F. Gamboa and A. Rouault (2011) “Large Deviations and Sum Rules,”
Applied Mathematics Research eXpress, Vol. 2011, No. 2, pp. 281–307
Advance Access publication July 22, 2011
doi:10.1093/amrx/abr009

Large Deviations for Random Spectral Measures and Sum Rules

Fabrice Gamboa[1] and Alain Rouault[2]

[1]Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31062 Toulouse Cedex 9, France and [2]LMV Bâtiment Fermat, Université Versailles-Saint-Quentin, F-78035 Versailles Cedex, France

Correspondence to be sent to: alain.rouault@math.uvsq.fr

We prove a Large Deviation Principle for the random spectral measure associated to the pair (H_N, e) where H_N is sampled in the GUE(N) and e is a fixed unit vector (and more generally in the β extension of this model). The rate function consists of two parts. The contribution of the absolutely continuous part of the measure is the reversed Kullback information with respect to the semicircle distribution and the contribution of the singular part is connected to the rate function of the extreme eigenvalue in the GUE(N). This method is also applied to the Laguerre and Jacobi ensembles, but in those cases, the expression of the rate function is not explicit.

1 Introduction

The aim of this paper is to study the asymptotic behavior of spectral measures in some classical self-adjoint random matrix models. To begin with, let us first clarify what we mean with spectral measure of a pair (and recall some asymptotic results) in the case of unitary operators.

Let U be a unitary operator on a Hilbert space $\mathcal{H}$, and e be a unit cyclic vector (the span generated by the iterates $(U^n x)$ is $\mathcal{H}$). The spectral measure associated with the pair (U, e) plays an important role and will be one of the object studied here. This

Received February 3, 2011; Revised May 26, 2011; Accepted June 2, 2011

measure is the unique probability measure (p.m.) μ on the unit circle $\mathbb{T}$ such that

$$\langle e, U^n e\rangle = \int_{\mathbb{T}} z^n \, \mathrm{d}\mu(z) \quad (n \geq 1).$$

This measure is a unitary invariant for the pair (U, e), which means that if V is unitary, then the pair (VUV^{-1}, Ve) yields the same measure μ. Assume further that $\dim \mathcal{H} = N$ and that the first vector of the canonical basis, e_1, is cyclic for U. Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of U (all lying on $\mathbb{T}$), and let $\psi_1, \ldots, \psi_N$ be a system of unit eigenvectors. The spectral measure is then

$$\mu_{\mathrm{w}}^{(N)} = \sum_{k=1}^{N} \pi_k \delta_{\lambda_k}, \tag{1}$$

with $\pi_k := |\langle \psi_k, e_1\rangle|^2 \;\, k = 1, \ldots, N$. Note that given λ_k, the vector ψ_k is determined up to a phase, but the number π_k is completely determined. To avoid confusion, we put an index w (for weight) to distinguish this measure from the classical empirical spectral distribution (ESD) defined by

$$\mu_{\mathrm{u}}^{(N)} = \frac{1}{N} \sum_{k=1}^{N} \delta_{\lambda_k}. \tag{2}$$

When U is uniformly sampled from $\mathbb{U}(N)$ (the unitary group of order N) with the Haar distribution, it is well known that the joint distribution of $(\lambda_1, \ldots, \lambda_N)$ has a density proportional to

$$|\Delta(\lambda_1, \ldots, \lambda_N)|^2,$$

where Δ is the Vandermonde determinant (see, e.g., [19]). Furthermore, e_1 is almost surely (a.s.) cyclic and $(\pi_1, \ldots, \pi_N)$ is independent of $(\lambda_1, \ldots, \lambda_N)$. Moreover, $(\pi_1, \ldots, \pi_N)$ is uniformly distributed on the simplex

$$\mathcal{S}_N = \{(\pi_1, \ldots, \pi_N) : \pi_k > 0, (k = 1, \ldots, N), \; \pi_1 + \ldots \pi_N = 1\}.$$

As N tends to infinity, both sequences of random measures $(\mu_{\mathrm{w}}^{(N)})$ and $(\mu_{\mathrm{u}}^{(N)})$ converge weakly to the equilibrium measure—that is, the uniform distribution on $\mathbb{T}$. In a previous work [10], we proved that the sequence $(\mu_{\mathrm{w}}^{(N)})$ satisfies a Large Deviation Principle (LDP), with speed N. The good rate function is the reversed Kullback entropy with respect to the equilibrium measure. Note that there is a quite important difference in the large deviation behavior of $(\mu_{\mathrm{w}}^{(N)})$ and $(\mu_{\mathrm{u}}^{(N)})$. Indeed, this last sequence of p.m.'s satisfies a LDP with speed N^2 and with a rate function connected to the Voiculescu entropy (see, e.g., [13]). To show a LDP for $(\mu_{\mathrm{w}}^{(N)})$, one may think of two kinds of proof. The first one, which could be called the direct way, uses the representation (1) [10]. Besides, it is possible to code a measure μ on $\mathbb{T}$ by the system of its Verblunsky (or Schur) coefficients, via the

Favard theorem [24]. Note that they are also called the canonical moments of μ (see [7] for the definition). The second method uses this coding. It turns out that, under the Haar distribution, the canonical moments $(c_1^{(N)}, \dots c_N^{(N)})$ of $\mu_{\mathrm{w}}^{(N)}$ are independent random variables with explicit distribution depending on N. It is then possible in a first step to check the LDP on these variables and in a second step to lift the LDP and the rate function on the space of measures [17].

The precise form of the rate function can be explained. In the first method, it follows from the Dirichlet weighting of the random measure, and in the second method from Szegö's formula, which enters in the class of the so-called sum rules. The same thing can be done for the Jacobi ensemble with the arcsine distribution (on $[0,1]$ or on $[-2,2]$) playing the role of the uniform distribution on $\mathbb{T}$. The speed N of the LDP comes essentially from the randomness of $(\pi_1, \dots, \pi_N)$; the ESD converging at speed N^2 to its equilibrium measure may be considered as deterministic.

In this paper, we will focus on models of self-adjoint matrices and their extensions. If H is a self-adjoint bounded operator on a Hilbert space $\mathcal{H}$ and e is a cyclic vector, the spectral measure is the unique p.m. μ on $\mathbb{R}$ such that

$$\langle e, H^n e\rangle = \int_{\mathbb{R}} x^n \,\mathrm{d}\mu(x) \quad (n \geq 1).$$

Here also, μ is an unitary invariant for the pair (H, e). Another invariant is the tridiagonal reduction recalled in Section 3. The coefficients of this reduction will play the role of the earlier-mentioned Verblunsky coefficients. If $\dim \mathcal{H} = N$ and e_1 is cyclic for H, the spectral measure is

$$\mu_{\mathrm{w}}^{(N)} = \sum_{k=1}^{N} \pi_k \delta_{\lambda_k}, \tag{3}$$

with the same notation as mentioned earlier, except that, now, the eigenvalues are real.

We will first study the β-Hermite ensemble. It is a family extending the Gaussian ensembles (GOE, GUE, GSE). The second model considered is the β-Laguerre ensemble that generalizes Wishart matrices. The third model is the β-Jacobi ensemble that generalizes MANOVA matrices. In these cases, we could expect that the sequence $(\mu_{\mathrm{w}}^{(N)})$ satisfies a LDP with speed N and with a rate function given by the reversed Kullback entropy with respect to the limit distribution (respectively semicircle, Marchenko–Pastur and Kesten–McKay distributions). Actually the difference with the unitary case comes from the behavior of extreme eigenvalues. Their LDP outside the support of the equilibrium measure is at speed N, and then cannot be neglected, except for the case of the arcsine distribution.

The paper is organized as follows. The next section is devoted to the introduction of notation and models: topology on moments spaces and real matrix models that we will study later. In Section 3, we discuss some relationships between the random spectral measures and coefficients appearing in the construction of the associated random orthogonal polynomials. The LDP for real matrix models are studied in the last two sections. The case of the β-Hermite ensemble is completely tackled in Section 4. Surprisingly, we manage to compute explicitly the rate function, with the help of a convenient sum rule. The β-Laguerre and β-Jacobi ensembles are studied in Sections 5 and 6. Here, the rate functions are not so explicit. All useful distributions we work with are defined in Section 7. After posting a previous version of this paper on arXiv, we have been aware of a paper of Dette and Nagel (see [20]) stating CLTs for moments of the random spectral measures studied here.

2 Notation and Models

2.1 Topology on moments spaces

Let $\mathcal{M}^1$ be the set of all p.m.'s on $\mathbb{R}$ and let $\mathcal{M}^1_m$ be the subset consisting of p.m.'s on $\mathbb{R}$ having finite moments of all orders. For $\mu \in \mathcal{M}^1_m$ we set

$$m_k(\mu) = \int_{\mathbb{R}} x^k \, \mathrm{d}\mu(x), \quad k \geq 1,$$

and $\mathbf{m}(\mu) = (m_k(\mu))_{k\geq 1} \in \mathbb{R}^{\mathbb{N}}$. If we equip $\mathbb{R}^{\mathbb{N}}$ with the product topology, the mapping $\mathbf{m}$ induces on $\mathcal{M}^1_m$ a topology that is not Hausdorff. So we have to consider the subset $\mathcal{M}^1_{m,d}$ of $\mathcal{M}^1_m$ consisting in all p.m.'s determined by their moments. We equip this space with the trace $\mathcal{T}_m$ of the earlier-mentioned topology—that is, the topology of convergence of moments. The mapping $\mathbf{m}$ is then injective and continuous from $\mathcal{M}^1_{m,d}$ to $\mathbb{R}^{\mathbb{N}}$. Notice that the topology $\mathcal{T}_m$ on $\mathcal{M}^1_{m,d}$ is stronger than the trace of the weak convergence topology.

In the next sections, we recall some classical ensembles of random matrices. We refer to [1] for a complete overview on this topic.

2.2 $\boldsymbol{\beta}$-Hermite ensemble

Let us begin with Gaussian matrix models and their extensions.

(1) GOE(N) The diagonal entries are independent and $\mathcal{N}(0; 2/N)$ distributed and the nondiagonal entries are independent up to symmetry and $\mathcal{N}(0; 1/N)$

distributed. The joint density on $\mathbb{R}^N$ of the eigenvalues is proportional to

$$\Delta(\lambda_1,\dots,\lambda_N)\exp -\frac{N}{4}\sum_j \lambda_j^2.$$

The matrix of eigenvectors is Haar-distributed in the orthogonal group of order N. So that, its first line is uniformly distributed on the N-dimensional sphere—that is, the vector $(\pi_1,\dots,\pi_N)$ has the distribution $\mathrm{Dir}_N(1/2)$ (see Section 7.3).

(2) GUE(N) The diagonal entries are independent and $\mathcal{N}(0;1/N)$ distributed and the nondiagonal entries are independent up to Hermitian conjugation and distributed as $\mathcal{N}(0;1/2N)+\sqrt{-1}\,\mathcal{N}(0;1/2N)$ where both normal variables are independent. The joint density of the eigenvalues is proportional to

$$\Delta(\lambda_1,\dots,\lambda_N)^2\exp -\frac{N}{2}\sum_j \lambda_j^2.$$

The matrix of eigenvectors is Haar distributed in the unitary group of order N; so the first line is uniformly distributed on the N-dimensional (complex) sphere—that is the vector $(\pi_1,\dots,\pi_N)$ has the distribution $\mathrm{Dir}_N(1)$.

If M is sampled from the GOE(N) or GUE(N), e_1 is a.s. cyclic, the eigenvalues are a.s. distinct and then we will consider the (random) spectral measure $\mu_{\mathrm{w}}^{(N)}$ given by (3).

We do not recall the definition of the symplectic ensemble GSE(N). Nevertheless, some of the previous objects may also be defined in this context.

(3) More generally, it is now classical to consider a parameter $\beta=2\beta'>0$, and a density in $\mathbb{R}^N$ proportional to

$$|\Delta(\lambda_1,\dots,\lambda_N)|^\beta\exp -\frac{N\beta}{4}\sum_j \lambda_j^2. \tag{4}$$

This expression extends the earlier-mentioned formulas ($\beta=1$ for the GOE, $\beta=2$ for the GUE and $\beta=4$ for the GSE). It is often called a Coulomb gas model and $(\lambda_1,\dots,\lambda_N)$ are called charges.

Dumitriu and Edelman [8, Theorem 2.12] gave a matrix model for this distribution—that is, a random real tridiagonal symmetric matrix whose eigenvalues follows the earlier-mentioned distribution, for every $\beta>0$. Moreover, they proved that the corresponding vector $(\pi_1,\dots,\pi_N)$ is independent of the eigenvalues and $\mathrm{Dir}_N(\beta')$ distributed. A specific description of the matrix will be given in the next section. This matrix model will be called the GβE(N) ensemble.

When $N \to \infty$, it is known that $(\mu_{\mathrm{u}}^{(N)})$ converges weakly to the semicircle distribution, and satisfies a LDP with speed N^2 and with a rate function connected to the Voiculescu entropy.

2.3 β-Laguerre and β-Jacobi ensembles

(1) The classical Wishart real ensemble is formed by $W = GG^{\mathrm{T}}$ with G a $m \times N$ matrix with independent $\mathcal{N}(0, 2/N)$ entries and G^{T} its transpose. When $m \le N$ the joint density of the eigenvalues of W in $\mathbb{R}_+^m$ is proportional to

$$|\Delta(\lambda)| \prod_{j=1}^m \lambda_j^{\frac{1}{2}(N-m+1)-1} \exp -\frac{N}{4} \sum_{j=1}^m \lambda_j$$

and the distribution of weights $(\pi_1, \ldots, \pi_m)$ is $\mathrm{Dir}_N(1/2)$.

This distribution of eigenvalues is classically extended to the β-Laguerre distribution of charges, with density on $\mathbb{R}_+^m$ proportional to

$$|\Delta(\lambda)|^\beta \prod_{j=1}^m \lambda_j^{\beta'(N-m+1)-1} \exp -\frac{N\beta}{4} \sum_{j=1}^m \lambda_j. \tag{5}$$

For this case, Dumitriu and Edelman [8, Theorem 3.4] also gave a random real tridiagonal symmetric matrix model and proved that the vector of weights $(\pi_1, \ldots, \pi_m)$ is also independent of the eigenvalues and is $\mathrm{Dir}_m(\beta')$ distributed. We will call this model the LβE(N, m) ensemble.

(2) The JβE(N; a, b) ensemble (with $\mathtt{a} > -1, \mathtt{b} > -1$) has been defined to extend the MANOVA ensemble known in statistics for $\beta = 1$ and $\beta = 2$. The starting point is a density on $(-2, 2)^N$ proportional to

$$|\Delta(x_1, \ldots, x_N)|^\beta \prod_{j=1}^N (2 - x_j)^{\mathtt{a}} (2 + x_j)^{\mathtt{b}}. \tag{6}$$

Killip and Nenciu [15] gave a (tridiagonal) matrix model and proved that the corresponding vector of weights is again independent of the eigenvalues and $\mathrm{Dir}_N(\beta')$ distributed. We will call it the JβE(N, a, b) ensemble. A variant is the $\widehat{J\beta E}(N, \mathtt{a}, \mathtt{b})$ ensemble where the charges are distributed on [0, 1] according to a density proportional to

$$|\Delta(x_1, \ldots, x_N)|^\beta \prod_{j=1}^{N} x_j^{\mathrm{a}} (1 - x_j)^{\mathrm{b}}. \tag{7}$$

In this ensemble, the weights have the same properties as mentioned previously.

3 Tridiagonal Representations

3.1 Spectral map

In this section, we will describe the Jacobi mapping between tridiagonal matrices and spectral measures. This mapping will be one of the key tools for our large deviations results. We consider finite size matrices corresponding to measures supported by a finite number of points and semi-infinite matrices corresponding to measures with bounded infinite support. The material of this section is largely borrowed from [22, 23, 26].

If μ is a p.m. with a finite support consisting of N points, the orthonormal polynomials (with positive leading coefficients) obtained by Gram–Schmidt procedure from the sequence $1, x, x^2, \ldots, x^{N-1}$ satisfy the recurrence relation

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1} p_{n-1}(x)$$

for $n \le N-1$, where $a_n > 0$ for those n. In the basis $\{p_0, p_1, \ldots, p_{n-1}\}$, the linear transform $f(x) \to xf(x)$ (multiplication by x) in $L^2(\mathrm{d}\mu)$ is represented by the matrix

$$J_\mu = \begin{pmatrix} b_0 & a_0 & 0 & \ldots & 0 \\ a_0 & b_1 & a_1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_{N-3} & b_{N-2} & a_{N-2} \\ 0 & \ldots & 0 & a_{N-2} & b_{N-1} \end{pmatrix}. \tag{8}$$

So, measures supported by N points lead to Jacobi matrices—that is, $N \times N$ symmetric tridiagonal matrices with subdiagonal positive terms. In fact, there is a one-to-one correspondence between them. If J is such a Jacobi matrix, then e_1 is cyclic. Let μ be the spectral measure associated to the pair (J, e_1), then J represents the multiplication by x in the basis of orthonormal polynomials associated to μ and $J = J_\mu$.

More generally, if μ is a p.m. on $\mathbb{R}$, with bounded infinite support, we may apply the same Gram–Schmidt process and consider the associated semi-infinite Jacobi matrix:

$$J_\mu = \begin{pmatrix} b_0 & a_0 & 0 & 0 & \dots \\ a_0 & b_1 & a_1 & 0 & \dots \\ 0 & a_1 & b_2 & a_3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \tag{9}$$

Note that again we have $a_k > 0$ for every k. The mapping $\mu \mapsto J_\mu$ (called here the Jacobi mapping) is a one-to-one correspondence between p.m's on $\mathbb{R}$ having compact infinite support and this kind of tridiagonal matrices with $\sup_n(|a_n| + |b_n|) < \infty$. This result is sometimes called Favard's theorem (see [23, p. 432]).

Furthermore, a compactly supported p.m. μ is completely determined by its moments $m_k(\mu)$ $(k \geq 1)$. So, an inversion formula for the Jacobi mapping may be performed by using J_μ to compute the moments of μ (see, e.g., [22]). Actually, it is a recursive procedure and it is possible to connect successive moments with successive sections of the matrix. For a general Jacobi semi-infinite (resp. $N \times N$) matrix A, let $A^{[j]}$ for $j \geq 1$ (resp. for $j \leq N$) the left top submatrix of A. It is known from [22] formula (5.37) that if A is semi-infinite, we have the identity

$$\langle e_1, A^k e_1 \rangle = \langle e_1, (A^{[j]})^k e_1 \rangle, \quad k = 1, \dots, 2j - 1. \tag{10}$$

It is straightforward that this formula holds true when A is a Jacobi $N \times N$ matrix, as soon as $j \leq N$ and $k \leq 2j - 2$. When $A = J_\mu$, the Jacobi matrix associated to a p.m. μ, we get, in terms of the moments:

$$m_k(\mu) = \langle e_1, (J_\mu^{[j]})^k e_1 \rangle, \quad k = 1, \dots, 2j - 1. \tag{11}$$

for every j if μ as an infinite support, and for $j \leq N$ if μ is supported by only N points. Note that this kind of formula leads to Gauss–Jacobi quadratures. It means that, there exists a sequence of polynomials f_r of $2[N/2] + 1$ variables, such that

$$m_r(\mu) = f_r(b_0, \dots, b_{[r/2]}; a_0, \dots, a_{[r/2]-1}), \tag{12}$$

for any r if μ as an infinite support, and for $r \leq 2N - 1$ if μ is supported by only N points.

Note that the inverse relations are quite intricated (see, for instance, Simon [22, Theorem A2]). Actually, a_n depends on $m_1, \dots, m_{2n+2}$ and b_n depends on $m_1, \dots, m_{2n+1}$.

3.2 Tridiagonal representations of β-ensembles

We now consider the tridiagonal matricial representations of the densities given in Sections 2.2 and 2.3. The first two are due to Dumitriu and Edelman [8] and the third to Killip and Nenciu [15].

(1) For the GβE(N), this representation is

$$H_\beta^{(N)} = \begin{pmatrix} b_0^{(N)} & a_0^{(N)} & 0 & \dots & 0 \\ a_0^{(N)} & b_1^{(N)} & a_1^{(N)} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_{N-3}^{(N)} & b_{N-2}^{(N)} & a_{N-2}^{(N)} \\ 0 & \dots & 0 & a_{N-2}^{(N)} & b_{N-1}^{(N)} \end{pmatrix},$$

where the variables $a_0^{(N)}, \dots, a_{N-2}^{(N)}, b_0^{(N)}, \dots, b_{N-1}^{(N)}$ are independent and

$$\begin{aligned} b_j^{(N)} &\overset{(d)}{=} \mathcal{N}(0; (\beta' N)^{-1}), \\ a_j^{(N)} &\overset{(d)}{=} \sqrt{\gamma(\beta'(N-1-j), (\beta' N)^{-1})}. \end{aligned} \tag{13}$$

It means that the joint density of the eigenvalues of $H_\beta^{(N)}$ is proportional to (2.2). Moreover the weights $(\pi_1, \dots, \pi_N)$, as defined in Section 1, are independent of the eigenvalues and are $\mathrm{Dir}_N(\beta')$ distributed.

We call GβE(N) ensemble the earlier-mentioned distribution on tridiagonal $N \times N$ matrices. It is clear that for $\beta = 1, 2, 4$ the GβE(N) ensemble is different from the GOE(N), GUE(N), GSE(N), respectively, but the random measure $\mu_{\mathrm{w}}^{(N)}$ has the same distribution in both frames.

(2) For the LβE(N, m) the representation is $L_\beta^{(N)} = B_\beta^{(N)} \, (B_\beta^{(N)})^T$

$$B_\beta^{(N)} = \begin{pmatrix} d_1^{(N)} & 0 & 0 & \dots & 0 \\ s_1^{(N)} & d_2^{(N)} & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & s_{m-2}^{(N)} & d_{m-1}^{(N)} & 0 \\ 0 & \dots & 0 & s_{m-1}^{(N)} & d_m^{(N)} \end{pmatrix},$$

where the variables $d_1^{(N)}, \dots, d_m^{(N)}, s_1^{(N)}, \dots, s_{m-1}^{(N)}$ are independent and

$$\begin{aligned} s_j^{(N)} &\overset{(d)}{=} \sqrt{\gamma(\beta'(m-j), (\beta' N)^{-1})}, \\ d_j^{(N)} &\overset{(d)}{=} \sqrt{\gamma(\beta'(N+1-j), (\beta' N)^{-1})}. \end{aligned} \tag{14}$$

It means that the joint density of the eigenvalues of $B_\beta^{(N)}$ is proportional to (5), and that the weights have the required distribution.

We call LβE(N, m) ensemble the above distribution on tridiagonal $m \times m$ matrices. It is clear that the LβE(N, m) ensemble for $\beta = 1$ is very different from the Wishart distribution, but the random measure $\mu_{\mathrm{w}}^{(N)}$ has the same distribution in both frames.

(3) The representation of the JβE$(N; \mathrm{a}, \mathrm{b})$ has been obtained by Killip and Nenciu [15]. Actually, they consider a measure μ on $[-2, 2]$ with finite support as the projection of a symmetric measure $\tilde{\mu}$ on the unit circle $\mathbb{T} = \{z : |z| = 1\}$ by the mapping $z \mapsto z + z^{-1}$. The Jacobi parameters $(a_0, \ldots ; b_0, \ldots)$ of μ are in bijection with the Verblunsky coefficients $(\alpha_0, \ldots)$ of $\tilde{\mu}$ by the Geronimus relations (this is also true for measures with infinite support, see [23, Section 11]). Note that choosing a probability distribution to sample Verblunsky coefficients leads to a probability distribution on Jacobi matrices.

Theorem 3.1 (Killip–Nenciu, Theorem 2). Given $\beta > 0$, let $\alpha_k^{(N)}, 0 \le k \le 2N-2$ be independent and distributed as follows:

$$\alpha_{2p}^{(N)} \stackrel{(d)}{=} \beta_s((N-p-1)\beta' + \mathrm{a} + 1, (N-p-1)\beta' + \mathrm{b} + 1), \tag{15}$$

$$\alpha_{2p-1}^{(N)} \stackrel{(d)}{=} \beta_s((N-p-2)\beta' + \mathrm{a} + \mathrm{b} + 2, (N-p)\beta'), \tag{16}$$

for $p = 0, \ldots, N-1$. Let $\alpha_{2N-1}^{(N)} = \alpha_{-1}^{(N)} = -1$ and define the Geronimus relations

$$\begin{aligned} b_k^{(N)} &= (1 - \alpha_{2k-1}^{(N)})\alpha_{2k}^{(N)} - (1 + \alpha_{2k-1}^{(N)})\alpha_{2k-2}^{(N)}, \\ a_k^{(N)} &= \sqrt{(1 - \alpha_{2k-1}^{(N)})(1 - (\alpha_{2k}^{(N)})^2)(1 + \alpha_{2k+1}^{(N)})}, \end{aligned} \tag{17}$$

for $0 \le k \le n-1$. Then the eigenvalues of the tridiagonal matrix $A^{(N)}$ built with these coefficients $a_k^{(N)}$ and $b_k^{(N)}$ have a joint density in $[-2, 2]$ proportional to

$$|\Delta(x_1, \ldots, x_N)|^\beta \prod_{j=1}^N (2 - x_j)^{\mathrm{a}} (2 + x_j)^{\mathrm{b}},$$

and the vector of weights $(\pi_1, \ldots, \pi_N)$ is $\mathrm{Dir}_N(\beta')$ distributed. □

We call JβE$(N, \mathrm{a}, \mathrm{b})$ ensemble the earlier-mentioned distribution on tridiagonal $N \times N$ matrices.

Since it is often convenient to work on $[0,1]$ instead of $[-2,2]$, let us introduce the affine mapping:

$$x \in [-2,2] \overset{s}{\mapsto} \frac{x+2}{4}. \tag{18}$$

We call $\widehat{J\beta E}(N,\mathrm{a},\mathrm{b})$ the image of $\mathrm{J}\beta\mathrm{E}(N,\mathrm{a},\mathrm{b})$ by s. The preceding result may be rephrased in the following way:

Corollary 3.2. If $A^{(N)}$ is sampled in the $\widehat{J\beta E}(N,\mathrm{a},\mathrm{b})$ ensemble, its eigenvalues have a joint density in $[0,1]^N$ proportional to

$$|\Delta(x_1,\ldots,x_N)|^\beta \prod_{j=1}^N x_j^{\mathrm{b}}(1-x_j)^{\mathrm{a}},$$

and the vector of weights $(\pi_1,\ldots,\pi_N)$ is $\mathrm{Dir}_N(\beta')$ distributed. □

4 Large Deviations in the β-Hermite Ensemble

4.1 Introduction

Recall that the sequence of ESD

$$\mu_{\mathrm{u}}^{(N)} = \frac{1}{N}\sum_{k=1}^N \delta_{\lambda_k}$$

satisfies the LDP in $\mathcal{M}_1$ equipped with the weak topology, with speed $\beta' N^2$ and good rate function

$$I^{\mathrm{u}}(\mu) = -\Sigma(\mu) + \int_{\mathbb{R}} \frac{x^2}{2}\,\mathrm{d}\mu(x) + K_H,$$

where K_H is a constant (see [3]) and

$$\Sigma(\mu) = \iint_{\mathbb{R}^2} \log|x-y|\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).$$

Here, I^{u} has a unique minimizer (the equilibrium measure): the semicircle distribution, denoted hereafter by SC (see Section 7). In particular, the sequence $(\mu_{\mathrm{u}}^{(N)})$ converges weakly in probability to SC.

To manage the large deviations of $(\mu_{\mathrm{w}}^{(N)})$, we will first tackle the large deviations of $(a_k^{(N)}, b_k^{(N)}, k \geq 0)$. It is important to notice already that, in view of (13), as $N \to \infty$, we have for fixed $k \geq 0$, $a_k^{(N)} \to 1$ and $b_k^{(N)} \to 0$ (in probability). The corresponding infinite Jacobi matrix satisfying

$$b_k = 0, \quad a_k = 1, \quad k \geq 0,$$

(often called the free Jacobi matrix, see Simon [24, p. 13] is J_μ with $\mu = \mathrm{SC}$).

In the large deviations properties of $(\mu_w^{(N)})$, the extreme eigenvalues will play also an important role. Let $\lambda_{\max}^{(N)}$ (resp. $\lambda_{\min}^{(N)}$) be the maximal (resp. minimal) eigenvalue. It is known that $\lim \lambda_{\max}^{(N)} = 2$ and $\lim \lambda_{\min}^{(N)} = -2$ (in probability). The following lemma gives the large deviations properties of $\lambda_{\max}^{(N)}$. To state it, we need the definition of a function that will appear also in the expression of the rate function for the random measure $\mu_w^{(N)}$. Notice that the result for $\lambda_{\min}^{(N)}$ can be easily deduced by symmetry. Let, for $x \geq 2$

$$\mathcal{F}_G(x) = \int_2^x \sqrt{t^2-4}\,\mathrm{d}t = \frac{x}{2}\sqrt{x^2-4} - 2\log\left(\frac{x+\sqrt{x^2-4}}{2}\right),$$

and for $x < -2$ set $\mathcal{F}_G(x) = \mathcal{F}_G(-x)$.

Lemma 4.1. For the GβE(N) ensemble the sequence $(\lambda_{\max}^{(N)})$ satisfies for $x \geq 2$

$$\lim_N \frac{1}{\beta' N} \log \mathbb{P}(\lambda_{\max}^{(N)} \geq x) = -\mathcal{F}_G(x). \tag{19}$$

□

The statement and proof for the GOE come from [2, Theorem 6.1], the GUE case is in [18, Proposition 3.1]. More generally, for a continuous potential V, the result is tackled in [1, Theorem 2.6.6] (the potential V in the last theorem is quadratic).

To prepare the statement of our main result, we need another definition, extending the one of Killip and Simon [16].

Definition 4.2. If $[s,t]$ is a compact interval, we say that a p.m. μ on $\mathbb{R}$ satisfies the Blumenthal–Weyl condition relatively to $[s,t]$ (in short BW(s,t)) if

(i) $\mathrm{Supp}(\mu) = [s,t] \cup \{E_j^-\}_{j=1}^{N^-} \cup \{E_j^+\}_{j=1}^{N^+}$ where N^+ (resp. N^-) is 0, finite or infinite,

$$E_1^- < E_2^- < \cdots < s \quad \text{and} \quad E_1^+ > E_2^+ > \cdots > t$$

are isolated points of the support.

(ii) If $N^+ = \infty$ (resp. $N^- = \infty$), then E_j^+ converges toward t (resp. E_j^- converges towards s). □

4.2 Main result

To formulate our main result, as well as to give further conjectures, let us recall the definition of the relative entropy, or Kullback–Leibler divergence. For P and $Q \in \mathcal{M}^1$, the

Kullback information between P and Q is defined by

$$\mathcal{K}(P \mid Q) = \begin{cases} \displaystyle\int_{\mathbb{R}} \log \frac{dP}{dQ}\, dP & \text{if } P \ll Q \text{ and } \log \dfrac{dP}{dQ} \in L^1(P), \\ \infty & \text{otherwise.} \end{cases} \tag{20}$$

In the famous Sanov theorem [5, Theorem 6.2.10], the argument is P and the reference measure is Q. Here, we are in the reverse situation: the argument will be Q and the reference measure will be P. It is nonnegative and zero only if $P = Q$. In both arguments, it is convex and upper semi-continuous. Here is our main result.

Theorem 4.3. For the GβE(N) ensemble, the sequence $(\mu_{\mathrm{w}}^{(N)})$ satisfies the LDP in $\mathcal{M}^1$ equipped with the weak topology, with speed $\beta' N$ and good rate function

$$I^{\mathrm{w}}(\nu) = \begin{cases} \mathcal{K}(\mathrm{SC} \mid \nu) + \sum_{n=1}^{N^+} \mathcal{F}_G(E_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_G(E_n^-) & \text{if } \nu \text{ satisfies BW}(-2, 2), \\ \infty & \text{otherwise.} \end{cases} \tag{21}$$

□

Hence, the rate function $I^{\mathrm{w}}(\nu)$ is finite if and only if

$$\nu(dx) = f_a(x) 1_{[-2,2]}(x)\, dx + \nu_s(dx) + \sum_{n=1}^{N^+} \kappa_n \delta_{E_n^+}(dx) + \sum_{n=1}^{N^-} \kappa_n \delta_{E_n^-}(dx),$$

where

(1) f_a satisfies

$$-\int_{-2}^{2} \log(f_a(x)) \sqrt{4 - x^2}\, dx < \infty,$$

(2) ν_s is singular (with respect to the Lebesgue measure) and is supported by a subset of $[-2, +2]$

(3) the E_n's satisfy

$$\sum_{n=1}^{N^+} \mathcal{F}_G(E_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_G(E_n^-) < \infty. \tag{22}$$

In this case

$$I^{\mathrm{w}}(\nu) = \int_{-2}^{2} \log\left(\frac{\sqrt{4 - x^2}}{2\pi f_a(x)} \right) \frac{\sqrt{4 - x^2}}{2\pi}\, dx + \sum_{n=1}^{N^+} \mathcal{F}_G(E_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_G(E_n^-). \tag{23}$$

Note that, of course, SC is the unique minimizer of this rate function, in accordance with the remark at the beginning of this section.

Proof. *Step* 1. The random measure $\mu_{\mathrm{w}}^{(N)}$ belongs to $\mathcal{M}^1_{m,d} \subset \mathcal{M}^1$. Since the topology $\mathcal{T}_m$ on $\mathcal{M}^1_{m,d}$ is stronger than the trace $\mathcal{T}_{\mathrm{w}}$ of the weak topology, it is enough to prove the LDP with respect to $\mathcal{T}_m$.

Step 2. For $k > 0$, the subset $M(k)$ of $\mathcal{M}^1_{m,d}$ of all p.m.'s supported by $[-k, +k]$ is compact for $\mathcal{T}_w$, and since on $M(k)$ both topologies are identical, it is also compact for $\mathcal{T}_m$.

Step 3. From Lemma 4.1 we know that

$$\lim_{k\to\infty} \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\lambda_{\max}^{(N)} > k) = -\infty.$$

By symmetry, we have also

$$\lim_{k\to\infty} \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\lambda_{\min}^{(N)} < -k) = -\infty.$$

This implies

$$\lim_{k\to\infty} \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mu_{\mathrm{w}}^{(N)} \notin M(k)) = -\infty;$$

hence the sequence $(\mu_{\mathrm{w}}^{(N)})$ is exponentially tight for $\mathcal{T}_m$.

Step 4. From the inverse contraction principle (see [5] Theorem 4.2.4 and Remark a), the LDP on $\mathcal{M}^1_{m,d}$ is then a consequence of a LDP for the sequence of moments (Theorem 4.4).

Step 5. The identification of the rate function as (21) is then a consequence of a *magic* formula (Theorem 4.5).

That ends the proof of Theorem 4.3. ∎

We now state the two missing ingredients of this proof. First define the functions $g(x) := x - 1 - \log x$ if $x > 0$ and $g(x) = \infty$ otherwise and let

$$G(x) := \begin{cases} g(x^2) & \text{if } x > 0, \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 4.4. The sequence $(\mathbf{m}(\mu_{\mathrm{w}}^{(N)}))$ satisfies the LDP in $\mathbb{R}^{\mathbb{N}}$ with speed $\beta' N$ and good rate function I defined as follows. This function is finite if and only if there exists $(b_0, \ldots; a_0, \ldots) \in \mathbb{R}^{\mathbb{N}} \times (0, \infty)^{\mathbb{N}}$ satisfying

$$\sum_{j=0}^{\infty} [b_j^2 + (a_j - 1)^2] < \infty, \tag{24}$$

such that $m_r = \langle e_1, A^r e_1\rangle$ for every $r \geq 1$ with A infinite tridiagonal matrix built with $(b_0,\ldots;a_0,\ldots)$ (see (9)). In that case

$$I(m_1,\ldots) = \frac{1}{2}\sum_{j=0}^{\infty} b_j^2 + \sum_{j=0}^{\infty} G(a_j) < \infty.$$

□

Theorem 4.5 (Killip–Simon [16, 25, Theorem 13.8.6]). Let J be a Jacobi matrix built with $(a_0,\ldots;b_0,\ldots) \in (0,\infty)^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ satisfying $\sup a_n + \sup |b_n| < \infty$. Let μ be the associated measure obtained by Favard's theorem. Then

$$\sum_k [b_k^2 + (a_k - 1)^2] < \infty, \tag{25}$$

if, and only if, the p.m. μ satisfies BW$(-2,2)$ and the two following conditions:

$$\sum_{j=1}^{N_+} (E_j^+ - 2)^{3/2} + \sum_{j=1}^{N_-} (-2 - E_j^-)^{3/2} < \infty \tag{26}$$

$$\int_{-2}^{2} \log(f_a(x))\sqrt{4 - x^2}\, dx > -\infty, \tag{27}$$

where f_a is the density of the absolutely continuous part of μ with respect to dx on $[-2,2]$. In that case

$$I^{\mathrm{w}}(\mu) = \sum_{n=0}^{\infty} \left[\frac{1}{2} b_n^2 + G(a_n)\right], \tag{28}$$

where both sides may be (simultaneously) infinite. □

Let us note that the condition (26), called Lieb–Thirring bound in [16], is equivalent to (22). The proof of Theorem 4.4 will use the following result.

Lemma 4.6. For fixed k, $(b_0^{(N)},\ldots,b_k^{(N)};a_0^{(N)},\ldots,a_{k-1}^{(N)})_{N\geq k}$ satisfies the LDP in $\mathbb{R}^{2k-1}$ with speed $\beta' N$ and good rate function

$$I_k(b_0,\ldots,b_k;a_0,\ldots,a_{k-1}) = \frac{1}{2}\sum_{j=0}^{k} b_j^2 + \sum_{j=0}^{k-1} G(a_j). \tag{29}$$

□

Proof. It is an immediate consequence of independence and the LDP for Gaussian and Gamma random variables recalled in the following lemma. ■

Lemma 4.7.

(1) The sequence of distributions $\mathcal{N}(0; n^{-1})$ satisfies the LDP in $\mathbb{R}$ with speed n and good rate function $x \mapsto x^2/2$.

(2) For $\alpha > 0$ and c fixed, the sequence of distributions $\gamma((n-c), (\alpha n)^{-1})$ satisfies the LDP in $\mathbb{R}$ with speed n and good rate function $x \mapsto g(\alpha x)$.

(3) For $u, v > 0$ and δ, δ' fixed, the sequence of distributions $\beta_s(un+\delta, vn+\delta')$ satisfies the LDP in $\mathbb{R}$ with speed n and good rate function:

$$h_{u,v}(q) = \begin{cases} -u \log\left[\frac{(u+v)}{2u}(1-q)\right] - v\log\left[\frac{(u+v)}{2v}(1+q)\right]; & q \in (-1,1) \\ \infty; \quad \text{otherwise}. \end{cases} \tag{30}$$

□

Proof. The points 1 and 2 are well known. For point 3, we use the representation

$$\beta_s(un+\delta, vn+\delta') \stackrel{(d)}{=} \frac{\gamma(vn+\delta, n^{-1}) - \gamma(un+\delta', n^{-1})}{\gamma(un+\delta, n^{-1}) + \gamma(vn+\delta', n^{-1})},$$

where the two gamma variables are independent. Using point 2 and the contraction principle, we get the rate function

$$h_{u,v}(q) = \inf\left\{ug(x/u) + vg(y/v);\ \frac{y-x}{x+y} = q\right\}.$$

A parametrization in $\lambda = y(1-q) = x(1+q)$ reveals that there is a minimum in $\lambda = (u+v)(1-q^2)/2$, which yields (30). ■

Remark 1. Note that $(u+v)^{-1}h_{u,v}(q)$ is the Kullback entropy of the Bernoulli distribution of parameter $u/(u+v)$ with respect to the Bernoulli distribution of parameter $(1-q)/2$. □

Proof of Theorem 4.4. Fix $\ell > 1$. By Lemma 4.6 and the contraction principle, the sequence $(m_1(\mu_{\mathrm{w}}^{(N)}), \dots, m_{2\ell-1}(\mu_{\mathrm{w}}^{(N)}))$ satisfies the LDP in $\mathbb{R}^{2\ell-1}$ with speed $\beta' N$ and rate function $\widetilde{I}_{2\ell-1}$ defined as follows. Note that there is at most only one tridiagonal matrix A_ℓ built from $(b_0, \dots, b_{\ell-1}; a_0, \dots, a_{\ell-2})$ as in (8) such that

$$m_r = \langle e_1, A_\ell^r e_1 \rangle, \quad r = 1, \dots, 2\ell - 1. \tag{31}$$

Hence, if $(m_1, \dots, m_{2\ell-1})$ satisfies (31), then

$$\widetilde{I}_{2\ell-1}(m_1, \dots, m_{2\ell-1}) = I_{\ell-1}(b_0, \dots, b_{\ell-1}; a_0, \dots, a_{\ell-2}). \tag{32}$$

Otherwise, $\widetilde{I}_{2\ell-1}(m_1,\ldots,m_{2\ell-1})$ is infinite. We do not consider the even case since there is no injectivity in that case.

We now apply the Dawson–Gärtner theorem (see [5, Theorem 4.6.1]). Recall that the space of moments is plugged in $\mathbb{R}^{\mathbb{N}}$. When equipped with the product topology, this latter set can be viewed as the projective limit

$$\mathbb{R}^{\mathbb{N}} = \varprojlim \mathbb{R}^n.$$

The family $\mathbf{m}(\mu_{\mathrm{w}}^{(N)})$ satisfies the LDP in $\mathbb{R}^{\mathbb{N}}$ with rate function

$$I(m_1,\ldots) = \sup\{\widetilde{I}_{2k+1}(m_1,\ldots,m_{2k+1}) : k \geq 0\}. \tag{33}$$

It is clear by (32) and (29) that

$$\begin{aligned}\sup\{\widetilde{I}_{2k+1}(m_1,\ldots,m_{2k+1}) : k \geq 0\} &= \sup_k \left\{\frac{1}{2}\sum_{j=0}^{k} b_j^2 + \sum_{j=0}^{k-1} G(a_j)\right\}\\ &= \frac{1}{2}\sum_{j=0}^{\infty} b_j^2 + \sum_{j=0}^{\infty} G(a_j) \leq \infty.\end{aligned}$$

This sum is finite if and only if the sums $\sum b_j^2$ and $\sum G(a_j)$ are finite. From the very definition of G, this latter sum is finite if and only if $\sum (a_j - 1)^2$ is finite. That ends the proof of Theorem 4.4. ■

4.3 Failure of the direct method

Mimicking the unitary case ([10]), it is tempting to define the random measure

$$\widetilde{\mu}_{\mathrm{w}}^{(N)} = \sum_{k=1}^{N} Y_k \delta_{\lambda_k},$$

with the Y_k independent and $\gamma(\beta', 1)$ distributed so that

$$\mu_{\mathrm{w}}^{(N)} = \frac{\widetilde{\mu}_{\mathrm{w}}^{(N)}}{\widetilde{\mu}_{\mathrm{w}}^{(N)}(1)}.$$

The problem is that the general method of Najim [21] cannot be applied. Indeed, the main assumption on the range of the eigenvalues is violated. As a matter of fact, not all the eigenvalues belong to the support of the semicircle law. Outliers give a contribution. So that, the conclusion given by this approach is not true. The rate function candidate only contains the Kullback part of the LDP but loose the *outer* part.

5 Large Deviations in the β-Laguerre Ensemble

In the Laguerre case, in the usual asymptotics $N \to \infty$, $m = m(N)$ with $m(N)/N \to \tau$ and $0 < \tau \leq 1$, we observe similar phenomena. Recall that the sequence of ESD

$$\mu_{\mathrm{u}}^{(N)} = \frac{1}{m(N)} \sum_{k=1}^{m(N)} \delta_{\lambda_k}$$

satisfies the LDP with speed $\beta' N^2$ and good rate function

$$I^{\mathrm{u}}(\mu) = -\tau^2 \Sigma(\mu) + \tau \int_0^\infty \left(\frac{x}{2} - (1-\tau)\log x\right) \mathrm{d}\mu(x) + K_L,$$

where K_L is a constant [12]. The equilibrium measure, unique minimizer of I^{u}, is the Marchenko–Pastur distribution of parameter τ (denoted hereafter by $\mathrm{MP}(\tau)$, see Section 7.5). In particular, the sequence $(\mu_{\mathrm{u}}^{(N)})$ converges weakly in probability to $\mathrm{MP}(\tau)$.

To manage the large deviations of $(\mu_{\mathrm{w}}^{(N)})$, we will first tackle the large deviations of $(s_k^{(N)}, 1 \leq k \leq m(N)-1; d_k^{(N)}, 1 \leq k \leq m(N))$. Recall that the elements of the tridiagonal matrix $L_\beta^{(N)}$ are

$$\begin{aligned} b_0^{(N)} = (d_1^{(N)})^2, b_k^{(N)} = (s_k^{(N)})^2 + (d_{k+1}^{(N)})^2 (1 \leq k \leq m(N)-1). \\ a_k^{(N)} = s_{k+1}^{(N)} d_{k+1}^{(N)} \quad (0 \leq k \leq m(N)-2). \end{aligned} \tag{34}$$

We can see already that, in view of (14), we have for fixed $k \geq 1$ and $N \to \infty$, $\lim d_k^{(N)} = 1$ and $\lim s_k^{(N)} = \sqrt{\tau}$ (in probability). From (34), this yields $\lim b_0^{(N)} = 1$ and for fixed $k \geq 1$, $\lim b_k^{(N)} = 1 + \tau$, $\lim a_{k-1}^{(N)} = \sqrt{\tau}$ (in probability). The corresponding infinite Jacobi matrix that satisfies

$$b_0 = 1, b_k = 1 + \tau, (k \geq 1); a_k = \sqrt{\tau}; (k \geq 0)$$

is J_μ with $\mu = \mathrm{MP}(\tau)$.

The extreme eigenvalues satisfy $\lim \lambda_{\max}^{(N)} = b(\tau)$, $\lim \lambda_{\min}^{(N)} = a(\tau)$ (in probability). For the large deviations, we need to define $\mathcal{F}_L$ by

$$\mathcal{F}_L(x) = \begin{cases} \displaystyle\int_{b(\tau)}^{x} \frac{\sqrt{(t-a(\tau))(t-b(\tau))}}{t\tau} \,\mathrm{d}t, & x \geq b(\tau), \\ \displaystyle\int_{x}^{a(\tau)} \frac{\sqrt{(a(\tau)-t)(b(\tau)-t)}}{t\tau} \,\mathrm{d}t, & 0 < x \leq a(\tau), \end{cases}$$

and $\mathcal{F}_L(x) = \infty$ if $x \leq 0$.

Lemma 5.1. For the LβE$(N, m(N))$ ensemble with the above notations

(1) the sequence $(\lambda_{\max}^{(N)})$ satisfies for $x \geq b(\tau)$

$$\lim_N \frac{1}{\beta' N} \log \mathbb{P}(\lambda_{\max}^{(N)} \geq x) = -\mathcal{F}_L(x). \tag{35}$$

(2) if $\tau < 1$, then $a(\tau) > 0$ and the sequence $(\lambda_{\min}^{(N)})$ satisfies for $0 < x \leq a(\tau)$

$$\lim_N \frac{1}{\beta' N} \log \mathbb{P}(\lambda_{\min}^{(N)} \leq x) = -\mathcal{F}_L(x). \tag{36}$$

□

Remark 2. As mentioned already, a LDP for a general continuous potential is proved in [1, Theorem 2.6.6; 9]. The knowledge of the Cauchy–Stieltjes transform of MP(τ) allows to recover the formula given in [9, p. 47]. Here, the potential is

$$V(x) = \tau \frac{x}{2} - \tau(1-\tau)\log x.$$

□

For a general double sequence of positive numbers $(d_k)_{k\geq 1}$ and $(s_k)_{k\geq 1}$, we set $d \circ s = (d_1, \ldots; s_1, \ldots)$. We deduce the elements

$$\begin{aligned} b_0 = d_1^2, b_k = s_k^2 + d_{k+1}^2 \, (k \geq 1), \\ a_k = s_{k+1} d_{k+1} \, (k \geq 0). \end{aligned} \tag{37}$$

Conversely, if $(a_0, \ldots; b_0, \ldots)$ is given in $(0, \infty)^{\mathbb{N}\times\mathbb{N}}$ such that the tridiagonal matrix is positive, there exists a unique $d \circ s$ satisfying (37).

Theorem 5.2. For the $L\beta E(N, m(N))$ ensemble, the sequence $(\mu_{\text{w}}^{(N)})$ satisfies the LDP in $\mathcal{M}^1$ equipped with the weak topology, with speed $\beta' N$ and good rate function I^{w} defined as follows. This function is finite at ν if and only if there exists $d \circ s \in [0, \infty)^{\mathbb{N}} \times [0, \infty)^{\mathbb{N}}$ (necessarily unique) satisfying

$$\sum_k G(d_k) + \tau \sum_k G(s_k/\sqrt{\tau}) < \infty,$$

such that $m_r(\nu) = \langle e_1, A^r e_1 \rangle$ for every $r \geq 1$ with A infinite tridiagonal matrix built with $(b_0, \ldots; a_0, \ldots)$ satisfying (37). In that case

$$I^{\text{w}}(\nu) = \sum_k G(d_k) + \tau \sum_k G(s_k/\sqrt{\tau}). \tag{38}$$

□

Remark 3.

(1) It is clear from (38) that the unique minimizer of I^{w} corresponds to $d_k \equiv 1$ and $s_k \equiv \sqrt{\tau}$ which corresponds to $\mathrm{MP}(\tau)$.

(2) When $\tau = 1$, we can write:

$$\begin{aligned} I^{\mathrm{w}}(\nu) &= \sum_{k\geq 1}[d_k^2 - 1 - \log d_k^2 + s_k^2 - 1 - \log s_k^2] \\ &= d_1^2 - 1 + \sum_{k\geq 1}[d_{k+1}^2 + s_k^2 - 2] - 2\sum_{k\geq 1}\log(d_k s_k) \\ &= b_0 - 1 + \sum_{k\geq 1}(b_k - 2) - 2\sum_{k\geq 0}\log a_k. \end{aligned}$$

This expression of I^{w} in terms of the Jacobi coefficients makes plausible the existence of a convenient sum rule, as in Section 4, and we propose the following conjecture. □

Conjecture 1. For the $L\beta E(N, m(N))$ ensemble, the rate function of the LDP of $(\mu_{\mathrm{w}}^{(N)})$ is given by

$$I^{\mathrm{w}}(\nu) = \mathcal{K}(\mathrm{MP}(\tau) \mid \nu) + \sum_j \mathcal{F}_L(E_j^{\pm}),$$

if ν satisfies $\mathrm{BW}(a(\tau), b(\tau))$ and $I^{\mathrm{w}}(\nu) = \infty$ otherwise. □

Note that if $\tau = 1$, $a(\tau) = 0$, there will be no contribution of $\mathcal{F}_L(E_j^-)$, since there is no negative eigenvalue.

We hope that part of the conjecture will be proved in a forthcoming paper, using an alternative parametrization proposed in Nagel and Dette [20].

Proof of Theorem 5.2. We follow the four steps of the proof of Theorem 4.3. First note that we are working with measures with support in $\mathbb{R}^+$ instead of $\mathbb{R}$. Steps 1 and 2 are still valid. The analogous of Step 3 comes from Lemma 5.1. In Step 4, owing to (14) and Lemma 4.7, we see that for k fixed, $(d_k^{(N)})$ (resp. $(s_k^{(N)})$) satisfies the LDP with good rate function $G(d_k)$ (resp. $\tau G(s_k/\sqrt{\tau})$). By independence we deduce the LDP for $(d_1^{(N)}, \ldots, d_k^{(N)}; s_1^{(N)}, \ldots, d_k^{(N)})$ and then for $(a_0^{(N)}, \ldots, a_{k-1}^{(N)}; b_0^{(N)}, \ldots, b_{k-1}^{(N)})$ (k fixed). To end, it is enough to follow the scheme of proof of Theorem 4.4. We get the rate function (38). ■

6 Large Deviations in the β-Jacobi Ensemble

Let us consider the $\widehat{J\beta E}(N, \mathrm{a}(N), \mathrm{b}(N))$ ensemble. The usual asymptotics is $N \to \infty$, $\mathrm{b}(N)/N \to \beta'\kappa_1$, $\mathrm{a}(N)/N \to \beta'\kappa_2$, with $\kappa_1 \geq 0, \kappa_2 \geq 0$. The sequence of ESD

$$\mu_N = \frac{1}{N}\sum_{k=1}^{N} \delta_{\lambda_k}$$

satisfies the LDP with speed $\beta' N^2$ and good rate function:

$$I^{\mathrm{u}}(\mu) = -\Sigma(\mu) - \int_0^1 (\kappa_1 \log x + \kappa_2 \log(1-x))\, \mathrm{d}\mu(x) + K_J, \tag{39}$$

where K_J is a constant (see [14]). The equilibrium measure, unique minimizer of I^{u}, is the Kesten–MacKay distribution of parameter $(u_-, u+)$, where

$$u_-, u_+ = u_\pm\left(\frac{1+\kappa_1}{2+\kappa_1+\kappa_2}, \frac{1+\kappa_1+\kappa_2}{2+\kappa_1+\kappa_2}\right), \tag{40}$$

denoted hereafter by $\mathrm{KMK}(u_-, u_+)$, see Section 7.6. In particular, the sequence $(\mu_{\mathrm{u}}^{(N)})$ converges weakly in probability to $\mathrm{KMK}(u_-, u_+)$.

To manage the large deviations of $(\mu_{\mathrm{w}}^{(N)})$, we will first tackle the large deviations of $(\alpha_k^{(N)}, k \geq 0)$. It is important to notice already that, in view of (16), we have for fixed $p \geq 0$,

$$\lim_N \alpha_{2p}^{(N)} = \frac{\kappa_1 - \kappa_2}{2+\kappa_1+\kappa_2} =: \alpha_{\mathrm{even}}, \quad \lim_N \alpha_{2p+1}^{(N)} = -\frac{\kappa_1+\kappa_2}{2+\kappa_1+\kappa_2} =: \alpha_{\mathrm{odd}}.$$

The symmetric measure $\nu_{\mathbb{T}}$ on the unit circle admitting these limiting Verblunsky coefficients is well understood by its Cauchy–Stieltjes transform since the work of Geronimus ([11], see also [25, Chapter 11]). We do not give details here to shorten the paper.

The Geronimus relations give the corresponding infinite Jacobi matrix which satisfies

$$a = \sqrt{(1-\alpha_{\mathrm{odd}}^2)(1-\alpha_{\mathrm{even}}^2)}, b = -2\alpha_{\mathrm{odd}}\alpha_{\mathrm{even}}.$$

With the notation of Section 3.1 it is $J_{\nu_{\mathbb{R}}}$ where $\nu_{\mathbb{R}}$ is the projection of $\nu_{\mathbb{T}}$. Then the image of $\nu_{\mathbb{R}}$ by the affine transformation s (18) is $\mathrm{KMK}(u_-, u_+)$. The extreme eigenvalues satisfy $\lim \lambda_{\max}^{(N)} = u_+$, $\lim \lambda_{\min}^{(N)} = u_-$ (in probability). For the large deviations, we need to define $\mathcal{F}_J$ by

$$\mathcal{F}_J(x) = \begin{cases} \displaystyle\int_{u_+}^{x} \frac{\sqrt{(t-u_+)(t-u_-)}}{t(1-t)}\, \mathrm{d}t, & u_+ \leq x < 1, \\ \displaystyle\int_{x}^{u_-} \frac{\sqrt{(u_- - t)(u_+ - t)}}{t(1-t)}\, \mathrm{d}t, & 0 < x \leq u_-, \end{cases}$$

and $\mathcal{F}_J(x) = \infty$ if $x \notin (0,1)$. The following lemma is a kin of Lemmas 4.1 and 5.1. Here, the potential is

$$V(x) = -\kappa_1 \log x - \kappa_2 \log(1-x).$$

Lemma 6.1. For the $\widehat{J\beta E}(N, \mathtt{a}(N), \mathtt{b}(N))$ ensemble with the above notations

(1) if $\kappa_2 > 0$, then $u_+ < 1$ and the sequence $(\lambda_{\max}^{(N)})$ satisfies for $x \in (u_+, 1)$

$$\lim_N \frac{1}{\beta' N} \log \mathbb{P}(\lambda_{\max}^{(N)} \geq x) = -\mathcal{F}_J(x). \tag{41}$$

(2) if $\kappa_1 > 0$, then $u_- > 0$ and the sequence $(\lambda_{\min}^{(N)})$ satisfies for $x \in (0, u_-)$

$$\lim_N \frac{1}{\beta' N} \log \mathbb{P}(\lambda_{\min}^{(N)} \leq x) = -\mathcal{F}_J(x). \tag{42}$$

□

Theorem 6.2.

(1) (Gamboa–Rouault [10]) For the $\widehat{J\beta E}(N, \mathtt{a}, \mathtt{b})$ ensemble, the sequence $(\mu_{\mathtt{w}}^{(N)})$ satisfies the LDP in $\mathcal{M}^1([0,1])$ endowed with the weak topology, with speed N and good rate function

$$I^{\mathtt{w}}(\nu) = \mathcal{K}(\mathrm{ARCSINE} \mid \nu). \tag{43}$$

(2) Let for $x \in (-1,1)$

$$\begin{aligned} I_{\mathrm{even}}(x) &:= -(1+\kappa_2) \log\left[\frac{(2+\kappa_1+\kappa_2)}{2(1+\kappa_2)}(1-x)\right] \\ &\quad - (1+\kappa_1) \log\left[\frac{(2+\kappa_1+\kappa_2)}{2(1+\kappa_1)}(1+x)\right] \\ I_{\mathrm{odd}}(x) &:= -(1+\kappa_1+\kappa_2) \log\left[\frac{(2+\kappa_1+\kappa_2)}{2(1+\kappa_1+\kappa_2)}(1-x)\right] \\ &\quad - \log\left[\frac{(2+\kappa_1+\kappa_2)}{2}(1+x)\right] \end{aligned}$$

and $I_{\mathrm{even}}(x) = I_{\mathrm{odd}}(x) = \infty$ for $x \notin (-1,1)$. For the $\widehat{J\beta E}(N, \mathtt{a}(N), \mathtt{b}(N))$ ensemble, the sequence $(\mu_{\mathtt{w}}^{(N)})$ satisfies the LDP in $\mathcal{M}^1([0,1])$ endowed with the weak topology, with speed $\beta' N$ and with a good rate function $I^{\mathtt{w}}$ defined as follows. This function is finite at ν if and only if there exists $\vec{\alpha} \in (-1,1)^{\mathbb{N}}$ (necessarily unique) such that

$$\mathcal{I}(\vec{\alpha}) := \sum_{k=0}^{\infty} I_{\mathrm{even}}(\alpha_{2k}) + \sum_{k=0}^{\infty} I_{\mathrm{even}}(\alpha_{2k+1})$$

is finite. Here $\vec{\alpha}$ is related to ν through Geronimus relation (see Equation (17)). In that case

$$I^{\mathrm{w}}(\nu) = \mathcal{I}(\vec{\alpha}). \qquad \square$$

Proof. Since the measures are supported by a subset of $[0, 1]$, the weak convergence and the convergence of moments are identical. To state the LDP for this latter topology, we first state the LDP for the Verblunsky coefficients $\alpha_{2p}^{(N)}$ and $\alpha_{2p+1}^{(N)}$ with p fixed. With this aim, we apply Lemma 4.7(3), with $n = \beta' N$, and for an even index we take $u = 1 + \kappa_2$, $v = 1 + \kappa_1$ and with odd index $u = 1 + \kappa_1 + \kappa_2, v = 1$. Then, the independence and Geronimus relations carry these LDP to the LDP for a sequence of a_k, b_k's of fixed length. We end as in the proof of Theorem 4.4. ■

In the particular case of a and b fixed, we have $\kappa_1 = \kappa_2 = 0$ and

$$I(\vec{\alpha}) = -\sum_0^{\infty} \log(1 - \alpha_k^2).$$

But Szegö's formula [24] says that it is exactly the reversed Kullback with respect to the ARCSINE distribution.

In the general case, there is, up to our knowledge, no known sum rule. Besides, it is very intricate to express the above sums in terms of the tridiagonal coefficients. Nevertheless, it is tempting to propose the following conjecture.

Conjecture 2. For the $\widehat{J\beta E}(N, \mathrm{a}(N), \mathrm{b}(N))$ ensemble, the rate function of the LDP of $(\mu_{\mathrm{w}}^{(N)})$ is given by

$$I^{\mathrm{w}}(\nu) = \mathcal{K}(\mathrm{KMK}(u_-, u_+) \mid \nu) + \sum_j \mathcal{F}_J(E_j^{\pm}),$$

if ν satisfies $\mathrm{BW}(u_-, u_+)$ and $I^{\mathrm{w}}(\nu) = \infty$ otherwise. $\square$

Notice again that in the case $\kappa_1 = 0$ (resp. $\kappa_2 = 0$), then $u_- = 0$ (resp. $u_+ = 1$) and there is no extra terms $\mathcal{F}_J(E_j^-)$ (resp. $\mathcal{F}_J(E_j^+)$). This is consistent with (43).

7 Some Distributions

7.1 Gamma distribution

For $a, b > 0$, the $\gamma(a, b)$ distribution is supported by $[0, \infty)$ with density

$$\frac{e^{-x/b} x^{a-1}}{b^a \Gamma(a)}.$$

Its mean is ab.

7.2 Beta distribution

For $a, b > 0$, the beta symmetric distribution of parameter (a, b), denoted by $\beta_s(a, b)$, is supported by $[-1, 1]$ and has density

$$2^{1-a-b}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(1-x)^{a-1}(1+x)^{b-1}.$$

Its mean is $\dfrac{b-a}{b+a}$.

7.3 Dirichlet distribution

For $k \geq 1$, we set

$$\mathcal{S}_k := \{(x_1, \ldots, x_k) : x_i > 0, (i = 1, \ldots, k),\ x_1 + \cdots + x_k = 1\},$$
$$\mathcal{S}_k^< := \{(x_1, \ldots, x_k) : x_i > 0, (i = 1, \ldots, k),\ x_1 + \cdots + x_k < 1\}.$$

Obviously, the mapping $(x_1, \ldots, x_{k+1}) \mapsto (x_1, \ldots, x_k)$ is a bijection from the simplex $\mathcal{S}_{k+1}$ onto $\mathcal{S}_k^<$.

For $a_j > 0,\ j = 1, \ldots, k+1$, the Dirichlet distribution $\mathrm{Dir}(a_1, \ldots, a_{k+1})$ on $\mathcal{S}_{k+1}$ has the density

$$\frac{\Gamma(a_1 + \cdots + a_{k+1})}{\Gamma(a_1) \cdots \Gamma(a_{k+1})} \quad x_1^{a_1-1} \cdots x_{k+1}^{a_{k+1}-1} \tag{44}$$

with respect to the Lebesgue measure on $\mathcal{S}_{k+1}$. When $a_1 = \cdots = a_{k+1} = a > 0$, we will denote the Dirichlet distribution by $\mathrm{Dir}_k(a)$. If $a = 1$ we recover the uniform distribution on $\mathcal{S}_k^<$.

7.4 Semicircle distribution

The semicircle distribution denoted by SC is supported by $[-2, 2]$ with density

$$\frac{\sqrt{4-x^2}}{2\pi}.$$

Its Cauchy–Stieltjes transform (throughout, all branches of the square roots are taken in accordance with the definition of Cauchy transform) is

$$m(z) = \int \frac{\mathrm{d}\mu(x)}{x-z} = \frac{-z+\sqrt{z^2-4}}{2}. \tag{45}$$

7.5 Marchenko–Pastur distribution

When $0 < \tau \le 1$, the Marchenko–Pastur distribution of parameter τ, denoted by $\mathrm{MP}(\tau)$ is supported by $(a(\tau), b(\tau))$ where $a(\tau) = (1-\sqrt{\tau})^2$, $b(\tau) = (1+\sqrt{\tau})^2$ with density

$$\frac{\sqrt{(x-a(\tau))(b(\tau)-x)}}{2\pi\tau x}. \tag{46}$$

Its Cauchy–Stieltjes transform is

$$m(z) = \frac{-z-1+\tau+\sqrt{(z-1-\tau)^2-4\tau}}{2\tau z}. \tag{47}$$

7.6 Kesten–McKay distribution

For $0 \le u_- < u_+ \le 1$, the Kesten–McKay distribution of parameters (u_-, u_+), denoted by $\mathrm{KMK}(u_-, u_+)$ is supported by $[u_-, u_+]$ and its density is

$$C_{u_-,u_+} \frac{\sqrt{(x-u_-)(u_+-x)}}{2\pi x(1-x)}, \tag{48}$$

where

$$C_{u_-,u_+}^{-1} := \tfrac{1}{2}[1-\sqrt{u_-u_+}-\sqrt{(1-u_-)(1-u_+)}].$$

To express its Cauchy–Stieltjes transform, let us give some notation. For $(b, c) \in (0,1) \times (0,1)$ we put

$$\sigma_\pm(b,c) = \tfrac{1}{2}[1+\sqrt{bc} \pm \sqrt{(1-b)(1-c)}], \tag{49}$$

and for $(x, y) \in (0,1) \times (0,1)$

$$\begin{aligned} u_\pm(x,y) &= (1-x-y+2xy) \pm 2\sqrt{x(1-x)y(1-y)} \\ &= (\sqrt{(1-x)(1-y)} \pm \sqrt{xy})^2. \end{aligned} \tag{50}$$

The mappings $\sigma_\pm$ and $u_\pm$ are inverse in the following sense:

$$\{(b,c) : 0 < b < c < 1\} \underset{(u_-,u_+)}{\overset{(\sigma_-,\sigma_+)}{\rightleftarrows}} \{(x,y) : 0 < x < y < 1 \text{ and } x+y > 1\}. \tag{51}$$

The Cauchy–Stieltjes is then (see for instance [6, p. 129] or [4, p. 425])

$$m(z) = \frac{(1-\sigma_+-\sigma_-)}{2(1-\sigma_+)z} + \frac{(\sigma_+-\sigma_-)}{2(1-\sigma_+)(1-z)} + \frac{\sqrt{(z-a_-)(z-a_+)}}{2z(1-z)}. \tag{52}$$

ARCSINE corresponds to $u_- = 0$ and $u_+ = 1$.

Acknowledgements

Many thanks are due to Professor Holger Dette for helpful discussions, and to the referee for a careful reading of the manuscript and for pointing out a computation mistake in a previous version.

Funding

A.R.'s work is supported by the ANR project Grandes Matrices Aléatoires ANR-08-BLAN-0311-01.

References

[1] Anderson, G., A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, 2nd ed. Cambridge: Cambridge University Press, 2011.

[2] Ben Arous, G., A. Dembo, and A. Guionnet. "Aging of spherical spin glasses." *Probability Theory and Related Fields* 120, no. 1 (2001): 1–67.

[3] Ben Arous, G. and A Guionnet. "Large deviations for Wigner's law and Voiculescu's non-commutative entropy." *Probability Theory and Related Fields* 108, no. 4 (1997): 517–42.

[4] Capitaine, M. and M. Casalis. "Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices." *Indiana University Mathematics Journal* 53, no. 2 (2004): 397–432.

[5] Dembo, A. and O. Zeitouni. *Large Deviations Techniques and Applications*, 2nd ed. Berlin: Springer, 1998.

[6] Demni, N. "Free Jacobi processes." *Journal of Theoretical Probability* 21, no. 1 (2006): 118–43.

[7] Dette, H. and W. Studden. *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*. New York: Wiley Series in Probability and Statistics, 1997.

[8] Dumitriu, I. and A. Edelman. "Matrix models for beta ensembles." *Journal of Mathematical Physics* 43, no. 11 (2002): 5830–47.

[9] Féral, D. "On Large Deviations for the Spectral Measure of Discrete Coulomb Gas." *Séminaire de Probabilités XLI*, 19–49. Lecture Notes in Mathematics 1934. Berlin: Springer, 2008.

[10] Gamboa, F. and A. Rouault. "Canonical moments and random spectral measures." *Journal of Theoretical Probability* 23, no. 4 (2010): 1015–38.

[11] Geronimus, J. "On polynomials orthogonal on the circle, on trigonometric moment-problem and on allied Carathéodory and Schur functions." *Matematicheskii Sbornik* 15, no. 57 (1944): 99–130.

[12] Hiai, F. and D. Petz. "Eigenvalue density of the Wishart matrix and large deviations." *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 1, no. 4 (1998): 633–46.

[13] Hiai, F. and D. Petz. "A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices." *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques* 36, no. 1 (2000): 71–85.

[14] Hiai, F. and D. Petz. "Large deviations for functions of two random projection matrices." *Acta Scientiarum Mathematicarum (Szeged)* 72, no. 3–4 (2006): 581–609.

[15] Killip, R. and I. Nenciu. "Matrix models for circular ensembles." *International Mathematics Research Notices* 50 (2004): 2665–701.

[16] Killip, R. and B. Simon. "Sum rules for Jacobi matrices and their applications to spectral theory." *Annals of Mathematics. Second Series* 158, no. 1 (2003): 253–321.

[17] Lozada-Chang, L. V. "Large deviations on moment spaces." *Electronic Journal of Probability* 10, no. 19 (2005): 662–90.

[18] Maida, M. "Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles." *Electronic Journal of Probability* 12. no. 41 (2007): 1131–15A52.

[19] Mehta, M. L. *Random Matrices*, 3rd ed. Pure and Applied Mathematics (Amsterdam) 142. Amsterdam: Elsevier/Academic Press, 2004.

[20] Nagel, J. and H. Dette. "Distributions on unbounded moment spaces and random moment sequences." (2010): preprint arXiv:1007.3369.

[21] Najim, J. "A Cramér type theorem for weighted random variables." *Electronic Journal of Probability* 7, no. 4 (2002): 1–32.

[22] Simon, B. "The classical moment problem as a self-adjoint finite difference operator." *Advances in Mathematics* 137, no. 1 (1998): 82–203.

[23] Simon, B. "OPUC on one foot." *American Mathematical Society. Bulletin. New Series* 42, no. 4 (2005): 431–60.

[24] Simon, B. *Orthogonal Polynomials on the Unit Circle. Part 1: Classical Theory*. Colloquium Publications. American Mathematical Society 54, Part 1. Providence, RI: American Mathematical Society (AMS), 2005.

[25] Simon, B. *Orthogonal Polynomials on the Unit Circle. Part 2: Spectral Theory*. Colloquium Publications. American Mathematical Society 51, Part 2. Providence, RI: American Mathematical Society, 2005.

[26] Simon, B. CMV matrices: five years after. *Journal of Computational and Applied Mathematics* 208, no. 1 (2007): 120–54.