Global spectrum fluctuations for the β-Hermite and β-Laguerre ensembles via matrix models

Ioana Dumitriu and Alan Edelman

October 10, 2005

Abstract

We study the global spectrum fluctuations for β-Hermite and β-Laguerre ensembles via the tridiagonal matrix models introduced in [11], and prove that the fluctuations describe a Gaussian process on monomials. We extend our results to slightly larger classes of random matrices.

1 Introduction

1.1 The Semicircle Law, deviations and fluctuations, numerically

The most celebrated theorem of random matrix theory, the Wigner semicircle law [44, 45], may be illustrated as in Figure 1 by histogramming the eigenvalues of a single random symmetric matrix using the simple MATLAB code (normalization omitted)

```
A = randn(n); S = (A + A′)/sqrt(8 ∗ n); a = hist(eig(S), [−1 : delta : 1]);
bar([−1 : delta : 1], a/(n ∗ delta))
```
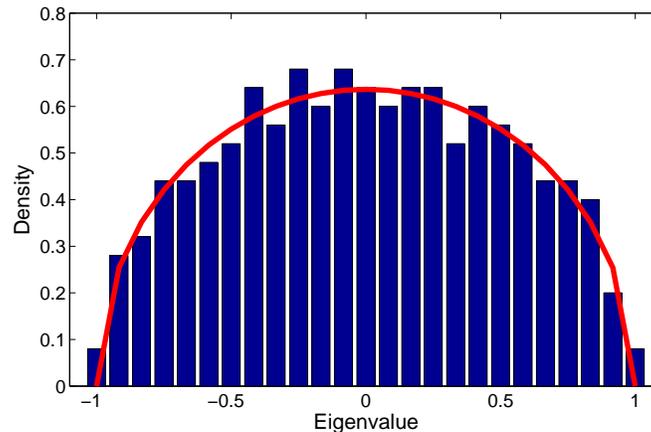

Figure 1: 25 bin-histogram of the eigenvalues of a 300×300 symmetric matrix from the real Gaussian distribution, versus the semicircle.

Every time we run this experiment, we obtain deviations from the semicircle. The difference from theory is readily explained computationally:

- we use a finite matrix size n, while the semicircle law is a theorem about the $n = \infty$ limit;

- histograms bin eigenvalues into boxes of finite width, while the semicircle density is a continuous function.

There can also be numerical error in the experiments due to finite precision computations and truncation error, but in practice this does not appear to be significant.

It is worth noting that the algorithm above is inefficient in two ways: first, it uses the full matrix A, rather than the equivalent tridiagonal matrix $H_{\beta,n}$ of Table 2 (with $\beta = 1$), and it calculates the eigenvalues to obtain the histogram. For more on how to obtain the histogram plot efficiently and without calculating the eigenvalues, see Section 5.

To study the next order behavior in the law, for large n, we can subtract away the semicircle and multiply by n. The next order *average* behavior is what we call the deviation and it was first computed by Johansson [21] to be

$$DEVIATION = \frac{1}{4}\delta_{-1}(x) + \frac{1}{4}\delta_1(x) - \frac{1}{2\pi}\sqrt{1-x^2} \; ; \tag{1}$$

here $\sim$ stands for "has the distribution of".

This expression for the deviation is the $\beta = 1$ (corresponding to real matrices) instance of the more general $\beta > 0$ case (which was also computed by Johansson, and will be explained in the next sections). The $\beta > 0$ deviation contains a $\left(\frac{2}{\beta} - 1\right)$ multiplicative factor in front of the expression on the right of (1), which disappears for $\beta = 1$.

One can see this deviation result as stating that as $n \to \infty$ the eigenvalues are decremented in the interior at a rate that is fastest at the center, pulling the eigenvalues toward the endpoints.

Upon further examination of the next-order term, one can observe a phenomenon not appearing in the leading order term; there are fluctuations around the mean.

Each time we run a trial we can compute this fluctuation. For example, if we have 25 bins, the random fluctuation vector $v = (v_1, \ldots, v_{25})^T$ is the difference between the count in each bin and the number of eigenvalues predicted by the semicircle plus the deviation. The entries of v can vary quite wildly, but inner products with discretized smooth functions result in normal distributions in the continuous limit. Specifically, if f_k is the vector of size 25 consisting of the evaluations of $f(x) = x^k$ on the centers of the bins, then the dot products $f_k^T v$ are heading towards Gaussians with covariances $E[(f_{k_1}^T v)(f_{k_2}^T v)]$.

Precise statements require the size of the matrix to go to infinity and the histogram to melt into a smooth density function so that the v vector becomes a Gaussian process.

Denote by $FLUCTUATION(f(x))$ the random quantity representing the limit obtained of $f^T v$ where, as above, f is the vector of function values and v is the vector of histogram differences. For smooth functions, $FLUCTUATION(f(x))$ converges to a normal distribution; for example, in the limit as $n \to \infty$,

$$FLUCTUATION(x^k) \sim 2N(0, \sigma_k^2) \; , \tag{2}$$

where

$$\sigma_k^2 = \begin{cases} \frac{k}{2^{2k}} \binom{k-1}{\frac{k-1}{2}}^2 \ , & \text{if } k = 1 \mod 2 \ , \\ \frac{k}{2^{2k+2}} \binom{k}{\frac{k}{2}}^2 \ , & \text{if } k = 0 \mod 2 \ . \end{cases}$$

For general β, the right side of (2) gains a multiplicative factor of $\frac{1}{\beta}$.

The limit of the entry (k_1, k_2) of the covariance matrix becomes the covariance between $FLUCTUATION(x^{k_1})$ and $FLUCTUATION(x^{k_2})$ (see Theorem 1.2 with $\beta = 1$).

1.2 β-Hermite and β-Laguerre ensembles

This paper studies deviations and fluctuations in a wider context than real symmetric matrices with semicircular asymptotic density: We consider Hermite and Laguerre matrices with general parameter $\beta > 0$. For a great reference for these ensembles, see Forrester's upcoming book [16].

Classical finite random matrix theory considers the study of eigenvalue ensembles with joint density

$$f(x_1, \ldots, x_n) = c_{w,n} \prod_{i<j} |x_i - x_j|^\beta \prod_{i=1}^n w(x_i) \ ,$$

with w a scalar weight function on an interval I. This interval may be a subinterval of the real line, or the unit circle in the complex plane; other possibilities have been considered, too, and generalizations are easily conceived. A good reference for these formulae can be found in Mehta's book [29].

Some of the most studied eigenvalue ensembles have Hermite, Laguerre, and Jacobi weight functions on the real line, or uniform weight on the unit circle. In this paper we will be examining the ensembles with Hermite and Laguerre weights on the real line (respectively, half-line); see Table 1.

For more references on Gaussian ensembles, see [29]; for Wishart and MANOVA ensembles, see [30]; for Hermite, Laguerre, and Jacobi ensembles, see [16].

For three particular values of β, namely $1, 2$, and 4, these ensembles have been studied since the birth of the field, as the Gaussian real, complex, and quaternion ensembles (Hermite with $\beta = 1, 2, 4$) of nuclear physics ([44], [45], [14], [1]). Similarly, the Wishart real and complex (Laguerre with $\beta = 1, 2$ and some restrictions on the Laguerre parameter) matrices emerged from the world of statistical multivariate analysis ([46], [5], [20], [24]).

The parameter β (making the connection to the Boltzmann factor of statistical physics) is seen by some communities (e.g. statistical mechanics) as an inverse temperature, or repulsion strength, of the ensemble of eigenvalues (the higher the β, the more separated the eigenvalues). It also has the advantage of the easy mnemonic of $1, 2$ and 4 corresponding to real, complex, and quaternion entries in the matrix models. However, some communities (like algebraic combinatorics) consider a different parameter, $\alpha = 2/\beta$, which tends to simplify certain formulas ([27], [36]). In this paper we will use both notations, for convenience, and make sure that the reader is informed when changes take place.

The reason for the attractivity and success that the study of Gaussian orthogonal, unitary, and symplectic (Hermite with $\beta = 1, 2, 4$), and the Wishart real and complex (Laguerre with $\beta = 1, 2$), ensembles have enjoyed lies in the existence of matrix models

Table 1: Random matrix ensembles with eigenvalue distribution proportional to $f(x_1,\ldots,x_n) \propto \prod_{i\neq j}|x_i - x_j|^\beta \prod w(x_i)$ with w defined on an interval $I \subseteq \mathbb{R}$. MANOVA stands for Multivariate ANalisys Of VAriance.

Name	*parameters*	*I*	*w*	*Historical name/constraints*
Hermite	$\boldsymbol{\beta > 0}$	$\mathbb{R}$	$w(x) = e^{-x^2/2}$	Gaussian $\beta = 1,2,4$
Laguerre	$\boldsymbol{\beta > 0}$ $a > (n-1)\frac{\beta}{2}$	$[0,\infty)$	$w(x) = x^p e^{-x/2}$ $p = a - (n-1)\frac{\beta}{2} - 1$	Wishart $\beta = 1,2$ $a = m\frac{\beta}{2},\, m \in \mathbb{N}$
Jacobi	$\boldsymbol{\beta > 0}$ $a_1, a_2 > (n-1)\frac{\beta}{2}$	$[0,1]$	$w(x) = x^{p_1}(1-x)^{p_2}$ $p_1 = a_1 - (n-1)\frac{\beta}{2} - 1$ $p_2 = a_2 - (n-1)\frac{\beta}{2} - 1$	MANOVA $\beta = 1,2$ $a_1 = m_1\frac{\beta}{2},\, m_1 \in \mathbb{N}$ $a_2 = m_2\frac{\beta}{2},\, m_2 \in \mathbb{N}$

with real, complex, and quaternion entries (see [29], [30]). These models have not only originated the study, but have also allowed for a relatively thorough analysis of the finite and asymptotical eigenstatistics (statistical properties of the eigenvalues).

One of the developments in the study of arbitrary β-Hermite, -Laguerre, and -Jacobi ensembles is the introduction of general *real* matrix models (see [11], [25], [38]). For every β, there are simple *real tridiagonal matrices* which model the corresponding eigenvalue distributions given by Table 1. For the β-Hermite and -Laguerre ensembles, we present these forms in Table 2. These matrix forms allow for efficient Monte Carlo experiments and an alternative representation for the study of eigenstatistics in the general β case.

This paper contains one example of such analysis. With the help these matrix models, we compute asymptotical global spectrum fluctuations for the β-Hermite and -Laguerre ensembles. For the latter, these results are new in the general β context; the global spectrum fluctuations of complex Wishart matrices (β-Laguerre with $\beta = 2$) were studied by Speicher et al. [26], and before by Cabanal-Duvillard [4]. Johansson's more extensive study [21] covers our results for the β-Hermite ensembles.

To prove our theorems, we use a very diverse set of methods and techniques from Jack Polynomial theory, special functions, perturbation theory, and combinatorial path-counting. Some of the techniques we used in studying the traces of powers of random matrices have been inspired by the work of Soshnikov and Sinai [33], [34], and by the study of traces of unitary random matrices by Diaconis and Shahshahani [9].

1.3   Statements of results

Among the asymptotical eigenstatistics one may distinguish two classes of properties: local (like the scaled distributions and fluctuations of extremal eigenvalues, or like the level spacing distribution, i.e., the distance between neighboring eigenvalues) and global (like

Table 2: **Unscaled and scaled** tridiagonal matrix models for the β-Hermite and β-Laguerre ensembles with any $\beta > 0$, $n \in \mathbb{N}$, $a \in \mathbb{R}$, and $a > \frac{\beta}{2}(n-1)$.

Hermite matrix, **unscaled**	$H_{\beta,n} \sim \frac{1}{\sqrt{2}} \begin{pmatrix} N(0,2) & \chi_{(n-1)\beta} & & & & \\ \chi_{(n-1)\beta} & N(0,2) & \chi_{(n-2)\beta} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \chi_{2\beta} & N(0,2) & \chi_{\beta} \\ & & & \chi_{\beta} & N(0,2) \end{pmatrix}$
Hermite matrix, **scaled**	$\tilde{H}_{\beta,n} = \frac{1}{\sqrt{2n\beta}} H_{\beta,n}$
Laguerre matrix, **unscaled**	$L^a_{\beta,n} = B^a_{\beta,n}(B^a_{\beta,n})^T$, where $$B^a_{\beta,n} \sim \begin{pmatrix} \chi_{2a} & & & \\ \chi_{\beta(n-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_{\beta} & \chi_{2a-\beta(n-1)} \end{pmatrix}$$
Laguerre matrix, **scaled**	$\tilde{L}^a_{\beta,n} = \frac{\gamma}{n\beta} L^a_{\beta,n}$

the limiting level density, i.e. the distribution of a random eigenvalue, and fluctuations thereof). The term global tends to refer to a property that involves all or a significant portion of the eigenvalues, while local tends to refer to a property that occurs near an individual or a constant number of eigenvalues. Among the more famous local properties we enumerate the level spacings for the Gaussian ensembles (see [29], [39]) and the extremal (largest, corresponding to "soft edge", smallest, to "hard edge") eigenvalue asymptotics (see [41], [40], [22], [23]). All these results relate to real, complex, or quaternion matrices; in the category of results relating to general β, we have to mention recent work by Desrosiers and Forrester [8], where they analyze the asymptotical corrections to the eigenvalue density, and, for $\beta \in 2\mathbb{N}$, obtain the expected $O(n^{2/3})$ order in the fluctuation of the largest eigenvalue in the case of both β-Hermite and -Laguerre ensembles.

For n (the size of the ensembles) finite, the n eigenvalues may be considered as n fluctuating particles. Roughly speaking, the larger the β, the less fluctuation there is in the particles (hence β is seen as an inverse temperature). As $\beta \to \infty$ the particle positions

behave like multivariate normals with variance $O(1/\beta)$ and means located at the roots of a Hermite (respectively Laguerre) polynomial (see [12]).

For a fixed β, as $n \to \infty$ and, in the case of Laguerre ensembles, $2a/(n\beta) \to \gamma$, the particles have an emerging (global) level density which is obeys a simple law (Wigner's semicircle law [45] for the Hermite ensembles, Marčenko-Pastur laws [28] for the Laguerre ensembles). The roots of the Hermite, respectively Laguerre, polynomial have this same asymptotical density – the fluctuations do not change the asymptotics, as they are on a smaller scale.

In [10], we have proved convergence almost surely (as $n \to \infty$) of the asymptotical eigenvalue distribution of the β-Hermite ensemble to the semicircle distribution S with density $\frac{2}{\pi}\sqrt{1-x^2}$, and of the asymptotical eigenvalue distribution of the β, a-Laguerre ensemble (with $n\beta/(2a) \to \gamma \leq 1$) to the Marčenko-Pastur E_γ distribution with density $\frac{1}{2\pi\gamma} \frac{\sqrt{(x-(\sqrt{\gamma}-1)^2)((\sqrt{\gamma}+1)^2-x)}}{x}$. We recall that convergence almost surely is stronger (and implies) convergence in distribution, a.k.a. convergence of moments.

We examine here the distribution of the statistic $\sum_{i=1}^{n} f(\lambda_i)$, where f is a function of the *scaled* eigenvalues λ_i. The scaling is $\lambda \to \sqrt{2n\beta}\lambda$ for the β-Hermite ensembles and $\lambda \to n\beta/\gamma\lambda$ for the β-Laguerre ensembles (see Table 2).

General β results for this kind of statistic can be found in [6], [21]; this or similar linear statistics have been considered also in [15] (for unitary matrices), [32] (for Wishart matrices), and, heuristically, in [31]. Another path of interest is represented by asymptotical large deviations from the density (spectral measure); we mention the results of [4], [2], [18] (which also covers global fluctuations for Wishart matrices) , [19], [7] (which covers moderate deviations).

For the linear statistic $\sum_{i=1}^{n} f(\lambda_i)$, the Wigner and Marčenko-Pastur laws for the β-Hermite and β-Laguerre ensembles state that for any "well-behaved" function f,

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \;\; \to \;\; \frac{2}{\pi}\int_{-1}^{1} f(t)\sqrt{1-t^2}\,dt \text{ , respectively,} \tag{3}$$

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \;\; \to \;\; \frac{1}{2\pi\gamma}\int_{(\sqrt{\gamma}-1)^2}^{(\sqrt{\gamma}+1)^2} f(t)\, \frac{\sqrt{(t-(\sqrt{\gamma}-1)^2)((\sqrt{\gamma}+1)^2-t)}}{t}\,dt \text{ ,} \tag{4}$$

where the convergence in the above is almost surely[1].

Examining the fluctuations from these laws takes us one step further. For f a polynomial, we prove that once we subtract the expected average over the limiting level density (i.e. the right hand sides of (3) and (4)), the rescaled resulting quantity tends asymptotically to a normal distribution with mean and variance depending on f. In other words, once the semicircle or Marčenko-Pastur distributions are subtracted, the fluctuations in the statistic $\sum_{i=1}^{n} f(\lambda_i)$ tend asymptotically to a Gaussian process on polynomials f.

How much more we can extend the class of functions that this process is well-defined on depends on the entries of the covariance matrix $C = \{C_{ij}\}$ (which can be expressed in any polynomial basis). One would be tempted to believe that a Gaussian process W

---

[1]Such a result is sometimes given the name of "Strong Law of Large Numbers"; if convergence is only in distribution, the name becomes "Law of Large Numbers".

defined on polynomials could, in principle, be extended to a class of continuous functions $h(x)$ with the property that, given a sequence of polynomials $p_n(x) \to h(x)$ in some norm, given $v_n = Var_W(p_n)$, $v_n \to \tilde{v} < \infty$, such that $\tilde{v} = Var_W(h)$.

Definition 1.1. *Let $\gamma \in [0, 1]$ be a real parameter, and let $a = (\sqrt{\gamma} - 1)^2$, $b = (\sqrt{\gamma} + 1)^2$. We define the following two measures:*

$$\mu_H(x) \;\; := \;\; \begin{cases} \frac{1}{4}\delta_1(x) + \frac{1}{4}\delta_{-1}(x) - \frac{1}{2\pi}\frac{1}{\sqrt{1-x^2}}, & \text{if } x \in [-1, 1] , \\ 0 , & \text{otherwise.} \end{cases} \;\; ,$$

$$\mu_L^\gamma(x) \;\; := \;\; \begin{cases} \frac{1}{4}\delta_b(x) + \frac{1}{4}\delta_a(x) - \frac{1}{2\pi}\frac{1}{\sqrt{(x-a)(b-x)}} , & \text{if } x \in [a, b] , \\ 0 , & \text{otherwise .} \end{cases}$$

Theorem 1.2. *Let $\tilde{H}_{\beta,n}$ be a scaled matrix from the β-Hermite ensemble of size n, with (scaled) eigenvalues $(\lambda_1, \ldots, \lambda_n)$, and let $k \geq 1$ be a positive integer. For all $i = 1, \ldots, k$, let*

$$X_i \;\; = \;\; tr((\tilde{H}_{\beta,n})^i) - n\,\frac{1}{4^{i/2}\left(\frac{i}{2}+1\right)}\binom{i}{\frac{i}{2}}\delta_{(i \bmod 2),0} - \left(\frac{2}{\beta}-1\right)\int_{-1}^1 t^i\mu_H(t)\; dt,$$

$$\equiv \;\; \sum_{j=1}^n \lambda_j^i \;\; - \;\; n\,\frac{2}{\pi}\int_{-1}^1 t^i\sqrt{1-t^2}\; dt \;\; - \;\; \left(\frac{2}{\beta}-1\right)\int_{-1}^1 t^i\mu_H(t)\; dt \;\; .$$

Let $(Y_1, Y_2, \ldots, Y_k)$ be a centered multivariate Gaussian with covariance matrix

$$Cov(Y_i, Y_j) = \begin{cases} \frac{1}{2^{i+j}}\frac{2ij}{i+j}\binom{i-1}{\frac{i-1}{2}}\binom{j-1}{\frac{j-1}{2}} , & \text{if } i = j = 1 \mod 2 ; \\ \frac{1}{2^{i+j+2}}\frac{2ij}{i+j}\binom{i}{\frac{i}{2}}\binom{j}{\frac{j}{2}} , & \text{if } i = j = 0 \mod 2 ; \\ 0 , & \text{otherwise.} \end{cases} \tag{5}$$

Then, as $n \to \infty$,

$$(X_1, X_2, \ldots, X_k) \Rightarrow \sqrt{\frac{2}{\beta}}\,(Y_1, Y_2, \ldots, Y_k) \; .$$

Remark 1.3. *For any size k, if we add a first row and a first column of zeros to the covariance matrix (5), the resulting $(k+1) \times (k+1)$ matrix C_k has a scaled Cholesky decomposition as $C_k = T_k D_k T_k^T$, where D_k is the diagonal matrix having on the diagonal the vector $(0, 1, 2, 3, \ldots, k)$, and T_k has an interpretation as the change-of-base matrix in the space of univariate polynomials from monomials basis to the Chebyshev polynomials basis. One can also look at the infinite version $C_\infty = T_\infty D_\infty T_\infty^T$.*

Remark 1.4. *This Gaussian process is extended to a larger class of continuous functions in [21]. Johansson conjectured that the regularity conditions imposed were purely technical, and that in fact the correct condition should be that the function h admits a Fourier-like expansion in the Chebyshev basis, which we write as an (infinite) vector $\vec{h}$, such that $\vec{h}^T D_\infty \vec{h}$, the variance of h under the process, is finite. This conjecture is equivalent to saying that the process could be extended to any function f such that $\vec{f}^T C_\infty \vec{f}$ is finite (since we have chosen the monomial basis as the representation, rather than the Chebyshev polynomial basis).*

Theorem 1.5. *Let $\tilde{L}^a_{\beta,n}$ be a scaled matrix from the β-Laguerre ensemble of parameter a and size n, with (scaled) eigenvalues $(\lambda_1, \ldots, \lambda_n)$, and let $k \geq 1$ be a positive integer. Assume that $n\beta/(2a) \to \gamma \leq 1$, and let $\gamma_{min} = (\sqrt{\gamma} - 1)^2$, $\gamma_{max} = (\sqrt{\gamma} + 1)^2$. For all $i = 1, \ldots, k$, let*

$$
\begin{aligned}
X_i &= tr((\tilde{L}^a_{\beta,n})^i) - n \sum_{r=0}^{i-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} \gamma^r - \left(\frac{2}{\beta} - 1\right) \int_{\gamma_{min}}^{\gamma_{max}} t^i \mu_L^\gamma(t) \; dt, \\
&\equiv \sum_{j=1}^n \lambda_j^i - n \frac{1}{2\pi\gamma} \int_{\gamma_{min}}^{\gamma_{max}} t^i \sqrt{(t - \gamma_{min})(\gamma_{max} - t)} \; dt - \left(\frac{2}{\beta} - 1\right) \int_{\gamma_{min}}^{\gamma_{max}} t^i \mu_L^\gamma(t) \; dt \; .
\end{aligned}
$$

Let $(Y_1, Y_2, \ldots, Y_k)$ be a centered multivariate Gaussian with covariance matrix

$$
Cov(Y_i, Y_j) = T_D(i,j) + T_S(i,j) \; , \tag{6}
$$

where

$$
T_D(i,j) = \sum_{q=1}^{i+j-1} (-1)^{q+1} \gamma^{i+j-q} \frac{\binom{i+j}{q}}{i+j} \sum_{l=q+1}^{i+j} \frac{(-1)^l}{\binom{i+j-1}{l-1}} \sum_{\substack{r+s=l \\ 1 \leq r \leq i \\ 1 \leq s \leq j}} rs \binom{i}{r}^2 \binom{j}{s}^2 \; ,
$$

and

$$
T_S(i,j) = \sum_{q=0}^{i+j-2} (-1)^q \gamma^{i+j-q} \frac{\binom{i+j}{q}}{i+j} \sum_{l=q}^{i+j-2} \frac{(-1)^l}{\binom{i+j-1}{l}} \sum_{\substack{r+s=l \\ 0 \leq r \leq i-1 \\ 0 \leq s \leq j-1}} (i-r)(j-s) \binom{i}{r}^2 \binom{j}{s}^2 \; .
$$

Then, as $n \to \infty$,

$$
(X_1, X_2, \ldots, X_k) \Rightarrow \sqrt{\frac{2}{\beta}} \; (Y_1, Y_2, \ldots, Y_k) \; .
$$

Remark 1.6. *For any size k, if we add a first row and a first column of zeros to the covariance matrix (6), the resulting matrix C should admit a scaled Cholesky decomposition of the form $C_k = T_k D_k T_k^T$, with D_k being the diagonal matrix with diagonal entries $\{j\gamma^{-j}\}_{0 \leq j \leq k}$, and T_k being the change-of-basis matrix in the space of polynomials from monomial basis to the* shifted Chebyshev *polynomials of the first kind (as defined by Cabanal-Duvillard [4] and used in [26]). Note that the constant c in [26] is our $1/\gamma$. It may also be useful to look at $C_\infty = T_\infty D_\infty T_\infty^T$.*

Remark 1.7. *It is worth noting that results like Theorems 1.2 and 1.5, where one averages over a set of quantities, then subtracts the mean and scales by the variance to obtain a limiting Gaussian, are sometimes called* **central limit theorems** *(see for example [32] and [33]). Free Probability uses this term as well, see for example [35], in a different context, namely, to express the fact that averaging over random matrices creates an eigenvalue distribution that approaches the semi-circular law. Both uses of the term "central limit theorem" draw different parallels to the classical case.*

Our approach to proving Theorems 1.2 and 1.5 consists of computing the first-order deviation from the mean (in Section 2), showing that the centered process is Gaussian on monomials (by the method of moments), and computing the covariance matrices (in Section 3).

Finally, in Section 4, we generalize our approach to two different classes of random matrices.

2   Deviation from the semicircle and Marčenko-Pastur laws

2.1   Dependence on β: symmetric functions and the "palindrome" effect

As stated in the introduction, we are interested in computing the deviation to the semicircle and Marčenko-Pastur laws (denoted below by $LAW(\infty)$, as opposed to the $LAW(n)$, which is the level density for finite n). These deviations have the form

$$LAW(n) \sim LAW(\infty) + \frac{1}{n}\left(1 - \frac{2}{\beta}\right) DEVIATION(\beta = \infty) + o\left(\frac{1}{n}\right) ,$$

as $n \to \infty$.

By integrating the above against x^k, we can write this in the moment form

$$moment_k(n) = moment_k(\infty) + \frac{1}{n}\left(1 - \frac{2}{\beta}\right) moment_k(DEVIATION(\beta = \infty)) + o\left(\frac{1}{n}\right) ,$$

again as $n \to \infty$.

We mention two interesting points, the first of which we prove in this section:

1. The factor $\frac{2}{\beta} - 1$ can be obtained from a symmetry principle alone. It is a direct consequence of Jack Polynomial theory that the coefficient of $1/n^j$ in $moment_k(n)$ is a *palindromic* polynomial (we define "palindromic" below) in $-\frac{2}{\beta}$, and from the tridiagonal matrix models it follows that the degree of this polynomial is j; thus when $j = 1$ this polynomial must be a multiple of $\frac{2}{\beta} - 1$. Mathematically, this is significant because because in order to study the deviations, it is sufficient to then study the non-random case, $\beta = \infty$. In summary, the powerful Jack Polynomial theory allows us to take a complicated random matrix problem and reduce it to an exercise on the properties of univariate Hermite and Laguerre polynomials.

2. With the Maple Library *MOPs* [13], we can compute symbolically the exact values of $moment_k(n)$ for small values of k, *as a function of n and β.* In other words, while this paper concerns itself with the constant and $O(1/n)$ behavior, it is worth remembering that higher order terms are in principle available to us.

To make notation a bit clearer, we have used the $\alpha = 2/\beta$ in the below; we also recall the scaled matrices $\tilde{H}_{\beta,n}$ and $\tilde{L}_{\beta,n}^a$, from Section 1.2. The following are the first three non-trivial moments for the traces of the scaled β-Hermite and β-Laguerre matrices with

$a = n\beta/(2\gamma) = n/(\alpha\gamma)$. We omit the n and a in the notation for reasons of space.

$$
\begin{aligned}
\frac{1}{n}\, E[\mathrm{tr}(H_{2/\alpha}^2)] &= \frac{1}{4} + \frac{\alpha-1}{4n}, \\
\frac{1}{n}\, E[\mathrm{tr}(H_{2/\alpha}^4)] &= \frac{2}{16} + \frac{5\alpha-5}{16n} + \frac{3\alpha^2-5\alpha+3}{16n^2}, \\
\frac{1}{n}\, E[\mathrm{tr}(H_{2/\alpha}^6)] &= \frac{5}{64} + \frac{11\alpha-11}{32n} + \frac{16\alpha^2-27\alpha+16}{32n^2} + \frac{15\alpha^3-32\alpha^2+32\alpha-15}{64n^3}, \\
\frac{1}{n}\, E[\mathrm{tr}(L_{2/\alpha})] &= 1\,, \\
\frac{1}{n}\, E[\mathrm{tr}(L_{2/\alpha}^2)] &= (1+\gamma) + \frac{\gamma(\alpha-1)}{n}\,, \\
\frac{1}{n}\, E[\mathrm{tr}(L_{2/\alpha}^3)] &= (1+3\gamma+\gamma^2) + \frac{3\gamma(\gamma+1)(\alpha-1)}{n} + \frac{\gamma^2(2\alpha^2-3\alpha+2)}{n^2}\,.
\end{aligned}
$$

Note that the $O(1)$ terms in the above correspond to the second, fourth, and sixth moments of the semicircle, in the Hermite case, respectively, to the first, second, and third moments of the Marčenko-Pastur distributions in the Laguerre case; these are Catalan numbers (scaled down by powers of 4 because of the semicircle $[-1, 1]$ normalization), respectively Narayana polynomials in γ.

The $O(1/n)$ terms, the moments of the deviation, are always multiplied by $\alpha - 1$, while the other coefficients of the negative powers of n in the above are "palindromic polynomials" of $(-\alpha)$; we recall the definition below.

Definition 2.1. *A classical "palindromic polynomial" is defined by the fact that its list of coefficients, is the same whether read from beginning to end or from end to beginning.*

Remark 2.2. *An odd-degree palindromic polynomial in x is a multiple of $(x + 1)$.*

To prove that the dependence of the first-order term in the deviation is indeed a multiple of $\alpha - 1$, i.e., of $\frac{2}{\beta} - 1$, we will use elements of Jack polynomial theory, and also a stronger form of a duality principle proved in [10].

We introduce below two notational conventions to be used throughout the rest of the paper.

Definition 2.3. *We denote by $E_\delta^H[P(x_1,\ldots,x_s)]$, respectively $E_{\delta,a}^L[P(x_1,\ldots,x_s)]$, the expectations of the polynomial P over the **scaled** $2/\delta$-Hermite, respectively $2/\delta, a$-Laguerre, ensembles of size s.*

*We denote by $\mathcal{E}_\delta^H[P(x_1,\ldots,x_s)]$, respectively $\mathcal{E}_{\delta,a}^L[P(x_1,\ldots,x_s)]$, the expectations of the polynomial P over the **unscaled** $2/\delta$-Hermite, respectively $2/\delta, a$-Laguerre, ensembles of size s.*

Let $\mathbb{R}[x_1,\ldots,x_n]$ be the space of symmetric polynomials in n variables (by symmetric we mean invariant under any permutation of the variables). A homogeneous basis for this vector space is a set of linearly independent, symmetric and homogeneous polynomials which generate $\mathbb{R}[x_1,\ldots,x_n]$. One such basis is given by the power-sum functions, defined multiplicatively below. For reference, see [37] and [27].

Definition 2.4. *Let $\lambda \equiv (\lambda_1, \lambda_2, \ldots, \lambda_n)$ denote an ordered partition $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n)$. We define the power sum functions by*

$$
\begin{aligned}
p_{\lambda_i} &= \sum_{j=1}^{n} x_j^{\lambda_i} , \quad\quad and \\
p_\lambda &= p_{\lambda_1} p_{\lambda_2} \ldots p_{\lambda_n} .
\end{aligned}
$$

The Jack polynomials J_λ^α constitute a parameter-dependent (the parameter being usually denoted by α) class of orthogonal multivariate polynomials; they are indexed by the powers of the highest-order term λ (in lexicographical ordering).

Throughout this section, we will think of the parameter α as a sort of inverse to β (recall that we have denoted $\alpha = 2/\beta$).

The Jack polynomials allow for several equivalent definitions (up to certain normalization constraints). We will work here with definition 2.5, which arose in combinatorics. We follow Macdonald's book [27].

Definition 2.5. *The Jack polynomials J_λ^α are orthogonal with respect to the inner product defined below on power-sum functions*

$$
\langle p_\lambda, p_\mu \rangle_\alpha = \alpha^{l(\lambda)} z_\lambda \delta_{\lambda\mu},
$$

where $z_\lambda = \prod_{i=1}^{l(\lambda)} a_i! i^{a_i}$, a_i being the number of occurrences of i in λ. In addition, the coefficient of the lowest-order term in J_λ^α, which corresponds to the partition $[1^{|\lambda|}] \equiv (1, 1, \ldots, 1)$ (of length $|\lambda|$), is $|\lambda|!$.

From now on, we will use the notations I_n for the vector of n ones and we will refer to the quantity $J_\kappa^\alpha(x_1, \ldots, x_n)/J_\kappa^\alpha(I_n)$ as the *normalized* Jack polynomial.

To prove our results, we will need the two lemmas below, the first of which is a stronger variant of the duality principle proved in [10] as Theorem 8.5.3 (the proof is virtually the same as in [10] and we will not repeat it here). The second one is a rewrite of a particular case ($z_a = 0$ and formula (4.14a)) of formula (4.36b) in [3].

Lemma 2.6. *Let κ' denote the conjugate partition to κ (obtained by transposing the rows and columns in the Young tableau). Then the following is true:*

$$
\mathcal{E}_\alpha^H \left[\frac{J_\kappa^\alpha(x_1, \ldots, x_n)}{J_\kappa^\alpha(I_n)} \right] = (-\alpha)^{-k/2} \mathcal{E}_{1/\alpha}^H \left[\frac{J_{\kappa'}^{1/\alpha}(y_1, \ldots, y_m)}{J_{\kappa'}^{1/\alpha}(I_m)} \right] .
$$

Lemma 2.7. *The following identity is true:*

$$
\mathcal{E}_{\alpha,a}^L \left[\frac{J_\kappa^\alpha(x_1, \ldots, x_n)}{J_\kappa^\alpha(I_n)} \right] = \prod_{(x,y)\in\kappa} \left(a - \frac{x}{\alpha} + y \right) .
$$

In particular, for $a = \frac{n}{\alpha\gamma}$,

$$
\mathcal{E}_{\alpha,\frac{n}{\gamma\alpha}}^L \left[\frac{J_\kappa^\alpha(x_1, \ldots, x_n)}{J_\kappa^\alpha(I_n)} \right] = \prod_{(x,y)\in\kappa} \left(\frac{n}{\alpha\gamma} - \frac{x}{\alpha} + y \right) .
$$

11

We can now prove the two main results of this section.

Theorem 2.8. *For k an even integer, let*

$$\mathcal{E}_\alpha^H \left[p_{[k]}(x_1, \ldots, x_n) \right] = \sum_{j=0}^{\frac{k}{2}} f(\alpha, j) \ n^{\frac{k}{2}+1-j} \ ,$$

with $p_{[k]}$ being the power-sum corresponding to partition $[k]$. Then $f(\alpha, j)$ is an integer-coefficient polynomial in $1/\alpha$ of degree at most $k/2$ such that

$$f(\alpha, j) = (-\alpha)^{-k+j} f(1/\alpha, j) \ .$$

Remark 2.9. *When we scale the ensembles, we have to multiply the expectation by $(2n\beta)^{-k/2} = \left(\frac{\alpha}{4n}\right)^{k/2}$, which means that*

$$\frac{1}{n} E_\alpha^H \left[p_{[k]}(x_1, \ldots, x_n) \right] = \frac{1}{4^{k/2}} \sum_{j=0}^{\frac{k}{2}} \alpha^{k/2} f(\alpha, j) \ n^{-j} \ .$$

The corollary below follows.

Corollary 2.10. *It follows that $g(\alpha, j) := \alpha^{k/2} f(\alpha, j)$ is an integer-coefficient polynomial of α for which*

$$g(\alpha, j) = (-\alpha)^j g(\frac{1}{\alpha}, j) \ ,$$

so the degree of $g(\alpha, j)$ is at most j. This yields, in particular, $g(\alpha, 1) = c_k(\alpha - 1)$, with c_k a constant depending on k.

Similarly, for the Laguerre ensembles, we have the following theorem.

Theorem 2.11. *Let $a = \frac{n}{\alpha\gamma}$, and*

$$\mathcal{E}_{\alpha,a}^L \left[p_{[k]}(x_1, \ldots, x_n) \right] = \sum_{j=1}^{k+1} \sum_{r=0}^{j-1} f(\alpha, j, r) \ \frac{n^j}{\gamma^r} \ ,$$

with $p_{[k]}$ being the power-sum corresponding to partition $[k]$. Then $f(\alpha, j, r)$ is a polynomial in $1/\alpha$ of degree at most k such that

$$f(\alpha, j, r) = (-\alpha)^{-k-j+1} f(1/\alpha, j, r) \ .$$

Remark 2.12. *When we scale the ensembles, we have to multiply the expectation by $\left(\frac{\gamma}{n\beta}\right)^{-k} = \left(\frac{\gamma\alpha}{2n}\right)^k$, which means that*

$$\frac{1}{n} E_{\alpha,a}^L \left[p_{[k]}(x_1, \ldots, x_n) \right] = \frac{1}{2^k} \sum_{j=1}^{k+1} \sum_{r=0}^{j-1} \alpha^k f(\alpha, j, r) \ n^{-k+j-1} \ \gamma^{k-r} \ .$$

The corollary below follows.

12

Corollary 2.13. *It follows that $g(\alpha, j, r) = \alpha^k f(\alpha, j, r)$ is an integer-coefficient polynomial of α for which*

$$g(\alpha, j, r) = (-\alpha)^{k-j+1} g(1/\alpha, j, r) ,$$

so that the degree of $g(\alpha, j, r)$ is at most $k - j + 1$. This yields, in particular, $g(\alpha, k, r) = c_{k,r}(\alpha - 1)$.

Proof of Theorem 2.8. Note that $1/\alpha = 2/\beta$ and that $p_{[k]}(x_1, \ldots, x_n) = \mathrm{tr}(X^k)$, for any matrix X with eigenvalues $x_1, \ldots, x_n$. By using the unscaled matrix model $H_{\beta,n}$ for the β-Hermite ensembles found in Table 2, one can obtain, as in Application 3 of [11] (more precisely, from Corollary 4.3), that $\mathcal{E}_\alpha^H[p_{[k]}(x_1, \ldots, x_n)]$ is an integer-coefficient polynomial in $1/\alpha$ of degree $k/2$, and a polynomial in n of degree $k/2 + 1$. Hence $f(\alpha, j)$ is an integer-coefficient polynomial in $1/\alpha$ of degree at most $k/2$.

Let us now express $p_{[k]}$ in Jack polynomial basis:

$$p_{[k]} = \sum_{\lambda \vdash k} c_\lambda(\alpha) \, J_\lambda^\alpha , \tag{7}$$

omitting the variables for simplicity.

Let $\mathbb{Q}(\alpha)$ be the field of all rational functions of α with rational coefficients.

Let $\Lambda \times \mathbb{Q}(\alpha)$ be the vector space of all symmetric polynomials of bounded degree with coefficients in $\mathbb{Q}(\alpha)$.

For every $0 \neq \theta \in \mathbb{Q}(\alpha)$, define the $\mathbb{Q}(\alpha)$-algebra automorphism $\omega_\theta : \Lambda \times \mathbb{Q}(\alpha) \to \Lambda \times \mathbb{Q}(\alpha)$ by the condition $\omega_\theta(p_k) = (-1)^{k-1} \theta p_k$, for all $k \geq 1$. This family of automorphisms appears in [27, Chapter 10], and similarly in [36]. In particular, $\omega = \omega_1$ is known as the Macdonald involution (and can be found in [27, Chapter 1]).

We will use the following formula due to Stanley [36] which can also be found as formula (10.24) in [27]:

$$\omega_\alpha J_\kappa^\alpha = \alpha^{|\kappa|} J_{\kappa'}^{1/\alpha} , \tag{8}$$

where λ' is the conjugate partition of λ (obtained from λ by transposing rows and columns in the Young tableau).

The first step of the proof is given by the following lemma.

Lemma 2.14. *The coefficients in equation (7) satisfy*

$$c_\lambda(\alpha) = (-\alpha)^{1-k} c_{\lambda'}(1/\alpha) .$$

Proof. We apply ω_α to both sides of (7), and use the fact that ω_α is linear, together with (8), to obtain that, on one hand,

$$\omega_\alpha p_{[k]} = (-1)^{k-1} \alpha p_{[k]} = (-1)^{k-1} \alpha \sum_{\lambda \vdash k} c_\lambda(1/\alpha) J_\lambda^{1/\alpha} ,$$

and

$$\omega_\alpha p_{[k]} = \sum_{\lambda \vdash k} c_\lambda(\alpha) \alpha^k J_{\lambda'}^{1/\alpha} ,$$

13

on the other hand.

Since there is a unique way of writing $\omega_\alpha p_{[k]}$ in Jack polynomial basis, it follows that

$$c_\lambda(\alpha) = (-\alpha)^{1-k} c_{\lambda'}(1/\alpha) \ .$$

$\square$

Remark 2.15. *Note that Lemma 2.14 does not say anything about expectations.*

We now write the expectation of $p_{[k]}$ over the unscaled Hermite ensemble using (7):

$$\mathcal{E}_\alpha^H \left[p_{[k]}(x_1, \ldots, x_n) \right] = \sum_{\lambda \vdash k} c_\lambda(\alpha) \ J_\lambda^\alpha(I_n) \ \mathcal{E}_\alpha^H \left[\frac{J_\lambda^\alpha(x_1, \ldots, x_n)}{J_\lambda^\alpha(I_n)} \right] \ .$$

We know (for example from [36, page]) that $J_\lambda^\alpha(I_n) = \prod_{(x,y)\in\lambda}(m - x + \alpha y)$; hence

$$\mathcal{E}_\alpha^H \left[p_{[k]}(x_1, \ldots, x_n) \right] = \sum_{\lambda \vdash k} c_\lambda(\alpha) \ \mathcal{E}_\alpha^H \left[\frac{J_\lambda^\alpha(x_1, \ldots, x_n)}{J_\lambda^\alpha(I_n)} \right] \prod_{(x,y)\in\lambda} (n - x + \alpha y) \ . \quad (9)$$

By Lemma 2.6, $\mathcal{E}_\alpha^H \left[\frac{J_\lambda^\alpha(x_1,\ldots,x_n)}{J_\lambda^\alpha(I_n)} \right]$ does *not* depend on n. Write

$$\prod_{(x,y)\in\lambda} (n - x + \alpha y) = \sum_{j=0}^{|\lambda|} b_\lambda(j, \alpha) n^j \ ;$$

we have

$$\prod_{(x,y)\in\lambda} (n - x + \alpha y) = \prod_{(y,x)\in\lambda'} (n + \alpha(y - \frac{x}{\alpha})) \ ,$$

and consequently

$$b_\lambda(j, \alpha) = (-\alpha)^{k-j} b_{\lambda'}(j, 1/\alpha) \ . \quad (10)$$

Using the above, (9), Lemma 2.14, Lemma 2.6, and substituting $\frac{k}{2} + 1 - j$ for j (in the power index of n), we obtain the statement of Theorem 2.8. $\square$

Proof of Theorem 2.11. The fact that $f(\alpha, j, r)$ is a polynomial in $1/\alpha$ of degree k follows similarly to Corollary 4.3 in Application 3 of [11].

We write

$$\begin{aligned} \mathcal{E}_{\alpha,a}^L \left[p_{[k]}(x_1, \ldots, x_n) \right] &= \sum_{\lambda \vdash k} c_\lambda(\alpha) \ \mathcal{E}_{\alpha,a}^L \left[\frac{J_\lambda^\alpha(x_1, \ldots, x_n)}{J_\lambda^\alpha(I_n)} \right] \prod_{(x,y)\in\lambda} (n - x + \alpha y) \ , (11) \\ &= \sum_{\lambda \vdash k} c_\lambda(\alpha) \prod_{(x,y)\in\lambda} \left(\frac{n}{\alpha\gamma} - \frac{x}{\alpha} + y \right) \prod_{(x,y)\in\lambda} (n - x + \alpha y) \ . (12) \end{aligned}$$

If we write

$$\prod_{(x,y)\in\lambda} \left(\frac{n}{\alpha\gamma} - \frac{x}{\alpha} + y \right) = \sum_{j=0}^{|\lambda|} \tilde{b}_\lambda(j, \alpha) n^j \gamma^{-j} \ ,$$

it is not hard to see that

$$\tilde{b}_\lambda(j, \alpha) = (-\alpha)^{-k-j} \ \tilde{b}_{\lambda'}(j, 1/\alpha) \ . \quad (13)$$

Using (10), (13), and Lemma 2.14, we obtain the statement of Theorem 2.11. $\square$

14

2.2   Computing the β-independent part of the deviation

In this section we will examine the deviation at $\beta = \infty$ for the Hermite and Laguerre ensembles. In proving this we employ a simple differential equations trick that will allow us to compute the zero and first order terms in the mean of the eigenvalue distributions at $\beta = \infty$.

Given a function $y(x)$ which satisfies the second order homogeneous differential equation

$$f(x)y''(x) + g(x)y'(x) + h(x)y(x) = 0 \ ,$$

denote by $m(x)$ the function $m(x) = y'(x)/y(x)$ (with poles at the zeroes of $y(x)$).

Proposition 2.16. *The function $m(x)$ satisfies the first-order differential algebraic equation*

$$m^2(x) + \frac{g(x)}{f(x)}m(x) + \frac{h(x)}{f(x)} + m'(x) = 0 \ .$$

The proof is immediate.

Remark 2.17. *If the function $y(x)$ is a polynomial with a finite number k of distinct roots, $m(x)$ is the generating function for the powers of y's roots.*

Let $\tilde{H}_{\beta,n}$ and $\tilde{L}^a_{\beta,n}$ denote matrices from the β-Hermite, respectively, β-Laguerre ensembles. As in [10], to obtain the deviation, we once again will examine the averaged traces of powers of the matrices $\tilde{H}_{\beta,n}$ and $\tilde{L}^a_{\beta,n}$, this time looking at the first-order terms.

As a consequence of Corollaries 2.10 and 2.13, for k an even positive integer in the Hermite case, and k an arbitrary positive integer in the Laguerre case with parameter $a = (n\beta)/(2\gamma)$,

$$\frac{1}{n}E^H_\beta\left[p_{[k]}(x_1,\ldots,x_n)\right] = c_k + c^1_k\frac{\frac{2}{\beta}-1}{n} + O(n^{-2}) \ ,$$

while

$$\frac{1}{n}E^{L,a}_\beta\left[p_{[k]}(x_1,\ldots,x_n)\right] = c_k(\gamma) + c^1_k(\gamma)\frac{\frac{2}{\beta}-1}{n} + O(n^{-2}) \ .$$

It follows that the deviation is given by the moments c^1_k, respectively, $c^1_k(\gamma)$, times the scaling factor $\frac{2}{\beta} - 1$. Equivalently, if we examine the generating functions

$$
\begin{aligned}
m(n,\beta,x) &= \frac{1}{n}E^H_\beta\left[\sum_{i=1}^n \frac{1}{x - \lambda_i}\right] \\
&= \frac{1}{n}\sum_{k=0}^\infty \frac{E^H_\beta\left[p_{[k]}(x_1,\ldots,x_n)\right]}{x^{k+1}} \\
&= \sum_{k=0}^\infty \left(\frac{c_k}{x^{k+1}} + \frac{1}{n}\left(\frac{2}{\beta}-1\right)\frac{c^1_k}{x^{k+1}} + \ldots\right)
\end{aligned}
$$

and

$$
\begin{aligned}
m(n, a, \beta, x) &= \frac{1}{n} E_{\beta,a}^L \left[\sum_{i=1}^{n} \frac{1}{x - \lambda_i} \right] \\
&= \frac{1}{n} \sum_{k=0}^{\infty} \frac{E_{\beta,a}^L \left[p_{[k]}(x_1, \ldots, x_n) \right]}{x^{k+1}} \\
&= \sum_{k=0}^{\infty} \left(\frac{c_k}{x^{k+1}} + \frac{1}{n} \left(\frac{2}{\beta} - 1 \right) \frac{c_k^1(\gamma)}{x^{k+1}} + \ldots \right)
\end{aligned}
$$

then in order to find the first order asymptotics of $m(n, \beta, x)$ and $m(n, a, \beta, x)$, it is enough to compute $\sum_{k=0}^{\infty} \frac{c_k^1}{x^{k+1}}$, respectively, $\sum_{k=0}^{\infty} \frac{c_k^1(\gamma)}{x^{k+1}}$.

We will do this by keeping n fixed, letting $\beta \to \infty$, and computing the zero and first order terms in $\frac{1}{n}$ in the generating function of the resulting matrix ensemble.

Remark 2.18. *Knowing that the quantities $\frac{1}{n} E_\beta^H \left[p_{[k]}(x_1, \ldots, x_n) \right]$ and $\frac{1}{n} E_\beta^{L,a} \left[p_{[k]}(x_1, \ldots, x_n) \right]$ are polynomial expressions in $\frac{1}{n}$ and $\frac{2}{\beta}$ (which follows immediately by rescaling the results of Theorems 2.8 and 2.11) is the key factor in allowing us to let $\beta \to \infty$ while keeping n fixed.*

To aid us in our calculations, we will make use of a beautiful property of the χ_r distribution, which will allow us to replace our tridiagonal models with even simpler ones. We give this property as a Proposition.

Proposition 2.19. *Let $\{X_n\}_{n \in \mathbb{N}}$ be a set of random variables with distributions χ_{r_n}, $n \in \mathbb{N}$, such that $r_n \to \infty$ as $n \to \infty$. Then the sequence $\{X_n - \sqrt{r_n}\}_{n \in \mathbb{N}}$ converges in distributions to a centered normal of variance $1/2$.*

It follows then that, as we have already shown in [12], that given n and $\gamma \leq 1$ fixed, the entries of the scaled random matrices $\tilde{H}_{\beta,n}$ and $\tilde{L}_{\beta,n}^a = B_{\beta,n}^a (B_{\beta,n}^a)^T$ (with $a = (n\beta)/(2\gamma)$) converge as $\beta \to \infty$ to the following models:

$$
\tilde{H}_{\beta,n} \to H = \frac{1}{2\sqrt{n}}
\begin{pmatrix}
0 & \sqrt{n-1} & & & & \\
\sqrt{n-1} & 0 & \sqrt{n-2} & & & \\
& \sqrt{n-2} & 0 & & & \\
& & & \ddots & & \\
& & & & 0 & \sqrt{1} \\
& & & & \sqrt{1} & 0
\end{pmatrix}.
\tag{14}
$$

respectively to

$$
\tilde{L}_{\beta,n}^a \to L_\gamma = \frac{\gamma}{n}
\begin{pmatrix}
\frac{n}{\gamma} & \sqrt{\frac{n}{\gamma}}\sqrt{n-1} & & & & \\
\sqrt{\frac{n}{\gamma}}\sqrt{n-1} & \frac{n}{\gamma}+n-2 & \sqrt{\frac{n}{\gamma}-1}\sqrt{n-2} & & & \\
& \sqrt{\frac{n}{\gamma}-1}\sqrt{n-2} & \frac{n}{\gamma}+n-3 & & & \\
& & & \ddots & & \\
& & & & \sqrt{\frac{n}{\gamma}-n+2}\sqrt{2} & \\
& & & \sqrt{\frac{n}{\gamma}-n+2}\sqrt{2} & \frac{n}{\gamma}-n+4 & \sqrt{\frac{n}{\gamma}-n+1}\sqrt{1} \\
& & & & \sqrt{\frac{n}{\gamma}-n+1}\sqrt{1} & \frac{n}{\gamma}-n+2
\end{pmatrix}
\tag{15}
$$

16

Note also that if we denote by $\tilde{B}^a_{\beta,n}$ the matrix such that $\tilde{L}^a_{\beta,n} = \tilde{B}^a_{\beta,n}(\tilde{B}^a_{\beta,n})^T$, then

$$
\tilde{B}^a_{\beta,n} \to B_\gamma = \sqrt{\frac{\gamma}{n}}
\begin{pmatrix}
\sqrt{\frac{n}{\gamma}} & & & & \\
\sqrt{n-1} & \sqrt{\frac{n}{\gamma}-1} & & & \\
& \sqrt{n-2} & \sqrt{\frac{n}{\gamma}-2} & & \\
& & \ddots & & \\
& & & \sqrt{2} & \sqrt{\frac{n}{\gamma}-n+2} \\
& & & & \sqrt{1} & \sqrt{\frac{n}{\gamma}-n+1}
\end{pmatrix}
\tag{16}
$$

and that $L_\gamma = B_\gamma B_\gamma^T$.

Since H, L_γ, and b_γ are **non-random** matrices, all expectations are exact.

2.2.1 $\beta = \infty$, Hermite case

The matrix H (see (14)) has as eigenvalues $h_1/\sqrt{2n}, \ldots, h_n/\sqrt{2n}$, where $h_1, \ldots, h_n$ are the roots of the nth Hermite polynomial $H_n(x)$ (this can be easily deduced from the three-term recurrence for the Hermite polynomials, see for example [42]). For a more detailed description of the properties of this matrix, see [12].

It follows that the generating function we need to compute is

$$
\tilde{m}(n, x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x - \frac{h_i}{\sqrt{2n}}} \ .
$$

We use the well-known identity

$$
\sum_{i=1}^{n} \frac{1}{x - x_i} = \frac{p'(x)}{p(x)} \ ,
\tag{17}
$$

where x_i are distinct values, and $p(x)$ is the polynomial whose roots the x_is are, to obtain that

$$
\tilde{m}(n, x) = \sqrt{\frac{n}{2}} \frac{H'_n(x\sqrt{2n})}{H_n(x\sqrt{2n})} \ ,
$$

and by applying Proposition 2.16, we get that $\tilde{m}(n, x)$ satisfies the differential equation

$$
(\tilde{m}(n, x))^2 - 4x\tilde{m}(n, x) + 4 + \frac{\tilde{m}'(n, x)}{n} = 0 \ .
\tag{18}
$$

Writing $\tilde{m}(n, x) = m_0(x) - \frac{1}{n}m_1(x) + O(n^{-2})$, we obtain from (18) that

$$
\begin{aligned}
m_0(x) &= 2\left(x - \sqrt{x^2 - 1}\right) \\
m_1(x) &= \frac{\left(x - \sqrt{x^2 - 1}\right)}{2(x^2 - 1)} \ .
\end{aligned}
$$

17

Computing the inverse Cauchy transform for $m_0(x)$ and $m_1(x)$ yields the semicircle distribution and, respectively

$$\mu_H(x) = \begin{cases} \frac{1}{4}(\delta_1(x) + \delta_{-1}(x)) - \frac{1}{2\pi}\frac{1}{\sqrt{1-x^2}}, & \text{if } x \in [-1,1] , \\ 0 , & \text{otherwise.} \end{cases} \tag{19}$$

We have thus proved the following result.

Lemma 2.20. *For any polynomial P,*

$$E_\beta^H\left[\sum_{i=1}^n P(x_i)\right] - n \int_{-1}^1 P(x)s(x)dx \longrightarrow \left(\frac{2}{\beta} - 1\right)\int_{-1}^1 P(x)\mu_H(x)dx ,$$

as $n \to \infty$.

Remark 2.21. *As a side note, in the computation above we have provided yet another way to obtain the semicircle law for all β.*

2.2.2 $\beta = \infty$, Laguerre case

The matrix L_γ has as eigenvalues $\gamma l_1/n, \ldots, \gamma l_n/n$, where $l_1, \ldots, l_n$ are the roots of the nth Laguerre polynomial $L_n^{n(1/\gamma - 1)}(x)$ (see [43]) (this can be easily deduced from one of the many recurrences for Laguerre polynomials, found for example as (26) in [43]). To get a more detailed description of the properties of this matrix, refer to [12], substituting $n(1/\gamma) - 1$ for γ.

Since we need to rescale the eigenvalues by an additional n, it follows that the quantity of interest is

$$\tilde{m}(n, \gamma, x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{x - \frac{\gamma l_i}{n}} ,$$

Once again we use identity 17 and obtain that

$$\tilde{m}(n, \gamma, x) = \frac{1}{n}\frac{\left(L_n^{n(1/\gamma - 1)}\right)'\left(\frac{xn}{\gamma}\right)}{L_n^{n(1/\gamma - 1)}\left(\frac{xn}{\gamma}\right)} ,$$

and by applying Proposition 2.16, we get that $\tilde{m}(n, \gamma, x)$ satisfies the algebraic differential equation

$$\gamma\left(\tilde{m}(n, \gamma, x)\right)^2 - \tilde{m}(n, \gamma, x)\left(1 - \frac{1}{x} + \frac{\gamma}{x}\right) + \frac{1}{x} + \frac{\gamma\tilde{m}'(n, \gamma, x) + \frac{\gamma\tilde{m}(n,\gamma,x)}{x}}{n} = 0 . \tag{20}$$

Writing $\tilde{m}(n, \gamma, x) = m_0(x) - \frac{1}{n}m_1(x) + O(n^{-2})$, we obtain from (20) that

$$m_0(x) = \frac{x + \gamma - 1 - \sqrt{(x - (\sqrt{\gamma} - 1)^2)(x - (\sqrt{\gamma} + 1)^2)}}{2\gamma x} ,$$

$$m_1(x) = \frac{x - \gamma - 1 - \sqrt{(x - (\sqrt{\gamma} - 1)^2)(x - (\sqrt{\gamma} + 1)^2)}}{2(x - (\sqrt{\gamma} - 1)^2)(x - (\sqrt{\gamma} + 1)^2)} .$$

By calculating the inverse Cauchy transforms of $m_0(x)$ and $m_1(x)$, one obtains the Marčenko-Pastur distribution, respectively

$$t_\gamma(x) = \begin{cases} \frac{1}{4}\delta_b(x) + \frac{1}{4}\delta_a(x) - \frac{1}{2\pi}\frac{1}{\sqrt{(x-a)(b-x)}} \,, & \text{if } x \in [a,b] \,, \\ 0 \,, & \text{otherwise} \,, \end{cases} \tag{21}$$

with $a = (\sqrt{\gamma} - 1)^2$, $b = (\sqrt{\gamma} + 1)^2$.

We have thus proved the following result.

Lemma 2.22. *For any polynomial P,*

$$E_{\beta,a}^L\left[\sum_{i=1}^n P(x_i)\right] - n \int_a^b P(x)e_\gamma(x)dx \longrightarrow \left(\frac{2}{\beta} - 1\right)\int_a^b P(x)t_\gamma(x)dx \,,$$

as $n \to \infty$.

3 Fluctuation of the semicircle and Marčenko-Pastur laws

In this section we compute the fluctuations terms for the β-Hermite and β-Laguerre ensembles; we show that the fluctuation of the trace of any given power of the matrix corresponding to the ensemble tends to a Gaussian.

The essence of the argument is simple. We will think of the random matrix as the sum between the non-random matrix of means (which we can also think about roughly as the $\beta = \infty$ non-random matrix), and a random matrix of the centered entries, and do some obvious computations of traces of powers. Much of the work goes into the technical carefulness to provide a complete argument, but the reader should not let this detract them from the simplicity of the idea, which is based on formula which gives, for a matrix T,

$$\text{tr}(T^k) = \sum_{i_1,\dots,i_k} t_{i_1 i_2} t_{i_2 i_3} \dots t_{i_k i_1} \,,$$

with the proviso that the above sum is especially simple when the matrix T is tridiagonal.

We provide here a heuristic explanation for the Hermite case, just for the purpose of emphasizing that the main idea is simple.

One can think of the random matrix, for all practical purposes, as

$$T \approx T_{\beta=\infty} + \frac{1}{\sqrt{\beta n}}G \,,$$

where G is symmetric tridiagonal matrix of $O(1)$-variance Gaussians (similar to the decomposition we used in [12]), with entries which are mutually independent, up to the symmetry condition.

Then for any polynomial h,

$$\text{tr}(h(T)) \approx \text{tr}(h(T_{\beta=\infty})) + \frac{1}{\sqrt{\beta n}}\text{tr}(h'(T_{\beta=\infty})G);$$

the second term is clearly normally distributed, and all we have to do is compute the variance and show it is finite (which we can achieve by examining the cases when h is a monomial).

It is worth noting that, in principle, one should be able to do this computation for continuously differentiable functions h with some additional conditions imposed by the fact that the variance needs to be finite.

The technicalities arise because $n \to \infty$ and the equalities above are just approximations, but this should not detract from the main idea. As we will see, using the method of moments will show that we do not need G's entries to be Gaussian (or even *approximately* Gaussian) in order for the *fluctuation* to monomials h to be Gaussian.

3.1 The β-Hermite case

We write the scaled matrix $\tilde{H}_{\beta,n}$ as

$$\tilde{H}_{\beta,n} = A + \frac{1}{\sqrt{n\beta}}Y \ , \tag{22}$$

where $A = E_\beta^H[\tilde{H}_{\beta,n}]$ is the symmetric tridiagonal matrix of mean entries ($A(i, i+1) = \frac{1}{2\sqrt{n\beta}}E[\chi_{(n-i)\beta}]$; all other entries are 0), and Y is symmetric tridiagonal matrix of centered variables (with a diagonal of independent Gaussians of variance $1/2$). Technically, A and Y depend both on n and on β; we drop these indices from notation for the sake of simplicity.

Remark 3.1. *From Proposition 2.19, we know that if $A = (a_{ij})_{1 \le i, j \le n}$, then for any $\epsilon > 0$, there exists $i_\epsilon \in \mathbb{N}$ such that*

$$\left(2\sqrt{n\beta} \ a_{i,i+1} - \sqrt{(n-i)\beta}\right) < \epsilon \ \ for \ any \ \ i \le (n - i_\epsilon) \ \ ,$$

and that $Y_{i,i+1} \to N(0, 1/8)$ in distribution as $i \to \infty$, while $Y_{i,i} \sim N(0, 1/2)$ for all i.

Remark 3.2. *(bounded moments) Note that the entries of A are **bounded**, both from below and from above; we will think of them as $O(1)$. Similarly, for any k and l finite, we know from the above that there exists an M such that*

$$E[\prod_{i=1}^{kl} Y_{j_i,j_i'}^{c_i}] \le M,$$

for all $0 \le c_i \le kl$, and for all $j_1, \ldots, j_{kl}$ and $j_1', \ldots, j_{kl}'$ such that $|j_i - j_i'| \le 1$.

Given integers k and n, consider the random variable

$$\omega_k(n) \quad = \quad \mathrm{tr}(\tilde{H}_{\beta,n}^k) - E_\beta^H[\mathrm{tr}(\tilde{H}_{\beta,n}^k)] \ .$$

Claim 3.2.1. *For any fixed integers k and l,*

$$\lim_{n \to \infty} E[(\omega_k(n))^l] = \begin{cases} \left(\frac{2}{\beta}\right)^{l/2} \frac{1}{2^{(k+1)l}} \ k^{l/2} \left(\frac{k}{2}\right)^l, & if \ k, \ l \ are \ even, \\ \left(\frac{2}{\beta}\right)^{l/2} \frac{1}{2^{kl}} \ k^{l/2} \left(\frac{k-1}{2}\right)^l, & if \ k \ is \ odd \ and \ l \ is \ even \ , \\ 0, & if \ l \ is \ odd. \end{cases}$$

Claim 3.2.2. *For any fixed integers k_1 and k_2,*

$$\lim_{n \to \infty} Cov(\omega_{k_1}(n), \omega_{k_2}(n)) = \begin{cases} \frac{2}{\beta} \frac{1}{2^{k_1+k_2}} \frac{2k_1 k_2}{k_1+k_2} \binom{k_1-1}{\frac{k_1-1}{2}} \binom{k_2}{\frac{k_2-1}{2}} , & \text{if } k_1, k_2 \text{ are odd,} \\ \frac{2}{\beta} \frac{1}{2^{k_1+k_2+2}} \frac{2k_1 k_2}{k_1+k_2} \binom{k_1}{\frac{k_1}{2}} \binom{k_2}{\frac{k_2}{2}} , & \text{if } k_1, k_2 \text{ are even,} \\ 0 , & \text{otherwise .} \end{cases}$$

Remark 3.3. *Claims 3.2.1 and 3.2.2 show that the centered fluctuation of the β-Hermite ensembles describes the Gaussian process on monomials defined in Theorem 1.2.*

For the remainder of this section, we will prove Claims 3.2.1 and 3.2.2. To give the reader a rough idea of where the calculations will lead, we provide below an intuition of what we will be doing.

Intuitive explanation. The first step is to note that $tr(T^k)$ is a sum of products of k entries of T; for a tridiagonal matrix $T = (t_{i,j})_{1 \le i,j \le n}$ with $t_{i,j} = 0$ if $|i - j| > 1$,

$$\text{tr}(T^k) = \sum_{1 \le i_1, \ldots, i_k \le n} t_{i_1, i_2} t_{i_2, i_3} \ldots t_{i_k, i_1} ,$$

where the sum needs to be taken only over the sequences $i_1, \ldots, i_k$ such that $|i_j - i_{j+1}| \le 1$, for all $j = 1, \ldots, k-1$, and also $|i_k - i_1| \le 1$.

We have a sliding "window" of size k down the diagonal of the matrix T in which we take products of powers of the elements. In particular, for the matrix A, this is easy to visualize, because with the exception of a finite bottom right corner, the entries of A in any finite window look roughly the same.

The second step is to identify the significant terms, i.e. the terms that have non-zero asymptotical contributions. Roughly speaking, these will be the terms which will contain *precisely* one element of Y, and all the others from A. Nothing surprising here, as Y is scaled by $\frac{1}{\sqrt{n\beta}}$ (see (22)).

Finally, we will compute the contribution from the significant terms and show it agrees with the result of Claim 3.2.1. Then we will note that the same reasoning yields the result of Claim 3.2.2.

We now proceed to make the above intuitive description rigorous. We need to introduce some notation.

Definition 3.4. *For given n and p, we denote by $\mathcal{S}_{n,p} \subset \{1, \ldots, n\}^p$ the set of sequences of integers $i_1, \ldots, i_p$ such that $(i_1, \ldots, i_p) \in \{1, \ldots, n\}^p$ and $|i_j - i_{j+1}| \le 1$ for all $j = 1, \ldots, p-1$, and also $|i_p - i_1| \le 1$.*

We denote by $\mathcal{I}$ an element of $\mathcal{S}_{n,p}$, and we denote by

$$(T)_{\mathcal{I}} := t_{i_1, i_2} t_{i_2, i_3} \ldots t_{i_p, i_1} ,$$

where $(i_1, i_2, \ldots, i_p) =: \mathcal{I}$.

For a given $\mathcal{I} \in \mathcal{S}_{n,p}$, note that we can break up the sequence $(i_1, \ldots, i_p)$ into concatenations of sequences

$$J = ((i_{p_0}, \ldots, i_{p_1}), (i_{p_2}, \ldots, i_{p_3}), (i_{p_4} \ldots i_{p_5}), \ldots, (i_{p_{2q}}, \ldots, i_{p_{2q+1}})) ,$$

and
$$R = ((i_{p_1}, \ldots, i_{p_2}), (i_{p_3}, \ldots, i_{p_4}), \ldots, (i_{p_{2q-1}}, \ldots, i_{p_{2q}})) \ ,$$

such that in each of the sequences $i_{p_{2k}}, i_{p_{2k}+1}, \ldots, i_{p_{2k+1}}$ (for $k = 0, \ldots, q$) which form J, consecutive indices differ by exactly 1, and in addition to this

$$(i_1, \ldots, i_p) = (i_{p_0}, \ldots, i_{p_1}, i_{p_1+1}, \ldots, i_{p_2}, i_{p_2+1}, \ldots, i_{p_3}, \ldots, i_{p_{2q+1}}) \ .$$

We allow for the possibility of having empty sequences $i_{p_0}, \ldots, i_{p_1}$ in the beginning and/or $i_{p_{2q}}, \ldots, i_{p_{2q+1}}$ in the end of J.

Remark 3.5. *To give an intuition for these sequences, note that any term in $tr((H_{\beta,n})^k)$ can be thought of in terms of products of entries from A and entries from Y; the sequences J and R will be overlapping "runs" recording the former, respectively the latter.*

Also note that, given a fixed $\mathcal{I}$, each J satisfying the requirements above has exactly one $R = R(J)$ corresponding to it, and that to different Rs correspond different Js. Furthermore, since k is finite, a given concatenation of sequences R may corresponds only to a finite number of sequences $\mathcal{I}$ (since all indices must be within k of each other!).

Definition 3.6. *We define the set $\mathcal{J}$ as the set of pairs (J, R) described above. For a tridiagonal matrix T, we define*

$$(T)_J = t_{i_{p_0}, i_{p_0}+1} \ldots t_{i_{p_1}-1, i_{p_1}} t_{i_{p_2} i_{p_2}+1} \ldots t_{i_{p_3}-1, i_{p_3}} \ldots t_{i_{2q}, i_{2q}+1} \ldots t_{i_{2q+1}-1, i_{2q+1}} \ ;$$

similarly,

$$(T)_R = t_{i_{p_1}, i_{p_1}+1} \ldots t_{i_{p_2}-1, i_{p_2}} t_{i_{p_3} i_{p_3}+1} \ldots t_{i_{p_4}-1, i_{p_4}} \ldots t_{i_{2q-1}, i_{2q-1}+1} \ldots t_{i_{2q}-1, i_{2q}}$$

For any sequence $\mathcal{I} \in \mathcal{S}_{n,p}$, we can write

$$(\tilde{H}_{\beta,n})_{\mathcal{I}} = (A + \frac{1}{\sqrt{n\beta}} Y)_{\mathcal{I}} = \sum_{(J,R) \in \mathcal{J}} \frac{1}{(n\beta)^{P/2}} (A)_J \ (Y)_R \ , \tag{23}$$

with P being the total length of the "runs" in the sequence R (i.e., $P = (p_2 - p_1 + 1) + (p_4 - p_3 + 1) + \ldots + (p_{2q} - p_{2q-1} + 1))$.

We have now enough information to start the proof of Claim 3.2.1.

Proof of Claim 3.2.1. First we examine

$$E[(\omega_k(n))^l] = E\left[\left(tr(\tilde{H}_{\beta,n}^k) - E_\beta^H[tr(\tilde{H}_{\beta,n}^k)]\right)^l\right] \ ;$$

note that

$$E[(\omega_k(n))^l] = E[\sum_{\mathcal{I}_1, \ldots, \mathcal{I}_l \in \mathcal{S}_{n,k}} \prod_{j=1}^{l} \left((\tilde{H}_{\beta,n})_{\mathcal{I}_j} - E[(\tilde{H}_{\beta,n})_{\mathcal{I}_j}]\right)] \ .$$

Using (23), we write

$$E[(\omega_k(n))^l] = E\left[\sum_{\substack{\mathcal{I}_j \in \mathcal{S}_{n,k} \\ 1 \le j \le l}} \sum_{(J_j, R_j) \in \mathcal{J}_j} \frac{1}{(n\beta)^{P_i/2}} \prod_{j=1}^{l} ((A)_{J_i}(Y)_{R_i} - E[(A)_{J_i}(Y)_{R_i}]) \right] \ . \tag{24}$$

22

Since A is a non-random matrix, it follows that we can rewrite (24) as

$$E[(\omega_k(n))^l] = \sum_{\substack{\mathcal{I}_j \in \mathcal{S}_{n,k} \\ 1 \leq j \leq l}} \sum_{(J_j,R_j) \in \mathcal{J}_j} \left(\prod_{j=1}^l \frac{1}{(n\beta)^{P_i/2}} (A)_{J_i} \right) E\left[\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}]) \right] \quad (25)$$

Denoting by $q := \sum_{i=1}^l P_i/2$, we obtain that

$$E[(\omega_k(n))^l] = \sum_{\substack{\mathcal{I}_j \in \mathcal{S}_{n,k} \\ 1 \leq j \leq l}} \sum_{(J_j,R_j) \in \mathcal{J}_j} \frac{1}{(n\beta)^q} \left(\prod_{j=1}^l (A)_{J_i} \right) E\left[\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}]) \right]. \quad (26)$$

Lemma 3.7. *The non-zero terms in (26) have $q \geq l/2$.*

Proof. If any of the l terms in the product $\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}])$ involves only variables that are independent from all other variables appearing in the remaining $l-1$ terms of the product, the expected value of the product is 0. This includes the case when at least of the R_is are empty. Hence, in all of the terms that have non-zero contribution to the expectation, each R_i must be nonempty, hence $2q = \sum_{i=1}^l P_i \geq l$. $\square$

Remark 3.8. *Note that in fact something stronger follows, namely, that for each $1 \leq i \leq l$, there exists an $1 \leq j \leq n$, $j \neq i$, such that the some variable appearing in $(Y)_{R_i}$ also appears in $(Y)_{R_j}$.*

Since the entries of A are $O(1)$, and k and l are fixed, the factors $\prod_{j=1}^l (A)_{J_i}$ are all $O(1)$. By Remark 3.8 and Lemma 3.7, it follows that every term with non-zero contribution in the double sum (26) corresponds to an l-tuple $(R_1, \ldots, R_l)$ with the property described in Remark 3.8. Each such l-tuple comes with a weight proportional to $1/n^q$.

To compute the asymptotics of the sum, we will do the following thought experiment: select a non-zero contribution term and draw the "correlation" graph with R_i as vertices, and an edge between R_i and R_j if and only if $(Y)_{R_i}$ and $(Y)_{R_j}$ are correlated. The resulting graph will have s connected components, with $1 \leq s \leq \lfloor \frac{l}{2} \rfloor$.

Call these connected components $C_1, \ldots, C_s$, and consider the set of variables $V_1, \ldots, V_s$, such that $v \in V_i$ if and only if there is an R_j in C_i such that v appears in $(Y)_{R_j}$. Select from each V_i a single variable; these variables will be independent. A variable corresponds to a choice of 1 index and 3 possibilities (since it will be of the form $Y_{i,i+1}$, $Y_{i+1,i}$, or $Y_{i,i}$).

If we were to choose a set of s independent variables from Y, roughly, to how many such l-tuples $(R_1, \ldots, R_l)$ would this choice correspond, and in turn, to how many sequences $\mathcal{I}$ do these correspond? In other words, to how many non-zero contribution terms in the sum (26) can a choice of s independent variables correspond?

The answer is $O(1)$.

Indeed, by the way we defined $\mathcal{I}$ and R, it follows that, once we have chosen a variable $v \in V_i$, for all other variables in V_i we have a finite number of corresponding indices to choose from. Indeed, this happens because the correlation of R_is induces a "clustering" of variables (since all indices must be within $|V_i| \times (k+1) \leq l(k+1)$ of each other).

Hence, for each of the possible $O(n^s)$ choices of s "representative" variables, we have only $O(1)$ possible non-zero contribution terms in the sum (26).

Going backwards, it follows that for any s, there are $O(n^s)$ terms for which the correlation graph has s components. Since each of these terms has weight at most $1/n^q \leq 1/n^{l/2}$, we have proved the following lemma.

Lemma 3.9. *The contribution to the expectation sum* (26) *from all terms with $s < l/2$ or $q > l/2$ is asymptotically negligible.*

Thus, the only terms of asymptotical significance are those for which $s = q = l/2$. If l is odd, this immediately implies

Lemma 3.10. *With the notations above, for k and l fixed, l odd,*

$$\lim_{n \to \infty} E[\omega_k(n)^l] = 0 \ .$$

Let us examine what happens when l is even and $s = q = l/2$. Such terms are easy to understand: they correspond precisely to l-tuples $(R_1, \ldots, R_l)$ for which $|R_i| = 1$ for all i, and for each $1 \leq i \leq l$ there exists a unique $1 \leq j \leq l$ such that $(Y)_{R_i} = (Y)_{R_j}$.

We make the following simple observation.

Lemma 3.11. *The number of diagonal terms $Y_{j,j}$ contained in each $(Y)_{R_i}$, counting multiplicities, has to have the same parity as k.*

Proof. Indeed, by the definition of any R, all the diagonal terms found in $(\tilde{H}_{\beta,n})_{\mathcal{I}}$ must be found in $(Y)_R$. The parity of these terms, counting multiplicities, has to be the same as the parity of k. This is easy to see; if $\mathcal{I} = (i_1, i_2, \ldots, i_k)$, then by Definition 3.6

$$i_1 - i_2 + i_2 - i_3 + \ldots + i_{k-1} - i_k + i_k - i_1 = 0 \ ,$$

and since each difference $i_j - i_{j+1}$ above is either 0, 1, or -1, it follows that the number of differences equal to 0 has the same parity as k. The number of differences equal to 0 is the number of diagonal terms. $\square$

It then follows that, for all l-tuples of $(R_1, \ldots, R_l)$ for which $s = q = l/2$,

- if k is odd, all variables present in the $(Y)_{R_i}$s are diagonal variables, and

- if k is even, all variables present in the $(Y)_{R_i}$s are off-diagonal variables.

We summarize here what we now know about the terms we need to study when l is even.

Lemma 3.12. *The only asymptotically relevant terms have the property that there exist $l/2$ distinct indices $i_1, \ldots, i_{l/2}$ such that for each i_j there exist precisely two values $j_1 < j_2$ for which*

1. *if k is odd, $R_{j_1} = R_{j_2} = \{(i_j, i_j)\}$.*

2. *if k is even, either one of these four possibilities:*

    - $R_{j_1} = R_{j_2} = \{(i_j, i_j + 1)\}$, *or*

24

- $R_{j_1} = R_{j_2} = \{(i_j + 1, i_j)\}$, *or*
- $R_{j_1} = \{(i_j, i_j + 1)\}$ *and* $R_{j_2} = \{(i_j + 1, i_j)\}$, *or*
- $R_{j_1} = \{(i_j + 1, i_j)\}$ *and* $R_{j_2} = \{(i_j, i_j + 1)\}$.

Note that in this case, $(Y)_{R_1} = (Y)_{R_2}$, *because the matrix is symmetric.*

We call all such terms significant.

We will now need a stronger result than Remark 3.2.

Lemma 3.13. *For any given* $\epsilon > 0$ *and* $k, l \in \mathbb{N}$, *with* l *even, there exists some* $i_\epsilon \in \mathbb{N}$ *such that for any significant term* $\left(\prod_{j=1}^l (A)_{J_i} \right) E\left[\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}]) \right]$ *and the corresponding* $l/2$-*tuplet* $(i_1, \ldots, i_{l/2})$, *if* $k \le i_1, \ldots, i_{l/2} \le n - i_\epsilon$, *then*

- *if* k *is odd,*

$$\left| \left(\prod_{j=1}^l (A)_{J_i} \right) E\left[\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}]) \right] - \prod_{m=1}^{l/2} \left(1 - \frac{i_m}{n} \right)^{k-1} \frac{1}{2^{kl}} \right| < \epsilon \ ;$$

- *if* k *is even,*

$$\left| \left(\prod_{j=1}^l (A)_{J_i} \right) E\left[\prod_{j=1}^l ((Y)_{R_i} - E[(Y)_{R_i}]) \right] - \prod_{m=1}^{l/2} \left(1 - \frac{i_m}{n} \right)^{k-1} \frac{1}{2^{kl+l}} \right| < \epsilon \ .$$

Proof. The lemma follows easily from Proposition 2.19 and Remark 3.1, together with the fact that if R_i contains the index j, then all indices present in J_i are within k of j. $\square$

We prove now that it is enough to look at the significant terms for which $n - i_\epsilon \ge i_1, \ldots, i_{l/2} \ge k$ (i.e., those covered by Lemma 3.13).

Lemma 3.14. *The contribution of significant terms for which some* $i_j > n - i_\epsilon$ *or* $i_j < k$ *is asymptotically negligible, i.e.* $o(1)$.

Proof. Each contribution from a significant term

$$\left(\prod_{i=1}^l (A)_{J_i} \right) E\left[\prod_{i=1}^l ((Y_{R_i} - E[(Y)_{R_i}]) \right]$$

is bounded by some constant $\tilde{M}$, by Lemmas 3.2 and 3.1. Since restricting a choice of i_j to be greater than $n - i_\epsilon$ or less than k yields a finite number of choices for that particular i_j, and since $j < l$ is finite, it follows that there are only $O(n^{l/2-1})$ such restricted terms. But since the contribution of any such term is weighed by $1/n^{l/2}$, the statement of the lemma follows. $\square$

So we have reduced the computation to examining the contribution from the terms for which $n - i_\epsilon \ge 1, \ldots, i_{l/2} \ge k$. Assume w.l.o.g $i_\epsilon > k$ (we can always choose a smaller ϵ).

Given an ordered $l/2$-tuplet of distinct indices $n - i_\epsilon \ge 1, \ldots, i_{l/2} \ge k$, how many terms can correspond to them? First, there are $(l-1)!!$ ways of pairing these indices to the R_is in this order. Second, once the pairing is given,

- for k odd, such a term must have the corresponding $\mathcal{I}_j$ sequence be a sequence where all but one consecutive difference are ± 1 (the one difference that is 0 corresponds to the insertion of the diagonal term). There are $\left(k\binom{k-1}{\frac{k-1}{2}}\right)$ such choices for each $\mathcal{I}_j$, for a total of $\left(k\binom{k-1}{\frac{k-1}{2}}\right)^l$ choices.

- for k even, such a term have the corresponding $\mathcal{I}_j$ sequence be a sequence where all consecutive difference are ± 1, and one of these differences corresponding to the "marked" term that belongs to R_j. Taking into account all the 4 possible case, we obtain a total number of $k^2\binom{k}{\frac{k}{2}}^2$ for each pair of matched $\mathcal{I}_j$s, and total number of $\left(k\binom{k}{\frac{k}{2}}\right)^{l/2}$ choices.

Note that in either one of the two cases above, all choices of sequences are valid, because the indices in each sequence will stay between 1 and n (this is where we need that all $k \le i_j \le n - i_\epsilon$).

Thus, the total number of significant terms which correspond to a given ordered $l/2$-tuplet of distinct indices $n - i_\epsilon \ge 1, \ldots, i_{l/2} \ge k$) is

- $(l-1)!! \left(k\binom{k-1}{\frac{k-1}{2}}\right)^l$ if k is odd, and

- $(l-1)!! \left(k\binom{k}{\frac{k}{2}}\right)^l$ if k is even.

From Lemmas 3.9, 3.12, 3.13 and 3.14, we obtain that for any given ϵ, if k is odd,

$$E[(\omega_k(n))^l] = \sum_{\substack{\text{all significant terms} \\ \text{with } k \le i_j \le n - i_\epsilon, \; \forall \, j}} \frac{1}{(n\beta)^q} \left(\prod_{i=1}^{l}(A)_{J_i}\right) E\left[\prod_{j=1}^{l}((Y)_{R_i} - E[(Y)_{R_i}])\right] + o(1) \;,$$

and so

$$\left| E[(\omega_k(n))^l] - (l-1)!! \left(k\binom{k-1}{\frac{k-1}{2}}\right)^l \sum_{\substack{\text{all significant terms} \\ \text{with } k \le i_j \le n - i_\epsilon, \; \forall \, j}} \frac{1}{(n\beta)^{l/2}} \prod_{j=1}^{l/2}\left(1 - \frac{i_j}{4n}\right)^{k-1} \right| \le$$

$$\le (l-1)!! \left(k\binom{k-1}{\frac{k-1}{2}}\right)^l \sum_{\substack{n - i_\epsilon \ge i_1, \ldots, i_{l/2} \ge k \\ \text{all } i_j \text{ distinct}}} \frac{1}{(n\beta)^{l/2}} \; \epsilon \; + o(1) \;,$$

$$= (l-1)!! \left(k\binom{k-1}{\frac{k-1}{2}}\right)^l \epsilon(1 + o(1)) + o(1) \;,$$

$$= (l-1)!! \left(k\binom{k-1}{\frac{k-1}{2}}\right)^l \; \epsilon \; (1 + o(1)) \;.$$

26

Since ϵ was arbitrarily small, it follows that if we can compute

$$S = (l-1)!! \left(k \binom{k-1}{\frac{k-1}{2}} \right)^l \sum_{\substack{n - i_\epsilon \geq i_1, \ldots, i_{l/2} \geq k \\ \text{all } i_j \text{ distinct}}} \frac{1}{(n\beta)^{l/2}} \prod_{j=1}^{l/2} \left(1 - \frac{i_j}{4n} \right)^{k-1} ,$$

we are done. But, since l and k are fixed, the sum in S is asymptotically the same as the value of the integral $\left(\frac{1}{\beta} \int_0^1 \left(\frac{1-x}{4} \right)^{k-1} dx \right)^{l/2}$, hence

$$\sum_{\substack{n - i_\epsilon \geq i_1, \ldots, i_{l/2} \geq k \\ \text{all } i_j \text{ distinct}}} \frac{1}{(n\beta)^{l/2}} \prod_{j=1}^{l/2} \left(1 - \frac{i_j}{4n} \right)^{k-1} \sim \frac{1}{2^{(k-1)l}} \frac{1}{\beta^{l/2}} \frac{1}{k^{l/2}} ,$$

hence

$$E[(\omega_k(n))^l] - \left(\frac{2}{\beta} \right)^{l/2} \frac{1}{2^{kl}} k^{l/2} \left(\binom{k-1}{\frac{k-1}{2}} \right)^l = O(\epsilon) ,$$

for arbitrarily small ϵ.

Similarly, for k even, we obtain through the same sort of calculation that

$$E[(\omega_k(n))^l] - \left(\frac{2}{\beta} \right)^{l/2} \frac{1}{2^{(k+1)l}} k^{l/2} \left(\binom{k}{\frac{k}{2}} \right)^l = O(\epsilon) ,$$

for arbitrarily small ϵ.

Claim 3.2.1 is thus proved. □

Proof of Claim 3.2.2. The proof is based on the same idea as the proof of Claim 3.2.1; the same reasoning applies to yield the asymptotical covariance result. □

Remark 3.15. *Note that we never actually used the full power of the fact that the entries of the tridiagonal symmetric matrix Y tend to independent centered **normal** variables. We only used the following three properties:*

- *$E[Y_{i,i}] = E[Y_{i+1,i}] = 0$;*

- *$Var[Y_{i,i}] = \frac{1}{2}$, while $\lim_{n \to \infty} Var[Y_{i+1,i}] = \frac{1}{8}$;*

- *for any k, there exists a number $M_k > 0$ such that $|E[(Y_{i,j})^k]| < M_k$, for all $1 \leq i, j \leq n$ (boundedness of moments).*

3.2 The β-Laguerre case

Given an integer k, consider the random variable

$$\eta_{k,\gamma}(n) = \text{tr}((\tilde{L}_{\beta,n}^a)^k) - E_{\beta,a}^L[\text{tr}((\tilde{L}_{\beta,n}^a)^k)] .$$

The main results of this section are given in the Claims below.

Claim 3.15.1. *For any fixed integers k and l,*

$$\lim_{n\to\infty} E[(\eta_{k,\gamma}(n))^l] = \begin{cases} \left(\frac{2}{\beta}\right)^{l/2} \left(Sum_1(k,\gamma) + Sum_2(k,\gamma)\right)^{l/2} (l-1)!! \,, & \text{if } l \text{ is even}\,, \\ 0\,, & \text{if } l \text{ is odd} \end{cases}$$

where

$$Sum_1(k,\gamma) = \sum_{q=1}^{2k-1} (-1)^{q+1} \gamma^{2k-q} \frac{\binom{2k}{q}}{2k} \sum_{j=q+1}^{2k} \frac{(-1)^j}{\binom{2k-1}{j-1}} \sum_{\substack{s_1 + s_2 = j \\ 1 \le s_1, s_2 \le k}} s_1 s_2 \binom{k}{s_1}^2 \binom{k}{s_2}^2,$$

$$Sum_2(k,\gamma) = \sum_{q=0}^{2k-2} (-1)^q \gamma^{2k-q} \frac{\binom{2k}{q}}{2k} \sum_{j=q}^{2k-2} \frac{(-1)^j}{\binom{2k-1}{j}} \sum_{\substack{s_1 + s_2 = j \\ 0 \le s_1, s_2 \le k-1}} (k-s_1)(k-s_2) \binom{k}{s_1}^2 \binom{k}{s_2}^2.$$

Claim 3.15.2. *For any fixed integers k and l,*

$$\lim_{n\to\infty} Cov(\eta_{i,\gamma}(n), \eta_{j,\gamma}(n)) = \frac{2}{\beta}\left(Sum_1(i,j,\gamma) + Sum_2(i,j,\gamma)\right),$$

where

$$Sum_1(i,j,\gamma) = \sum_{q=1}^{i+j-1} (-1)^{q+1} \gamma^{i+j-q} \frac{\binom{i+j}{q}}{i+j} \sum_{j=q+1}^{i+j} \frac{(-1)^j}{\binom{i+j-1}{j-1}} \sum_{\substack{r+s = j \\ 1 \le r \le i \\ 1 \le s \le j}} rs \binom{i}{r}^2 \binom{j}{s}^2,$$

$$Sum_2(i,j,\gamma) = \sum_{q=0}^{i+j-2} (-1)^q \gamma^{i+j-q} \frac{\binom{i+j}{q}}{i+j} \sum_{j=q}^{i+j-2} \frac{(-1)^j}{\binom{i+j-1}{j}} \sum_{\substack{r+s = j \\ 0 \le r \le i-1 \\ 0 \le s \le j-1}} (i-r)(j-s) \binom{i}{r}^2 \binom{j}{s}^2.$$

The method we employ for proving these claims is basically the same as in Section 3.1; the only things that change are the details of the sequences we will deal with, which in turn effect a change in the calculations, and yield the results of Claims 3.15.1 and 3.15.2. In the following we will point out where definitions and calculations differ from before, but we will not go over the reduction arguments again, for the sake of brevity.

We write the scaled matrix $\tilde{B}^a_{\beta,n}$ as

$$\tilde{B}^a_{\beta,n} = D + \frac{1}{\sqrt{\beta n}} Z\,,$$

where $D = E^L_{\beta,a}[\tilde{B}^a_{\beta,n}]$ is the bidiagonal matrix of mean entries $D(i,i) = \frac{\sqrt{\gamma}}{\sqrt{n\beta}} E[\chi_{2a-i\beta}]$ and $D(i+1,i) = \frac{\sqrt{\gamma}}{\sqrt{n\beta}} E[\chi_{(n-i)\beta}]$, and Z is the lower bidiagonal matrix of centered variables; we drop the dependence of D and Z on β, a, and n, for simplicity.

Remark 3.16. *From Proposition 2.19, if $D = (d_{ij})_{1 \leq i,j \leq n}$, given any $\epsilon > 0$, there is an $i_\epsilon \in \mathbb{N}$ such that*

$$\begin{aligned}|\sqrt{n\beta/\gamma}\, d_{i,i} - \sqrt{n\beta/\gamma - i\beta}| &\leq \epsilon\,,\ \text{and}\\ |\sqrt{n\beta/\gamma}\, d_{i+1,i} - \sqrt{(n-i)\beta}| &\leq \epsilon\,,\end{aligned}$$

for any $i \leq n - i_\epsilon$.

Here we also used the fact that $2a/(n\beta) \sim 1/\gamma$.

Similarly, again from Proposition 2.19, we know that $Z_{i,i} \to N(0, \gamma/2)$ and $Z_{i+1,i} \to N(0, \gamma/2)$ in distribution.

As in Section 3.1, we start from the expression for $\mathrm{tr}((BB^T)^k)$, for B a lower bidiagonal matrix:

$$\mathrm{tr}((BB^T)^k) = \sum_{1 \leq i_1, i_2, \ldots, i_{2k} \leq n} b_{i_1,i_2} b_{i_2,i_3} \ldots b_{i_{2k-1},i_{2k}} b_{i_{2k},i_1}\,,$$

where the sum is taken over sequences $(i_1, \ldots, i_{2k})$ with the property that $i_{2j-1} - i_{2j} \in \{0,1\}$, for all $1 \leq j \leq k$, and $i_{2j} - i_{2j+1} \in \{0,-1\}$, for all $1 \leq j \leq k-1$, and also $i_{2k} - i_1 \in \{0,-1\}$.

Just as before, we will introduce a few notations (we "recycle" some of the notations we used before; note that the quantities change).

Definition 3.17. *We denote by $S_{n,k} \in \{1, \ldots, n\}^{2k}$ the set of sequences of integers $i_1, \ldots, i_{2k}$ such that $i_{2j-1} - i_{2j} \in \{0,1\}$, for all $1 \leq j \leq k$, and $i_{2j} - i_{2j+1} \in \{0,-1\}$, for all $1 \leq j \leq k-1$, and also $i_{2k} - i_1 \in \{0,-1\}$. We denote by $\mathcal{I}$ an element in $S_{n,k}$.*

For each such $\mathcal{I}$, we consider all the ways in which we can "break up" $\mathcal{I} := (i_1, \ldots, i_{2k})$ into overlapping "runs" J and R, i.e.

$$J = ((i_{p_0}, \ldots, i_{p_1}), (i_{p_2}, \ldots, i_{p_3}), (i_{p_4} \ldots i_{p_5}), \ldots, (i_{p_{2q}}, \ldots, i_{p_{2q+1}}))\,,$$

and

$$R = ((i_{p_1}, \ldots, i_{p_2}), (i_{p_3}, \ldots, i_{p_4}), \ldots, (i_{p_{2q-1}}, \ldots, i_{p_{2q}}))\,,$$

with

$$(i_1, \ldots, i_{2k}) = (i_{p_0}, \ldots, i_{p_1}, i_{p_1+1}, \ldots, i_{p_2}, i_{p_2+1}, \ldots, i_{p_3}, \ldots, i_{p_{2q+1}})\,.$$

Note that this preserves the requirement that $i_j - i_{j+1} \in \{0, (-1)^{1+j \mod 2}\}$ for all j.

We allow for the possibility of having empty sequences $i_{p_0}, \ldots, i_{p_1}$ in the beginning and/or $i_{p_{2q}}, \ldots, i_{p_{2q+1}}$ in the end of J.

Definition 3.18. *For any $\mathcal{I}$, we introduce a set $\mathcal{J}$ of pairs (J, R), as described above. For a bidiagonal matrix B, we define*

$$\begin{aligned}(BB^T)_{\mathcal{I}} &= b_{i_1,i_2} b_{i_2,i_3} \ldots b_{i_{2k-1},i_{2k}} b_{i_{2k},i_1}\,,\\ (BB^T)_J &= b_{i_{p_0},i_{p_0}+1} \ldots b_{i_{p_1}-1,i_{p_1}} b_{i_{p_2} b_{p_2}+1} \ldots b_{i_{p_3}-1,i_{p_3}} \ldots b_{i_{2q},i_{2q}+1} \ldots b_{i_{2q+1}-1,i_{2q+1}}\,;\\ (BB^T)_R &= b_{i_{p_1},i_{p_1}+1} \ldots b_{i_{p_2}-1,i_{p_2}} b_{i_{p_3} i_{p_3}+1} \ldots b_{i_{p_4}-1,i_{p_4}} \ldots b_{i_{2q-1},i_{2q-1}+1} \ldots b_{i_{2q}-1,i_{2q}}\,.\end{aligned}$$

Remark 3.19. *Note that any term in $\mathrm{tr}((L_{\beta,n}^a)^k)$ will consists of terms in D and terms in Z, with a sequence of runs J recording the former, and a sequence of runs R recording the latter.*

Proof of Claim 3.15.2. As before, we note that

$$(\tilde{L}^a_{\beta,n})_{\mathcal{I}} = \left(\left(D + \frac{1}{\sqrt{n\beta}}Z\right)\left(D + \frac{1}{\sqrt{n\beta}}Z\right)^T\right)_{\mathcal{I}} = \sum_{(J,R)\in\mathcal{J}}\frac{1}{(n\beta)^{P/2}}(D)_J(Z)_R \ , (27)$$

with $P = p_2 - p_1 + 1 + \ldots + p_{2q} - p_{2q-1} + 1$.
    Similarly with (26), write

$$E[(\eta_{k,\gamma}(n))^l] = \sum_{\substack{\mathcal{I}_j \in \mathcal{S}_{n,k} \\ 1 \le j \le l}} \sum_{(J_j,R_j)\in\mathcal{J}_j} \frac{1}{(n\beta)^q}\left(\prod_{j=1}^l (D)_{J_i}\right) E\left[\prod_{j=1}^l ((Z)_{R_i} - E[(Z)_{R_i}])\right] \ , (28)$$

with $q = \sum_i P_i/2$.
    The rest of the argument follows in the footsteps of the proof of Claim 3.2.1. Just as before, it can be shown that the only terms with significant contribution are those for which, for each i, $(Z)_{R_i}$ consists of a single term, and, in addition to that, the set of R_is can be split in pairs (R_{i_1}, R_{i_2}) such that $(Z)_{R_{i_1}} = (Z)_{R_{i_2}}$. This yields

$$\lim_{n\to\infty} E[(\eta_{k,\gamma}(n))^{2p+1}] = 0 \ .$$

Also, if l is even, the argument that we can consider only the terms for which there is an $l/2$-tuple $(i_1, \ldots, i_{l/2})$ which "avoid" the upper and lower corners of the matrix (like in Lemmas 3.13 and 3.14) still applies.
    The one way in which this computation will differ from the one we made for the proof of Claim 3.2.1 lies in the fact that approximating $(D)_J$, given an index i_1 present in J, becomes a little trickier, since the diagonal and off-diagonal elements will approximate respectively to $\sqrt{\gamma}\sqrt{1/\gamma - \frac{i}{n}}$ and $\sqrt{\gamma}\sqrt{1 - \frac{i}{n}}$.
    Thus, one more parameter will become important, namely, the number of off-diagonal terms in each $(D)_{J_i}$. Note that in each sequence $\mathcal{I}$ we must have an *even* number $2s$ of off-diagonal terms (either from D or from Z), since $i_1 - i_2 + \ldots + i_{2k} - i_1 = 0$, and this also implies that we have an even number $2(k - s)$ of diagonal terms.

1. Suppose we fix the term in Z_{R_i} to be the diagonal term z_{i_1,i_1}; to how many sequences $\mathcal{I}$ with a fixed number $2s$ of off-diagonal terms can this correspond? The answer is $2(k - s)\binom{k}{s}^2$; we have $\binom{k}{s}$ ways of picking the off-diagonal terms (because of the alternating property), and once those are picked we have $2(k - s)$ choices for the location of z_{i_1,i_1} among the diagonal terms remaining.

   Each such sequence will have asymptotical weight

$$(D)_J \sim \gamma^{k-1/2}\left(\frac{1}{\gamma} - \frac{i}{n}\right)^{k-s-1/2}\left(1 - \frac{i}{n}\right)^s \ .$$

2. Suppose we now fix the term in Z_{R_i} to be the off-diagonal term z_{i_1+1,i_1}; to how many sequences $\mathcal{I}$ with a fixed number $2s$ of off-diagonal terms can this correspond? The answer is $2s\binom{k}{s}^2$; we have $\binom{k}{s}$ ways of picking the off-diagonal terms (because

30

of the alternating property), and once those are picked we have $2s$ choices for the location of z_{i_1+1,i_1} among them.

Each such sequence will have asymptotical weight

$$(D)_J \sim \gamma^{k-1/2} \left(\frac{1}{\gamma} - \frac{i}{n} \right)^{k-s} \left(1 - \frac{i}{n} \right)^{s-1/2} .$$

Finally, using the binomial formula

$$\left(\frac{1}{\gamma} - \frac{t}{n} \right)^{s-1} = \sum_{i=0}^{s-1} (-1)^i \binom{s-1}{i} \gamma^{-s+1-i} \left(\frac{t}{n} \right)^i ,$$

and after some processing and use of the Riemann-sum and Beta-function formula

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} \left(1 - \frac{t}{n} \right)^{2k-r-s} \left(\frac{t}{n} \right)^{r+s-1} = \int_0^1 (1-x)^{2k-r-s} x^{r+s-1} \, dx$$

$$= \frac{(2k-r-s)! \, (r+s-1)!}{(2k)!} ,$$

combined with all the possible pairings of the R_is (which yields the necessary $(l-1)!!$), we obtain the result of claim 3.15.1. $\qquad\square$

Claim 3.15.2 has a similar proof.

Remark 3.20. *As in Section 3.1, we never actually use the full power of the fact that the entries of the bidiagonal matrix Z tend to independent centered **normal** variables. We only used the following three properties:*

- $E[Z_{i,i}] = E[Z_{i+1,i}] = 0$;

- $\lim_{n\to\infty} Var[Z_{i+1,i}] = \lim_{n\to\infty} Var[Z_{i,i}] = \frac{\gamma}{2}$;

- *for any k, there exists a number $M_k > 0$ such that $|E[(Z_{i,j})^k]| < M_k$, for all $1 \le i, j \le n$ (boundedness of moments).*

4   A more general setting

We present here a way to generalize the "non-random matrix + small random fluctuation" decompositions we have used in Section 3 in order to analyze the deviation and fluctuation from the asymptotical law for the eigenvalue density in the case of the Hermite and Laguerre β-ensembles.

4.1   Tridiagonals

Let f and g be two integrable functions $f, g : [0, 1] \to \mathbb{R}$ such that the integrals $\int_0^1 f^a(x)g^b(x)dx$ exist for all $a, b \in \mathbb{N}$, and such that the quantities

$$m_k := \sum_{r=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{r, r, k-2r} \int_0^1 f^{k-2r}(x) \, g^{2r}(x) \, dx \tag{29}$$

are the moments of a (uniquely determined) distribution ρ, defined on a compact $[a, b]$.

For any $n \in \mathbb{R}$, we then consider the $n \times n$ matrix F_T:

$$F_T = \begin{pmatrix} f\left(\frac{n}{n}\right) & g\left(\frac{n-1}{n}\right) & & & & \\ g\left(\frac{n-1}{n}\right) & f\left(\frac{n-1}{n}\right) & g\left(\frac{n-2}{n}\right) & & & \\ & g\left(\frac{n-2}{n}\right) & \ddots & \ddots & & \\ & & & \ddots & \ddots & g\left(\frac{1}{n}\right) \\ & & & & g\left(\frac{1}{n}\right) & f\left(\frac{1}{n}\right) \end{pmatrix},$$

so the diagonal of F_T is an equidiscretization of f with step $1/n$ (with the exception of the endpoint at 0, which is missing), and the off-diagonal is an equidiscretization of g with step $1/n$ (missing both the endpoint at 1 and the one at 0).

We consider the tridiagonal symmetric matrix

$$R_T = \begin{pmatrix} x_n & y_{n-1} & & & & \\ y_{n-1} & x_{n-1} & y_{n-2} & & & \\ & y_{n-2} & \ddots & \ddots & & \\ & & \ddots & \ddots & y_2 & \\ & & & y_2 & x_2 & y_1 \\ & & & & y_1 & x_1 \end{pmatrix},$$

with the variables x_i and y_j being mutually independent and satisfying the following three properties:

- $E[x_i] = E[y_j] = 0$, for all $1 \le i \le n$, $1 \le j \le n - 1$,

- $\mathrm{Var}[x_i] = \sigma^2$, for all $1 \le i \le n$, and $\mathrm{Var}[y_j] = \eta^2$, for all $1 \le j \le n - 1$,

- for all k there exists a $M_k > 0$ such that $|E[(x_i)^k]| < M_k$ and $|E[(y_j)^k]| < M_k$, for all $1 \le i \le n$, $1 \le j \le n - 1$.

Note that F_T is a non-random matrix, while R_T is a random one.

Now we consider the matrix model

$$M_T = F_T + \frac{1}{\sqrt{n}} R_T \ . \tag{30}$$

We will compute the asymptotical eigenvalue distribution, the first-order deviation from it, and the first-order fluctuation, for the random matrix M_T.

We need two more definitions.

Definition 4.1. *Let P be a path on the lattice $\mathbb{Z}^2$, starting at $(0, 0)$ and ending at $(k, 0)$, with up $((x, y) \to (x+1, y+1))$, down $((x, y) \to (x+1, y-1))$, and level $((x, y) \to (x+1, y))$ steps. For each level $j \in \mathbb{Z}$, we define the quantities $a_j(P)$ and $b_j(p)$, as follows:*

$$\begin{aligned} a_j(P) &:= \quad \#\textit{of level steps from } j \textit{ to } j; \\ b_j(P) &:= \quad \#\textit{of down steps from } j \textit{ to } j - 1 \ . \end{aligned}$$

Note that, since the path P ends at $(k, 0)$, the number of up steps it takes must always equal the number of down steps it takes.

Also let

$$\begin{aligned}
\mathcal{P}_{r,k} &:= \{paths\ from\ (0,0)\ to\ (k,0)\ with\ exactly\ r\ down\ steps\}, \\
\mathcal{P}_{r,k,i} &:= \{paths\ in\ \mathcal{P}_{r,k}\ which\ descend\ to,\ but\ not\ below,\ y=-i\}, \\
p_{r,k,i} &:= |\mathcal{P}_{r,k,i}|, \\
\mathcal{P}_k &:= \cup_{r=0}^{\lfloor\frac{k}{2}\rfloor}\mathcal{P}_{r,k}\ .
\end{aligned}$$

Theorem 4.2. *Let M_T be a matrix from the ensemble defined by (30), of size n, with eigenvalues $(\lambda_1,\ldots,\lambda_n)$, and let $k \geq 1$ be a positive integer. For all $i = 1,\ldots,k$, let*

$$\begin{aligned}
\mu_i = & \sum_{r=0}^{\lfloor\frac{i}{2}\rfloor}\left(f(1)^{i-2r}g(1)^{2r}\left(\sum_{j=0}^{r-1}(r-j)p_{k,j,r} - \frac{1}{2}\binom{i}{r,r,i-2r}\right) \right. + \\
& \left. + \quad f(0)^{i-2r}g(0)^{2r}\left(\sum_{j=0}^{r-1}(r-j)p_{i,j,r} - \frac{3}{2}\binom{i}{r,r,i-2r}\right)\right) \\
& - \sum_{P\in\mathcal{P}_{r,i}}\sum_{j\in\mathbb{Z}}ja_j(P)\int_0^1 f^{i-2r-1}(t)g^{2r}(t)f'(t)dt \quad - \\
& - 2\sum_{P\in\mathcal{P}_{r,i}}\sum_{j\in\mathbb{Z}}jb_j(P)\int_0^1 f^{i-2r}(t)g^{2r-1}(t)g'(t)dt \ + \\
& + \sigma^2\left(\sum_{P\in\mathcal{P}_{r,i}}\sum_{j\in\mathbb{Z}}\binom{a_j(P)}{2}\right)\int_0^1 f^{i-2r-2}(t)g^{2r}(t)dt \quad + \\
& + \eta^2\left(\sum_{P\in\mathcal{P}_{r,i}}\sum_{j\in\mathbb{Z}}\binom{2b_j(P)}{2}\right)\int_0^1 f^{i-2r}(t)g^{2r-2}(t)dt\ .
\end{aligned}$$

Also, for any $1 \leq i \leq k$, let

$$\begin{aligned}
X_i &= tr((M_T)^i) - n\int_a^b x^i\rho(x)\ dx - \mu_i, \\
&\equiv \sum_{j=1}^n\lambda_j^i - n\int_a^b x^i\rho(x)\ dx - \mu_i
\end{aligned}$$

Let $(Y_1,Y_2,\ldots,Y_k)$ be a centered multivariate Gaussian with covariance matrix

$$\begin{aligned}
Cov(Y_i,Y_j) = & \ \sigma^2\sum_{\substack{0\leq r\leq\lfloor\frac{i}{2}\rfloor-1 \\ 0\leq s\leq\lfloor\frac{j}{2}\rfloor-1}}(i-2r)(j-2s)\binom{i}{r,r,i-2r}\binom{j}{s,s,j-2s}\int_0^1 f^{A(r,s)}(x)g^{B(r,s)}(x)\ dx\ + \\
& + \eta^2\sum_{\substack{1\leq r\leq\lfloor\frac{i}{2}\rfloor \\ 1\leq s\leq\lfloor\frac{j}{2}\rfloor}}4rs\binom{i}{r,r,i-2r}\binom{j}{s,s,j-2s}\int_0^1 f^{C(r,s)}(x)g^{D(r,s)}(x)\ dx\ ,
\end{aligned}$$

where $A(r, s) = i - 2r + j - 2s - 2$, $B(r, s) = 2r + 2s$, $C(r, s) = 2r + 2s - 2$, and $D(r, s) = i - 2r + j - 2s$.

Then, as $n \to \infty$,

$$(X_1, X_2, \ldots, X_k) \Rightarrow (Y_1, Y_2, \ldots, Y_k) .$$

Proof. A technical but simple calculation in the spirit of the ones performed in Section 3.1 shows that m_k are the moments of the asymptotic level density, via the expansion of $\mathrm{tr}(M_T^k)$ as a sum of products of the entries of M_T and separation of the zero-order term. Furthermore, examining the first-order terms in this expansion yields the covariance result, using the same counting techniques as in Section 3.1.

Computing the first-order deviation from the mean is slightly more complicated, as first-order terms in the expansion come from four sources; we only enumerate these sources here and indicate how they play a part in the total sum.

We begin with expressing

$$
\begin{aligned}
E[\mathrm{tr}(M_T^k)] &= E\left[\mathrm{tr}\left(\left(F_T + \frac{1}{\sqrt{n}} R_T\right)^k\right)\right] \\
&= E\left[\sum_{\mathcal{I} \in S_{n,k}} (M_T)_{\mathcal{I}}\right] \\
&= E\left[\sum_{\mathcal{I} \in S_{n,k}} \sum_{(J,R) \in \mathcal{J}} \frac{1}{n^{P/2}} (F_T)_J (R_T)_R\right] ,
\end{aligned}
$$

where we have used the notation of Section 3.1.

It is easy to prove, like we did in Section 3.1, that the zero-order terms are given by the the pairs (J, R) where $R = \emptyset$, that the terms $(R_T)_R$ which contain a single variable (at first power) are annihilated by the expectation (since all variables in R_T are centered), and that the terms where $(R_T)_R$ contains three or more variables (counting multiplicities) do not contribute to the first-order deviation.

To each sequence $\mathcal{I}$ (recall that $\mathcal{I} = (i_1, \ldots, i_k)$ with $|i_j - i_{j+1}| \in \{0, 1\}$ for $1 \leq i \leq k-1$ and $|i_k - i_1| \in \{0, 1\}$) we associate in a one-to-one fashion, a path from $(0, 0)$ to $(k, 0)$ taking steps up, down, or level (depending on the nest term being larger, smaller, or equal to the current one). The zero-order terms sum asymptotically to m_k (with the integral being obtained from the Riemann sum, and $\binom{k}{r, r, k-2r} = |\mathcal{P}_k|$).

Three of the first-order term sources come from those terms that have $|R| = \emptyset$, while the fourth comes from the terms for which $(R_T)_R$ contains a single variable, at the second power. Note also that the terms for which $(R_T)_R$ contains two different variables will be annihilated by the expectation.

Source 1. In the zero-order count, we ignore the fact that at the "edges", i.e. upper left corner, corresponding to $i_1 \in \{1, \ldots, \lfloor \frac{k}{2} \rfloor\}$, and lower right corner, corresponding to $i_1 \in \{n - \lfloor \frac{k}{2} \rfloor, \ldots, n\}$, not all paths in $\mathcal{P}_{k,r}$ can appear in the sum. This approximation yield a first-order term which is asymptotically equal to

$$
S_1 := \sum_{r=0}^{\lfloor \frac{k}{2} \rfloor} \left(\left(f(1)^{k-2r} g(1)^{2r} + f(0)^{k-2r} g(0)^{2r} \right) \left(\sum_{i=0}^{r-1} (r-i) p_{k,i,r} - \binom{k}{r, r, k-2r} \right) \right) .
$$

Source 2. In the zero-order count, we approximate the value of the integral $\int_0^1 f^{k-2r}(x)\, g^{2r}(x)\, dx$ by the Riemann sum $\frac{1}{n}\sum_{i=0}^n f\left(\frac{i}{n}\right)^{k-2r} g\left(\frac{i}{n}\right)^{2r}$. This yields a first-order term asymptotically equal to[2]

$$S_2 := \sum_{r=0}^{\lfloor \frac{k}{2}\rfloor} \frac{1}{2}\binom{k}{r,r,k-2r}\left(f(1)^{k-2r}g(1)^{2r} - f(0)^{k-2r}g(0)^{2r}\right) .$$

Source 3. Finally, in the zero-order approximation, we replace $f\left(\frac{i-j}{n}\right)$ and $g\left(\frac{i-j}{n}\right)$ by $f\left(\frac{i}{n}\right)$, respectively $g\left(\frac{i}{n}\right)$ for each $j \in \{-k,\ldots,k\}$. A Taylor series approximation and the counting of terms shows that this way we need to subtract off have a first-order term which is asymptotically

$$S_3 := \sum_{r=0}^{\lfloor \frac{k}{2}\rfloor} \sum_{P\in\mathcal{P}_{r,k}} \sum_{j\in\mathbb{Z}} \left(ja_j(P)\int_0^1 f^{k-2r-1}(t)g^{2r}(t)f'(t)dt - 2jb_j(P)\int_0^1 f^{k-2r}(t)g^{2r-1}(t)g'(t)dt\right) .$$

Source 4. The last source of first-order terms comes from the terms in which $(R_T)_R$ contains a single variable at the second power, and its contribution is asymptotically equal to

$$
\begin{aligned}
S_4 \;:=\; & \sigma^2\left(\sum_{P\in\mathcal{P}_{r,k}}\sum_{j\in\mathbb{Z}}\binom{a_j(P)}{2}\right)\int_0^1 f^{k-2r-2}(t)g^{2r}(t)dt \;+\; \\[2mm]
& + \; \eta^2\left(\sum_{P\in\mathcal{P}_{r,k}}\sum_{j\in\mathbb{Z}}\binom{2b_j(P)}{2}\right)\int_0^1 f^{k-2r}(t)g^{2r-2}(t)dt .
\end{aligned}
$$

Finally, adding S_1, S_2, S_3, and S_4 yields the statement of Theorem 4.2. $\square$

Remark 4.3. *Note that when $f(x) = \sqrt{x}/2$, $g(x) = 0$ for all $x \in [0,1]$, $\sigma^2 = \frac{1}{2\beta}$, and $\eta^2 = \frac{1}{8\beta}$, both the level density asymptotics **and** the covariance matrix for the fluctuations are the same as for the β-Hermite ensemble of Section 3.1.*

*Crucially, the deviation is **different**. The reason is that in the approximation*

$$E[\frac{1}{2}\chi_{(n-i)\beta}] \sim \frac{1}{2}\sqrt{(n-i)\beta} ,$$

the next order term is of order $\frac{1}{\sqrt{(n-i)\beta}}$, which plays a part in computing the deviation.

We remind the reader that we computed the deviation for the β-Hermite ensembles by using the palindromic property of expectations of trace, thus reducing the problem to computing the deviation for the "$\beta = \infty$" case, for which we used Hermite polynomials properties. This allowed us to find the distribution behind the moments of the deviation.

---

[2]One can for example use the Euler-Maclauren approximation formulas to obtain this.

4.2 Bidiagonals

Let f and g be two integrable functions, $f, g, : [0, 1] \to \mathbb{R}$, such that the integrals $\int_0^1 f^a(x) \, g^b(x) \, dx$ exist for all $a, b \in \mathbb{N}$, and such that the quantities

$$\tilde{m}_k := \sum_{r=0}^{k} \binom{k}{r}^2 \int_0^1 f^{2r}(x) \, g^{2(k-r)}(x) \, dx$$

are the moments of a (uniquely determined) distribution ν defined on $[a, b]$.

For any $n \in \mathbb{N}$, we consider the $n \times n$ matrix F_B, defined below:

$$F_B = \begin{pmatrix} f\left(\frac{n}{n}\right) & & & \\ g\left(\frac{n-1}{n}\right) & f\left(\frac{n-1}{n}\right) & & \\ & \ddots & \ddots & \\ & & g\left(\frac{1}{n}\right) & f\left(\frac{1}{n}\right) \end{pmatrix},$$

in other words the diagonal of F_B has an equidiscretization of f with step $1/n$ (with the exception of the endpoint at 0, which is missing), and the subdiagonal is an equidiscretization of g with step $1/n$ (missing both the endpoint at 0 and the one at 1).

We next consider the random bidiagonal matrix

$$R_B = \begin{pmatrix} x_n & & & \\ y_{n-1} & x_{n-1} & & \\ & \ddots & \ddots & \\ & & y_1 & x_1 \end{pmatrix},$$

where the variables x_i, y_j are mutually independent and satisfying the following properties:

- $E[x_i] = E[y_j] = 0$, for all $1 \leq i \leq n$, $1 \leq j \leq n-1$,

- $\mathrm{Var}[x_i] = \sigma^2$ for all $1 \leq i \leq n$ and $\mathrm{Var}[y_j] = \eta^2$, for all $1 \leq j \leq n-1$,

- for all k there is a constant $M_k > 0$ such that $|E[(x_i)^k]| < M_k$ and $|E[(y_j)^k]| < M_k$ for all $1 \leq i \leq n$ and $1 \leq j \leq n-1$.

Finally, consider the matrix

$$M_L = M_B M_B^T , \tag{31}$$

with

$$M_B = F_B + \frac{1}{\sqrt{n}} R_B .$$

Note that while F_B is a non-random matrix, R_B, M_B, and M_L are random.

We will compute the asymptotical level density and the first-order deviation and fluctuation for the random matrix M_L.

We need to define the following quantities.

Definition 4.4. *Let Q be a path on the lattice $\mathbb{Z}^2$, starting at $(0,0)$ and ending at $(2k,0)$, with up $((x,y) \to (x+1,y+1))$, down $((x,y) \to (x+1,y-1))$, and level $((x,y) \to (x+1,y))$ steps. In addition, we require that the path is alternating, i.e. on each odd-numbered step (first, third, etc.) the path is only allowed to go down or stay at the same level, whereas on each even-numbered step (second, fourth, etc.), the path is allowed only to go up or stay at the same level.*

For each level $j \in \mathbb{Z}$, we define the quantities $c_j(Q)$ and $d_j(Q)$, as follows:

$$
\begin{aligned}
c_j(Q) &:= \#\text{of level steps from } j \text{ to } j; \\
d_j(Q) &:= \#\text{of down steps from } j \text{ to } j - 1 \ .
\end{aligned}
$$

Note that, since the path Q ends at $(2k,0)$, the number of up steps it takes must always equal the number of down steps it takes.

Also let

$$
\begin{aligned}
\mathcal{Q}_{r,k} &:= \{\text{alternating paths from } (0,0) \text{ to } (2k,0) \text{ with exactly } r \text{ down steps}\}, \\
\mathcal{Q}_{r,k,i} &:= \{\text{alternating paths in } \mathcal{Q}_{r,k} \text{ which descends to, but not below, } y = -i\}, \\
q_{r,k,i} &:= |\mathcal{Q}_{r,k,i}|, \\
\mathcal{Q}_k &:= \cup_{r=0}^{k} \mathcal{Q}_{r,k} \ .
\end{aligned}
$$

Theorem 4.5. *Let M_L be the matrix from the ensembles defined by (31), of size n, with eigenvalues $(\lambda_1, \ldots, \lambda_n)$, and let $k \geq 1$ be a positive integer. For all $1 \leq i \leq k$, let*

$$
\begin{aligned}
\tilde{\mu}_i \;=\; & \sum_{r=0}^{i} \left(f(1)^{2i-2r} g(1)^{2r} \left(\sum_{j=0}^{r-1} (r-i) q_{i,j,r} - \frac{1}{2}\binom{i}{r}^2 \right) \right. \\
& \left. + \; f(0)^{2i-2r} g(0)^{2r} \left(\sum_{j=0}^{r-1} (r-j) q_{i,j,r} - \frac{3}{2}\binom{i}{r,r}^2 \right) \right) \\
& - \sum_{Q \in \mathcal{Q}_{r,i}} \sum_{j \in \mathbb{Z}} j c_j(Q) \int_0^1 f^{2i-2r-1}(t) g^{2r}(t) f'(t) dt \quad - \\
& - 2 \sum_{Q \in \mathcal{Q}_{r,i}} \sum_{j \in \mathbb{Z}} j d_j(Q) \int_0^1 f^{2i-2r}(t) g^{2r-1}(t) g'(t) dt \;+ \\
& + \sigma^2 \left(\sum_{Q \in \mathcal{Q}_{r,i}} \sum_{j \in \mathbb{Z}} \binom{c_j(Q)}{2} \right) \int_0^1 f^{2i-2r-2}(t) g^{2r}(t) dt \quad + \\
& + \eta^2 \left(\sum_{Q \in \mathcal{Q}_{r,i}} \sum_{j \in \mathbb{Z}} \binom{2d_j(Q)}{2} \right) \int_0^1 f^{2i-2r}(t) g^{2r-2}(t) dt \ .
\end{aligned}
$$

Also, for for any $1 \leq i \leq k$, let

$$
\begin{aligned}
X_i &= tr((M_l)^i) - n \int_a^b x^i \nu(x) \ dx - \tilde{\mu}_i \ , \\
&\equiv \sum_{j=1}^{i} \lambda_j^i - n \int_a^b x^i \nu(x) \ dx - \tilde{\mu}_i
\end{aligned}
$$

37

Let $(Y_1, Y_2, \ldots, Y_k)$ be a centered multivariate Gaussian with covariance matrix

$$Cov(x^i, x^j) = 4\sigma^2 \sum_{\substack{1 \le r \le i \\ 1 \le s \le j}} rs \binom{i}{r}^2 \binom{j}{s}^2 \int_0^1 f^{A(r,s)}(x) g^{B(r,s)}(x) \, dx \; +$$

$$+ \; 4\eta^2 \sum_{\substack{0 \le r \le i-1 \\ 0 \le s \le j-1}} (i-r)(j-s) \binom{i}{r}^2 \binom{j}{s}^2 \int_0^1 f^{A(r,s)+2}(x) g^{B(r,s)-2}(x) \, dx \; ,$$

with $A(r,s) = 2(r+s-1)$ and $B(r,s) = 2(i+j-r-s)$.

Then, as $n \to n$,

$$(X_1, X_2, \ldots, X_k) \Rightarrow (Y_1, Y_2, \ldots, Y_k) \;.$$

Proof. The proof of Theorem 4.5 is based on the same calculations as Theorem 4.2. The covariance can be computed using the same general principles as in Section 3.2, and the examination of the zero-order and first-order terms in the mean can be done as in Section 4.2. Moreover, the sources of the first-order terms are the same as in Section 4.2; it is only the type of path we are counting that changes (from paths of length k to alternating paths of length $2k$). $\qquad\square$

Remark 4.6. *Note that when $f = \sqrt{\frac{1}{\gamma} - 1 + x}$, $g = \sqrt{x}$, $\sigma^2 = \eta^2 = \frac{\gamma}{2\beta}$, both the level density asymptotics **and** the covariance matrix for the fluctuations are the same as for the β-Laguerre ensemble of Section 3.2.*

*Once again, the deviation is **different**, for the same reason as in Section 4.2: in the approximation $E[\frac{1}{2}\chi_{(n-i)\beta}] \sim \frac{1}{2}\sqrt{(n-i)\beta}$ the next order term is of order $\frac{1}{\sqrt{(n-i)\beta}}$, which plays a part in computing the deviation.*

We computed the deviation for the β-Laguerre ensembles by using the palindromic property of expectations of trace, thus reducing the problem to computing the deviation for the "$\beta = \infty$" case, for which we used Laguerre polynomials properties. This allowed us to find the distribution behind the moments of the deviation.

5 Histogramming eigenvalues efficiently

We propose a very effective numerical trick for counting the number of eigenvalues in an interval numerically. This method does not require the computation of eigenvalues and requires a number of operations that is $O(n)$, rather than $O(n^2)$, which allows for counts for matrices of a very large size. The method is the standard Sturm sequence method for tridiagonal symmetric matrices. We take as input D, a vector of length n, which is the diagonal of the matrix, and E, a vector of length $n-1$, the squares of the elements on the super or subdiagonal. This avoids unnecessary square roots in the formation of the matrix which can slow down computation.

The algorithm is remarkably simple. For a given value σ, which is not an eigenvalue of the matrix, compute

$$t_i := D_i - \sigma - E_{i-1}/t_{i-1}$$

for $t = 1$ through n with $t_0 = 1$ and $E_0 = 0$.

From Sturm's theorem (see for example [17, Theorem 13]) we know that the number of eigenvalues of the tridiagonal matrix less than or equal to σ is the number of t_is that are negative, with $1 \le i \le n$.

To histogram eigenvalues of β-ensembles for n's into the millions or even the billions, one can simply compute E from a chi-square random number generator (implemented in MATLAB's Statistics Toolbox as `chi2rnd`), and D from a normal random number generator for the β-Hermite case (`randn`) or from another chi-square in the β-Laguerre case.

An interesting special case is the β-Hermite case with $\beta = \infty$, which computes the roots of the corresponding scaled Hermite polynomial. This may be performed by taking D to be the zero vector of length n and $E = (n-1, n-2, \dots, 2, 1)/(4n)$.

When $\beta = \infty$, there is no fluctuation, but there are deviations for large n. Roughly the theorem states that the number of eigenvalues in an interval I is linear of the form

$$(\# \text{ eigenvalues}) = n(\text{area under the semicircle}) + DEVIATION , \qquad (32)$$

where

$$DEVIATION = \frac{1}{2\pi} arcsin(x)\Big|_I \qquad (33)$$

and we subtract $1/4$ if I contains $+1$ and $1/4$ if I contains -1.

In one numerical experiment (see Figure 2), we took $n = 1,000,000 + i$ for $i = 0, 1, ..., 99$ and computed the deviation from the mean, i.e. the number of eigenvalues in the interval I minus the area $A = n\frac{1}{2\pi} \int_I \sqrt{1-x^2} \, dx$; we did this arbitrarily for the interval $I = [.2 \ .8]$. Since this experiment is non-random it is repeatable without any reference to a random number generator. We found the experimental deviation of 0.1167 which is close to the theoretical deviation of $.1155$, given by (33).
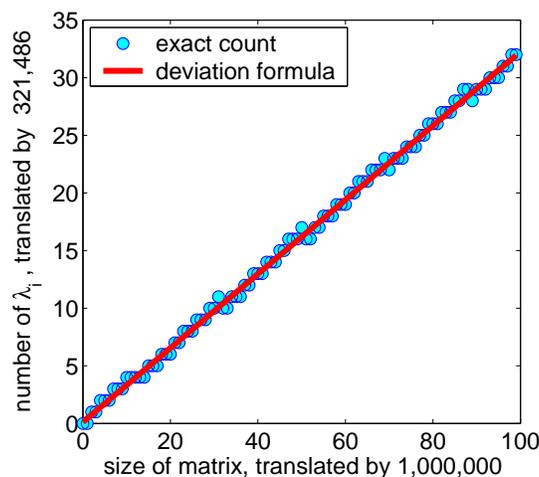
Figure 2: Calculating the deviation in the interval $I = [.2 \ .8]$ for the $\beta = \infty$ Hermite ensemble with sizes $10^6 : 1 : (10^6 + 99)$; circles represent the eigenvalue count minus the area under the semicircle over I; the solid line is the theoretical deviation given in (33)

Figure 2 plots the results of the experiment. The red line which contains the theoretical value of the intercept represents the best fit line to the data in the sense that the average vertical deviation is minimized.

6   Remarks and open problems

There are many object-counting (combinatorial) approaches to the study of traces of powers of random matrices; they depend on the matrix model, and on the polynomial whose trace is being computed. For example, the counting approach of [34] uses full matrix models (all entries are non-zero variables), and traces of powers (thus using the monomial basis, like we have done here), and counts paths in the complete graph of size n. By contrast, in [26], the polynomials used are the shifted Chebyshev polynomials, and the objects counted are non-crossing annular partitions; the matrix models are still full. Here, we use tri/bidiagonal matrix models, consider the monomials, and count essentially paths with three types of steps (up, down, level) in the plane.

Though the objects we count here are simpler than in [26], our counting technique expresses the results in a less compact form than in [21] and [26]. In the latter two papers, the covariance matrix is diagonalized by the choice of polynomial basis, whereas in our paper it is obtained as full because we work with the monomials. There seems to be a trade-off between the simplicity of the object to be counted and the simplicity of the form in which the covariance matrix is expressed.

We would like to conjecture that by using a hybrid way of counting, for example, using the tridiagonal matrices and some of the techniques of [26], both the counting process and the resulting format of the answer could be simplified. The development of such a technique would be of great interest.

7   Acknowledgments

References

[1] Ludwig Arnold. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrsch. Verw. Gebiete*, 19:191–198, 1971.

[2] G. Ben Arous and A. Guionnet. Large Deviations for Wigner's law and Voiculescu's non-communtative entropy. *PTRF*, 108:517–548, 1997.

[3] T. Baker and Peter Forrester. The Calogero-Sutherland model and generalized classical polynomials. *Commun.Math.Phys.*, 188:175–216, 1997.

[4] T. Cabanal-Duvillard. Fluctuations de la Loi Empirique de Grandes Matrices Aléatoires. *Annales de l'Institut Henri Poincaré (B) Probab. Statist.*, 37:373–402, 2001.

[5] A.G. Constantine. Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.*, 34:1270–1285, 1963.

[6] A. Boutet de Montvel, L. Pastur, and M. Shcherbina. On the statistical mechanics approach in the random matrix theory:Integrated density of states. *J. Stat. Phys.*, 79:585–611, 1995.

[7] A. Dembo, A. Guionnet, and O. Zeitouni. Moderate deviations for the spectral measure of certain random matrices. *Ann. Inst. H. Poincaré(B) Probab. Statist.*, 39 (6):1013–1042, 2003.

[8] Patrick Desrosiers and Peter Forrester. Hermite and Laguerre β-ensembles: asymptotic corrections to the eigenvalue density. Technical report, 2005. arxiv.org/PS_cache/math-ph/pdf/0509/0509021.pdf.

[9] Persi Diaconis and Mehrdad Shahshahani. On the eigenvalues of random matrices. *J. Appl. Prob*, 31:49–61, 1994.

[10] Ioana Dumitriu. *Eigenvalue Statistics for the Beta-Ensembles*. PhD thesis, Massachusetts Institute of Technology, 2003.

[11] Ioana Dumitriu and Alan Edelman. Matrix models for beta-ensembles. *J. Math. Phys.*, 43:5830–5847, 2002.

[12] Ioana Dumitriu and Alan Edelman. Eigenvalues of Hermite and Laguerre ensembles: Large Beta Asymptotics. *Ann. Inst. H. Poincaré (B) Probab. Statist.*, 41 (6):1083–1099, 2005.

[13] Ioana Dumitriu, Alan Edelman, and Gene Shuman. MOPS: Multivariate Orthogonal Polynomials (symbolically). 2004. Preprint found at lanl.arxiv.org/abs/math-ph/0409066.

[14] Freeman J. Dyson. The threefold way. Algebraic structures of symmetry groups and ensembles in Quantum Mechanics. *J. Math. Phys.*, 3:1199–1215, 1963.

[15] Peter Forrester. Global fluctuation formulas and universal correlations for random matrices and log-gas systems at infinite density. *Nuclear Phys. B*, 435:421–429, 1995.

[16] Peter Forrester. *Random Matrices*. 2001. Preprint.

[17] F.R. Gantmacher and M.G. Krein. *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems (Revised English Edition)*. Amer. Math. Soc. Chelsea Publ., Amer. Math. Soc., Providence, R.I., 2002.

[18] A. Guionnet. Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. H. Poincaré (B) Probab. Statist.*, 38:341–348, 2002.

[19] A. Guionnet and O. Zeitouni. Large deviations asymptotics for spherical integrals. *J. Funct. Anal.*, 188 (2):461–515, 2002.

[20] Alan T. James. Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.*, 35:475–501, 1964.

[21] Kurt Johansson. On fluctuations of random hermitian matrices. *Duke Math. J.*, 91:151–203, 1998.

[22] Kurt Johansson. Shape fluctuations and random matrices. *Commun. Math. Phys.*, 209:437–476, 2000.

[23] Iain M. Johnstone. On the distribution of the largest principal component. *Ann. of Stat.*, 29(2):295–327, 2001.

[24] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate. Anal.*, 12:1–38, 1982.

[25] Rowan Killip and Irina Nenciu. Matrix Models for Circular Ensembles. *International Mathematics Research Notices*, 50:2665–2701, 2004.

[26] Timothy Kusalik, James A. Mingo, and Roland Speicher. Orthogonal Polynomials and Fluctuations of Random Matrices. Technical report, 2005. arxiv.org/abs/math.OA/0503169.

[27] I.G. Macdonald. *Symmetric Functions and Hall Polynomials.* Oxford University Press Inc, New York, 1995.

[28] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sbornik*, 1:457–483, 1967.

[29] Madan Lal Mehta. *Random Matrices.* Academic Press, Boston, second edition, 1991.

[30] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, 1982.

[31] H.D. Politzer. Random-matrix description of the distribution of mesoscopic conductance. *Phys. Review B*, 40:11917–11919, 1989.

[32] Jack W. Silverstein and Z.D. Bai. CLT of linear spectral statistics of large dimensional sample covariance matrices. preprint, 2003. Accepted for publication in *Annals. of Probab.*

[33] Alexander Soshnikov and Yakov Sinai. Central limit theorem for traces of large symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat., Nova Sr.*, 29:1–24, 1998.

[34] Alexander Soshnikov and Yakov Sinai. A Refinement of Wigner's Semicircle Law in a Neighborhood of the Spectrum Edge for Random Symmetric Matrices. *Func. Anal. and Appl.*, 32 (2):114–131, 1998.

[35] Roland Speicher. Combinatorial Aspects of Free Probability Theory. Lecture Notes, Freie Wahrscheinlichkeitstheorie, Goettingen, August 2005.

[36] Richard P. Stanley. Some combinatorial properties of Jack symmetric functions. *Adv. in Math.*, 77:76–115, 1989.

[37] Richard P. Stanley. *Enumerative Combinatorics, vol. 2.* Cambridge University Press, 1999.

[38] Brian Sutton. *The Stochastic Operator Approach to Random Matrix Theory.* PhD thesis, Massachusetts Insitute of Technology, 2003.

[39] Craig A. Tracy and Harold Widom. Fredholm determinants, differential equations and matrix models. *Commun. Math. Phys.*, 163:33–72, 1994.

[40] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *J. Stat. Phys.*, 92:809–835, 1996.

[41] Craig A. Tracy and Harold Widom. The distribution of the largest eigenvalue in the Gaussian ensembles. In *Calogero-Moser-Sutherland Models, CRM Series in Mathematical Physics*, volume 4, pages 461–472. Springer-Verlag, 2000.

[42] Eric W. Weisstein. "Hermite Polynomial." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/HermitePolynomial.html.

[43] Eric W. Weisstein. "Laguerre Polynomial." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LaguerrePolynomial.html.

[44] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, 62:548–564, 1955.

[45] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, 67:325–327, 1958.

[46] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika A*, 20:32–43, 1928.