Sum rules via large deviations

Fabrice Gamboa, Jan Nagel, Alain Rouault

Sum rules via large deviations

Fabrice Gamboa
Université Paul Sabatier
Institut de Mathématiques de Toulouse
118 route de Narbonne
31062 Toulouse Cedex 9, France
e-mail: gamboa@math.univ-toulouse.fr

Jan Nagel
Technische Universität München
Fakultät für Mathematik
Boltzmannstr. 3
85748 Garching, Germany
e-mail: jan.nagel@tum.de

Alain Rouault
Université Versailles-Saint-Quentin
LMV UMR 8100
45 Avenue des Etats-Unis
78035-Versailles Cedex France
e-mail: alain.rouault@uvsq.fr

June 21, 2015

Abstract

In the theory of orthogonal polynomials, sum rules are remarkable relationships between a functional defined on a subset of all probability measures involving the reverse Kullback-Leibler divergence with respect to a particular distribution and recursion coefficients related to the orthogonal polynomial construction. Killip and Simon (Killip and Simon (2003)) have given a revival interest to this subject by showing a quite surprising sum rule for measures dominating the semicircular distribution on $[-2, 2]$. This sum rule includes a contribution of the atomic part of the measure away from $[-2, 2]$. In this paper, we recover this sum rule by using probabilistic tools on random matrices. Furthermore, we obtain new (up to our knowledge) *magic* sum rules for the reverse Kullback-Leibler divergence with respect to the Marchenko-Pastur or Kesten-McKay distributions. As in the semicircular case, these formulas include a contribution of the atomic part appearing away from the support of the reference measure.

Keywords: Sum rules, Jacobi matrix, Kullback-Leibler divergence, orthogonal polynomials, spectral measures, large deviations, random matrices.

1 Introduction

1.1 Szegő-Verblunsky theorem and sum rules

A very famous result in the theory of orthogonal polynomial on the unit circle (OPUC) is the Szegő-Verblunsky theorem (see Simon (2011) Theorem 1.8.6 p. 29). It concerns a deep relationship between the coefficients involved in the construction of the orthogonal polynomial sequence of a measure supported by the unit circle and its logarithmic entropy. More precisely, the inductive relation between two successive monic orthogonal polynomials ϕ_{n+1} and ϕ_n ($\deg \phi_n = n$, $n \geq 0$) associated with a probability measure μ on the unit circle $\mathbb{T}$ supported by at least $n+1$ points involves a complex number α_n and may be written as

$$\phi_{n+1}(z) = z\phi_n(z) - \overline{\alpha}_n \phi_n^*(z) \text{ where } \phi_n^*(z) := z^n \overline{\phi_n(1/\bar{z})}. \tag{1.1}$$

The complex number $\alpha_n = -\overline{\phi_{n+1}(0)}$ is the so-called Verblunsky coefficient. In other contexts, it is also called Schur, Levinson, Szegő coefficient or even canonical moment (Dette and Studden (1997)).

The Szegő-Verblunsky theorem is the identity

$$\frac{1}{2\pi}\int_0^{2\pi} \log g_\mu(\theta) d\theta = \sum_{n\geq 0} \log(1 - |\alpha_n|^2)\,, \tag{1.2}$$

where the Lebesgue decomposition of μ with respect to the uniform measure $d\theta/2\pi$ on $\mathbb{T}$ is

$$d\mu(\theta) = g_\mu(\theta)\frac{d\theta}{2\pi} + d\mu_s(\theta)\,,$$

and where both sides of (1.2) are simultaneously finite or infinite. An exhaustive overview and the genesis tale of this crucial theorem may be found in the very nice book of Simon (Simon (2011)). The identity (1.2) is one of the most representative example of a sum rule (or trace formula): it connects the coefficients of an operator (Killip (2007)) to its spectral data. There are various analytical methods of proof (see Chapter 1 in Simon (2011)) and a probabilistic one (see section 5.2 of Gamboa and Rouault (2010)).

In the theory of orthogonal polynomials on the real line (OPRL), given a probability measure μ with an infinite support, a.k.a. nontrivial case (resp. with a finite support consisting of n points, a.k.a. trivial case), the orthonormal polynomials (with positive leading coefficients) obtained by applying the orthonormalizing Gram-Schmidt procedure to the sequence $1, x, x^2, \dots$ obey the recursion relation

$$xp_k(x) = a_{k+1}p_{k+1}(x) + b_{k+1}p_k(x) + a_k p_{k-1}(x) \tag{1.3}$$

for $k \geq 0$ (resp. for $0 \leq k \leq n-1$) where the Jacobi parameters satisfy $b_k \in \mathbb{R}, a_k > 0$. Notice that here the orthogonal polynomials are not monic but normalized in $L^2(\mu)$.

The sum rule analogous to (1.2) in the OPRL case is given by the Killip-Simon theorem (Killip and Simon (2003)). It relates the sum of functions of a_k and b_k to a spectral expression involving

μ. Like the Szegő-Verblunsky formula, the spectral side of the sum rule equation measures in some sense the deviation from a reference measure, the semi-circle law

$$\mathrm{SC}(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbb{1}_{[-2,2]}(x)\,dx \tag{1.4}$$

and gives on the "sum-side" the corresponding contribution by the sequence of recursion coefficients. We restate the sum rule of Killip and Simon (2003) in Section 2.1 in full detail. Again, an exhaustive discussion and history of this sum rule can be found in Section 1.10 of the book of Simon (2011). The deep analytical proof is in Chapter 3 of the book.

In both models, the Kullback-Leibler divergence or relative entropy between two probability measures μ and ν plays a major role. When the probability space is $\mathbb{R}$ endowed with its Borel σ-field it is defined by

$$\mathcal{K}(\mu \mid \nu) = \begin{cases} \displaystyle\int_{\mathbb{R}} \log\frac{d\mu}{d\nu}\,d\mu & \text{if } \mu \text{ is absolutely continuous with respect to } \nu \\ \infty & \text{otherwise.} \end{cases} \tag{1.5}$$

Usually, ν is the reference measure. Here the spectral side will involve the reversed Kullback-Leibler divergence, where μ is the reference measure and ν is the argument.

Later, Nazarov et al. (2005) obtained a more general sum rule, when the reference measure is $A(x)\,\mathrm{SC}(dx)$ with A a nonnegative polynomial (see also Kupin (2005) for other generalizations). We will discuss this point in Section 3.3.

1.2 Our main results with hands: outline of the paper

The contribution of this paper is twofold. On the one hand, we show two new sum rules. One for measures on the positive half line and one for measures restricted to a compact interval. In each case, the reference measure is different and the sum involves a function of specific coefficients related to the sequences $(a_n)_n$ and $(b_n)_n$. On the other hand, we also show new large deviation theorems for spectral measures of random operators. In fact, this probabilistic result yields the new sum rules as a direct consequence and also allows for an alternative probabilistic proof of the Killip-Simon sum rule. Notice that as pointed out by Simon in Simon (2007): *"The gems of spectral theory are ones that set up one-one correspondences between classes of measures and coefficients with some properties."* In Section 2.4, we will discuss the underlying gems deduced from the new sum rules.

Large deviations for these random spectral measures arising from the classical ensembles of random matrix theory have been considered before in Gamboa and Rouault (2011). Therein, the main tool was the study of large deviation properties of the recursion coefficients. This method yields as rate function precisely the *sum side* of the new sum rules. Furthermore, the rate function in our new result is the *spectral side*. Since the rate function in large deviations is unique, both sides must be equal. This is, in a nutshell, our proof for sum rules. Our method of proof also stress why both sides of the sum rule equations gives a measure for the *divergence* to some reference measure as they are large deviation rate functions.

To build a comfortable common ground for both mathematical analyst and probabilist reader, we will recap in the further course of this section some useful facts on spectral measures, on their tridiagonal representations and on their randomization. We will also recall the definition of large deviations. In Section 2, we restate the sum rule showed in Killip and Simon (2003) and give our new sum rules. For the convenience of the reader, we formulate the sum rules without mentioning the underlying randomization. In Section 3 we give the main large deviation result and we explain why the sum rules are a consequence of this theorem. We also give a conjecture for a more general sum rule going away from the frame of classical ensembles. The proof of the main large deviation theorem can be found in Section 4. Finally, some technical details are referred to the Appendix.

Let us notice that three extensions of the above method are quite natural and will appear in further work.

1. A matricial version of the Killip-Simon sum rule is the due to Damanik, Killip and Simon (see Damanik et al. (2010) or Theorem 4.6.3 in Simon (2011)). We will extend the results of the present paper to block Laguerre and block Jacobi random matrices. As a matter of fact, we will lean on the large deviation results proved in Gamboa et al. (2012).
2. In the unit circle case, there is a natural model having a limit measure supported by a proper arc of $\mathbb{T}$. In this frame, the random Verblunsky coefficients have a nice independence structure (see Bourgade et al. (2009)). This allows to extend the sum rules developed here.
3. All along this paper, we consider measures with essential support consisting in a single interval (so-called one-cut assumption). We will later consider equilibrium measures supported by several intervals. For this task, the probabilistic tools may be found in Borot and Guionnet (2013a) and the analytic ones are in Chapter 9 of Simon (2011).

1.3 OPRL and tridiagonal matrices

If H is a self-adjoint bounded operator on a Hilbert space $\mathcal{H}$ and e is a cyclic vector, the spectral measure of the pair (H, e) is the unique probability measure μ on $\mathbb{R}$ such that

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

Actually, μ is a unitary invariant for (H, e). Another invariant is the tridiagonal reduction whose coefficients will play the role of the earlier-mentioned Verblunsky coefficients for unitary operators. If $\dim \mathcal{H} = n$ and e is cyclic for H, let $\lambda_1, \dots, \lambda_n$ be the (real) eigenvalues of H and let $\psi_1, \dots, \psi_n$ be a system of orthonormal eigenvectors. The spectral measure of the pair (H, e) is then

$$\mu^{(n)} = \sum_{k=1}^{n} \mathtt{w}_k \delta_{\lambda_k}\,, \tag{1.6}$$

with $\mathtt{w}_k = |\langle \psi_k, e\rangle|^2$. This measure is a weighted version of the empirical eigenvalue distribution

$$\mu_{\mathtt{u}}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k}\,. \tag{1.7}$$

$\mu^{(n)}$ is called *eigenvector empirical distribution function* in a recent paper of Xia et al. (2013). Let us now describe shortly the Jacobi mapping between tridiagonal matrices and spectral measures. We consider $n \times n$ matrices corresponding to measures supported by n points (trivial case) and semi-infinite matrices corresponding to measures with bounded infinite support (non-trivial case). In the basis $\{p_0, p_1, \dots, p_{n-1}\}$, the linear transform $f(x) \to xf(x)$ (multiplication by x) in $L^2(d\mu)$ is represented by the matrix

$$J_\mu = \begin{pmatrix} b_1 & a_1 & 0 & \dots & 0 \\ a_1 & b_2 & a_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_{n-2} & b_{n-1} & a_{n-1} \\ 0 & \dots & 0 & a_{n-1} & b_n \end{pmatrix} \tag{1.8}$$

So, measures supported by n points lead to Jacobi matrices, i.e. $n \times n$ symmetric tridiagonal matrices with subdiagonal positive terms. In fact, there is a one-to-one correspondence between such a matrix and such a measure. If J is a Jacobi matrix, we can take the first vector of the canonical basis as the cyclic vector e. 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 probability measure on $\mathbb{R}$ with bounded infinite support, we may apply the same Gram-Schmidt process and consider the associated semi-infinite Jacobi matrix:

$$J_\mu = \begin{pmatrix} b_1 & a_1 & 0 & 0 & \cdots \\ a_1 & b_2 & a_2 & 0 & \cdots \\ 0 & a_2 & b_3 & a_3 & \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix} \tag{1.9}$$

Notice 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 probability measures 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.

1.4 Randomization: gas distribution and random matrices

In this paper we consider distributions of log-gases and random matrices. In the sequel, n is the number of particles (or eigenvalues), denoted by $\lambda_1, \dots, \lambda_n$, with the joint distributions $\mathbb{P}_V^n$ on $\mathbb{R}^n$ having the density

$$\frac{d\mathbb{P}_V^n(\lambda)}{d\lambda} = \frac{1}{Z_V^n} e^{-n\beta' \sum_{k=1}^n V(\lambda_k)} \prod_{1 \le i < j \le n} |\lambda_i - \lambda_j|^\beta. \tag{1.10}$$

with respect to the Lebesgue measure $d\lambda = d\lambda_1 \cdots d\lambda_n$. The potential $V : \mathbb{R} \to (-\infty, +\infty]$ is supposed to be continuous real valued on the interval (b^-, b^+) $(-\infty \le b^- < b^+ \le +\infty)$, infinite outside of $[b^-, b^+]$ and $\lim_{x\to b^\pm} V(x) = V(b^\pm)$ with possible limit $V(b^\pm) = +\infty$. Let $\beta = 2\beta' > 0$ be the inverse temperature. Under the assumption

(A1) Confinement: $$\liminf_{x\to b^\pm} \frac{V(x)}{2\log|x|} > \max(1, \beta^{-1}),$$

the empirical distribution $\mu_{\mathtt{u}}^{(n)}$ of eigenvalues $\lambda_1, \dots, \lambda_n$ has a limit μ_V (in probability)[1], which is the unique minimizer of

$$\mu \mapsto \mathcal{E}(\mu) := \int V(x)d\mu(x) - \iint \log|x-y|d\mu(x)d\mu(y). \tag{1.11}$$

μ_V has compact support (see Johansson (1998) or Anderson et al. (2010)). Indeed, this is a consequence of the large deviations of the empirical spectral measure (see Theorem 1.2). We will make the following assumptions on μ_V:

(A2) One-cut regime: the support of μ_V is a single interval $[\alpha^-, \alpha^+] \subset [b^-, b^+]$ ($\alpha^- < \alpha^+$).

(A3) Control (of large deviations): the effective potential

$$\mathcal{J}_V(x) := V(x) - 2\int \log|x-\xi|\, d\mu_V(\xi) \tag{1.12}$$

achieves its global minimum value on $(b^-, b^+) \setminus (\alpha^-, \alpha^+)$ only on the boundary of this set.

Furthermore, to obtain a non-variational expression for the rate we need the following conditions:

(A4) Offcriticality: We have

$$d\mu_V(x) = \frac{1}{2\pi} S(x) \sqrt{\frac{\Pi_{\tau\in\mathrm{Soft}}|x-\alpha^\tau|}{\Pi_{\tau'\in\mathrm{Hard}}|x-\alpha^{\tau'}|}}\, dx$$

where $S > 0$ on $[\alpha^-, \alpha^+]$ and $\tau \in \mathrm{Hard}$ iff $b^\tau = \alpha^\tau$, otherwise $\tau \in \mathrm{Soft}$ ($\mathrm{Hard} \cap \mathrm{Soft} = \emptyset$ and $\mathrm{Hard} \cup \mathrm{Soft} = \{-, +\}$).

(A5) Analyticity: V can be extended as an holomorphic function is some open neighborhood of $[\alpha^-, \alpha^+]$.

We remark that for V strictly convex, the assumptions (A2), (A3) and (A4) are fulfilled (see Borot and Guionnet (2013b) and Johansson (1998)).

Hereafter, we discuss the classical models with their potentials and their domains, and their equilibrium measure.

[1] Various authors used to say *almost surely*, but since the probability spaces are not embedded, it seems more convenient to keep *in probability*.

1. Hermite ensemble:

$$V(x) = \frac{x^2}{2} \quad , \quad (b^-, b^+) = \mathbb{R} \text{ and the equilibrium measure is } \mathrm{SC}(dx),\ \alpha^\pm = \pm 2\,.$$

2. Laguerre ensemble of parameter $\tau \in (0, 1]$:

$$V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log x \quad , \quad [b^-, b^+) = [0, \infty)$$

with equilibrium measure the Marchenko-Pastur law with parameter τ,

$$\mathrm{MP}_\tau(dx) = \frac{\sqrt{(\tau^+ - x)(x - \tau^-)}}{2\pi\tau x}\ \mathbb{1}_{(\tau^-,\tau^+)}(x)dx \quad , \quad \alpha^\pm = \tau^\pm = (1 \pm \sqrt{\tau})^2$$

3. Jacobi ensemble of parameters $\kappa_1, \kappa_2 \geq 0$:

$$V(x) = -\kappa_1 \log x - \kappa_2 \log(1 - x) \quad , \quad [b^-, b^+] = [0, 1]\,.$$

The equilibrium measure is the Kesten-McKay distribution

$$\mathrm{KMK}_{\kappa_1,\kappa_2}(dx) = \frac{(2 + \kappa_1 + \kappa_2)}{2\pi} \frac{\sqrt{(u^+ - x)(x - u^-)}}{x(1 - x)}\ \mathbb{1}_{(u^-,u^+)}(x)dx\,,$$

where

$$\alpha^\pm = u^\pm := \frac{1}{2} + \frac{\kappa_1^2 - \kappa_2^2 \pm 4\sqrt{(1 + \kappa_1)(1 + \kappa_2)(1 + \kappa_1 + \kappa_2)}}{2(2 + \kappa_1 + \kappa_2)^2}. \tag{1.13}$$

Let us start with a crash recall in random matrix theory. The GOE of order n is a probability distribution $P_n^{(1)}$ on the set of all symmetric real $n \times n$ matrices, obtained by assuming that the diagonal entries are distributed as $\mathcal{N}(0, 2)$ and the non-diagonal ones as $\mathcal{N}(0, 1)$ and that entries are independent up to symmetry. Taking on-or-above-diagonal entries as coordinates of the random matrix H, this gives a density with respect to the Lebesgue measure proportional to $\exp -\mathrm{tr}H^2/4$. The distribution of eigenvalues of $\frac{1}{\sqrt{n}}H$ is given by (1.10) with $\beta = 1$ and $V(x) = x^2/2$. Besides, Trotter (1984) proved that the coefficients b_k are Gaussian and the coefficients a_k^2 are gamma distributed, with convenient parameter. Furthermore, by the invariance of the Gaussian distribution under rotation, the first row of the eigenvector matrix is independent of the eigenvalues and uniformly distributed on the sphere. Dumitriu and Edelman (2002) also proved that conversely, if we take an array $(\lambda_1, \ldots, \lambda_n)$ distributed as in (1.10), with general $\beta > 0$ and an independent array of weights $(\mathtt{w}_1, \ldots, \mathtt{w}_n)$ sampled with the Dirichlet distribution $\mathrm{Dir}_n(\beta')$ of order n and parameter β' on the simplex $\sum_i \mathtt{w}_i = 1$, i.e. with density proportional to

$$(\mathtt{w}_1 \cdots \mathtt{w}_n)^{\beta'-1}\,,$$

then the coefficients of the tridiagonal matrix are independent Gaussian and gamma variables, respectively. In the case of Laguerre and Jacobi ensembles, other systems of auxiliary variables with nice structure of independence were introduced in Dumitriu and Edelman (2002) and in Killip and Nenciu (2004). We will use these parametrizations in the next sections.

The case of a general potential V is not so easy. Nevertheless, the correspondence is ruled by the following result. Let $A^{(n)} = (a_1, \dots, a_{n-1})$ and $B^{(n)} = (b_1, \dots, b_n)$ and let $T^{(n)}$ be the symmetric tridiagonal matrix with $T^{(n)}_{k,k} = b_k$ for $k \le n$ and $T^{(n)}_{k,k+1} = T^{(n)}_{k+1,k} = a_k$ for $k \le n-1$. Then we may state the following theorem.

Theorem 1.1 (Krishnapur et al. (2013) Prop.2) *Let $(A^{(n)}, B^{(n)})$ sampled from the density proportional to*

$$\exp -n\beta' \left[\operatorname{tr} V(T^{(n)}) - 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} - \frac{1}{n\beta}\right) \log a_k \right] .$$

Then the eigenvalues $\lambda_1, \dots, \lambda_n$ have joint density proportional to

$$e^{-n\beta' \sum_1^n V(\lambda_k)} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|^\beta$$

and the weights $\mathtt{w}_1, \dots, \mathtt{w}_n$ are independent with distribution $\mathrm{Dir}_n(\beta')$.

1.5 Large deviations

In order to be self-contained, let us recall the definition of a large deviation principle. For a general reference of large deviation statements we refer to the book of Dembo and Zeitouni (1998) or to the Appendix D of Anderson et al. (2010).

Let U be a topological Hausdorff space with Borel σ-algebra $\mathcal{B}(U)$. We say that a sequence $(P_n)_n$ of probability measures on $(U, \mathcal{B}(U))$ satisfies a large deviation principle (LDP) with speed a_n and rate function $\mathcal{I} : U \to [0, \infty]$ if:

(i) $\mathcal{I}$ is lower semicontinuous.

(ii) For all closed sets $F \subset U$:

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(F) \le -\inf_{x\in F} \mathcal{I}(x)$$

(iii) For all open sets $O \subset U$:

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(O) \ge -\inf_{x\in O} \mathcal{I}(x)$$

The rate function $\mathcal{I}$ is good if its level sets $\{x \in U |\ \mathcal{I}(x) \le a\}$ are compact for all $a \ge 0$. If in the conditions above, we replace *closed sets* by *compact sets*, we say that $(P_n)_n$ satisfies a weak LDP. In this case, we can recover a LDP if the additional condition of exponential tighness is fulfilled:

For every $M > 0$ there exists a compact set $K_M \subset U$ such that

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(U \setminus K_M) \leq -M \,.$$

In our case, the measures P_n will be the distributions of the random spectral measures μ_n and we will say that the sequence of measures μ_n satisfies a LDP.

The most famous LDP in random matrix theory is for the sequence of empirical spectral measures. We let $\mathcal{P}_1$ denote the set of all probability measures on $\mathbb{R}$.

Theorem 1.2 *If the potential V satisfies assumption (A1), and if $(\lambda_1, \dots, \lambda_n)$ is distributed according to $\mathbb{P}_V^n$ (see (1.10)), then the sequence of random probability measures $(\mu_{\mathbf{u}}^{(n)})$ satisfies in $\mathcal{P}_1$ equipped with the weak topology, a LDP with speed $\beta' n^2$ and good rate function*

$$\mu \mapsto \mathcal{E}(\mu) - \mathcal{E}(\mu_V)$$

where $\mathcal{E}$ is defined in (1.11).

Let us recall the definition of convex duality, used several times in the proofs. If R is a function defined on a topological vector space $\mathcal{H}$ and valued in $(-\infty, \infty]$, then its convex dual R^* is a function defined on the topological dual space $\mathcal{H}^*$ by

$$R^*(x) = \sup_{\theta\in\mathcal{H}} \left[\langle \theta, x\rangle - R(\theta)\right] \quad (x \in \mathcal{H}^*).$$

Here, $\langle\ ,\ \rangle$ is the duality bracket. For $\mathcal{H} = \mathbb{R}$, two examples are meaningful in our context

1. Gaussian case

$$L_0(\theta) = \frac{\theta^2}{2} \Rightarrow L_0^*(x) = \frac{x^2}{2} \tag{1.14}$$

2. Exponential case

$$L(\theta) = \begin{cases} -\log(1-\theta) & \text{if } \theta < 1\,, \\ \infty & \text{otherwise,} \end{cases} \tag{1.15}$$

then

$$L^*(x) = G(x) := \begin{cases} x - 1 - \log x & \text{if } x > 0\,, \\ \infty & \text{otherwise.} \end{cases} \tag{1.16}$$

Let us denote respectively by $\mathcal{N}(0, \sigma^2)$ the centered Gaussian distribution and Gamma(a, b) the gamma distribution of order $a > 0$ and scale factor $b > 0$ with density

$$x \mapsto \frac{x^{a-1}}{b^a \Gamma(a)} e^{-\frac{x}{b}} \quad x > 0\,.$$

L_0^* is the rate function of the LDP satisfied by $(\mathcal{N}(0, n^{-1}))_n$ at speed n and L^* is the rate function of the LDP satisfied by $(\mathrm{Gamma}(n, n^{-1}))_n$, also at speed n. Besides, $(\mathrm{Gamma}(a, n^{-1}))_n$ satisfies a LDP at speed n with rate function

$$x \mapsto \begin{cases} x & \text{if } x \geq 0, \\ \infty & \text{otherwise.} \end{cases} \tag{1.17}$$

2 Sum rules from large deviations

Let $\mathcal{S} = \mathcal{S}(\alpha^-, \alpha^+)$ be the set of all bounded positive measures μ on $\mathbb{R}$ with

(i) $\mathrm{supp}(\mu) = J \cup \{\lambda_i^-\}_{i=1}^{N^-} \cup \{\lambda_i^+\}_{i=1}^{N^+}$, where $J \subset I = [\alpha^-, \alpha^+]$, $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$ and
$$\lambda_1^- < \lambda_2^- < \cdots < \alpha^- \quad \text{and} \quad \lambda_1^+ > \lambda_2^+ > \cdots > \alpha^+.$$

(ii) If N^- (resp. N^+) is infinite, then λ_j^- converges towards α^- (resp. λ_j^+ converges to α^+).

Such a measure μ will be written as

$$\mu = \mu_{|I} + \sum_{i=1}^{N^+} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \gamma_i^- \delta_{\lambda_i^-} \tag{2.1}$$

Further, we define $\mathcal{S}_1 = \mathcal{S}_1(\alpha^-, \alpha^+) := \{\mu \in \mathcal{S} \,|\, \mu(\mathbb{R}) = 1\}$. We endow $\mathcal{S}_1$ with the weak topology and the corresponding Borel σ-algebra.

2.1 Hermite case revisited

We start by stating the classical sum rule (due to Killip and Simon (2003) and explained in Simon (2011) p.37), the new probabilistic proof using large deviations is tackled in Section 3.2. The sum rule gives two different expressions for the *distance* to the semicircle law SC. Its Jacobi coefficients are

$$a_k^{\mathrm{SC}} = 1,\ b_k^{\mathrm{SC}} = 0 \ \text{ for all } k \geq 1\,. \tag{2.2}$$

For a probability measure μ on $\mathbb{R}$ with recursion coefficients $(a_k)_k$, $(b_k)_k$ as in (1.3), define the sum

$$\mathcal{I}_H(\mu) = \sum_{k\geq 1} \big(\frac{1}{2} b_k^2 + G(a_k^2)\big) = \sum_{k \geq 1} \big(L_0^*(b_k) + G(a_k^2)\big), \tag{2.3}$$

where G and L_0^* have been defined in the previous section. Further, let

$$\mathcal{F}_H^+(x) := \begin{cases} \displaystyle\int_2^x \sqrt{t^2-4}\, dt = \tfrac{x}{2}\sqrt{x^2-4} - 2\log\left(\tfrac{x+\sqrt{x^2-4}}{2}\right) & \text{if } x \geq 2 \\ \infty \quad \text{otherwise,} & \end{cases}$$

and, for $x \in \mathbb{R}$, set $\mathcal{F}_H^-(x) := \mathcal{F}_H^+(-x)$. Then we have the following theorem.

Theorem 2.1 (Killip and Simon (2003)) *Let J be a Jacobi matrix with diagonal entries $b_1, b_2, \ldots \in \mathbb{R}$ and subdiagonal entries $a_1, a_2, \ldots > 0$ satisfying $\sup_k a_k + \sup_k |b_k| < \infty$ and let μ be the associated spectral measure. Then $\mathcal{I}_H(\mu)$ is infinite if $\mu \notin \mathcal{S}_1(-2,2)$ and for $\mu \in \mathcal{S}_1(-2,2)$,*

$$\mathcal{I}_H(\mu) = \mathcal{K}(\mathrm{SC}|\, \mu) + \sum_{n=1}^{N^+} \mathcal{F}_H^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_H^-(\lambda_n^-)$$

where both sides may be infinite simultaneously.

2.2 New magic sum rule: the Laguerre case

For our first new sum rule, we consider probability measures μ supported on $[0,\infty)$. In this case, the recursion coefficients can be decomposed as

$$
\begin{aligned}
b_k =& z_{2k-2} + z_{2k-1}, \\
a_k^2 =& z_{2k-1} z_{2k},
\end{aligned}
\tag{2.4}
$$

for $k \geq 1$, where $z_k \geq 0$ and $z_0 = 0$. In fact, by Favard's Theorem a measure μ is supported on $[0,\infty)$ if and only if the decomposition as in (2.4) holds. The central probability measure is the Marchenko-Pastur law MP_τ defined in Section 1.4, whose Jacobi coefficients are

$$
a_k^{\mathrm{MP}} = \sqrt{\tau}\ (k \geq 1) \quad , \quad b_1^{\mathrm{MP}} = 1\ ,\ b_k^{\mathrm{MP}} = 1+\tau \quad (k \geq 2)
\tag{2.5}
$$

and correspond to $z_{2n-1}^{\mathrm{MP}} = 1$ and $z_{2n}^{\mathrm{MP}} = \tau$ for all $n \geq 1$. For a measure μ supported on $[0,\infty)$, let

$$
\mathcal{I}_L(\mu) := \sum_{k=1}^{\infty} \tau^{-1} G(z_{2k-1}) + G(\tau^{-1} z_{2k}).
\tag{2.6}
$$

For the new sum rule, we have to replace $\mathcal{F}_H^{\pm}$ by

$$
\mathcal{F}_L^+(x) = \begin{cases} \displaystyle\int_{\tau^+}^{x} \frac{\sqrt{(t-\tau^-)(t-\tau^+)}}{t\tau}\, dt & \text{if } x \geq \tau^+, \\ \infty \quad \text{otherwise,} \end{cases}
$$

$$
\mathcal{F}_L^-(x) = \begin{cases} \displaystyle\int_{x}^{\tau^-} \frac{\sqrt{(\tau^- - t)(\tau^+ - t)}}{t\tau}\, dt & \text{if } x \leq \tau^-, \\ \infty \quad \text{otherwise.} \end{cases}
$$

Then we have the following magic sum rule for probability measures on $[0,\infty)$. The probabilistic proof can be found in Section 3.2.

Theorem 2.2 *Assume the entries of the Jacobi matrix J can be decomposed as in* (2.4) *with $\sup_k z_k < \infty$ and let μ be the spectral measure of J. Then for all $\tau \in (0,1]$, $\mathcal{I}_L(\mu) = \infty$ if $\mu \notin \mathcal{S}_1(\tau^-,\tau^+)$. If $\mu \in \mathcal{S}_1(\tau^-,\tau^+)$, we have*

$$
\mathcal{I}_L(\mu) = \mathcal{K}(\mathrm{MP}_\tau \,|\, \mu) + \sum_{n=1}^{N^+} \mathcal{F}_L^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_L^-(\lambda_n^-)
$$

where both sides may be infinite simultaneously.

Note that if $\tau = 1$, the support of the limit measure is $[0,4]$, so that we have a hard edge at 0 with $N^- = 0$ and no contribution of outliers to the left.

2.3 New magic sum rule: the Jacobi case

Our second new sum rule is a generalization of the Szegő theorem for probability measures on the unit circle. The classical Szegő mapping is a correspondence between a probability measure ν on $\mathbb{T}$ invariant by $\theta \mapsto 2\pi - \theta$ and a probability measure μ on $[-2, 2]$ obtained by pushing forward ν by the mapping $\theta \mapsto 2\cos\theta$. In this case the Verblunsky coefficients $(\alpha_k)_{k\geq 0}$ of ν (they all belong to $[-1, 1]$ by symmetry) are by extension called the Verblunsky coefficients of μ. For $k \geq 1$, the recursion coefficients associated with μ are connected with the Verblunsky coefficients by the Geronimus relations:

$$\begin{aligned} b_{k+1} &= (1-\alpha_{2k-1})\alpha_{2k} - (1+\alpha_{2k-1})\alpha_{2k-2} \\ a_{k+1} &= \sqrt{(1-\alpha_{2k-1})(1-\alpha_{2k}^2)(1+\alpha_{2k+1})} \end{aligned} \tag{2.7}$$

where $\alpha_k \in [-1, 1]$ and $\alpha_{-1} = -1$. While these recursion coefficients give a measure μ on $[-2, 2]$, it is more convenient for our approach to consider the measure $\tilde{\mu}$ on $[0, 1]$ obtained by pushing forward μ by the affine mapping $x \mapsto \frac{1}{2} - \frac{1}{4}x$. We keep calling $(\alpha_k)_k$ the Verblunsky coeffcients of $\tilde{\mu}$. The Jacobi coefficients of $\widetilde{\mu}$ are

$$\widetilde{b}_k = \frac{2-b_k}{4}\ ,\ \widetilde{a}_k = \frac{a_k}{4}\quad (k \geq 1)\,.$$

Here, the important probability measure is the Kesten-McKay distribution $\mathrm{KMK}_{\kappa_1,\kappa_2}$ on $[0, 1]$ with parameters $\kappa_1, \kappa_2 \geq 0$. The associated Verblunsky coefficients are, for $k \geq 0$,

$$\alpha_{2k}^{KMK} = \frac{\kappa_1 - \kappa_2}{2+\kappa_1+\kappa_2}, \quad \alpha_{2k+1}^{KMK} = -\frac{\kappa_1+\kappa_2}{2+\kappa_1+\kappa_2}.$$

and the corresponding Jacobi coefficients are

$$\widetilde{a}_1^{KMK} = \frac{\sqrt{(1+\kappa_1)(1+\kappa_2)}}{(2+\kappa_1+\kappa_2)^{3/2}}\ , \quad \widetilde{b}_1^{KMK} = \frac{1+\kappa_2}{2+\kappa_1+\kappa_2}\,, \tag{2.8}$$

and for $k \geq 2$

$$\widetilde{a}_k^{KMK} = \frac{\sqrt{(1+\kappa_1+\kappa_2)(1+\kappa_1)(1+\kappa_2)}}{(2+\kappa_1+\kappa_2)^2}\ , \quad \widetilde{b}_k^{KMK} = \frac{1}{2}\left[1 - \frac{\kappa_1^2-\kappa_2^2}{(2+\kappa_1+\kappa_2)^2}\right]\,.$$

Set

$$\mathcal{I}_J(\tilde{\mu}) = \sum_{k=0}^{\infty} H_1(\alpha_{2k+1}) + H_2(\alpha_{2k}), \tag{2.9}$$

where for $x \in [-1, 1]$

$$H_1(x) = -(1+\kappa_1+\kappa_2)\log\left[\frac{2+\kappa_1+\kappa_2}{2(1+\kappa_1+\kappa_2)}(1-x)\right] - \log\left[\frac{2+\kappa_1+\kappa_2}{2}(1+x)\right]$$

$$H_2(x) = -(1+\kappa_1)\log\left[\frac{(2+\kappa_1+\kappa_2)}{2(1+\kappa_1)}(1+x)\right] - (1+\kappa_2)\log\left[\frac{(2+\kappa_1+\kappa_2)}{2(1+\kappa_1)}(1-x)\right].$$

Let $\mathcal{F}_J^+$ be defined by

$$\mathcal{F}_J^+(x) = \begin{cases} \displaystyle\int_{u^+}^{x} \frac{\sqrt{(t-u^+)(t-u^-)}}{t(1-t)}\,dt & \text{if } u^+ \le x \le 1 \\ \infty & \text{otherwise.} \end{cases}$$

Similarly, let

$$\mathcal{F}_J^-(x) = \begin{cases} \displaystyle\int_{x}^{u^-} \frac{\sqrt{(u^- - t)(u^+ - t)}}{t(1-t)}\,dt & \text{if } 0 \le x \le u^- \\ \infty & \text{otherwise.} \end{cases}$$

Then the following magic sum rule for probability measures on $[0,1]$ holds.

Theorem 2.3 *Let $\tilde{\mu}$ be a probability measure on $[0,1]$ and let α_k be the Verblunsky coefficients of $\tilde{\mu}$. Then for any $\kappa_1, \kappa_2 \ge 0$, $\mathcal{I}_J(\tilde{\mu}) = \infty$ if $\tilde{\mu} \notin \mathcal{S}_1(u^-, u^+)$. If $\tilde{\mu} \in \mathcal{S}_1(u^-, u^+)$, then*

$$\mathcal{I}_J(\tilde{\mu}) = \mathcal{K}(\mathrm{KMK}_{\kappa_1,\kappa_2} \,|\, \tilde{\mu}) + \sum_{n=1}^{N^+} \mathcal{F}_J^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_J^-(\lambda_n^-)$$

and both sides may be infinite simultaneously.

Similar to the Laguerre case, if $\kappa_1 = 0$ or $\kappa_2 = 0$, then $u^- = 0$ or $u^+ = 1$, respectively, and we have no contribution coming from respective outliers. In particular, if $\kappa_1 = \kappa_2 = 0$, the Kesten-McKay distribution reduces to the arcsine distribution

$$d\mu_0(x) = \frac{1}{\pi\sqrt{x(1-x)}} \mathbb{1}_{(0,1)}(x)\,dx$$

and then the sum rule reads

$$\mathcal{K}(\mu_0 \,|\, \tilde{\mu}) = -\sum_{n=0}^{\infty} \log(1-\alpha_n^2), \tag{2.10}$$

which is nothing more than the classical Szegő sum rule written for probability measures pushed forward by the application $\theta \to \frac{1}{2} - \frac{\cos\theta}{2}$. Notice also that in this frame we may rewrite this sum rule in terms of a cousin parametrization. Namely, by the way of the so-called canonical moments defined for a measure supported on $[0,1]$ (see the excellent book of Dette and Studden (1997)). More precisely, let for $k \ge 1$, p_k denote the canonical moment of order k of $\tilde{\mu}$. Recall that p_k

may be defined browsing different paths, the straightest is from the ordinary moments. Indeed, assuming that $\tilde{\mu}$ is not supported by a finite number of points, we have

$$p_1 := \int_0^1 x\tilde{\mu}(dx),$$
$$p_{n+1} := \frac{\int_0^1 x^{n+1}\tilde{\mu}(dx) - c_{n+1}^-}{c_{n+1}^+ - c_{n+1}^-} \quad \text{for } n \geq 1.$$

Here, c_{n+1}^+ (resp. c_{n+1}^-) is the maximum (resp. minimum) possible value for the $(n+1)$-th moment of a probability measure supported by $[0,1]$ having the same n first moments as $\tilde{\mu}$. With this parametrisation, as $\alpha_n = 2p_{n+1} - 1$ for all integer n (see Dette and Studden (1997) p. 287), the sum rule (2.10) becomes

$$\mathcal{K}(\mu_0 \,|\, \tilde{\mu}) = -\sum_{n=0}^{\infty} \log(4p_n(1-p_n)).$$

Notice that the last expression is also the functional obtained in Gamboa and Lozada-Chang (2004) in the study of large deviations for a random Hausdorff moment problem.

2.4 Semiprecious gems

In the introduction of the paper we pointed out that the gem of spectral theory is to set up one-one correspondences between classes of measures and coefficients with some properties. More precisely, a gem (see Simon (2011) Section 1.4) gives a one-one correspondence between properties on the sequences (a_n) and (b_n) and the associated spectral measure. The gem corresponding to the Hermite case has been proved by Killip and Simon (see Simon (2011) Section 1.10). We discuss here such correspondences for the two sum rules given in Theorems 2.2 and 2.3. Although we do not succeed to find the holy grail of such a correspondence between classes relying on (a_n) and (b_n), we set it in terms of the sequences (z_n) (Laguerre case) and (α_n) (Jacobi case).

Corollary 2.4 *Assume the entries of the Jacobi matrix J can be decomposed as in* (2.4) *with* $\sup_k z_k < \infty$ *and let μ be the spectral measure of J. Then*

$$\sum_{k=1}^{\infty} [(z_{2k-1} - 1)^2 + (z_{2k} - \tau)^2] < \infty \tag{2.11}$$

(that is, $\mathcal{I}_L(\mu) < \infty$) if and only if

1. $\mu \in \mathcal{S}_1(\tau^-, \tau^+)$
2. $\sum_{i=1}^{N^+} (\lambda_i^+ - \tau^+)^{3/2} + \sum_{i=1}^{N^-} (\tau^- - \lambda_i^-)^{3/2} < \infty$ *and if* $N^- > 0$*, then* $\lambda_1^- > 0$.
3. *the spectral measure μ of J with decomposition $d\mu(x) = f(x)dx + d\mu_s(x)$ with respect to the Lebesgue measure satisfies*

$$\int_{\tau^-}^{\tau^+} \frac{\sqrt{(\tau^+ - x)(x - \tau^-)}}{x} \log(f(x))dx > -\infty.$$

Proof: It is enough to notice that $\mathcal{F}_L^-(0) = \infty$ and

$$\mathcal{F}_L^\pm(\tau^\pm \pm h) = \frac{4}{3\tau^{3/4}(1+\pm\sqrt{\tau})^2} h^{3/2} + o(h^{3/2}) \quad (h \to 0^+)$$

and

$$G(1+h) = \frac{h^2}{2} + o(h) \quad (h \to 0)$$

Remark 2.5 *Comparing with the Hermite gem, it would be most desirable to obtain a purely spectral criterion for when J is a Hilbert-Schmidt operator relative to the Jacobi operator of the equilibrium measure, that is, in the Laguerre case*

$$\sum_{k=1}^{\infty} [(b_k - 1 - \tau)^2 + (a_k - \sqrt{\tau})^2] < \infty. \tag{2.12}$$

Unfortunately, Theorem 2.2 will not yield such a criterion. (2.11) implies (2.12) but the converse is not true. As an example, set $z_{2k-1} = \tau$, $z_{2k} = 1$ for all $k \geq 1$, then (2.12) is clearly satisfied, but $\mathcal{I}_L(\mu) = \infty$ for μ the spectral measure of J and $\tau \neq 1$. Actually, this system of coefficients correspond to the measure

$$\mu(dx) = (1-\tau)\delta_0 + \tau \, \mathrm{MP}_\tau(dx)$$

(see Saitoh and Yoshida (2001) for the identification); the extra mass in 0 gives a contribution $\mathcal{F}_L^-(0) = \infty$, the condition 2 is not fulfilled, although the conditions 1 and 3 are fulfilled.

Corollary 2.6 *Assume the entries of the Jacobi matrix J can be decomposed as in (2.7) and let μ be the spectral measure of J with pushforward $\tilde{\mu}$ under the mapping $x \mapsto \frac{1}{2} - \frac{1}{4}x$. Then, for any $\kappa_1, \kappa_2 > 0$,*

$$\sum_{k=1}^{\infty} \left[\left(\alpha_{2k-1} + \frac{\kappa_1 + \kappa_2}{2 + \kappa_1 + \kappa_2)} \right)^2 + \left(\alpha_{2k} - \frac{\kappa_1 - \kappa_2}{2 + \kappa_2 + \kappa_2} \right)^2 \right] < \infty \tag{2.13}$$

(that is, $\mathcal{I}_J(\tilde{\mu}) < \infty$) if and only if

1. $\tilde{\mu} \in \mathcal{S}_1(u_-, u_+)$
2. $\sum_{i=1}^{N^-} (\frac{1}{2} - \frac{1}{4}\lambda_i^- - u^+)^{3/2} + \sum_{i=1}^{N^+} (u^- - \frac{1}{2} + \frac{1}{4}\lambda_i^+)^{3/2} < \infty$ *and* $\lambda_1^- > -2$ *if* $N^- > 0$ *and* $\lambda_1^+ < 2$ *if* $N^+ > 0$.
3. *when* $d\tilde{\mu}(x) = f(x)dx + d\tilde{\mu}_s(x)$ *is the decomposition of* $\tilde{\mu}$ *with respect to the Lebesgue measure, then*

$$\int_{u^-}^{u^+} \frac{\sqrt{(u^+ - x)(x - u^-)}}{x(1-x)} \log(f(x))dx > -\infty.$$

The proof is similar to the Laguerre case.

Remark 2.7 *We can argue as in Remark 2.5. In particular (2.13) implies*

$$\sum_{k=1}^{\infty}[(\tilde{b}_k - \tilde{b}_k^{\mathrm{KMK}})^2 + (\tilde{a}_k - \tilde{a}_k^{\mathrm{KMK}})^2] < \infty\,, \tag{2.14}$$

but it is not equivalent.

3 Large deviations main theorem

3.1 The main result

Our large deviation result will hold for general eigenvalue distributions $\mathbb{P}_V^n$ defined in (1.10). The corresponding spectral measure $\mu^{(n)}$ is then defined by (1.6), where the weights $\mathtt{w}_1, \dots, \mathtt{w}_n$ are $\mathrm{Dir}_n(\beta')$ distributed and independent of the eigenvalues. We regard $\mu^{(n)}$ as a random element of $\mathcal{P}_1$, the set of all probability measures on $\mathbb{R}$, endowed with the weak topology and the corresponding σ-algebra. We need one more definition in order to formulate the general result.

Recall that $\mathcal{J}_V$ has been defined in assumption (A3). We define, in the general case, the rate function for the extreme eigenvalues,

$$\mathcal{F}_V^+(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi\in\mathbb{R}} \mathcal{J}_V(\xi) & \text{if } \alpha^+ \le x \le b^+, \\ \infty & \text{otherwise,} \end{cases} \tag{3.1}$$

$$\mathcal{F}_V^-(x) = \begin{cases} \mathcal{J}_\mathcal{V}(x) - \inf_{\xi\in\mathbb{R}} \mathcal{J}_V(\xi) & \text{if } b^- \le x \le \alpha^-, \\ \infty & \text{otherwise.} \end{cases} \tag{3.2}$$

Theorem 3.1 *Assume that the potential V satisfies the assumptions (A1), (A2) and (A3). Then the sequence of spectral measures $\mu^{(n)}$ under $\mathbb{P}_V^n \otimes \mathrm{Dir}_n(\beta')$ satisfies the LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}_V(\mu) = \mathcal{K}(\mu_V \,|\, \mu) + \sum_{n=1}^{N^+} \mathcal{F}_V^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_V^-(\lambda_n^-)$$

if $\mu \in \mathcal{S}_1(\alpha^-, \alpha^+)$ and $\mathcal{I}_V(\mu) = \infty$ otherwise.

Additionally, we have an alternative expression for $\mathcal{F}_V^\pm$, given by the following proposition. This result is more or less classical. It may be found in Deift et al. (1999) (proof of Theorem 3.6) or in Albeverio et al. (2001) (Equation (1.13)).

Proposition 3.2 *If moreover, the conditions of analyticity (A5) and off-criticality (A4) are satisfied, then*

$$\mathcal{F}_V^+(x) = \int_{\alpha^+}^{x} S(t)\sqrt{(t-\alpha^-)(t-\alpha^+)}\,dt \qquad \text{if } x \ge \alpha^+, \tag{3.3}$$

$$\mathcal{F}_V^-(x) = \int_{x}^{\alpha^-} S(t)\sqrt{(\alpha^- - t)(\alpha^+ - t)}\,dt \qquad \text{if } x \le \alpha^-. \tag{3.4}$$

3.2 From large deviations to sum rules

As described in the introduction, the sum rules of Section 2 are a consequence of two different proofs of a LDP, one leading to our main result Theorem 3.1, giving the spectral side and another one yielding the sum side. Let us explain this in detail for the Hermite case.

For the Hermite case with probability measures on the whole real line, the correct randomization on the set of probability measures is the Hermite ensemble, defined by the eigenvalue density

$$c_\beta \prod_{i<j} |\lambda_j - \lambda_i|^\beta \prod_{i=1}^n e^{-\frac{\beta' n}{2}\lambda_i^2}$$

corresponding to the potential $V(x) = x^2/2$, and with weights following the Dirichlet distribution independent of the eigenvalues. Wigner's famous theorem states that the weak limit of the empirical eigenvalue distribution is then the semicircle law SC. Indeed, here the potential V satisfies all assumptions in Section 3.1 with $\mu_V = \mathrm{SC}$ and $S(x) = \frac{1}{2}$ on $[-2, 2]$. Thus, by Theorem 3.1 the LDP for the measure $\mu^{(n)}$ holds. Further, by Proposition 3.2, we may calculate the rate for the outliers as $\mathcal{F}_V^\pm = \mathcal{F}_H^\pm$. The rate function $\mathcal{I}_V$ is therefore precisely the right hand side in Theorem 2.1.
On the other hand, the recursion coefficients $(a_k)_k, (b_k)_k$ of the measure $\mu^{(n)}$ are independent with respectively gamma and normal distributions. Using this representation for the spectral measure, Gamboa and Rouault (2011) proved that $\mu^{(n)}$ satisfies an LDP, again with speed $\beta' n$, and with rate function $\mathcal{I}_H$ the left hand side in Theorem 2.1. Since the rate function is unique, we must have $\mathcal{I}_V = \mathcal{I}_H$.

For the new sum rules, the arguments are similar. In the Laguerre case, the eigenvalue distribution of the spectral measure is

$$c_{\tau,\beta} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n \lambda_i^{\beta' n(\tau^{-1}-1)} e^{-\beta' n \tau^{-1}\lambda_i} \mathbb{1}_{\{\lambda_i > 0\}}$$

with $\tau \in (0, 1]$ and independent Dirichlet distributed weights. The potential of the Laguerre ensemble is $V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log x$ on $(0, \infty)$. As $n \to \infty$, the empirical eigenvalue distribution and the weighted spectral measure $\mu^{(n)}$ converge to the Marchenko-Pastur law MP_τ. Moreover, the assumptions of Theorem 3.1 are satisfied and we have an LDP with speed $\beta' n$ and rate function $\mathcal{I}_V$ the right hand side in the new sum rule, as $\mathcal{F}_V^\pm = \mathcal{F}_L^\pm$. As for the Hermite ensemble, Gamboa and Rouault (2011) proved an LDP for $\mu^{(n)}$ in the subset of probability measures on $[0, \infty)$ with speed $\beta' n$ and rate function $\mathcal{I}_L(\mu)$ (note that Gamboa and Rouault (2011) consider the speed $\beta' n\tau$). The uniqueness of the rate function implies $\mathcal{I}_V = \mathcal{I}_L$.

In the Jacobi case, the eigenvalue density is

$$c_{\kappa_1,\kappa_2,\beta} \cdot \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n \lambda_i^{\kappa_1 \beta' n}(1-\lambda_i)^{\kappa_2 \beta' n} \mathbb{1}_{\{0<\lambda_i<1\}}$$

corresponding to the potential $V(x) = -\kappa_1 \log(x) - \kappa_2 \log(1-x)$ on $(0,1)$ for parameters $\kappa_1, \kappa_2 \geq 0$. The equilibrium measure is then the Kesten-McKay distribution $\mathrm{KMK}_{\kappa_1,\kappa_2}$. By Theorem 3.1, $\mu^{(n)}$ satisfies the LDP with rate function $\mathcal{I}_V$, where additionally $\mathcal{F}_V^{\pm} = \mathcal{F}_J^{\pm}$. On the other hand, we know from the paper of Gamboa and Rouault (2011), that the LDP with rate function $\mathcal{I}_H$ holds. The combination of these two results yields Theorem 2.3.

3.3 Conjecture for a general sum rule

3.3.1 The probabilistic point of view

We know from Section 3.1 that under some assumptions on the potential V, the random spectral measure sequence $(\mu^{(n)})_n$ satisfies the LDP with rate function $\mathcal{I}_V$. Besides, owing to Theorem 1.1, we can hope to compute the rate function of the encoding by Jacobi coefficients directly from the expression of the density. For a semi-infinite Jacobi matrix $T = T((a_k)_k, (b_k)_k)$ with upper left $n \times n$ block $T^{(n)}$ set

$$\mathcal{H}(T^{(n)}) = \mathrm{tr}\ V(T^{(n)}) - 2\sum_{k=1}^{n-1} \log a_k$$

It would give for the rate function of the LDP at the speed $n\beta'$

$$T \longmapsto \lim_{n\to\infty} \left[\mathcal{H}(T^{(n)}) - \inf_S \mathcal{H}(S^{(n)})\right]$$

So, we conjecture the following identity, as soon as V is a polynomial with even degree and positive leading coefficient :

$$\mathcal{K}(\mu_V \,|\, \mu) + \sum_{k=1}^{N^+} \mathcal{F}_V(\lambda_k^+) + \sum_{k=1}^{N^-} \mathcal{F}_V(\lambda_k^-) = \lim_{n\to\infty} \left[\mathcal{H}(J_\mu^{(n)}) - \inf_S \mathcal{H}(S^{(n)})\right] \tag{3.5}$$

Let us show that this is in agreement with the sum rules proven in this paper. In the Hermite case when $V(x) = x^2/2$, we get

$$\mathrm{tr}\ V(T^{(n)}) = \frac{1}{2}\sum_{k=1}^{n} b_k^2 + \sum_{k=1}^{n-1} a_k^2$$

and then

$$\mathcal{H}(T^{(n)}) = \frac{1}{2}\sum_{k=1}^{n} b_k^2 + \sum_{k=1}^{n-1} a_k^2 - \log a_k^2$$

Now, $\inf_S \mathcal{H}(S^{(n)})$ is achieved for $b_k(S) \equiv 0, a_k(S) \equiv 1$, so that

$$\lim_{n\to\infty} \left[\mathcal{H}(J_\mu^{(n)}) - \inf_S \mathcal{H}(S^{(n)})\right] = \sum_{k=1}^{\infty} \frac{1}{2} b_k^2 + G(a_k^2),$$

where $G(x) = x - 1 - \log x$ and this is exactly the rate function of (2.3).

In the Laguerre case when $V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log x$,

$$\mathrm{tr}\ V(T^{(n)}) = \tau^{-1}\mathrm{tr}\ T^{(n)} - (\tau^{-1} - 1)\log\det T^{(n)}.$$

But with the notation of Section 2.2,

$$b_k = z_{2k-2} + z_{2k-1} \quad \text{and} \quad a_k = \sqrt{z_{2k-1}z_{2k}}$$

so that $\mathrm{tr}\ T^{(n)} = \sum_{k=1}^{2n-1} z_k$ and $T^{(n)} = B^{(n)}(B^{(n)})^*$, with

$$B^{(n)}_{k,k} = \sqrt{z_{2k-1}}, \quad B^{(n)}_{k+1,k} = \sqrt{z_{2k}}$$

and $B^{(n)}_{i,j} = 0$ for other entries. Then we have $\det T^{(n)} = (\det B^{(n)})^2 = \prod_{k=1}^n z_{2k-1}$ and

$$\begin{aligned}\mathcal{H}(T^{(n)}) &= \tau^{-1}\sum_{k=1}^{2n-2} z_k - (\tau^{-1} - 1)\sum_{k=1}^{n}\log z_{2k-1} - \sum_{k=1}^{2n-1}\log z_k \\ &= \tau^{-1}\sum_{k=1}^{n}(G(z_{2k-1}) + 1) + \sum_{k=1}^{n-1}(G(\tau^{-1}z_{2k}) - \log(\tau) + 1) + \log z_{2n-1}.\end{aligned}$$

Of course, the infimum is achieved for $z_{2k-1} \equiv 1, z_{2k} \equiv \tau$, which gives exactly the expression of (2.6).

Finally, let us look at the Jacobi case, transformed onto the interval $[-2, 2]$, with

$$V(x) = -\kappa_1\log(2 - x) - \kappa_2\log(2 + x)$$

and

$$\mathrm{tr}\ V(T^{(n)}) = -\kappa_1\log\det(2I^{(n)} - T^{(n)}) - \kappa_2\log\det(2I^{(n)} + T^{(n)}).$$

In formula (5.2) of Killip and Nenciu (2004) we see

$$\Phi_{2n}(1) = \prod_{k=0}^{2n-1}(1 - \alpha_k) = \prod_{k=1}^{n}(2 - \lambda_j) = \det(2I^{(n)} - T^{(n)})$$

Similarly, from formula (5.3) of the same paper we get

$$\Phi_{2n}(-1) = \prod_{k=0}^{2n-1}(1 + (-1)^k\alpha_k) = \prod_{k=1}^{n}(2 + \lambda_j) = \det(2I^{(n)} + T^{(n)})$$

such that we obtain

$$\begin{aligned}\mathrm{tr}\ V(T^{(n)}) &= -\kappa_1\log\prod_{k=0}^{n-1}(1 - \alpha_{2k}) - \kappa_2\log\prod_{k=0}^{n-1}(1 + \alpha_{2k}) \\ &\quad - \kappa_1\log\prod_{k=0}^{n-1}(1 - \alpha_{2k+1}) - \kappa_2\log\prod_{k=0}^{n-1}(1 - \alpha_{2k+1})\end{aligned}$$

Recall that $a_{k+1}^2 = (1-\alpha_{2k-1})(1-(\alpha_{2k})^2)(1+\alpha_{2k+1})$, and then

$$\begin{aligned}\mathcal{H}(T^{(n)}) &= \operatorname{tr} V(T^{(n)}) - \sum_{k=0}^{n-1} \log(1-\alpha_k^2) + A_n \\ &= -\sum_{k=0}^{n-1}(1+\kappa_1)\log(1-\alpha_{2k}) + (1+\kappa_2)\log(1+\alpha_{2k}) \\ &\quad -\sum_{k=0}^{n-1}(1+\kappa_1+\kappa_2)\log(1-\alpha_{2k+1}) + \log(1+\alpha_{2k+1}) \quad + A_n\end{aligned}$$

with $A_n = \log(1-\alpha_{2n-3}) + \log(1-\alpha_{2n-2}^2) + \log(1-\alpha_{2n-1}^2)$. Again, in the limit $n \to \infty$ this leads exactly to the rate function (2.9).

3.3.2 Mathematical analysis point of view

In Nazarov et al. (2005), a sum rule is given when the reference measure may be written as

$$\sigma(dx) = A(x)\,\mathrm{SC}(dx) \tag{3.6}$$

with A a nonnegative polynomial (Theorem 1.5 therein). Under appropriate conditions, the sum rule is

$$\mathcal{K}(\sigma \,|\, \mu) + \sum_{k=1}^{N^+} \mathcal{F}(\lambda_k^+) + \sum_{k=1}^{N^-} \mathcal{F}(\lambda_k^-) = \lim_{n\to\infty} \left(-2\sum_{1}^{n-1} \log a_k + \operatorname{tr}\left(\Phi(T^{(n)}) - \Phi(T_0^{(n)}) \right) \right) \tag{3.7}$$

where

$$\mathcal{F}(x) = \begin{cases} \displaystyle\int_2^x A(t)\sqrt{t^2-4}\,dt & \text{if } x \geq 2\,, \\ \displaystyle\int_x^{-2} A(t)\sqrt{t^2-4}\,dt & \text{if } x \leq -2\,. \end{cases}$$

Actually, these conditions warrant the existence of this limit. Set

$$\Phi'(z) = zA(z) - \frac{1}{\pi}\int \frac{A(x)-A(z)}{x-z}\,\mathrm{SC}(dx) \tag{3.8}$$

which leads for $z \notin [-2,2]$ to

$$\Phi'(z) = \mathcal{F}'(z) - \int \frac{\sigma(dx)}{x-z}\,.$$

The function Φ was defined in Nazarov et al. (2005) as an auxiliary function. In view of our Proposition 3.2 (up to an affine change of variables), it appears then that the triple $(\sigma, \mathcal{F}, \Phi)$ is actually identical to the triple $(\mu_V, \mathcal{F}_V, V)$.

Nazarov et al. claimed that the scope of generality was not clear to them, so that they use the polynomial nature of A. Actually, our classical ensembles are, again up to an affine change of variables, of the form (3.6) with A analytic in a neighboorhood of $[-2,2]$, namely with $1/A$ polynomial of degree at most 2.

To end this section let us give some concluding remarks.

1. The sum rule given by Nazarov et al. should be true beyond the polynomial case for more general functions A.

2. The function Φ is nothing else than a potential and $\mathcal{F}$ the corresponding effective potential.

3. When A is a polynomial the underlying potential is also a polynomial.

4. In this latter case, Section 3.3.1 gives a probabilistic interpretation of (3.7) and a draft for a probabilistic proof.

We also refer to Kupin (2005) for other extensions.

4 Proofs

This section is devoted to the proof of Theorem 3.1. The main idea is to apply the projective limit method to reduce the spectral measure to a measure with only a fixed number of eigenvalues outside the limit support $[\alpha^-, \alpha^+]$. For this we need to consider a topology on $\mathcal{S}$ different from the weak topology. Recall that measures $\mu \in \mathcal{S}$ are written as

$$\mu = \mu_{|I} + \sum_{i=1}^{N^+} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \gamma_i^- \delta_{\lambda_i^-} \tag{4.1}$$

and we associate the measure (4.1) with

$$\big(\mu_{|I}, (\lambda_i^+)_{i\geq 1}, (\lambda_i^-)_{i\geq 1}, (\gamma_i^+)_{i\geq 1}, (\gamma_i^-)_{i\geq 1}\big) \tag{4.2}$$

with $\lambda_i^+ = \alpha^+$ and $\gamma_i^+ = 0$ if $i > N^+$ and $\lambda_i^- = \alpha^-$ and $\gamma_i^- = 0$ if $i > N^-$. The topology on $\mathcal{S}$ is then defined by the vector (4.2): we say that μ_n converges to μ in $\mathcal{S}$ if:

$$\begin{gathered} \mu_{n|I} \xrightarrow[n\to\infty]{} \mu_{|I} \text{ weakly and for every } i \geq 1 \\ \big(\lambda_i^+(\mu_n), \lambda_i^-(\mu_n), \gamma_i^+(\mu_n), \gamma_i^-(\mu_n)\big) \xrightarrow[n\to\infty]{} \big(\lambda_i^+(\mu), \lambda_i^-(\mu), \gamma_i^+(\mu), \gamma_i^-(\mu)\big) \end{gathered} \tag{4.3}$$

We will show in Section 4.4 that on the smaller set $\mathcal{S}_1 = \{\mu \in \mathcal{S} |\, \mu(\mathbb{R}) = 1\}$, this convergence implies weak convergence, but we remark that $\mu_n \to \mu$ weakly does not imply convergence in our topology. For example, the merging of two atoms outside of I is no continuous operation, while it is continuous in the weak topology. The σ-algebra on $\mathcal{S}$ is then the corresponding Borel-algebra. On $\mathcal{S}$ we define a family of projections $(\pi_j)_{j\in\mathbb{N}}$, where for a measure μ as in (4.1),

$$\pi_j(\mu) = \mu_{|I} + \sum_{i=1}^{N^+\wedge j} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-\wedge j} \gamma_i^- \delta_{\lambda_i^-}\ , \tag{4.4}$$

that is, we keep μ on I but delete all but up to j point masses left of α^- and right of α^+. Note that the projections are continuous in our topology, but they are not in the weak topology.

4.1 LDP for a finite collection of extreme eigenvalues

The study of LDP for (one) extreme eigenvalue of random matrices began in Ben Arous et al. (2001) and in Albeverio et al. (2001). For detailed comments on the assumptions see Section 5.1.

4.1.1 Notation

Under the probability measures considered, there are almost surely no ties among eigenvalues, so that we may reorder $\lambda = (\lambda_1, \dots, \lambda_n)$ as $\hat{\lambda} = (\lambda_1^n, \dots, \lambda_n^n)$ such that $\lambda_1^n > \lambda_2^n > \cdots > \lambda_n^n$. Let $\lambda_i^+ = \lambda_i^n$ and $\lambda_i^- = \lambda_{n-i+1}^n$ and for j a fixed integer and $n > 2j$

$$\lambda^+(j) = (\lambda_1^+, \dots, \lambda_j^+)\ ,\ \lambda^-(j) = (\lambda_1^-, \dots, \lambda_j^-)\,.$$

For the sake of simplicity, we denote by $\mathbb{R}^{\uparrow j}$ (resp. $\mathbb{R}^{\downarrow j}$) the subset of $\mathbb{R}^j$ of all vectors with non decreasing (resp. non increasing) components.

4.1.2 Main result

Theorem 4.1 *Let j and ℓ be fixed integers. Assume that V is continuous and satisfies (A1), (A2) and the control condition (A3).*

1. *If $b^- < \alpha^-$ and $\alpha^+ < b^+$, then the law of $(\lambda^+(j), \lambda^-(\ell))$ under $\mathbb{P}_V^n$ satisfies a LDP in $\mathbb{R}^{j+\ell}$ with speed $\beta' n$ and rate function*

$$\mathcal{I}_{\lambda^\pm}(x^+, x^-) := \begin{cases} \sum_{k=1}^j \mathcal{F}_V^+(x_k^+) + \sum_{k=1}^\ell \mathcal{F}_V^-(x_k^-) & \text{if } (x_1^+, \dots, x_j^+) \in \mathbb{R}^{\downarrow j} \text{ and } (x_1^-, \dots, x_\ell^-) \in \mathbb{R}^{\uparrow \ell} \\ \infty & \text{otherwise.} \end{cases}$$

2. *If $b^- = \alpha^-$, but $\alpha^+ < b^+$, the law of $\lambda^+(j)$ satisfies the LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}_{\lambda^+}(x^+) = \mathcal{I}_{\lambda^\pm}(x^+, \alpha^-) = \begin{cases} \sum_{k=1}^j \mathcal{F}_V^+(x_k^+) & \text{if } (x_1^+, \dots, x_j^+) \in \mathbb{R}^{\downarrow j} \\ \infty & \text{otherwise.} \end{cases}$$

3. *If $b^- < \alpha^-$, but $\alpha^+ = b^+$, the law of $\lambda^-(\ell)$ satisfies the LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}_{\lambda^-}(x^-) = \mathcal{I}_{\lambda^\pm}(\alpha^+, x^-) = \begin{cases} \sum_{k=1}^\ell \mathcal{F}_V^-(x_k^-) & \text{if } (x_1^-, \dots, x_\ell^-) \in \mathbb{R}^{\uparrow \ell} \\ \infty & \text{otherwise.} \end{cases}$$

The same statement is Theorem 2.10 in Benaych-Georges et al. (2012), but with an extra technical assumption that is not easy to check. Besides, after the above mentioned publications, recently Borot and Guionnet (2013b) and Borot and Guionnet (2013a) provided other sketch of proofs for the case where $\ell = 0, j = 1$ and without this assumption. For the sake of completeness, we give a proof for the general case. As this proof is technical, we postpone it to the Appendix (Section 5).

4.2 Joint LDP for the restricted measure and a finite collection of extreme eigenvalues

Recall that $((\lambda_1, \dots, \lambda_n), (\mathtt{w}_1, \dots, \mathtt{w}_n))$ is distributed according to $\mathbb{Q}_n^V = \mathbb{P}_V^n \otimes \mathrm{Dir}_n(\beta')$. The two sources of randomness are not at the same scale. On the one hand, the eigenvalues are ruled by a LDP at speed n^2. On the other hand, the weights are ruled by a factor n. Hence, it is natural to consider the eigenvalues as quasi-deterministic, and to begin by a conditioning upon these variables. As in a previous paper (Gamboa and Rouault (2010)), it is then convenient to decouple the weights by introducing independent random variables. We know that

$$(\mathtt{w}_1, \dots, \mathtt{w}_n) \stackrel{(d)}{=} \left(\frac{\gamma_1}{\gamma_1 + \dots + \gamma_n}, \dots, \frac{\gamma_n}{\gamma_1 + \dots + \gamma_n}\right) \tag{4.5}$$

where $\stackrel{(d)}{=}$ means equality in distribution, and $\gamma_1, \dots, \gamma_n$ are independent variables with distribution $\mathrm{Gamma}(\beta', (\beta' n)^{-1})$ and mean n^{-1}. We enlarge the probability space to define such variables γ_i's and denote by $\widetilde{\mathbb{Q}_n^V}$ the corresponding probability measure. With this notation, we can rewrite the spectral measure $\mu^{(n)}$ as

$$\mu^{(n)} = \frac{\widetilde{\mu}^{(n)}}{\widetilde{\mu}^{(n)}(\mathbb{R})} \tag{4.6}$$

where

$$\widetilde{\mu}^{(n)} := \sum_{k=1}^{n} \gamma_k \delta_{\lambda_k} \tag{4.7}$$

is a random measure with independent masses $\gamma_1, \dots, \gamma_n$ and $\widetilde{\mu}^{(n)}(\mathbb{R})$ is its total mass $\sum_{k=1}^n \gamma_k$. We denote by $\mu_I^{(n)}$ the restriction of $\mu^{(n)}$ to the interval I. Similarly, for $I(j) = I \setminus \{\lambda_1^+, \lambda_1^-, \dots, \lambda_j^+, \lambda_j^-\}$, $\mu_{I(j)}^{(n)}$ is the restriction of $\mu^{(n)}$ to $I(j)$ and we use the analogous notation for the restrictions of the empirical measure $\mu_{\mathtt{u}}^{(n)}$. Notice that we choose $j = \ell$ for the sake of simplicity. The aim of this subsection is to prove the following joint LDP for the restricted spectral measure and a collection of largest and/or smallest eigenvalues.

Theorem 4.2

1. *If $b^- < \alpha^- < \alpha^+ < b^+$, then for any fixed $j \in \mathbb{N}$ and under $\widetilde{\mathbb{Q}_n^V}$, the sequence of random objects $\big(\tilde{\mu}_{I(j)}^{(n)}, \lambda^+(j), \lambda^-(j)\big)$ satisfies the joint LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}(\mu, x^+, x^-) = \mathcal{K}(\mu_V \mid \mu) + \mu(I) - 1 + \mathcal{I}_{\lambda^\pm}(x^+, x^-)$$

2. *If $b^- = \alpha^-$, but $\alpha^+ < b^+$ (or $b^+ = \alpha^+$, but $\alpha^- > b^-$), then, with the same notation as in the previous section, $\big(\mu_{I(j)}^{(n)}, \lambda^+(j)\big)$ (or $\big(\mu_{I(j)}^{(n)}, \lambda^-(j)\big)$ respectively,) satisfies the LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}^+(\mu, x^+) = \mathcal{I}(\mu, x^+, \alpha^-) \quad (\text{or } \ \mathcal{I}^-(\mu, x^-) = \mathcal{I}(\mu, \alpha^+, x^-) \ \text{respectively}) .$$

Proof: We only prove here the first point of the theorem. The second claim can be shown in the same way. We first show a joint LDP when the eigenvalues are truncated. For $M > \max\{|\alpha^+|, |\alpha^-|\}$, let $\lambda_M^+(j)$ (resp.$\lambda_M^-(j)$) be the collection of truncated eigenvalues

$$\lambda_{M,i}^+ = \min\{\lambda_i^+, M\} \quad (\text{resp. } \lambda_{M,i}^- = \max\{\lambda_i^-, -M\}) \, ,$$

for $i = 1, \dots, j$. To further simplify notation, let $\lambda_M^\pm(j) = (\lambda_{M,1}^+, \dots, \lambda_{M,j}^+, \lambda_{M,1}^-, \dots, \lambda_{M,j}^+)$.

Exponential tightness:
In a first step, we will obtain the joint LDP for $(\tilde{\mu}_{I(j)}^{(n)}, \lambda_M^\pm(j))$ by applying Theorem 1.1 of Baldi (1988). For this, we need to check that this sequence is exponentially tight. For M as above, define the set

$$K_M = \left\{(\mu, \lambda) \in \mathcal{S} \times \mathbb{R}^{2j} | \, \mu(I) \le M, \mu(I^c) = 0, \lambda \in [-M, M]^{2j}\right\} .$$

Indeed, K_M is a compact set in the topology (4.3) and

$$\mathbb{P}((\tilde{\mu}_{I(j)}^{(n)}, \lambda_M^\pm(j)) \notin K_M) = \mathbb{P}(\tilde{\mu}_{I(j)}^{(n)}(I) > M) \le \mathbb{P}\left(\sum_{k=1}^n \gamma_k > M\right) .$$

The sum in the last probability is Gamma$(\beta' n, (\beta' n)^{-1})$ distributed. By the LDP for the Gamma-distribution with rate G,

$$\mathbb{P}\left(\sum_{k=1}^n \gamma_k > M\right) \le e^{-\beta' n G(M)}$$

Therefore,

$$\limsup_{n\to\infty} \frac{1}{\beta' n} \log \mathbb{P}((\tilde{\mu}_{I(j)}^{(n)}, \lambda_M^\pm(j)) \notin K_M) \le -G(M),$$

which can be chosen to be arbitrarily small, i.e., the sequence $(\tilde{\mu}_{I(j)}^{(n)}, \lambda_M^\pm(j))$ is exponentially tight.

Joint LDP for measure and truncated eigenvalues:
Let f be a continuous function from $\mathbb{R}$ to $\mathbb{R}$ such that $\log(1-f)$ is bounded. For $s^\pm \in \mathbb{R}^{2j}$, we calculate the joint moment generating function

$$\mathcal{G}_n(f, s^\pm) = \mathbb{E}\left[\exp\left\{n\beta' \left(\tilde{\mu}_{I(j)}^{(n)}(f) + \langle s^\pm, \lambda_M^\pm(j)\rangle\right)\right\}\right]$$

under $\widetilde{\mathbb{Q}_n^V}$. First recall that γ_i is Gamma$(\beta', (\beta' n)^{-1})$ distributed, so that

$$\frac{1}{\beta'} \log \mathbb{E} e^{\beta' n \gamma_i t} = L(t) \tag{4.8}$$

(see (1.16)) and then, integrating with respect to the γ_i's we get

$$\begin{aligned}\mathcal{G}_n(f, s^\pm) &= \mathbb{E}\left[\exp\left(n\beta'\langle s^\pm, \lambda^\pm_M(j)\rangle\right) \prod_{i\in I(j)} \mathbb{E}\left[e^{n\beta'\gamma_1 f(\lambda_i)}|\lambda_1, \dots, \lambda_n\right]\right] \\ &= \mathbb{E}\left[\exp\left\{n\beta'\left(\mu^{(n)}_{\mathtt{u},I(j)}(L\circ f) + \langle s^\pm, \lambda^\pm_M(j)\rangle\right)\right\}\right],\end{aligned}$$

This expectation only involves $\mathbb{P}^n_V$. Set

$$D_n(s^\pm) := \mathbb{E}\left[\exp\left\{n\beta'\langle s^\pm, \lambda^\pm_M(j)\rangle\right\}\right].$$

By Theorem 4.1 we have a LDP for the extremal eigenvalues $\lambda^\pm(j)$ of the spectral measure with rate function $\mathcal{I}_{\lambda^\pm}$. By the contraction principle (see Dembo and Zeitouni (1998) p.126), the truncated eigenvalues satisfy the LDP with rate function

$$\mathcal{I}_{M,\lambda^\pm}(x^\pm) = \begin{cases}\mathcal{I}_{\lambda^\pm}(x^+, x^-) & \text{if } x^\pm = (x^+, x^-) \in [-M, M]^{2j}, \\ \infty & \text{otherwise.}\end{cases}$$

Since the truncated eigenvalues are bounded, Varadhan's Integral Lemma (Dembo and Zeitouni (1998) p. 137) implies

$$\lim_{n\to\infty} \frac{1}{\beta' n} \log D_n(s^\pm) = \mathcal{I}^*_{M,\lambda^\pm}(s^\pm), \tag{4.9}$$

where

$$\mathcal{I}^*_{M,\lambda^\pm}(s^\pm) = \sup_{x^\pm\in\mathbb{R}^{2j}} \left\{\langle s^\pm, x^\pm\rangle - \mathcal{I}_{M,\lambda^\pm}(x^\pm)\right\}$$

is the convex dual of $\mathcal{I}_{M,\lambda^\pm}$. To control $n\mu^{(n)}_{\mathtt{u},I(j)}(L\circ f)$, let for $\eta > 0$

$$A(\eta) = \left\{d(\mu^{(n)}_{\mathtt{u},I(j)}, \mu_V) < \eta\right\},$$

with a metric d inducing weak convergence. Since $\mu^{(n)}_{\mathtt{u},I(j)}$ and $\mu^{(n)}_{\mathtt{u},I}$ differ only by at most $2j$ support points, their total variation distance is bounded by $2j/n$. For n large enough this implies

$$\left\{d(\mu^{(n)}_{\mathtt{u},I}, \mu_V) < \eta/2\right\} \subset A(\eta).$$

Now,

$$\mathbb{P}(A(\eta)^c) \le \mathbb{P}(d(\mu^{(n)}_{\mathtt{u},I}, \mu_V) \ge \eta/2) \le \mathbb{P}(d(\mu^{(n)}_{\mathtt{u}}, \mu_V) \ge \eta/2)$$

and then, since $\mu^{(n)}_{\mathtt{u}}$ satisfies an LDP with speed n^2 and a rate which is good with unique minimizer μ_V (Theorem 1.2) we have for n large enough

$$\mathbb{P}(A(\eta)^c) \le e^{-n^2\delta} \tag{4.10}$$

with a $\delta = \delta(\eta) > 0$. Writing $\mathcal{G}_n(f, s^\pm) = \mathcal{G}_{n,A}(f, s^\pm) + \mathcal{G}_{n,A^c}(f, s^\pm)$ with

$$\mathcal{G}_{n,A}(f, s^\pm) = E\left[\exp\left\{n\beta'\left(\mu^{(n)}_{\mathbf{u},I(j)}(L \circ f) + \langle s^\pm, \lambda^\pm_M(j)\rangle\right)\right\} \mathbb{1}_{A(\eta)}\right]$$

we can bound

$$C_n(s^\pm) \exp\left\{n\beta'\left(\mu_V(L \circ f) - \eta\right)\right\} \leq \mathcal{G}_{n,A}(f, s^\pm) \leq C_n(s^\pm) \exp\left\{n\beta'\left(\mu_V(L \circ f) + \eta\right)\right\} \tag{4.11}$$

where

$$C_n(s^\pm) := \mathbb{E}\left[\exp\left\{n\beta'\left(\langle s^\pm, \lambda^\pm_M(j)\rangle\right)\right\} \mathbb{1}_{A(\eta)}\right] \leq D_n(s^\pm)\,,$$

and then from (4.9)

$$\limsup_{n\to\infty} \frac{1}{\beta' n} \log \mathcal{G}_{n,A}(f, s^\pm) \leq \mu_V(L \circ f) + \eta + \mathcal{I}^*_{M,\lambda^\pm}(s^\pm)\,. \tag{4.12}$$

For the complimentary event, we have the upper bound

$$\mathcal{G}_{n,A^c}(f, s^\pm) \leq \left(D_n(s^\pm) - C_n(s^\pm)\right) \exp\left\{n\beta' ||L \circ f||_\infty\right\}\,.$$

By the Cauchy-Schwarz inequality and (4.10) we get

$$\left(D_n(s^\pm) - C_n(s^\pm)\right) \leq D_n(2s^\pm) e^{-n^2\delta}\,,$$

and then, using (4.9) we get

$$\limsup_{n\to\infty} \frac{1}{n\beta'} \log(D_n(s^\pm) - C_n(s^\pm)) = -\infty \tag{4.13}$$

which eventually leads to

$$\limsup_{n\to\infty} \frac{1}{n\beta'} \log \mathcal{G}_{n,A^c}(f, s^\pm) = -\infty. \tag{4.14}$$

Combining (4.12) and (4.14), we get

$$\limsup_{n\to\infty} \frac{1}{n\beta'} \log \mathcal{G}_n(f, s^\pm) \leq \mu_V(L \circ f) + \eta + \mathcal{I}^*_{M,\lambda^\pm}(s^\pm). \tag{4.15}$$

For the lower bound, we have

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n\beta'} \log \mathcal{G}_n(f, s^\pm) &\geq \liminf_{n\to\infty} \frac{1}{n\beta'} \log \mathcal{G}_{n,A}(f, s^\pm) \\ &\geq \mu_V(L \circ f) - \eta + \liminf_{n\to\infty} \frac{1}{n\beta'} \log C_n(s^\pm)\,,\end{aligned}$$

and from (4.9) and (4.13)

$$\lim_{n\to\infty} \frac{1}{n\beta'} \log C_n(s^\pm) = \mathcal{I}^*_{M,\lambda^\pm}(s^\pm)\,,$$

so that

$$\liminf_{n\to\infty} \frac{1}{n\beta'} \log \mathcal{G}_n(f, s^{\pm}) \geq \mu_V(L \circ f) - \eta + \mathcal{I}^*_{M,\lambda^{\pm}}(s^{\pm})\,. \tag{4.16}$$

From (4.15) and (4.16) and since $\eta > 0$ was arbitrary, we can conclude:

$$\lim_{n\to\infty} \frac{1}{\beta' n} \log \mathcal{G}_n(f, s^{\pm}) = \mu_V(L \circ f) + \mathcal{I}^*_{M,\lambda^{\pm}}(s^{\pm}) =: \mathcal{G}(f, s^{\pm}).$$

The convex dual of $\mathcal{G}$ is

$$\begin{aligned}
\mathcal{G}^*(\mu, \lambda^{\pm}) &= \sup_{f\in C_b(I), s^{\pm}\in\mathbb{R}^{2j}} \left(\int f d\mu + \langle \lambda^{\pm}, s^{\pm}\rangle - \mathcal{G}(f, s^{\pm}) \right) \\
&= \sup_{f\in C_b(I)} \left(\int f d\mu - \mu_V(L \circ f) \right) + \sup_{s^{\pm}\in\mathbb{R}^{2j}} \left(\langle \lambda^{\pm}, s^{\pm}\rangle - \mathcal{I}^*_{M,\lambda^{\pm}}(s^{\pm}) \right) \\
&= \Lambda^*(\mu) + \mathcal{I}_{M,\lambda^{\pm}}(\lambda^{\pm}),
\end{aligned}$$

where Λ^* is the convex dual of $\mu_V(L \circ \cdot)$. The LDP follows now from Theorem 1.1 of Baldi (1988) with rate function given by $\mathcal{G}^*$, provided that $\mathcal{G}^*$ is strictly convex on a set of points that is dense in the set of all points where $\mathcal{G}^*$ is finite.

Identification of Λ^*:
By Theorem 5 of Rockafellar (1971), we can write Λ^* as

$$\Lambda^*(\mu) = \mu_V(L^* \circ h_\mu) + r(1)\mu_s(\mathbb{R}),$$

where L^* is the convex dual of L and r its recession function and $d\mu = h_\mu \cdot d\mu_V + d\mu_s$ is the Lebesgue-decomposition of μ with respect to μ_V. The expression of L^* is given in (1.16). The recession function is

$$r(x) = \sup\{xy \mid L(y) < \infty\} = x.$$

for nonnegative x. We obtain

$$\begin{aligned}
\Lambda^*(\mu) &= -\int \log(h_\mu)\, d\mu_V - 1 + \int h_\mu\, d\mu_V + \mu_s(\mathbb{R}) \\
&= -\int \log(h_\mu)\, d\mu_V - 1 + \mu(I) \\
&= \mathcal{K}(\mu_V|\mu) - 1 + \mu(I)\,.
\end{aligned}$$

Now Λ^* is strictly convex at μ if there exists a $f \in C_b(I)$, called an exposing hyperplane, such that

$$\Lambda^*(\mu) - \int f d\mu < \Lambda^*(\nu) - \int f d\nu \tag{4.17}$$

for any $\nu \neq \mu$. Suppose $d\mu = h_\mu \cdot d\mu_V$ is absolutely continuous with respect to μ_V with a density h_μ positive on I and choose $f = 1 - h_\mu^{-1}$. Then (4.17) is equivalent to

$$\int \log(h_\mu/h_\nu) d\mu_V > \int d\mu_V - \int h_\mu^{-1} d\nu,$$

The last inequality follows from $\log(x) > 1 - x^{-1}$, $(x > 0, x \neq 1)$. Indeed,

$$\int \log(h_\mu/h_\nu) d\mu_V \geq \int (1 - h_\mu^{-1} \cdot h_\nu) d\mu_V \geq \int d\mu_V - \int h_\mu^{-1} d\nu,$$

where the first inequality is strict unless μ_V-almost everywhere $h = h_\nu$ and the second one is strict unless $\nu_s = 0$. So if $\nu \neq \mu$ at least one inequality is strict, i.e. Λ^* is strictly convex at all points $d\mu = h \cdot d\mu_V$, which are dense in the set of nonnegative measures on I. Consequently, $(\tilde{\mu}^{(n)}_{I(j)}, \lambda^\pm_M(j))$ satisfies an LDP with speed $\beta' n$ and rate function

$$\mathcal{I}(\mu, x^\pm) = \mathcal{K}(\mu_V \,|\, \mu) + \mu(I) - 1 + \mathcal{I}_{M,\lambda^\pm}(x^\pm).$$

Extending the LDP to untruncated eigenvalues:
From the LDP for $(\lambda^+(j), \lambda^-(j))$ the exponential tightness of the (unrestricted) extremal eigenvalues holds (see Section 5.2.2). This implies

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(\lambda^\pm_M(j) \neq \lambda^\pm(j)\big) = -\infty,$$

so that as $M \to \infty$, the truncated eigenvalues are exponentially good approximation of the unrestricted ones. In fact, $(\tilde{\mu}^{(n)}_{I(j)}, \lambda^\pm_M(j))$ are exponentially good approximations of $(\tilde{\mu}^{(n)}_{I(j)}, \lambda^\pm(j))$. Since the rate function of the untruncated eigenvalues can be recovered as the pointwise limit

$$\mathcal{I}_{\lambda^\pm}(\lambda^\pm) = \lim_{M\to\infty} \mathcal{I}_{M,\lambda^\pm}(\lambda^\pm),$$

we get from Theorem 4.2.16 in Dembo and Zeitouni (1998) that $(\tilde{\mu}^{(n)}_{I(j)}, \lambda^\pm(j))$ satisfies the LDP with speed $\beta' n$ and rate function

$$\mathcal{I}(\mu, x^\pm) = \mathcal{K}(\mu_V \,|\, \mu) + \mu(I) - 1 + \mathcal{I}_{\lambda^\pm}(x^\pm) = \mathcal{K}(\mu_V \,|\, \mu) + \mu(I) - 1 + \sum_{i=1}^{j} \mathcal{F}^+(x_i^+) + \mathcal{F}^-(x_i^-)\,,$$

which ends the proof of Theorem 4.2. ■

4.3 LDP for the projected measure

Recall the definition of the projections π_j in (4.4).

Theorem 4.3 *For any fixed j, the sequence of projected spectral measures $\pi_j(\tilde{\mu}^{(n)})$ as elements of $\mathcal{S}$ with topology* (4.3) *satisfies under $\widetilde{\mathbb{Q}_n^V}$ the LDP with speed $\beta' n$ and rate function*

$$\tilde{\mathcal{I}}_j(\tilde{\mu}) = \mathcal{K}(\mu_V \mid \tilde{\mu}) + \tilde{\mu}(I) - 1 + \sum_{i=1}^{N^+\wedge j} \left(\mathcal{F}_V^+(\lambda_i^+) + \gamma_i^+\right) + \sum_{i=1}^{N^-\wedge j} \left(\mathcal{F}_V^-(\lambda_i^-) + \gamma_i^-\right).$$

Proof: This result is a direct consequence of Theorem 4.2 and the contraction principle. Again, suppose that the first case of Theorem 4.2 holds, otherwise N^+ or N^- is 0 and we may omit the largest and/or smallest eigenvalues and weights. By the independence of the weights γ_k and independence of weights and eigenvalues, the collection of weights $(\gamma^+(j), \gamma^-(j))$ is independent of $(\tilde{\mu}^{(n)}_{I(j)}, \lambda^+(j), \lambda^-(j))$. Recall that $(\gamma^+(j), \gamma^-(j))$ contains a collection of independent Gamma$(\beta', (\beta' n)^{-1})$ distributed random variables and then satisfies an LDP with speed $\beta' n$ and rate function

$$\mathcal{I}_\gamma(y^+, y^-) = \sum_{i=1}^{j} (y_i^+ + y_i^-)$$

for $y_i^+, y_i^- \geq 0$ and $\mathcal{I}_\gamma(y^+, y^-) = \infty$ otherwise (see (1.17)). Thus,

$$\big(\tilde{\mu}^{(n)}_{I(j)}, \lambda^+(j), \lambda^-(j), \gamma^+(j), \gamma^-(j)\big) \tag{4.18}$$

satisfies the LDP with speed $\beta' n$ and rate function

$$\mathcal{I}(\mu, x^+, x^-, y^+, y^-) = \mathcal{K}(\mu_V \,|\, \mu) + \mu(I) - 1 + \sum_{i=1}^{j} \mathcal{F}(x_i^+) + \mathcal{F}(x_i^-) + y_i^+ + y_i^- \,.$$

By definition of the projections π_j, we can write

$$\pi_j(\tilde{\mu}^{(n)}) = \mathcal{C}(\tilde{\mu}^{(n)}_{I(j)}, \lambda^+(j), \lambda^-(j), \gamma^+(j), \gamma^-(j)) = \tilde{\mu}^{(n)}_{|I} + \sum_{i=1}^{N^+\wedge j} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-\wedge j} \gamma_i^- \delta_{\lambda^-{}_i}$$

with a continuous $\mathcal{C}$, defined by

$$\mathcal{C}(\mu, x^+, x^-, y^+, y^-) = \mu + \sum_{i=1}^{j} y_i^+ \delta_{x_i^+} + \sum_{i=1}^{j} y_i^- \delta_{x_i^-}$$

Note that $\mathcal{C}$ is not a bijection: point masses in I may come from μ or from the points x^+, x^-. However, for a given $\tilde{\mu} \in \mathcal{S}$, we may still easily calculate

$$\tilde{\mathcal{I}}_j(\tilde{\mu}) = \inf \big\{ \mathcal{I}(\mu, x^+, x^-, y^+, y^-) \,|\, \mathcal{C}(\mu, x^+, x^-, y^+, y^-) = \tilde{\mu} \big\}$$

as the minimum is attained by choosing $\mu = \tilde{\mu}_{|I(j)}$ and $\lambda_i^+ = \alpha^+$ and $\gamma_i^+ = 0$ if $i > N^+$ and $\lambda_i^- = \alpha^-$ and $\gamma_i^- = 0$ if $i > N^-$. Therefore

$$\tilde{\mathcal{I}}_j(\tilde{\mu}) = \mathcal{K}(\mu_V \,|\, \tilde{\mu}) + \tilde{\mu}(I) - 1 + \sum_{i=1}^{N^+\wedge j} \big(\mathcal{F}(\lambda_i^+) + \gamma_i^+\big) + \sum_{i=1}^{N^-\wedge j} \big(\mathcal{F}(\lambda_i^-) + \gamma_i^-\big)$$

which, by the contraction principle, is the rate function of $\pi_j(\tilde{\mu}^{(n)})$. ∎

4.4 Projective limit and normalization

From Theorem 4.3, the projective method of the Dawson-Gärtner theorem, p. 162 in the book of Dembo and Zeitouni (1998), yields the LDP for $\tilde{\mu}^{(n)}$ under $\widetilde{\mathbb{Q}_n^V}$ with speed $\beta' n$ and rate function

$$\tilde{\mathcal{I}}(\tilde{\mu}) = \sup_j \tilde{\mathcal{I}}_j(\tilde{\mu}) = \mathcal{K}(\mu_V \mid \tilde{\mu}) + \tilde{\mu}(\mathbb{R}) - 1 + \sum_{i=1}^{N^+} \mathcal{F}(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}(\lambda_i^-), \tag{4.19}$$

defined for $\tilde{\mu} \in \mathcal{S}$. Recalling (4.6), we want to come back to a normalized measure $\mu \in \mathcal{S}_1$. It would be natural to apply the mapping $\tilde{\mu} \mapsto \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{R})}$ but unfortunately, this mapping is not continuous in our topology induced by (4.3). As a workaround, note that from the LDP for $\pi_j(\tilde{\mu}^{(n)})$, we also get the joint LDP of

$$\left(\pi_j(\tilde{\mu}^{(n)}), \pi_j(\tilde{\mu}^{(n)})(\mathbb{R})\right)$$

as the mapping

$$\pi_j(\tilde{\mu}) \longmapsto \pi_j(\tilde{\mu}^{(n)})(\mathbb{R}) = \tilde{\mu}^{(n)}(I) + \sum_{i=1}^{N^+\wedge j} \gamma_i^+ + \sum_{i=1}^{N^-\wedge j} \gamma_i^-$$

is continuous in our topology for any j. Thus, applying the projective method to $(\pi_j(\tilde{\mu}^{(n)}), \pi_j(\tilde{\mu}^{(n)})(\mathbb{R}))$, we get the LDP for the pair $(\tilde{\mu}^{(n)}, \tilde{\mu}^{(n)}(\mathbb{R}))$ with rate function

$$\overline{\mathcal{I}}(\tilde{\mu}, \kappa) = \begin{cases} \tilde{\mathcal{I}}(\tilde{\mu}) & \text{if } \tilde{\mu}(\mathbb{R}) = \kappa \\ \infty & \text{otherwise.} \end{cases}$$

Now we are able to recover the original spectral probability measures $\mu^{(n)}$ from the unnormalized measures $\tilde{\mu}^{(n)}$, by applying to the pair $(\tilde{\mu}^{(n)}, \tilde{\mu}^{(n)}(\mathbb{R}))$ the (continuous) mapping $(\tilde{\mu}, \kappa) \mapsto \kappa^{-1}\tilde{\mu}$. The contraction principle yields then the LDP for $\mu^{(n)}$ under $\widetilde{\mathbb{Q}_n^V}$ (hence under $\mathbb{Q}_n^V$) with rate function

$$\mathcal{I}(\mu) = \inf_{\nu = \kappa\cdot\mu,\, \kappa>0} \tilde{\mathcal{I}}(\nu) = \inf_{\kappa>0} \tilde{\mathcal{I}}(\kappa \cdot \mu)$$

By (4.19), we need to minimize over κ the function

$$-\int \log\left(\kappa \frac{d\mu}{d\mu_V}\right) d\mu_V - 1 + \kappa + \sum_{i=1}^{N^+} \mathcal{F}(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}(\lambda_i^-).$$

The term $\kappa - 1 - \log\kappa = L^*(\kappa)$ attains its minimal value 0 for $\kappa = 1$. We obtain therefore the following LDP.

Theorem 4.4 *The sequence of spectral measures $\mu^{(n)}$ under $\mathbb{Q}_n^V$, as a random element of $\mathcal{S}_1$ equipped with the topology induced by (4.3), satisfies the LDP with speed $\beta' n$ and rate function*

$$\mathcal{I}_V(\mu) = \mathcal{K}(\mu_V \,|\, \mu) + \sum_{i=1}^{N^+} \mathcal{F}(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}(\lambda_i^-).$$

4.5 Topological considerations

It remains to show that the LDP in Theorem 4.4 holds in the weak topology and can be extended from $\mathcal{S}_1$ to $\mathcal{P}_1$. This step is a consequence of the following two lemmas. Their proofs are postponed to the appendix.

Lemma 4.5 *The weak topology on $\mathcal{S}_1$ is coarser than the topology induced by* (4.3)*.*

Lemma 4.6 *The function $\mathcal{I}_V$, extended to $\mathcal{P}_1$ by setting $\mathcal{I}_V(\mu) = \infty$ if $\mu \notin \mathcal{S}_1$, is lower semi-continuous in the weak topology.*

By Lemma 4.5, the LDP of Theorem 4.4 holds with the weak topology on $\mathcal{S}_1$, and by Lemma 4.6, the LDP can by extended to $\mathcal{P}_1$. This completes the proof of Theorem 3.1.

5 Appendix 1: Proof of Theorem 4.1

5.1 Comments on the assumptions

In Benaych-Georges et al. (2012), the result is proved under their assumption 2.9, which consists of three requirements: confinement, technical condition and convergence of extreme eigenvalues. We may also refer to Auffinger et al. (2013) for a general result in the same vein. For the sake of completeness let us shortly discuss the three requirements.

1. Confinement

   (B1) $$\liminf_{|x|\to\infty} \frac{V(x)}{2\log|x|} > 1$$

   which differs from (A1) in the case $\beta < 1$. Actually, the proof of the LDP for the empirical spectral distribution $\mu_{\mathtt{u}}^{(n)}$ of Anderson et al. (2010) does require $\max(\beta^{-1}, 1)$ instead of 1 since it is used to warrant the finiteness of $\int e^{-\beta' V(x)}dx$. Recently, Serfaty (2014) proved the LDP for $\mu_{\mathtt{u}}^{(n)}$ under the assumption

   (S1) $$\lim_{|x|\to\infty} V(x) - 2\log|x| = \infty\,.$$

   The proof is completely different, using the notion of Γ-convergence. Besides, with another method (carrying everything on the unit circle by the Cayley transform), the LDP for $\mu_{\mathtt{u}}^{(n)}$ was proved in Hardy (2012) under the weak confinement assumption

   (H1) $$\liminf_{|x|\to\infty} V(x) - 2\max(1, \beta^{-1})\log|x| > -\infty\,.$$

Of course, when $\beta \geq 1$
$$(A1) = (B1) \Rightarrow (S1) \text{ and } (H1)\,,$$
but when $\beta < 1$ and $\liminf \frac{V(x)}{2\log|x|} \in (1, \beta^{-1})$, (S1) is satisfied but not (H1).

Since we use several times arguments taken from the proof of the LDP for $\mu_{\mathbf{u}}^{(n)}$ in Anderson et al. (2010), we did not weaken our hypothesis (A1) into (B1) to avoid a complete rewriting, though we conjecture that this β^{-1} is an artefact.

2. Technical condition.

(B2) For every $p \geq 1$, the limit $$\lim_{n\to\infty} \frac{1}{n} \log \frac{Z^{n-p}_{\frac{n}{n-p}V}}{Z^n_{nV}} \quad \text{exists}\,.$$

3. Convergence of extreme eigenvalues.

(AGZ) Under $\mathbb{P}^n_V$, the largest (resp. lowest) eigenvalue converges to $\alpha_\pm$ almost surely[2].

In Anderson et al. (2010) the LDP for $\lambda_{\min}$ is proved under (B2) with $p = 1$. Later, in an erratum[3], the authors claimed that the proof of the LDP needs one more assumption: either a slight modification of (AGZ) (replace $\mathbb{P}^n_V$ by $\mathbb{P}^n_{nV/(n-1)}$) or (A3) (control of large deviations). Later again, in Borot and Guionnet (2013b) a proof for $p = 1$ is given under assumption (A3) alone, without (AGZ). It is worthwhile to mention the papers Borot and Guionnet (2013a) and Fan et al. (2014) on connected topics.

To update the proof of Benaych-Georges et al. (2012) with the tools of Borot and Guionnet (2013b) adapted to the case $p > 1$ and for the sake of completeness, we give now the detailed scheme. We will use three lemmas whose proofs are postponed.

The first statement is a fact often mentioned (for instance Anderson et al. (2010) pp. 83-84 or Benaych-Georges et al. (2012) p.744) but (as far as we know) never checked explicitly. We set it as a lemma and for which we give a complete proof in Section 5.3.1 for convenience of the reader.

Lemma 5.1 *Let V be a potential satisfying the confinement condition (A1) and let r be a fixed integer. If $\mathbb{P}^n_{V_n}$ is the probability measure associated to the potential $V_n = \frac{n+r}{n}V$, then the law of $\mu_{\mathbf{u}}^{(n)}$ under $\mathbb{P}^n_{V_n}$ satisfies the LDP with speed $\beta' n^2$ with good rate function*

$$\mu \mapsto \mathcal{E}(\mu) - \inf_\nu \mathcal{E}(\nu) \tag{5.1}$$

where $\mathcal{E}$ is defined in (1.11).

Lemma 5.2 *If the potential V is finite and continuous on a compact set and infinite outside, we have, for every $p \geq 1$*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{Z^n_V}{Z^{n-p}_{\frac{n}{n-p}V}} = -\inf_{x_1,\dots,x_k} \sum_{k=1}^p \mathcal{J}_V(x_k) = -p \inf_x \mathcal{J}_V(x)\,. \tag{5.2}$$

[2] see footnote 1

[3] available online http://www.wisdom.weizmann.ac.il/~zeitouni/cormat.pdf.

Lemma 5.3 *Under Assumption (A1) and (A3), the largest (resp. lowest) eigenvalue converges in probability to α^+ (resp. α^-).*

5.2 Proof

5.2.1 Outline

First notice that $\mathcal{I}_{\lambda^\pm}$ is a good rate function: it is lower semicontinuous as proved in Borot and Guionnet (2013b) A.1.p.478. From the same reference, $\mathcal{F}_V^+$ and $\mathcal{F}_V^-$ have compact level sets, so that $\mathcal{I}_{\lambda^\pm}$ has compact level set by the union bound. The exponential tightness proved below implies that the weak LDP for the extreme eigenvalues will be sufficient for the statement of Theorem 4.1. Throughout this proof, we will assume $\ell = j$, the generalization is straightforward. The weak LDP will follow from the upper bound

$$\limsup_{n\to\infty} (\beta' n)^{-1} \log \mathbb{P}_V^n(\lambda^\pm(j) \in F^+ \times F^-) \leq - \inf_{(x^+,x^-)\in F^+\times F^-} \mathcal{I}_{\lambda^\pm}(x^+, x^-) \tag{5.3}$$

for sets $F^+ = F_1^+ \times \cdots \times F_j^+$ and idem for F^-, which generate the topology on $\mathbb{R}^{2j}$ and from the lower bound

$$\liminf_{n\to\infty} (\beta' n)^{-1} \log \mathbb{P}_V^n(\lambda^\pm(j) \in G) \geq -\mathcal{I}_{\lambda^\pm}(x^+, x^-), \tag{5.4}$$

for open sets G containing (x^+, x^-).

5.2.2 Exponential tightness

We define the compact set

$$H_M = \{x \in [-M, M] : V(x) \leq M\} \tag{5.5}$$

and have to show that for every j

$$\limsup_{M\to\infty} \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_V^n\big((\lambda^-(j), \lambda^+(j)) \notin H_M^{2j}\big) = -\infty. \tag{5.6}$$

But as

$$\mathbb{P}_V^n\big((\lambda^-(j), \lambda^+(j)) \notin H_M^{2j}\big) \leq \mathbb{P}_V^n\big(\lambda_1^+ > M\big) + \mathbb{P}_V^n\big(\lambda_1^- < -M\big),$$

exponential tightness reduces to the case $j = 1$ and further (by symmetry), we only have to prove

$$\limsup_{M\to\infty} \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_V^n\big(\lambda_1^+ > M\big) = -\infty\,. \tag{5.7}$$

By exchangeability, we have,

$$\mathbb{P}_V^n(\lambda_1^+ \geq M) = n\mathbb{P}_V^n(\lambda_1 \geq M \text{ and } \lambda_k \leq \lambda_1 \text{ for } k = 2, \dots, n) \leq n\mathbb{P}_V^n(\lambda_1 \geq M)$$

Now

$$(5.8)\qquad \mathbb{P}_V^n(\lambda_1 \geq M) \leq \frac{Z_V^{n-1}}{Z_V^n}\int_{\{x\geq M\}} e^{-n\beta' V(x)}\int \prod_{k=2}^n \left(|x-\lambda_k|^\beta e^{-\beta' V(\lambda_k)}\right) d\mathbb{P}_V^{n-1}(\lambda_2,\dots,\lambda_n)dx$$

Since $y \mapsto y^\beta$ is convex on $[0,\infty)$ for $\beta > 1$ and concave and subadditive for $\beta < 1$, we have

$$|x-\lambda|^\beta \leq (|x|+|\lambda|)^\beta \leq a(\beta)(|x|^\beta + |\lambda|^\beta)$$

with $a(\beta) = 2^{(\beta-1)_+}$, and then

$$|x-\lambda|^\beta e^{-\beta' V(\lambda)} \leq a(\beta)(|x|^\beta + |\lambda|^\beta) e^{-\beta' V(\lambda)}\,.$$

From the confinement assumption (A1) there exists $\eta > 0$ such that

$$\liminf_{|\lambda|\to\infty} \frac{V(\lambda)}{2\log|\lambda|} = 1 + 2\eta$$

and then there exists M_0 such that for $|\lambda| \geq M_0$ we have

$$(5.9)\qquad \frac{V(\lambda)}{2\log|\lambda|} \geq 1+\eta$$

so that

$$(5.10)\qquad |\lambda|^\beta e^{-\beta' V(\lambda)} \leq |\lambda|^{-\beta\eta}$$

(for $|\lambda| \geq M_0$) and then $|\lambda|^\beta e^{-\beta' V(\lambda)}$ is bounded by a constant M_1 uniformly in λ. Now, V is bounded from below, say by M_2, and then

$$|x-\lambda|^\beta e^{-\beta' V(\lambda)} \leq a(\beta)(e^{-M_2} x^\beta + M_1)$$

and since $1 \leq x^\beta M_0^{-\beta}$ for $x \geq M_0$, we get

$$|x-\lambda|^\beta e^{-\beta' V(\lambda)} \leq b(\beta) x^\beta \leq b(\beta) e^{\frac{\beta'}{1+\eta} V(x)}$$

with $b(\beta) = a(\beta)(e^{-M_2} + M_1 M_0^{-\beta})$ (the last inequality follows from (5.10)).
Plugging this bound into (5.8) we get

$$\mathbb{P}_V^n(\lambda_1 \geq M) \leq \frac{Z_V^{n-1}}{Z_V^n}[b(\beta)]^{(n-1)}\int_M^\infty \exp\left(-\beta'\left(\frac{\eta n}{1+\eta}\right)V(x)\right)dx$$

Since for n large, $\beta'\eta n > (1+\eta)/2$ we may write

$$\int_M^\infty \exp\left(-\beta'\left(\frac{\eta n}{1+\eta}\right)V(x)\right)dx \leq \exp\left(-\left(\beta'\frac{\eta n}{1+\eta} - \frac{1}{2}\right)C(M)\right)\times \int_M^\infty e^{-\frac{V(x)}{2}}dx$$

where $C(M) = \inf\{V(x) : |x| > M\}$. This last integral is finite in view of (5.9).
We need the following lemma.

Lemma 5.4

$$\limsup_{n\to\infty} n^{-1}\log\frac{Z_V^{n-1}}{Z_V^n} := c_1 < \infty\,.$$

Assuming the result of this lemma, we may write

$$\limsup_{n\to\infty} n^{-1}\log\mathbb{P}_V^n(\lambda_1^+ \geq M) = \limsup_{n\to\infty} n^{-1}\log\mathbb{P}_V^n(\lambda_1 > M) \leq c_1 + \log b(\beta) - C(M)\frac{\beta'\eta}{1+\eta}$$

which ends the proof of exponential tightness since $C(M)\to\infty$ when $M\to\infty$.
The proof of Lemma 5.4 is in Borot and Guionnet (2013b) p. 478 and makes use of the exponential tightness of $(\mu_{\mathtt{u}}^{(n)})_n$ under assumption (A1).

5.2.3 Upper bound

From the above paragraph, we may assume that all the $F_k^\pm$ are subsets of H_M. Furthermore, it suffices to consider F_k^+ subsets of $[\alpha^+, M]\cap H_M$ and F_k^- subsets of $[-M,\alpha^-]\cap H_M$, as an extreme eigenvalue contained in $(\alpha^-+\varepsilon, \alpha^+-\varepsilon)$ implies that the distance of $\mu_{\mathtt{u}}^{(n)}$ to μ_V in the weak topology is at least some $\delta>0$. From the LDP for $\mu_{\mathtt{u}}^{(n)}$ with speed n^2 (Theorem 1.2), this event has negligible probability on our scale of speed n.
As before, we write $x^\pm$ for the vector $(x_1^+,\dots,x_j^+,x_1^-,\dots,x_j^-)$. After permutation, we may assume that the extreme eigenvalues are not among the eigenvalues $\lambda_0 = (\lambda_1,\dots,\lambda_{n-2j})$. Again to simplify notation and to omit several indicator functions, we will not assume throughout this proof that $x_1^+,\dots,x_j^+$ or $x_1^-,\dots x_j^-$ are in the right order. The assertion for the ordered eigenvalues follows directly from the contraction principle.
We have the representation:

$$\mathbb{P}_V^n(\lambda^\pm(j)\in F^+\times F^-) = \frac{1}{Z_V^n}\frac{n!}{(j!)^2(n-2j)!}\int_{F^+\times F^-}\Upsilon_{n,j}(x^\pm)\,dx^\pm \tag{5.11}$$

where

$$\Upsilon_{n,j}(x^\pm) = H(x^\pm)\Xi_{n,j}(x^\pm)e^{-\beta' n\sum_1^{2j}V(x_k^\pm)} \tag{5.12}$$

with

$$H(x^\pm) = \prod_{1\leq r<s\leq 2j}|x_r^\pm - x_s^\pm|^\beta$$

and, setting $\Delta(x^\pm) = (\max x_i^-, \min x_i^+)^{n-2j}$,

$$\Xi_{n,j}(x^\pm) = \int_{\Delta(x^\pm)}\prod_{r=1}^{2j}\prod_{s=1}^{n-2j}|x_r^\pm-\lambda_s|^\beta\prod_{r=1}^{n-2j}e^{-n\beta' V(\lambda_r)}\prod_{1\leq r<s\leq n-2j}|\lambda_r-\lambda_s|^\beta d\lambda_0$$

For M large enough we may replace V by V_M defined by

$$V_M = \begin{cases} V(x) & \text{if } x\in H_M,\\ \infty & \text{otherwise.}\end{cases}$$

without any change (note that V is necessarily bounded on $[\alpha^-,\alpha^+]$). We have then

$$\Xi_{n,j}(x^\pm) = Z^{n-2j}_{\frac{n}{n-2j}V_M}\int_{\Delta(x^\pm)}\prod_{r=1}^{2j}\prod_{s=1}^{n-2j}|x_r^\pm-\lambda_s|^\beta d\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(\lambda_0) \tag{5.13}$$

Set finally

$$Y^M_{n,n-2j} = \frac{Z^n_V}{Z^{n-2j}_{\frac{n}{n-2j}V_M}}\,.$$

We first find an upper bound for $\Upsilon_{n,j}(x^\pm)$. Let $B_\kappa = \{\lambda_0\in\mathbb{R}^{n-2j} : d(\mu_{\mathtt{u}}^{(n-2j)},\mu_V)\le\kappa\}$. On $\Delta(x^\pm)$ the integrand in (5.13) is bounded by $e^{c_1 n}$ for some $c_1=c_1(M)\ge 0$, so that

$$\begin{aligned}\left(Z^{n-2j}_{\frac{n}{n-2j}V_M}\right)^{-1}\Xi_{n,j}(x^\pm) \le &\int_{\Delta(x^\pm)\cap B_\kappa}\prod_{r=1}^{2j}\prod_{s=1}^{n-2j}|x_r^\pm-\lambda_s|^\beta d\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(\lambda_0)\\ &+ e^{c_1 n}\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(B_\kappa^c)\,.\end{aligned} \tag{5.14}$$

One has to use Lemma 5.1. Since the rate function of the LDP for $\mu_{\mathtt{u}}^{(n-2j)}$ has a unique minimizer, Lemma 5.1 yields for the second term in the bound (5.14)

$$e^{c_1 n}\,\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(B_\kappa^c)\le c_2 e^{-c_3 n^2}$$

for some positive constants c_2, c_3. The right hand side of the first line in (5.14) is bounded by

$$\exp\left\{\beta(n-2j)\sup_{\mu:\ d(\mu,\mu_V)\le\kappa}\sum_{r=1}^{2j}\int\log|x_r^\pm-\eta|d\mu(\eta)\right\}$$

and then

$$\begin{aligned}&\left(Z^{n-2j}_{\frac{n}{n-2j}V_M}\right)^{-1}\Upsilon_{n,j}(x^\pm)\\ &\le H(x^\pm)\exp\left\{\beta' n\left(-\sum_{r=1}^{2j}V(x_r^\pm)+2\sup_{\mu:\ d(\mu,\mu_V)\le\kappa}\int\log|x_r^\pm-\eta|d\mu(\eta)\right)\right\}+\ e^{-c_4 n^2}\,.\end{aligned} \tag{5.15}$$

Recall the expression (1.12) of the effective potential

$$\mathcal{J}_V(x) = V(x) - 2\int\log|x-\xi|d\mu_V(\xi)$$

and use the bound

$$\limsup_{\kappa\downarrow 0}\sup_{\xi\in F}\sup_{d(\mu,\mu_V)\le\kappa}\left(2\int\log|\xi-\eta|d\mu(\eta)-V(x)\right)\le-\inf_{\xi\in F}\mathcal{J}_V(\xi) \tag{5.16}$$

(see Borot and Guionnet (2013b) p. 480) we get, for any $\eta>0$ and n large enough

$$\left(Z^{n-2j}_{\frac{n}{n-2j}V_M}\right)^{-1}\sup_{x^\pm\in F^\pm}\Upsilon_{n,j}(x^\pm)\le\exp\beta' n\left(\eta-\inf_{x^\pm\in F^\pm}\sum_{r=1}^{2j}\mathcal{J}_V(x_r^\pm)\right)\,, \tag{5.17}$$

and then, owing to (5.11) and (5.13), we get for any $\eta > 0$

$$\limsup_{n\to\infty} (\beta' n)^{-1} \log \mathbb{P}_V^n(\lambda^{\pm}(j) \in F^{\pm}) \leq \beta'\eta - \beta' \inf_{x^{\pm}\in F^{\pm}} \sum_{r=1}^{2j} \mathcal{J}_V(x_r^{\pm}) - \liminf_{n\to\infty} (\beta' n)^{-1} \log Y_{n,n-2j}^M \,.$$

so that, since η is arbitrary

$$\limsup_{n\to\infty} (\beta' n)^{-1} \log \mathbb{P}_V^n(\lambda^{\pm}(j) \in F^{\pm}) \leq - \inf_{x^{\pm}\in F^{\pm}} \sum_{r=1}^{2j} \mathcal{J}_V(x_r^{\pm}) - \liminf_{n\to\infty} (\beta' n)^{-1} \log Y_{n,n-2j}^M \,. \tag{5.18}$$

It remains to find a lower bound for $Y_{n,n-2j}^M$. We start from

$$Y_{n,n-2j}^M = \frac{Z_V^n}{Z_{\frac{n}{n-2j}V_M}^{n-2j}} \geq \frac{Z_{V_M}^n}{Z_{\frac{n}{n-2j}V_M}^{n-2j}}$$

and we use the result of the Lemma 5.2, noticing that that for M large enough the equilibrium measure is still μ_V and also that $\inf \mathcal{J}_V(x) = \inf \mathcal{J}_{V_M}(x)$ for M large enough.
Coming back to (5.18) yields to the expected upperbound (5.3).

5.2.4 Lowerbound for large deviations

We start from an open ball $B = B(\xi^{\pm}, \varepsilon)$ centered at $\xi^{\pm} \in \mathbb{R}^{2j}$ with radius ε in the sup-norm. Without loss of generality we may assume that it is included in $(\alpha^+, M)^j \times (-M, \alpha^-)^j$ as well as in $\{x \in \mathbb{R}^{2j} | \, V(x_i) \leq M \text{ for all } i\}$. We have again

$$Z_V^n \, \mathbb{P}_V^n(\lambda^{\pm}(j) \in B) = \int_B \Upsilon_{n,j}(x^{\pm}) dx^{\pm}$$

Let us consider the probability measure χ_j^M on $\mathbb{R}^n$, defined by

$$d\chi_j^M(x^{\pm}, \lambda) := (\kappa_{n,j}^M)^{-1} \mathbb{1}_B(x^{\pm}) \mathbb{1}_{\Delta(x^{\pm})}(\lambda_0) dx^{\pm} d\mathbb{P}_{\frac{n}{n-2j}V_M}^{n-2j}(\lambda_0)$$

where $\kappa_{n,j}^M$ is the normalizing constant. We have

$$\int_B \Upsilon_{n,j}(x^{\pm}) dx^{\pm} = Z_{\frac{n}{n-2j}V_M}^{n-2j} \kappa_{n,j} I_{n,j}^M$$

where

$$I_{n,j}^M := \int H(x^{\pm}) e^{-\beta' n \sum_{k=1}^{2j} V_M(x_k^{\pm})} \left(\prod_{k=1}^{n-2j} |x_r^{\pm} - \lambda_k|^{\beta} \right) d\chi_j^M(x^{\pm}, \lambda)$$

Jensen's inequality gives,

$$\frac{1}{\beta'} \log I_{n,j}^M \geq n I_n^{(1)} + 2 I_n^{(2)} + 2(n-2j) I_n^{(3)},$$

where

$$I_n^{(1)} = -\int \sum_{k=1}^{2j} V_M(x_k^\pm) d\chi_j^M(x^\pm, \lambda),$$

$$I_n^{(2)} = \int \sum_{1\le r<s\le 2j} \log|x_r^\pm - x_s^\pm| d\chi_j^M(x^\pm, \lambda),$$

$$I_n^{(3)} = \frac{1}{n-2j} \int \sum_{r=1}^{n-2j} \sum_{k=1}^{j} \log|x_k^\pm - \lambda_r| d\chi_j^M(x^\pm, \lambda)\,.$$

Lemma 5.3 implies that for $x^\pm \in B$ (recall that points in B are bounded away from the support of μ_V),

$$\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(\lambda_0 \in \Delta(x^\pm)) \xrightarrow[n\to\infty]{} 1 \tag{5.19}$$

and then

$$\kappa_{n,j}^M \xrightarrow[n\to\infty]{} \int_B dx^\pm = (2\varepsilon)^{2j} \tag{5.20}$$

We have that $\sum_k V(x_k^\pm)$ is uniformly bounded on B. So that, from (5.19) and (5.20)

$$\lim_{n\to\infty} \frac{1}{\beta' n} I_n^{(1)} = (2\varepsilon)^{-2j} \int_B \sum_{k=1}^{2j} V(x_k^\pm)\, dx^\pm\,. \tag{5.21}$$

For the second term we may assume without loss of generality that B is tie-free. Hence, we get a similar conclusion

$$\lim_{n\to\infty} I_n^{(2)} = (2\varepsilon)^{-2j} \int_B \sum_{1\le r<s\le 2j} \log|x_r^\pm - x_s^\pm|\, dx^\pm\,. \tag{5.22}$$

Now for the third term, we first bound by below $\mathbb{1}_{z>0}\log z$ by $\log_M(z) := (\log z)\mathbb{1}_{0<z\le M}$ and set

$$\ell_{M,B}(t) := \int_B \sum_{k=1}^{2j} \log_M|x_k^\pm - t|\, dx^\pm\,,$$

so that

$$I_n^{(3)} \ge (\kappa_{n,j}^M)^{-1} \int\int \ell_{M,B}(z) d\mu_{\mathrm{u}}^{(n-2j)}(z) d\mathbb{P}^{n-2j}_{\frac{n}{n-2j}V_M}(\lambda_0)\,.$$

Since $\ell_{M,B}$ is continuous and bounded as the convolution of a L^1 and a L^∞ function, the above bound converges to $(2\varepsilon)^{-2j}\int \ell_{M,B}(z)d\mu_V(z)$ as $n\to\infty$. Let us notice that since the support of μ_V is compact, and B is fixed, we can choose M large enough so that $x - M \le t \le x + M$ for every $x \in B$ and t in the support of μ_V. We get $\int \ell_{M,B}(z)d\mu_V(z) = \int \ell_B(z)d\mu_V(z)$ where

$$\ell_B(t) = \int_B \sum_{k=1}^{2j} \log|x_k^\pm - t| dx^\pm\,.$$

At this stage, we have :

$$\liminf_{n\to\infty} \frac{1}{\beta' n} \log I^M_{n,j} \geq -(2\varepsilon)^{-2j} \int_B \sum_{k=1}^{2j} V(x_k^\pm) dx^\pm + 2(2\varepsilon)^{-2j} \int \ell_B(z) d\mu_V(z)\,.$$

Going back to $\mathbb{P}_V^n(\lambda^\pm(j) \in B)$ we get

$$\liminf_{n\to\infty} \frac{1}{\beta' n} \log \mathbb{P}_V^n(\lambda^\pm(j) \in B) \geq -\limsup_{n\to\infty} \frac{1}{\beta' n} \log \frac{Z_V^n}{Z^{n-2j}_{\frac{n}{n-2j}V_M}} - (2\varepsilon)^{-2j} \int_B \sum_{k=1}^{2j} V(x_k^\pm) dx^\pm$$

$$+ 2(2\varepsilon)^{-2j} \int \ell_B(z) d\mu_V(z)\,. \tag{5.23}$$

Splitting into two parts the integral defining Z_V^n we have

$$Z_V^n = Z_{V_M}^n + Z_V^n\,\mathbb{P}_V^n(\lambda_1^\pm \notin H_M)$$

and from the exponential tightness,

$$\frac{Z_V^n}{Z^{n-p}_{nV_M/n-p}} \leq \frac{1}{1-e^{-nC(M)}} \frac{Z_{V_M}^n}{Z^{n-p}_{nV_M/n-p}}$$

Now we apply Lemma 5.2

$$\limsup_{n\to\infty} \frac{1}{\beta' n} \log \frac{Z_V^n}{Z^{n-2j}_{\frac{n}{n-2j}V_M}} \leq -\inf_{x^\pm} \sum_{k=1}^{2j} \mathcal{J}(x_k^\pm)$$

which, plugged into (5.23), yields

$$\liminf_{n\to\infty} \frac{1}{\beta' n} \log \mathbb{P}_V^n(\lambda^\pm(j) \in B) \geq \inf_{x^\pm} \sum_{k=1}^{2j} \mathcal{J}(x_k^\pm) - (2\varepsilon)^{-2j} \int_B \sum_{k=1}^{2j} V(x_k^\pm) dx^\pm$$

$$+ 2(2\varepsilon)^{-2j} \int \ell_B(z) d\mu_V(z)\,. \tag{5.24}$$

Remembering that B has volume $(2\varepsilon)^{2j}$, and letting $\varepsilon \to 0$ we obtain

$$\liminf_{n\to\infty} \frac{1}{\beta' n} \log \mathbb{P}_V^n(\lambda^\pm(j) \in B) \geq \inf_{x^\pm} \sum_{k=1}^{2j} \mathcal{J}(x_k^\pm) - \sum_{k=1}^{2j} V(\xi_k^\pm) + 2 \int \sum_{k=1}^{2j} \log|\xi_k^\pm - t| d\mu_V(t).$$

This yields the lower bound (5.4) and completes the proof of Theorem 4.1. ∎

5.3 Proofs of Lemmas of Section 5.1

5.3.1 Proof of Lemma 5.1

Notice that

$$\frac{d\mathbb{P}^n_{V_n}}{d\mathbb{P}^n_V} = \frac{Z^n_V}{Z^n_{V_n}} \exp\left(-\beta' rn\mu^{(n)}_{\mathtt{u}}(V)\right) \tag{5.25}$$

We need the following lemma.

Lemma 5.5

$$\lim_{n\to\infty} \frac{1}{n^2} \log \frac{Z^n_{V_n}}{Z^n_V} = 0$$

Admitting the result of this lemma, we follow the steps of Anderson et al. (2010) Section 2.6.

Exponential tightness:
We have

$$\begin{aligned}\mathbb{P}^n_{V_n}(\mu^{(n)}_{\mathtt{u}}(V) > t) &= \frac{Z^n_V}{Z^n_{V_n}} \int_{\{\mu^{(n)}_{\mathtt{u}}(V)>t\}} \exp\left(-\beta' rn\mu^{(n)}_{\mathtt{u}}(V)\right) d\mathbb{P}^n_V \\ &\leq \frac{Z^n_V}{Z^n_{V_n}} e^{-n\beta' rt}\mathbb{P}^n_V\left(\mu^{(n)}_{\mathtt{u}}(V) > t)\right)\end{aligned}$$

which yields

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_{V_n}(\mu^{(n)}_{\mathtt{u}}(V) > t)) \leq \limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_V(\mu^{(n)}_{\mathtt{u}}(V) > t))$$

and we may refer to the classical case.

Large Deviation upper bound:
Let μ be any probability measure on $\mathbb{R}$. We start from

$$\mathbb{P}^n_{V_n}(d(\mu^{(n)}_{\mathtt{u}}, \mu) \leq \varepsilon) = \frac{Z^n_V}{Z^n_{V_n}} \int_{\{d(\mu^{(n)}_{\mathtt{u}},\mu)\leq\varepsilon\}} \exp\left(-\beta' rn\mu^{(n)}_{\mathtt{u}}(V)\right) d\mathbb{P}^n_V \tag{5.26}$$

Since V is bounded below by $V_{\min}$, we get the upper bound

$$\mathbb{P}^n_{V_n}(d(\mu^{(n)}_{\mathtt{u}}, \mu) \leq \varepsilon) \leq \frac{Z^n_V}{Z^n_{V_n}} e^{-\beta' rnV_{\min}}\mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \leq \varepsilon)$$

so that

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_{V_n}(d(\mu^{(n)}_{\mathtt{u}}, \mu) \leq \varepsilon) \leq \limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \leq \varepsilon)$$

and we may refer to the classical case.

Large deviations lower bound:
We start again from (5.26) and get, for every $t > 0$, the bound

$$\mathbb{P}^n_{V_n}(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon) \ge \frac{Z^n_V}{Z^n_{V_n}} e^{-\beta' rnt} \mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon,\, \mu^{(n)}_{\mathtt{u}}(V) \le t)$$

Now,

$$\mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon,\, \mu^{(n)}_{\mathtt{u}}(V) \le t) \ge \mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon) - \mathbb{P}^n_V(\mu^{(n)}_{\mathtt{u}}(V) > t)$$

From the previous consideration of exponential tightness, it is possible to choose t large enough so that

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon,\, \mu^{(n)}_{\mathtt{u}}(V) \le t) \ge \liminf_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}^n_V(d(\mu^{(n)}_{\mathtt{u}}, \mu) \le \varepsilon)$$

and refer to the classical case. This ends the proof of Lemma 5.1.

Proof of Lemma 5.5:
We have

$$\frac{Z^n_{V_n}}{Z^n_V} = \int \exp\left(-\beta' rn\mu^{(n)}_{\mathtt{u}}(V)\right) d\mathbb{P}^n_V$$

On the one hand, we observe that since V is bounded from below by a constant $V_{\min}$, we have

$$\frac{Z^n_{V_n}}{Z^n_V} \le \exp\left(-\beta' rnV_{\min}\right)$$

so that

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \frac{Z^n_{V_n}}{Z^n_V} \le 0\,. \tag{5.27}$$

On the other hand, for every $t > 0$

$$\frac{Z^n_{V_n}}{Z^n_V} \ge e^{-\beta' rnt} \mathbb{P}^n_V\left(\mu^{(n)}_{\mathtt{u}}(V) \le t\right)$$

From (2.6.21) in Anderson et al. (2010), we know that $\lim_{n\to\infty} \mathbb{P}^n_V\left(\mu^{(n)}_{\mathtt{u}}(V) \le t\right) = 1$ for t large enough. We easily deduce

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \frac{Z^n_{V_n}}{Z^n_V} \ge 0\,, \tag{5.28}$$

which ends the proof of Lemma 5.5.

5.3.2 Proof of Lemma 5.2

Let

$$Y_{n,n-p} := \frac{Z_V^n}{Z^{n-p}_{\frac{n}{n-p}V}}\,.$$

Lower bound:

$$\begin{aligned}\frac{Z_V^n}{Z^{n-p}_{\frac{n}{n-p}V}} &= \int\int\prod_{r=1}^{p}\left(e^{-\beta' nV(x_r)}\prod_{k=1}^{n-p}|x_r-\lambda_k|^\beta\right)H(x)\,dx\,d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)\\ &\geq \int_{B(\xi,\varepsilon)}\int\prod_{r=1}^{p}\left(e^{-\beta' nV(x_r)}\prod_{k=1}^{n-p}|x_r-\lambda_k|^\beta\right)H(x)\,dx\,d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)\end{aligned}$$

where ξ is a point in $\mathbb{R}^p$ with tie-free entries and $B(\xi,\varepsilon)$ a ball of radius ε in the sup-norm. We may assume that $B(\xi,\varepsilon)\subset\{x\in\mathbb{R}^{2j}|\,V(x_i)<\infty \text{ for all } i\}$. On $B(\xi,\varepsilon)$ the potential V is uniformly continuous and we may replace each x_k by ξ_k and get for ε small enough

$$H(x)\prod_{r=1}^{p}e^{-\beta' nV(x_r)}\geq\prod_{r=1}^{p}e^{-\beta' nV(\xi_r)-\beta' n\delta_\varepsilon}$$

for some $\delta_\varepsilon>0$. Set

$$\ell_\varepsilon(y,\lambda)=(2\varepsilon)^{-1}\int_{y-\varepsilon}^{y+\varepsilon}\log|t-\lambda|dt$$

Applying Jensen's inequality (the exponential is convex and we integrate over $(2\varepsilon)^p\prod_{r=1}^{p}\mathbb{1}_{|x_r-\xi_r|\leq\varepsilon}dx_k$) we get

(5.29)

$$Y_{n,n-p}\geq(2\varepsilon)^p\left(\prod_{r=1}^{p}e^{-\beta' nV(\xi_r)-\beta' n\delta_\varepsilon}\right)\int\exp\left(2\beta'(n-p)\sum_{r=1}^{p}\int\ell_\varepsilon(\xi_r,z)d\mu_{\mathtt{u}}^{(n-p)}(z)\right)d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0).$$

The function $\lambda\mapsto\ell_\varepsilon(y,\lambda)$ is continuous and bounded (all variables live in a compact set). We can bound the integral in (5.29) from below by

$$\begin{aligned}&\int_{\{\lambda_0:d(\mu_{\mathtt{u}}^{(n-2j)},\mu_V)\leq\kappa\}}\exp\left(2\beta'(n-p)\sum_{r=1}^{p}\int\ell_\varepsilon(\xi_r,z)d\mu_{\mathtt{u}}^{(n-p)}(z)\right)d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)\\ &\geq\mathbb{P}^{n-p}_{\frac{n}{n-p}V}\left(d(\mu_{\mathtt{u}}^{(n-p)},\mu_V)\leq\varepsilon\right)\exp\left(2\beta'(n-p)\sum_{r=1}^{p}\int\ell_\varepsilon(\xi_r,z)d\mu_V(z)-\beta'\delta'_\varepsilon n\right)\end{aligned}$$

with $\delta'_\varepsilon>0$, hence

$$\begin{aligned}Y_{n,n-p}\geq(2\varepsilon)^p&\left(\prod_{r=1}^{p}e^{-\beta' nV(\xi_r)-\beta' n\delta_\varepsilon}\right)\mathbb{P}^{n-p}_{\frac{n}{n-p}V}\left(d(\mu_{\mathtt{u}}^{(n-p)},\mu_V)\leq\varepsilon\right)\\ &\times\exp\left(2\beta'(n-p)\sum_{r=1}^{p}\int\ell_\varepsilon(\xi_r,z)d\mu_V(z)\right)\end{aligned}$$

According to Lemma 5.1, we know that $\mathbb{P}^{n-p}_{\frac{n}{n-p}V}\left(d(\mu_{\mathbf{u}}^{(n-p)},\mu_V)\leq\varepsilon\right)\to 1$. Since the logarithmic potential $t\mapsto\int\log|t-\lambda|d\mu_V(\lambda)$ is continuous, we may write

$$\int\ell_\varepsilon(\xi_r,z)d\mu_V(z)\geq\int\log|\xi_r-z|d\mu_V(z)-\delta''_\varepsilon\,,$$

where δ''_ε may depend on ξ but tends to zero with ε. We have then, for ξ fixed, for every $\varepsilon>0$

$$\liminf_{n\to\infty}\,(\beta' n)^{-1}\log Y_{n,n-p}\geq-\sum_{r=1}^{p}\mathcal{J}(\xi_r)-2p(\delta'_\varepsilon+\delta''_\varepsilon)$$

Since it is true for every ε, after optimizing in ξ we may conclude

$$\liminf_{n\to\infty}\,(\beta' n)^{-1}\log Y_{n,n-p}\geq-\inf_{\xi}\sum_{r=1}^{p}\mathcal{J}(\xi_r)\,. \tag{5.30}$$

Upper bound:
We have

$$Y_{n,n-p}=\int H(x)\prod_{r=1}^{p}e^{-\beta' n\sum_1^p V(x_r)}\left(\int\prod_{k=1}^{n-p}|x_r-\lambda_k|^\beta\,d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)\right)dx$$

Recall the definition $B_\kappa=\{\lambda_0\in\mathbb{R}^{n-2j}:d(\mu_{\mathbf{u}}^{(n-2j)},\mu_V)\leq\kappa\}$. Since all variables live on a compact set, $\prod_k|x_r-\lambda_k|^\beta$ is bounded by $e^{c_1 n}$ for some $c_1>0$, and then

$$\int\prod_{k=1}^{n-p}|x_r-\lambda_k|^\beta\,d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)\leq\int_{B_\kappa}\prod_{k=1}^{n-p}|x_r-\lambda_k|^\beta d\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(\lambda_0)+e^{c_1 n}\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(B_\kappa^c)\,. \tag{5.31}$$

Since the rate function of the LDP has a unique minimizer, Proposition 5.1 yields

$$e^{c_1 n}\,\mathbb{P}^{n-p}_{\frac{n}{n-p}V}(B_\kappa^c)\leq c_2 e^{-c_3 n^2}$$

for some positive constants c_2,c_3. The integral on the right hand side of (5.31) is bounded by

$$\exp\left\{\beta(n-p)\sup_{\mu:\;d(\mu,\mu_V)\leq\kappa}\sum_{r=1}^{p}\int\log|x_r-\eta|d\mu(\eta)\right\}$$

and then, since we integrate over a compact set,

$$\begin{aligned}Y_{n,n-p}\leq&\int H(x)\exp\left\{\beta' n\left(-\sum_{r=1}^{p}V(x_r)+2\sup_{\mu:\;d(\mu,\mu_V)\leq\kappa}\int\log|x_r-\eta|d\mu(\eta)\right)\right\}dx\\&+\;c_4 e^{-c_5 n^2}\,.\end{aligned} \tag{5.32}$$

If we use again the bound (5.16) we get, for any $\eta>0$ and n large enough

$$Y_{n,n-p}\leq\exp\beta' n\left(\eta-\inf_x\sum_{r=1}^{p}\mathcal{J}_V(x_r)\right)\,, \tag{5.33}$$

and then

$$\limsup_{n\to\infty}\frac{1}{\beta' n}\log Y_{n,n-p}\leq-\inf_x\sum_{r=1}^{p}\mathcal{J}_V(x_r)\,.$$

5.3.3 Proof of Lemma 5.3

From the LDP for extreme value of Borot and Guionnet (2013b), the rate function is $\mathcal{J}_V - \inf_x \mathcal{J}_V(x)$. So, if this rate function which vanishes on the support of μ_V does not vanish outside, that means that the probability that λ_1^+ is greater than $\alpha+\varepsilon$ is exponentially small, and similarly for λ_1^-.

6 Appendix 2 : Proof of Lemma 4.5 and Lemma 4.6

6.1 Proof of Lemma 4.5

Let $\mu_n \to \mu$ in $\mathcal{S}_1$ equipped with the topology induced by (4.3). Let f be continuous and bounded and $\varepsilon > 0$. Since μ is normalized, we may choose N so large that

$$\mu(I) + \sum_{i=1}^{N\wedge N^+} \gamma_i^+ + \sum_{i=1}^{N\wedge N^-} \gamma_i^- > 1 - \varepsilon.$$

Note that N may be 0. Given this N, choose n_0 so large such that for all $n \geq n_0$

$$d_n := \left| \int g d\mu_{n|I} - \int g d\mu_{|I} \right| + \sum_{i=1}^{N\wedge N^+} |\gamma_{i,n}^+ g(\lambda_{i,n}^+) - \gamma_i^+ g(\lambda_i^+)| + \sum_{i=1}^{N\wedge N^-} |\gamma_{i,n}^- g(\lambda_{i,n}^-) - \gamma_i^- g(\lambda_i^-)| < \varepsilon$$

for $g \in \{1, f\}$, which is possible thanks to our topology on $\mathcal{S}$. This implies in particular

$$\sum_{i=N\wedge N^+ +1}^{N^+} |\gamma_{i,n}^+| + \sum_{i=N\wedge N^- +1}^{N^-} |\gamma_{i,n}^-| \leq 2\varepsilon.$$

Then we have

$$\begin{aligned}
&\left| \int f d\mu_n - \int f d\mu \right| \\
&\leq d_n + \sum_{i=N\wedge N^+ +1}^{N^+} |\gamma_{i,n}^+ f(\lambda_{i,n}^+)| + \sum_{i=N\wedge N^- +1}^{N^-} |\gamma_{i,n}^- f(\lambda_{i,n}^-)| + \sum_{i=N\wedge N^+ +1}^{N^+} |\gamma_i^+ f(\lambda_i^+)| + \sum_{i=N\wedge N^- +1}^{N^-} |\gamma_i^- f(\lambda_i^-)| \\
&\leq d_n + 2\varepsilon ||f||_\infty + \varepsilon ||f||_\infty \leq \varepsilon + 3\varepsilon ||f||_\infty
\end{aligned}$$

for all $n \geq n_0$. ■

6.2 Proof of Lemma 4.6

We need to show that for measures $\mu_n \in \mathcal{S}_1$ with $\mu_n \to \mu \in \mathcal{P}_1 \setminus \mathcal{S}_1$ weakly, we have $\mathcal{I}_V(\mu) \to \infty$. If $\mu \notin \mathcal{S}_1$, then either μ has a nondiscrete part outside of $[\alpha_-, \alpha_+]$ or an infinite number of atoms outside of $[\alpha_- - \varepsilon, \alpha_+ + \varepsilon]$ for some $\varepsilon > 0$.

Now, the only way for μ_n to be arbitrarily close to such a measure μ in the weak topology is if there exists a number $\ell(n)$ of atoms $x_1, \ldots, x_{\ell(n)}$ that are not lying in $[\alpha_- - \varepsilon, \alpha_+ + \varepsilon]$ for some $\varepsilon > 0$ and $\ell(n) \to \infty$. This implies $\mathcal{F}(x_i) > \delta$ for some positive δ for all $i \leq \ell(n)$ and then $\mathcal{I}_V(\mu) \to \infty$. ■

Acknowledgments

Many thanks are due both to Stanislas Kupin and Gérard Ben Arous for helpful discussions at the time of the early genesis of this work. The spectral point of view of Stanislas on sum rules has strongly enlighten the darkness of our research paths on this subject. During the conference for the birthday of one author of this work, Gérard gave us the keys for the study for the large deviations of the spectral measure part supported by the extreme eigenvalues. At the time of the first revision, we warmly thank Barry Simon for its insight remarks and suggestions, especially on the gems.

References

Albeverio, S., Pastur, L., and Shcherbina, M. (2001). On the $1/n$ expansion for some unitary invariant ensembles of random matrices. *Comm. Math. Phys.*, 224(1):271–305.

Anderson, G., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices.* Cambridge University Press, Cambridge.

Auffinger, A., Arous, G. B., and Černỳ, J. (2013). Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201.

Baldi, P. (1988). Large deviations and stochastic homogenization. *Annali di Matematica Pura ed Applicata*, 151(1):161–177.

Ben Arous, G., Dembo, A., and Guionnet, A. (2001). Aging of spherical spin glasses. *Probab. Theory Relat. Fields*, 120:1–67.

Benaych-Georges, F., Guionnet, A., and Maïda, M. (2012). Large deviations of the extreme eigenvalues of random deformations of matrices. *Probability Theory and Related Fields*, 154(3-4):703–751.

Borot, G. and Guionnet, A. (2013a). Asymptotic expansion of β matrix models in the multi-cut regime. arXiv preprint arXiv:1303.1045.

Borot, G. and Guionnet, A. (2013b). Asymptotic expansion of β matrix models in the one-cut regime. *Comm. Math. Phys.*, 317(2):447–483.

Bourgade, P., Nikeghbali, A., and Rouault, A. (2009). Circular Jacobi ensembles and deformed Verblunsky coefficients. *Int. Math. Res. Notices*, (23):4357–4394.

Damanik, D., Killip, R., and Simon, B. (2010). Perturbations of orthogonal polynomials with periodic recursion coefficients. *Annals of mathematics*, 171(3):1931–2010.

Deift, P., Kriecherbauer, T., McLaughlin, K., Venakides, S., and Zhou, X. (1999). Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52(11):1335–1425.

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer.

Dette, H. and Studden, W. (1997). *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics,.

Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847.

Fan, C., Guionnet, A., Song, Y., and Wang, A. (2014). Convergence of eigenvalues to the support of the limiting measure in critical β matrix models. arXiv preprint arXiv:1402.1796.

Gamboa, F. and Lozada-Chang, L.-V. (2004). Large deviations for random power moment problem. *The Annals of Probability*, 32(3B):2819–2837.

Gamboa, F., Nagel, J., Rouault, A., and Wagener, J. (2012). Large deviations for random matricial moment problems. *J. Multivariate Anal.*, 106:17–35.

Gamboa, F. and Rouault, A. (2010). Canonical moments and random spectral measures. *J. Theoret. Probab.*, 23:1015–1038.

Gamboa, F. and Rouault, A. (2011). Large deviations for random spectral measures and sum rules. *Appl. Math. Res. Express AMRX*, (2):281–307.

Hardy, A. (2012). A note on large deviations for 2D Coulomb gas with weakly confining potential. *Electron. Commun. Probab*, 17(19):1–12.

Johansson, K. (1998). On fluctuations of eigenvalues of random hermitian matrices. *Duke Math. J.*, 91(1):151–204.

Killip, R. (2007). Spectral theory via sum rules. In *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, volume 76 of *Proc. Sympos. Pure Math.*, pages 907–930. Amer. Math. Soc., Providence, RI.

Killip, R. and Nenciu, I. (2004). Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701.

Killip, R. and Simon, B. (2003). Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math.*, 158(1):253–321.

Krishnapur, M., Rider, B., and Virag, B. (2013). Universality of the Stochastic Airy Operator. arXiv 1306.4832v1.

Kupin, S. (2005). Spectral properties of Jacobi matrices and sum rules of special form. *Journal of Functional Analysis*, 227(1):1 – 29.

Nazarov, F., Peherstorfer, F., Volberg, A., and Yuditskii, P. (2005). On generalized sum rules for Jacobi matrices. *Int. Math. Res. Not.*, (3):155–186.

Rockafellar, R. (1971). Integrals which are convex functionals, II. *Pacific J. Math.*, 39(2):439–469.

Saitoh, N. and Yoshida, H. (2001). The infinite divisibility and orthogonal polynomials with a constant recursion formula in free probability theory. *Probab. Math. Statist.*, 21(1, Acta Univ. Wratislav. No. 2298):159–170.

Serfaty, S. (2014). Coulomb gases and Ginzburg-Landau vortices. arXiv 1403.6860v1.

Simon, B. (2007). Orthogonal polynomials with exponentially decaying recursion coefficients. In *Probability and mathematical physics*, pages 453–463.

Simon, B. (2011). *Szegő's theorem and its descendants*. M. B. Porter Lectures. Princeton University Press, Princeton, NJ.

Trotter, H. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. in Math.*, 54(1):67–82.

Xia, N., Qin, Y., and Bai, Z. (2013). Convergence rates of eigenvector empirical spectral distribution of large dimensional sample covariance matrix. *Ann. Statist.*, 41(5):2572–2607.