1/\alpha}_\lambda(x)}{P^{1/\alpha}_\lambda(I_n)} \, d\mu_f(x) = \prod_{i=1}^n \frac{\Gamma(1 + \frac{1}{\alpha}) \Gamma(\lambda_i + r + \frac{n-i}{\alpha}) \Gamma(s + \frac{n-i}{\alpha})}{\Gamma(1 + \frac{1}{\alpha}) \Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})} \, J W(n, \alpha, r, s) dx
\]

\[
= \prod_{i=1}^n \frac{\Gamma(\lambda_i + r + \frac{n-i}{\alpha}) \Gamma(r + \frac{n-i}{\alpha})}{\Gamma(r + \frac{n-1}{\alpha}) \Gamma(\lambda_i + r + s + \frac{2n-i-1}{\alpha})} \left( \frac{r + \frac{n-1}{\alpha}}{r + s + \frac{2n-2}{\alpha}} \right)_{\lambda}.
\]

(59)

A.2. Palindromy.

Lemma A.3. Let

\[
\int_{[0,1]^n} \frac{P^{1/\alpha}_\lambda(x)}{P^{1/\alpha}_\lambda(I_n)} \, d\mu_f(x) = \sum_{k=0}^{\infty} \rho(k, \lambda, \alpha)n^{-k}
\]

be the series expansion about \( n = \infty \). The coefficients \( \rho(k, \lambda, \alpha) \) are skew-palindromic in that

\[
\rho(k, \lambda, \alpha) = (-\alpha)^k \rho(k, \lambda', \frac{1}{\alpha}).
\]

Proof. In the calculation that follows, let \( f(\lambda, \alpha, t) = f(t) = \alpha a'(t) - l'(t) \), for tableau block \( t \in \lambda \). Starting from the formula computed in [52], and applying formula [52] gives

\[
\left( \frac{nb}{\alpha a} \right)_\lambda = \prod_{t \in \lambda} \frac{nb + af(t)}{n + af(t)} = \prod_{t \in \lambda} \frac{b + \frac{a}{n} f(t)}{1 + \frac{a}{n} f(t)} = \prod_{t \in \lambda} \left( b (1 + \frac{a}{n} f(t)) \sum_{k=0}^{\infty} \left( -\frac{a}{n} f(t) \right)^k \right) = b^{|\lambda|} \prod_{t \in \lambda} \left( 1 + \left( 1 - \frac{1}{n} \right) \sum_{k=1}^{\infty} \left( -\frac{a}{n} f(t) \right)^k \right).
\]

Let \( M(\lambda, k) \) be the collection of all \( k \)-element multisets sampled from \( \lambda \). If \( \tau \in M(\lambda, k) \) is such a multiset, let \( m_\tau(t) \) denote the multiplicity of \( t \in \tau \) and let \( \epsilon_\tau(t) \) be the characteristic function for \( t \in \tau \). The sum can be written as:

\[
= b^{|\lambda|} \sum_{k=0}^{\infty} n^{-k} \left[ \sum_{\tau \in M(\lambda, k)} \prod_{t \in \lambda} \left( -af(t) \right)^{m_\tau(t)} (1 - \frac{1}{\alpha})^{\epsilon_\tau(t)} \right].
\]

This gives an explicit form for the coefficients \( f(k, \lambda, \alpha) \). Mapping \( \lambda \) to \( \lambda' \) induces a bijection mapping the collection \( M(\lambda, k) \) to \( M(\lambda', k) \). In the conjugate, the arm co-length \( a' \) and leg co-length \( l' \) are reversed, so that \( f(t) \) becomes \( \alpha l'(t) - a'(t) \). Thus \( f(\lambda, \alpha, t) = -\alpha f(\lambda', \alpha^{-1}, t) \), so that

\[
\sum_{\tau \in M(\lambda, k)} \prod_{t \in \lambda} \left( -af(\lambda, \alpha, t) \right)^{m_\tau(t)} (1 - \frac{1}{\alpha})^{\epsilon_\tau(t)} = (-\alpha)^k \sum_{\tau \in M(\lambda', k)} \prod_{t \in \lambda'} \left( -af(\lambda', \alpha^{-1}, t) \right)^{m_\tau(t)} (1 - \frac{1}{\alpha})^{\epsilon_\tau(t)}.
\]

Let \( J^{1/\alpha}_\lambda \) be the Jack functions renormalized by

\[
J^{1/\alpha}_\lambda = c(\lambda, \alpha) P^{1/\alpha}_\lambda.
\]

(60)
Expand the symmetric power function $p_k$ as

$$p_k = \sum_{\lambda \vdash k} \xi(\lambda, \alpha) J^{1/\alpha}_\lambda.$$

By applying Stanley’s formula (see (48)), it follows (see [17]) that

(61) $$\xi(\lambda, \alpha) = (-\alpha)^{1-|\lambda|} \xi(\lambda', \alpha^{-1}).$$

One last piece is needed. The normalization factor $J^{1/\alpha}_{\lambda/\alpha}(I_n)$ can be computed by relating (53) and the definition of $J^{\alpha}_{\lambda}$ in (60). These two combined give that

$$J^{1/\alpha}_{\lambda/\alpha}(I_n) = (\frac{n}{\alpha})_{\lambda}^{1/\alpha} = \prod_{t \in \lambda} (n + \alpha a'(t) - l'(t));$$

expand this as a polynomial in $n$, i.e. put

$$\prod_{t \in \lambda} (n + \alpha a'(t) - l'(t)) = \sum_{j=0}^{|\lambda|} \zeta(j, \lambda, \alpha) n^j.$$

Because the product can be expressed as

$$\prod_{t \in \lambda} (n + \alpha a'(t) - l'(t)) = \prod_{t \in \lambda'} (n + (-\alpha) (\alpha^{-1} a'(t) - l'(t))),$$

it follows that

(62) $$\zeta(j, \lambda, \alpha) = (-\alpha)^{|\lambda|-j} \zeta(j, \lambda', \alpha^{-1}).$$

**Proof of Theorem 5.1.** Expand $p_{[k]}$ in the Jack function basis:

$$\frac{1}{n} E_\alpha p_{[k]} = \frac{1}{n} \sum_{\lambda \vdash k} \xi(\lambda, \alpha) E_\alpha J^{1/\alpha}_\lambda$$

$$= \frac{1}{n} \sum_{\lambda \vdash k} \xi(\lambda, \alpha) J^{1/\alpha}_{\lambda/\alpha}(I_n) E_\alpha J^{1/\alpha}_{\lambda/\alpha}(I_n).$$

Apply Lemma [5.3] and expand $J^{1/\alpha}_{\lambda/\alpha}(I_n)$. Note that the alternative normalization used in the Lemma cancels out.

$$\frac{1}{n} E_\alpha p_{[k]} = \frac{1}{n} \sum_{\lambda \vdash k} \xi(\lambda, \alpha) \left( \sum_{j=0}^k \zeta(j, \lambda, \alpha) n^j \right) \left( \sum_{j=0}^\infty \rho(j, \lambda, \alpha) n^{-j} \right)$$

$$= \sum_{j=-\infty}^k n^{j-1} \left( \sum_{\lambda \vdash k} \xi(\lambda, \alpha) \sum_{l=0}^k \zeta(l, \lambda, \alpha) \rho(l-j, \lambda, \alpha) \right).$$

with $\rho(l-j, \lambda, \alpha) = 0$ for negative $l-j$.

This gives a formula for $\eta_k(j, \alpha)$, namely that

$$\eta_k(j, \alpha) = \sum_{\lambda \vdash k} \xi(\lambda, \alpha) \sum_{l=0}^k \zeta(l, \lambda, \alpha) \rho(l+j-1, \lambda, \alpha).$$

The $j < 0$ terms vanish, which can be seen because the trace can naturally be bounded as

$$\frac{1}{n} |\xi_{[k]}| \leq \frac{1}{n} E_\alpha \sum_{i=0}^n |x_i|^k \leq \frac{1}{n} n = 1,$$

as the Jacobi distribution is supported on $[0, 1]^n$. 35
We will show that each $\eta_k(j, \alpha)$ is palindromic. Applying Lemma A.3 (1), and (2), these can be written as

$$\eta_k(j, \alpha) = \sum_{\lambda \vdash k} \xi(\lambda, \alpha) \sum_{l=0}^{k} \zeta(l, \lambda, \alpha) \rho(l + j - 1, \lambda, \alpha)$$

$$= \sum_{\lambda \vdash k} (-\alpha)^l \sum_{l=0}^{k} (-\alpha)^{k-l} \zeta(l, \lambda, \alpha^{-1}) (-\alpha)^{l+j-1} \rho(l + j - 1, \lambda, \alpha^{-1})$$

$$= (-\alpha)^j \sum_{\lambda \vdash k} \xi(\lambda, \alpha^{-1}) \sum_{l=0}^{k} \zeta(l, \lambda, \alpha^{-1}) \rho(l + j - 1, \lambda, \alpha^{-1})$$

The sum is over all partitions of $k$, so taking conjugates makes no difference. Thus,

$$\eta_k(j, \alpha) = (-\alpha)^j \eta_k(j, \alpha^{-1}).$$

The last claim we make is that $\eta_k(j, \alpha)$ is a polynomial in $\alpha$ of degree $j$. This is more involved, and requires that we appeal to Edelman and Sutton’s tridiagonal matrix model (see the start of Section 3). The moment $\frac{1}{n} E \rho_k = \frac{1}{n} E(\text{tr}(A^k))$ can be written in terms of a sum over alternating bridges (see Section 3.1),

$$\frac{1}{n} E(\text{tr}(A^k)) = \frac{1}{n} \sum_{\bar{w} \in A_{2k}} E(\bar{B}_\beta)_{\bar{w}+i}.$$ 

A priori, these expectations are moments of random variables distributed as the square root of a Beta random variable. However, by Lemma 2.6 the alternating bridge visits each matrix entry an even number of times. Thus, any term in the sum takes the form

$$E \left( \begin{subarray}{c}
\sum_{i=1}^{k} \omega_i^{2m_i - 1} \omega_i^{2n_i - 1}
\end{subarray} \; \begin{subarray}{c}
\omega_i^{2m_i} \omega_i^{2n_i}
\end{subarray} \right),$$

where $\omega_i$ ranges over the matrix entries referenced by the bridge $\bar{w}$ and $\sum_{i=1}^{2k} m_i = k$. By independence, this expectation is a product of terms of the form

$$E \left( \begin{subarray}{c}
\omega_i^{2m} \omega_i^{2n}
\end{subarray} \; \begin{subarray}{c}
\omega_i^{2m} \omega_i^{2n}
\end{subarray} \right).$$

By Lemma A.4 each such Beta moment admits a series expansion around $n = \infty$ and a $K$ so that

$$E(\bar{B}_\beta)_{\bar{w}+i} = \sum_{m=0}^{\infty} n^{-m} \alpha^m \Omega_{\bar{w}+i,m}(n),$$

where $0 < \Omega_{\bar{w}+i,m}(n) < K^m$ for all $n$. Moreover, this constant $K$ can be chosen independently of $\bar{w} + i$. Thus the entire trace admits such a series expansion,

$$\frac{1}{n} E(\text{tr}(A^k)) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\bar{w} \in A_{2k}} \sum_{m=0}^{\infty} n^{-m} \alpha^m \Omega_{\bar{w}+i,m}(n)$$

$$= \sum_{m=0}^{\infty} n^{-m} \alpha^m \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{\bar{w} \in A_{2k}} \Omega_{\bar{w}+i,m}(n) \right).$$

Because the cardinality of $A_{2k}$ is at most $\binom{2k}{k}$, the sum $\Omega_m(n) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\bar{w} \in A_{2k}} \Omega_{\bar{w}+i,m}(n)$ satisfies an estimate $0 < \Omega_m(n) < (\frac{2k}{k}) K^m = CK^m$. Thus there are two expansions for the trace, valid for all $n$ sufficiently large, i.e.

$$\sum_{j=0}^{\infty} \eta(j, \alpha)n^{-j} = \frac{1}{n} E(\text{tr}(A^k)) = \sum_{j=0}^{\infty} \Omega_j(n) \alpha^j n^{-j}$$

The left hand side expansion shows that the $n \to \infty$ limit must exist. Thus

$$\eta(0, \alpha) = \lim_{n \to \infty} \sum_{j=0}^{\infty} \eta(j, \alpha)n^{-j} = \lim_{n \to \infty} \sum_{j=0}^{\infty} \Omega_j(n) \alpha^j n^{-j} = \lim_{n \to \infty} \Omega_0(n).$$
In particular, \( \eta(0, \alpha) \) has no \( \alpha \) dependence. The proof now proceeds by induction. Suppose that for all \( j < l \), the term \( \eta(j, \alpha) \) is a polynomial in \( \alpha \) of degree \( j \). It should be shown that \( \eta(l, \alpha) \) is a polynomial in \( \alpha \) of degree \( l \). The limit

\[
\lim_{n \to \infty} n^l \left[ \frac{1}{l} \mathbb{E} \text{tr}(A^k) - \sum_{j=0}^{l-1} \eta(j, \alpha)n^{-j} \right] = \lim_{n \to \infty} \sum_{j=0}^{\infty} \eta(j + l)n^{-j} = \eta(l, \alpha)
\]

exists by virtue of the \( \eta \) expansion, and by substituting the right hand side of (63), it follows that

\[
\eta(l, \alpha) = \lim_{n \to \infty} n^l \left[ \sum_{j=0}^{\infty} \Omega_j(n)\alpha^j n^{-j} - \sum_{j=0}^{l-1} \eta(j, \alpha)n^{-j} \right] = \lim_{n \to \infty} \sum_{j=0}^{l-1} \left[ \Omega_j(n)\alpha^j - \eta(j, \alpha) \right] n^{l-j} + \Omega_l(n)\alpha^l.
\]

By the inductive hypothesis, this limit can be written in the form

\[
\eta(l, \alpha) = \lim_{n \to \infty} f_0(n) + f_1(n)\alpha + f_2(n)\alpha^2 + \cdots + f_l(n)\alpha^l,
\]

and the limit exists for each fixed \( \alpha \). Take \( l + 1 \) distinct values of \( \alpha \). The convergence is uniform on this finite set \( \alpha_0, \ldots, \alpha_l \), and so each \( f_l(n) \) converges, where \( 0 \leq i \leq l \). Thus \( \eta(j, \alpha) \) is a polynomial of degree \( l \) in \( \alpha \), concluding the proof. \( \square \)

**Lemma A.4.** Let \( f_r(n) \) and \( f_s(n) \) be positive real-valued functions defined on \( \mathbb{N} \) so that

\[
0 < f_r(n) \leq C_2, \ 0 < f_s(n) \leq C_2i, \ C_1 < f_r(n) + f_s(n),
\]

where \( C_i \) are some positive constants. Let \( r = \alpha^{-1}f_r(n)n \), \( s = \alpha^{-1}f_s(n)n \), and let \( X \sim \text{Beta}(r, s) \). There is an asymptotic expansion

\[
\mathbb{E} [X^k(1-X)^l] = \sum_{m=0}^{\infty} n^{-m} \alpha^m p_m(n),
\]

and a constant \( K \) depending only on \( k, l, C_1, \) and \( C_2 \) so that \( 0 < p_m(n) < K^n \).

**Proof.** The expectation, which can be computed using Euler’s Beta integral formula, gives that

\[
\mathbb{E} [X^k(1-X)^l] = \frac{(r)_k(s)_l}{(r+s)_{k+l}}.
\]

Substituting in the definitions for \( r \) and \( s \) and writing out the Pochhammer symbols gives

\[
= \prod_{i=0}^{k-1} \frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n + i} \prod_{i=0}^{l-1} \frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n + k + i}
\]

All rational terms in this product produce similar asymptotic series expansions, and so we will only examine one. Working with a term from the left hand product,

\[
\frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n + i} = \left( \frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n} \right) \left( \frac{1}{1 + \frac{i}{\alpha^{-1}(f_r + f_s)n}} \right)
\]

Provided that \( n \) is sufficiently large (depending on \( C_1 \) and \( \alpha \)), this can be expanded as a series.

\[
\begin{align*}
\frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n + i} &= \frac{\alpha^{-1}f_r n + i}{\alpha^{-1}(f_r + f_s)n} \sum_{m=0}^{\infty} (f_r + f_s)^{-m} n^{-m} \alpha^m \\
&= \frac{f_r}{f_r + f_s} + \sum_{m=1}^{\infty} \left( \frac{f_r}{f_r + f_s} + i \right) (f_r + f_s)^{-m} n^{-m} \alpha^m \\
&= \sum_{m=1}^{\infty} \hat{p}_m(n)n^{-m} \alpha^m.
\end{align*}
\]
The coefficients \( \hat{p}_m(n) \) satisfy an estimate
\[
0 < \hat{p}_m(n) < (C_1C_2 + k)(C_1)^{-m}.
\]

\[\square\]

**B. Poincaré Inequality for Beta**

**Lemma B.1.** Let \( Y \sim \text{Beta}(p, q) \). For any Lipschitz function \( f \) on \([0, 1]\),
\[
\text{Var}(f(Y)) \leq \frac{1}{4(p+q)}E|f'(Y)|^2.
\]

We note that in the case that both \( p \) and \( q \) are greater than 1, the density is log-concave, and it is possible to use the general theory outlined by Bobkov in [8] to produce an equivalent bound, but we require the inequality to hold for all \( p \) and \( q \) positive, and thus we use an alternative technique.

**Proof.** We begin by showing the analogous bound for the translated random variable \( X = 2Y - 1 \), and write \( Y = T(X) := \frac{1}{2}(X + 1) \). The density of \( Y \) is given by
\[
\frac{d\mu_\beta}{dx} = \frac{1}{Z_{p,q}}(1-x)^{p-1}(1+x)^{q-1}.
\]

We will show that for any Lipschitz function \( f \) on \([-1, 1]\), that
\[
(64) \quad \text{Var}(f(X)) \leq \frac{1}{p+q}E[(1-X^2)|f'(X)|^2].
\]

As will be seen in the proof, this inequality is attained taking \( f \) to be a multiple of the linear Jacobi polynomial (for definitions, see [17]). The proof follows from (64), as
\[
\text{Var}(f(Y)) = \text{Var}(f \circ T)(X)
\]
\[
\leq \frac{1}{p+q}E[(1-X^2)|(f \circ T)'(X)|^2]
\]
\[
\leq \frac{1}{p+q}E|f'(T)(x)|^2
\]
\[
= \frac{1}{p+q}E|f'(X)|^2 \frac{12}{2}
\]
\[
= \frac{1}{4(p+q)}E|f'(Y)|^2.
\]

The method of proof follows the general outline in the notes of Bakry [5]. Define the Jacobi differential operator \( L \) to be
\[
Lf = (1-x^2)f''(x) + (q-p-(p+q)x)f'(x),
\]
and define the \textit{carré du champ} operator \( \Gamma \) by
\[
\Gamma(f, g) = (1-x^2)f'(x)g'(x).
\]

It can be checked by integration by parts that for all \( C^2 \) functions on \([-1, 1]\) that the Dirichlet form \( \mathcal{E}(f, g) \) associated to \( L \) satisfies
\[
\mathcal{E}(f, g) := -\int_{-1}^1 f(x)(Lg)(x)d\mu_\beta(x) = \int_{-1}^1 \Gamma(f(x), g(x))d\mu_\beta(x).
\]

The spectrum of \( L \) restricted to \( L^2(\mu_\beta) \) is non-positive, with eigenvalues \( y_n = -n(n+p+q-1) \) for non-negative integers \( n \). Further, its eigenfunctions are given by the Jacobi polynomials \( P_n^{p-1,q-1}(x) \), which, when normalized form a complete orthonormal system for \( L^2(\mu_\beta) \). From the density of the polynomials in \( L^2(\mu_\beta) \), it is an immediate consequence that
\[
p + q = -y_1 = \inf_{f \in L^2(\mu_\beta)} \frac{\mathcal{E}(f, f)}{\text{Var}(f(X))} = \inf_{f \in L^2(\mu_\beta)} \frac{\int_{-1}^1 \Gamma(f(x), g(x))d\mu_\beta(x)}{\text{Var}(f(X))},
\]

which upon rewriting, gives (64).
C. Coupling Bound for $\sqrt{\text{Beta}}$

We provide an auxiliary lemma regarding the square root of Beta variables that appear in the matrix entries. Note that because one of the parameters of the $c_i'$ family is not $\Omega(n)$ for all $i$, this approximation can not be applied to every matrix entry with uniform error.

**Lemma C.1.** If $Y$ is distributed as $\sqrt{\text{Beta}(np, nq)}$, then

$$2\left(\frac{p+q}{q}\right)\sqrt{n} \left( Y - \sqrt{\frac{p}{p+q}} \right) \Rightarrow N(0,1),$$

as $n \to \infty$, where $p, q$ are fixed positive constants. Moreover, it is possible to couple $Y$ to a standard normal $X$ so that

$$\text{Var} \left( Y - \sqrt{\frac{q}{2(p+q)\sqrt{n}}} X \right) \leq \frac{K_{p,q}}{n^2},$$

for some $K_{p,q} > 0$, independent of $n$, and continuous in $p, q$ positive, provided that $n > \max\left\{\frac{1}{p}, \frac{1}{q}\right\}$.

**Proof of C.1.** Let $Y$ be distributed as $\sqrt{\text{Beta}(np, nq)}$. Put

$$\mu = \sqrt{\frac{p}{p+q}}, \quad \sigma = \sqrt{\frac{q}{2np+2nq}}.$$

Note that these are not exactly the mean or standard deviation of $Y$, however,$$
\tilde{Y} = \frac{Y - \mu}{\sigma} \Rightarrow N(0,1).$$

Moreover, it will be shown that there is an $X$ distributed as $N(0,1)$ so that

$$E(\tilde{Y} - X)^2 \leq \frac{K}{n},$$

for some $K = K(p, q)$ depending continuously on $p, q$ positive. Note that this implies Lemma [C.1] after dividing through by $\sigma$.

The primary machinery here is Talagrand’s transport inequality, which bounds the square $L^2$-Wasserstein distance of $\tilde{Y}$ and $X$, with $X$ distributed as $N(0,1)$. We use a special case of Theorem 1.1 of [48], which states

**Proposition C.2 (Talagrand).** Let $\tilde{Y}$ be a random variable given by probability measure $\tilde{\nu}$, which is absolutely continuous with Lebesgue measure, and let $\gamma$ be a standard Gaussian measure. There is a standard normal random variable $X$ so that

$$E(\tilde{Y} - X)^2 \leq 2 \int \log \frac{d\tilde{\nu}}{d\gamma} d\gamma.$$

The density $\frac{d\nu}{dy}$ of $Y$ can be computed to be

$$\frac{d\nu}{dy} = 2y(y^2)^{np-1}(1-y^2)^{nq-1} \frac{\Gamma(np+1)\Gamma(nq)}{\Gamma(np+1)\Gamma(nq)}$$

for $y \in [0,1]$. It follows that density of $\tilde{Y}$ is given by

$$\frac{d\tilde{\nu}}{dy} = 2\sigma(\mu + y\sigma)^{2np-1}(1-(\mu + y\sigma)^2)^{nq-1} \frac{\Gamma(np+1)\Gamma(nq)}{\Gamma(np+1)\Gamma(nq)}$$

and thus the Radon-Nikodym derivative $\frac{d\tilde{\nu}}{d\gamma}(y)$ is a product of four terms

$$\frac{d\tilde{\nu}}{d\gamma}(y) = (\mu + y\sigma)^{2np-1}(1-(\mu + y\sigma)^2)^{nq-1} e^{y^2/2} \frac{\Gamma(np+1)\Gamma(nq)}{\Gamma(np+1)\Gamma(nq)} \sqrt{2\pi}.$$

The logs of terms (i) and (ii) can be controlled by Taylor expansion. Explicitly,

$$\ln[1+y]^g = g \ln(1+y) \leq g \left[ y - \frac{y^2}{2} + \frac{y^3}{3} \right],$$

for some $K_{p,q} > 0$, independent of $n$, and continuous in $p, q$ positive, provided that $n > \max\left\{\frac{1}{p}, \frac{1}{q}\right\}$. 


for all $y > -1$, and all $g > 0$. Note that both produce a nonzero constant term, by virtue of the relationship \( \ln(a+y) = \ln(a) + \ln(1+y/a) \). This bound is applied to the logs of both \((i)\) and \((ii)\) after suitable rearrangement. This bounds the sum of the logs by a polynomial in $y$ of degree 6. We can bound the log of term \((i)\) as

\[
\ln [(\mu + y\sigma)^{2np-1}] = (2np - 1) \ln \mu + (2np - 1) \ln \left[1 + \frac{y\sqrt{n}}{2\sqrt{2np}}\right] \\
\leq (2np - 1) \left[ \ln \mu + \frac{y\sqrt{n}}{\sqrt{2np}} - \frac{1}{2} \left( \frac{y\sqrt{n}}{\sqrt{2np}} \right)^2 \right] + \frac{1}{3} \left( \frac{y\sqrt{n}}{\sqrt{2np}} \right)^3.
\]

Applying the same to term \((ii)\),

\[
\ln [(1 - (\mu + y\sigma)^{2np-1}] \leq (nq - 1) \left[ \ln(1 - \mu^2) - \left( \frac{2y\sqrt{n}}{\sqrt{2np}} \right)^2 - \frac{1}{2} \left( \frac{2y\sqrt{n}}{\sqrt{2np}} + \left[ \frac{y\sqrt{n}}{\sqrt{2np}} \right]^2 \right)^2 - \frac{1}{3} \left( \frac{2y\sqrt{n}}{\sqrt{2np}} + \left[ \frac{y\sqrt{n}}{\sqrt{2np}} \right]^2 \right)^3 \right].
\]

From this form, it is easy to see that the coefficients of this polynomial depend continuously on $p$ and $q$. Further, the coefficients of $y^4, y^5,$ and $y^6$ already decay at least as fast as $1/n$. The coefficient of $y^3$ decays like $n^{-1/2}$, so some amount of control over $\mathbb{E}Y^3$ will need to be gained. The coefficients of the lower order terms do not \textit{a priori} decay at all, but there is strong cancellation. The constant term is

\[
C_0(p,q) := (2np - 1) \ln (\mu) + (nq - 1) \ln (1 - \mu^2),
\]

the linear term has coefficient

\[
\frac{C_1(p,q)}{\sqrt{n}} := \frac{(2np - 1) \sigma}{\mu} - 2 \frac{(nq - 1) \mu \sigma}{1 - \mu^2},
\]

and the quadratic term has coefficient

\[
-\frac{1}{2} + \frac{C_2(p,q)}{n} := -1/2 \frac{(2np - 1) \sigma^2}{\mu^2} + (nq - 1) \left( \frac{\sigma^2}{1 - \mu^2} - 2 \frac{\mu^2 \sigma^2}{(1 - \mu^2)^2} \right).
\]

The $-\frac{1}{2}$ in the quadratic term represents the asymptotically Gaussian portion, and it annihilates term \((iii)\). This leaves four sources of error that need to be controlled to show the desired $O(n^{-1})$ bound:

1. $|\mathbb{E}Y| \leq C(p,q)n^{-\frac{1}{2}}$ to control the linear term.
2. $|\mathbb{E}Y^3| \leq C(p,q)n^{-\frac{1}{2}}$ to control the cubic term.
3. $|\mathbb{E}Y^4| \leq C(p,q)$ to control the second, fourth, fifth, and sixth terms.
4. The constants from the Taylor approximation and the constants from part \((iv)\) of the Radon-Nikodym derivative need to cancel to order $O(n^{-1})$.

The raw moments of $Y$ are easily computable, and their formula follows immediately from Euler’s Beta integral,

\[
\mathbb{E}(Y)^k = \frac{\Gamma(n(p+q)) \Gamma(np + \frac{k}{q})}{\Gamma(n(p+q) + \frac{k}{q}) \Gamma(np)}.
\]

Appropriate control over the first 6 raw moments could be achieved by taking sufficiently many terms from the Stirling approximation and canceling terms. To some extent, doing such a procedure is necessary, as this is necessary to get the precise control over the first and third raw moments. However, we will not need to do this for all 6 moments, because we can appeal to a Poincaré inequality. Provided that $n > \max\left\{ \frac{1}{p}, \frac{1}{q} \right\}$, the density $\frac{dc}{dy}$ is log-concave. Thus if it can be shown that $\tilde{Y}$ has constant order variance, we can use the Poincaré inequality to bound higher moments by lower moments, i.e.

\[
\text{Var} f(\tilde{Y}) \leq C \mathbb{E}|f'(\tilde{Y})|^2,
\]

applied to $f(\tilde{Y}) = (\tilde{Y})^k$, gives

\[
\mathbb{E} \tilde{Y}^{2k} \leq \left( \mathbb{E} \tilde{Y}^k \right)^2 + Ck^2 \mathbb{E} \tilde{Y}^{2k-2}.
\]
Because of the log-concavity, $C$ can be taken to be $12\E|\hat{Y}|^2$ (see Corr 4.3 of [8]), which is continuous in $p$ and $q$. Thus provided that $\E|\hat{Y}|$ can be bounded by some continuous function in $p$ and $q$, iterating the Poincaré inequality gives constant order bounds that are continuous in $p$ and $q$ for all absolute moments. Further,  
\[
\E|\hat{Y}| \leq \sqrt{\E \left| \hat{Y} \right|^2},
\]
so the problem has been reduced to finding good bounds for the first three raw moments of $\hat{Y}$.

By appealing to Stirling’s formula, and using that the error-in-approximation is bounded by the first omitted term in the asymptotic expansion, the first three moments of $\hat{Y}$ can be bounded by  
\[
\begin{align*}
\left| \E \left| \hat{Y} \right| \right| & \leq \frac{\sqrt{q}}{4\sqrt{p(p+q)}} n^{-\frac{1}{2}}, \\
\left| \E \left| \hat{Y} \right|^2 \right| & \leq 1, \\
\left| \E \left| \hat{Y} \right|^3 \right| & \leq \frac{8p+q}{4\sqrt{pq(p+q)}} n^{-\frac{1}{2}}.
\end{align*}
\]

It only remains to control the constant terms. The log of (iv) can be approximated by Stirling’s formula:  
\[
\ln \left[ \frac{2\pi n^p}{\Gamma\left(\frac{np}{2}\right)} \sqrt{\frac{2\pi}{np}} \right] - \left[ -np\ln\mu^2 - nq\ln(1-\mu^2) + \ln \frac{q\sqrt{q}}{(p+q)^{\frac{1}{2}}} \right] \leq \frac{1}{12} \frac{1}{n} \frac{1}{\sqrt{p}\sqrt{q}} \sqrt{p+q}.
\]
Comparing this with the constants produced by the Taylor approximation on terms (i) and (ii), it is seen that only the $O(n^{-1})$ term remains.

\[\Box\]

References

[1] Milton Abramowitz and Irene A. Stegun (eds.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1984. Reprint of the 1972 edition, Selected Government Publications. MR757537 (85j:00005a)

[2] Greg W. Anderson and Ofer Zeitouni, *A CLT for a band matrix model*, Probab. Theory Related Fields 134 (2006), no. 2, 283–338. MR2222385 (2007j:60031)

[3] Z. D. Bai and Jack W. Silverstein, *CLT for linear spectral statistics of large-dimensional sample covariance matrices*, Ann. Probab. 32 (2004), no. 1A, 553–605. MR2040792 (2005b:60046)

[4] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, Second, Springer Series in Statistics, Springer, New York, 2010. MR2567175 (2011d:60014)

[5] Dominique Bakry, *Functional inequalities for Markov semigroups*, Probability measures on groups: recent directions and trends, 2006, pp. 91–147. MR2213477 (2007g:60086)

[6] C. W. J. Beenakker, *Random-matrix theory of quantum transport*, Rev. Mod. Phys. 69 (1997Jul), 731–808.

[7] G. Ben Arous and K. Dang, *On fluctuations of eigenvalues of random permutation matrices*, ArXiv e-prints (June 2011), available at [1106.2108](http://arxiv.org/abs/1106.2108)

[8] S. G. Bobkov, *Isoperimetric and analytic inequalities for log-concave probability measures*, Ann. Probab. 27 (1999), no. 4, 1903–1921. MR1742893 (2001h:60026)

[9] S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal. 10 (2000), no. 5, 1028–1052. MR1800062 (2002k:60028)

[10] A. A. Borovkov and S. A. Utev, *An inequality and a characterization of the normal distribution connected with it*, Teor. Veroyatnost. i Primenen. 28 (1983), no. 2, 209–218. MR700206 (84j:60029)

[11] Thierry Cabanal-Duvillard, *Fluctuations de la loi empirique de grandes matrices aléatoires*, Ann. Inst. H. Poincaré Probab. Statist. 37 (2001), no. 3, 373–402. MR1831988 (2002c:15040)

[12] M. Capitaine and M. Casalis, *Asymptotic freeness by generalized moments for Gaussian and Wishart matrices*. Application to beta random matrices, Indiana Univ. Math. J. 53 (2004), no. 2, 397–431. MR2060407 (2007k:60111)

[13] Sourav Chatterjee, *Fluctuations of eigenvalues and second order Poincaré inequalities*, Probab. Theory Related Fields 143 (2009), no. 1-2, 1–40. MR2449121 (2009j:60044)

[14] Benoît Collins, *Product of random projections, Jacobi ensembles and universality problems arising from free probability*, Probab. Theory Related Fields 133 (2005), no. 3, 315–344. MR2198815 (2007b:60046)

[15] Persi Diaconis and Steven N. Evans, *Linear functionals of eigenvalues of random matrices*, Trans. Amer. Math. Soc. 353 (2001), no. 7, 2615–2633. MR1828463 (2002d:60003)
[49] Hale F. Trotter, *Eigenvalue distributions of large Hermitian matrices; Wigner’s semicircle law and a theorem of Kac, Murdock, and Szegő*, Adv. in Math. 54 (1984), no. 1, 67–82. MR761763 (86c:60055)

[50] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR1964483 (2004e:90003)

[51] Eugene P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) 62 (1955), 548–564. MR0077805 (17,1097c)

[52] *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) 67 (1958), 325–327. MR0095527

**Department of Mathematics, University of Washington**

*E-mail address:* dumitriu@math.washington.edu

**Department of Mathematics, University of Washington**

*E-mail address:* paquette@math.washington.edu