



**HAL**  
open science

# Sum rules and large deviations for spectral measures on the unit circle

Fabrice Gamboa, Jan Nagel, Alain Rouault

► **To cite this version:**

Fabrice Gamboa, Jan Nagel, Alain Rouault. Sum rules and large deviations for spectral measures on the unit circle. 2016. hal-01300205v2

**HAL Id: hal-01300205**

**<https://hal.science/hal-01300205v2>**

Preprint submitted on 22 Apr 2016 (v2), last revised 30 Jan 2017 (v6)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sum rules and large deviations for spectral measures on the unit circle

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

Laboratoire de Mathématiques de Versailles

UVSQ, CNRS

Université Paris-Saclay

78035-Versailles Cedex France

e-mail: alain.rouault@uvsq.fr

April 22, 2016

## Abstract

This work is a companion paper of [20] and [19]. We continue to explore the connections between large deviations for random objects issued from random matrix theory and sum rules. Here, we are concerned essentially with measures on the unit circle whose support is an arc that is possibly proper. We particularly focus on two matrix models. The first one is the Gross-Witten ensemble. In the gapped regime we give a probabilistic interpretation of a Simon sum rule. The second matrix model is the Hua-Pickrell ensemble. Unlike the Gross-Witten ensemble the potential is here infinite at one point. Surprisingly, but as in [20] and [19], we obtain a completely new sum rule for the deviation to the equilibrium measure of the Hua-Pickrell ensemble. The extension to matrix measure is also studied.

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

# 1 Introduction

Two of the most famous sum rules are Szegő's formula and Killip-Simon's sum rule. They are related respectively to the theory of orthogonal polynomials on the unit circle (OPUC) and to the theory of orthogonal polynomials on the real line (OPRL).

In the OPUC frame, the Szegő-Verblunsky theorem (see [41], Theorem 1.8.6, p. 29) 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 equation between two successive monic orthogonal polynomials  $\phi_{n+1}$  and  $\phi_n$  (where  $\deg \phi_n = n$ ,  $n \geq 0$ ) associated with a probability measure  $\mu$  on the unit circle  $\mathbb{T}$  supported by at least  $n + 1$  points involves a complex number  $\alpha_n$  and may be written as

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

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 ([16]).

The Szegő-Verblunsky theorem is the identity

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

where the Lebesgue decomposition of  $\mu$  with respect to the normalized Lebesgue measure 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. Changing the signs in both side in this equation leads to

$$(1.3) \quad \mathcal{K}(\text{UNIF} | \mu) = - \sum_{n \geq 0} \log(1 - |\alpha_n|^2)$$

where, for probability measures  $\nu$  and  $\mu$ ,  $\mathcal{K}(\nu | \mu)$  denotes the Kullback-Leibler divergence or relative entropy of  $\nu$  with respect to  $\mu$  (see (4.1)), and UNIF is the normalized Lebesgue measure on  $\mathbb{T}$ .

In the OPRL frame, for a probability measure  $\mu$  having an infinite support, a.k.a. nontrivial case (resp. with a finite support consisting of  $n > 0$  points, a.k.a. trivial case), the orthonormal polynomials associated to  $\mu$  (with positive leading coefficients) obtained by applying the orthonormalizing Gram-Schmidt procedure to the sequence  $1, x, x^2, \dots$  obey the recursion relation

$$(1.4) \quad xp_k(x) = a_{k+1}p_{k+1}(x) + b_{k+1}p_k(x) + a_k p_{k-1}(x)$$

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

To describe the Killip-Simon sum rule, we need some more notations. Let  $\mathcal{M}_1(I)$  denote the set of all probability measures on  $I$  a subset of  $\mathbb{R}$  or of the unit the circle  $\mathbb{T}$ . For  $\alpha^- < \alpha^+$ , let  $\mathcal{S}_1^{\mathbb{R}}(\alpha^-, \alpha^+)$  be the set of all probability measures  $\mu$  on  $\mathbb{R}$  with

(i)  $\text{supp}(\mu) = J \cup \{\lambda_i^-\}_{i=1}^{N^-} \cup \{\lambda_i^+\}_{i=1}^{N^+}$ , where  $J \subset [\alpha^-, \alpha^+]$ ,  $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$  and

$$\lambda_1^- < \lambda_2^- < \dots < \alpha^- \quad \text{and} \quad \lambda_1^+ > \lambda_2^+ > \dots > \alpha^+.$$

(ii) If  $N^-$  (resp.  $N^+$ ) is infinite, then  $\lambda_j^-$  converges towards  $\alpha^-$  (resp.  $\lambda_j^+$  converges to  $\alpha^+$ ).

Such a measure  $\mu$  will be written as

$$(1.5) \quad \mu = \mu|_I + \sum_{i=1}^{N^+} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \gamma_i^- \delta_{\lambda_i^-}$$

The reference probability measure is now the semicircle law

$$(1.6) \quad \text{SC}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2} 1_{[-2,2]}(x) dx.$$

Additionally, we set

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

and  $\mathcal{F}_H^-(x) := \mathcal{F}_H^+(-x)$  for  $x \in \mathbb{R}$ . Then, the Killip-Simon sum rule is the following equation (1.7) (see [41] Theorem 3.5.5):

For a probability measure  $\mu \in \mathcal{S}_1^{\mathbb{R}}(-2, 2)$  with recursion coefficients  $(a_k)_k, (b_k)_k$  as in (1.4),

$$(1.7) \quad \mathcal{K}(\text{SC} | \mu) + \sum_{n=1}^{N^+} \mathcal{F}_{\text{SC}}^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_{\text{SC}}^-(\lambda_n^-) = \sum_{k \geq 1} \left( \frac{1}{2} b_k^2 + G(a_k^2) \right),$$

where  $G(x) = x - 1 - \log x$ , and where both sides may be infinite simultaneously.

The common feature of formulas (1.3) and (1.7) is that they state equalities between non-negative functionals. We can consider them as equalities of two discrepancies. On the left side it is the reverse relative entropy with respect to some reference probability measure plus eventually a contribution of the outlying point masses. On the right side it is a sum vanishing only when the *coefficients* involved are those of the reference probability measure.

In [21] and [20], we revisit these results by giving a probabilistic proof based on large deviations (as we will explain below). This allowed in the OPRL case to discover new sum rules, corresponding to the Marchenko-Pastur and Kesten-McKay measures, respectively. The main interesting feature of (1.7) is the role played by the outliers of the measure  $\mu$ , i.e. its discrete masses located out of the support of the reference measure. Coming back to the the OPUC case, in the Szegő-Verblunsky theorem (1.2) there are no outliers as the reference probability measure is supported by the full unit circle. Nevertheless, in the context of the unit circle, there are some very interesting probability measures supported by a proper arc. In this paper, we give and prove new original sum rules for families of reference probability measures that are possibly supported

by a proper arc of the unit circle. One of our new sum rule (see Theorem 5.1) concerns the reference probability measure  $\text{HP}_d$  (see (3.17),  $d$  is a positive parameter) that is supported by the a proper arc depending on  $d$  included in  $(0, \pi)$  and whose density is proportional to  $[\sin(2^{-1}\theta)]^{-1}$ . Up to our knowledge, Theorem 5.1 is completely new.

Our method for finding and showing a sum rule relies on the large deviations properties for a sequence of random measures built on random matrices. Let us give in a nutshell the scheme of our probabilistic method. We interpret the measure  $\mu$  as the realization of a (random) spectral measure of a pair  $(M, e)$  where  $M$  is a random normal operator (unitary or Hermitian) and  $e$  a fixed vector in a Hilbert space  $\mathcal{H}$ .

Let assume that  $\dim \mathcal{H} = n$  ( $n \in \mathbb{N}_*$ ). Then,  $\mu$  is a discrete probability measure which can be encoded as

$$(1.8) \quad \mu = \sum_{k=1}^n w_k \delta_{\lambda_k} .$$

A classical assumption is the invariance by any unitary conjugations of the law of  $M$ . Under this assumption, the joint density of  $(\lambda_1, \dots, \lambda_n)$  is proportional to the square of the Vandermonde determinant multiplied by the exponential of some potential. Furthermore, the distribution of the weights  $(w_1, \dots, w_n)$  is uniform on the simplex. This allows in the cases studied in [21] and [20], with convenient assumptions on the potential, to show that the random measure defined in (1.8) satisfies, as  $n$  grows, a large deviation principle (LDP). The speed of the LDP is  $n$  and the rate function is the left hand side of (1.3) or (1.7) or its analogue. In the sum rules, the right hand side is obtained as the rate function seeing the random measure as encoded by its Verblunsky (OPUC) or Jacobi (OPRL) coefficients. Since a rate function is unique, the equality of both sides is a straightforward conclusion.

Of course, there is a natural way to travel in both directions from  $\mathbb{T}$  to  $\mathbb{R}$ . This is the so-called Cayley transform. So that, the results obtained for random measures on  $\mathbb{R}$  may be carried to random measures on  $\mathbb{T}$ . Nevertheless, the confinement assumption made on the potential in [20] is not always true in all the interesting cases on  $\mathbb{T}$ . Two examples are particularly representative and more or less emblematic of studies in OPUC and in equilibrium measures on  $\mathbb{T}$ . The first one is the Gross-Witten (GW) ensemble (gapped/ungapped regime), corresponding to a potential continuous on  $\mathbb{T}$ . The second one is the Hua-Pickrell (HP) ensemble, corresponding to a potential infinite at one point. Both are distributions on the set (group)  $\mathbb{U}(n)$  of unitary  $n \times n$  matrices, (see [36]) or pure Fisher-Hartwig symbol (Section 2 (for HP) and Section 5 in [3]). In the HP case, the potential on the real line satisfies the confinement assumption. It is then possible to use the results of [20] to state directly a LDP principle for the spectral measure. Moreover, since the *deformed* Verblunsky coefficients are independent with known distributions given in [9], the coding with these coefficients gives rise to a LDP and by uniqueness, we conclude with a new sum rule. This method is robust enough to be extended to the matrix case.

In the GW case, the potential on the real line satisfies only a weak growth assumption and we cannot use previous results. Nevertheless, we may work directly on  $\mathbb{T}$ , copying the scheme of proof of the real case, looking carefully at the differences. We do not have exponential tightness for the extremal eigenvalues anymore, but since the potential is finite everywhere, we take benefit

of the compactness of  $\mathbb{T}$ . It was the secret of the paper of [25] to get the LDP for empirical spectral distribution under the weak growth assumption. Besides and to be complete, we revisit the gapped case, for which the LDP is a direct consequence of [21] and we give some probabilistic evidence for the celebrated sum rule due to [39] (Theorem 2.8.1 therein).

The paper is organised as follows. In the next section we give some necessary notations and assumptions. In Section 3 we describe the two main matrix models studied. Section 4 is devoted to our large deviation results for random spectral measure. The sum rules obtained from large deviation considerations are setted in Section 5 while extensions to matrix spectral measure are discussed in Section 6. All technical proofs are postponed to the last section.

## 2 Notations, assumptions and tools

### 2.1 Two encodings of a probability measure on $\mathbb{T}$

If  $U$  is a unitary operator on a Hilbert space  $\mathcal{H}$  and  $e$  a cyclic vector for  $H$ , the spectral measure of the pair  $(U, e)$  is the unique probability measure  $\mu$  on  $\mathbb{T}$  such that

$$(2.1) \quad \langle e, U^k e \rangle = \int_{\mathbb{T}} z^k d\mu(z) \quad (k \in \mathbb{Z}).$$

Actually,  $\mu$  is a unitary invariant for  $(U, e)$ . If the dimension of  $\mathcal{H}$  is  $n$  and  $e$  is cyclic for  $U$ , let  $\lambda_1 = e^{i\theta_1}, \dots, \lambda_n = e^{i\theta_n}$  be the eigenvalues of  $U$  and let  $\psi_1, \dots, \psi_n$  be a system of orthonormal eigenvectors. The spectral measure of the pair  $(U, e)$  is then

$$(2.2) \quad \mu^{(n)} = \sum_{k=1}^n \mathbf{w}_k \delta_{\lambda_k},$$

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

$$(2.3) \quad \mu_{\mathbf{u}}^{(n)} = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k}.$$

Another invariant is the CMV (or 5-diagonal) reduction of  $U$ . Let us now describe shortly the CMV mapping between 5-diagonal 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  $(\chi_k)_{k \geq 0}$  obtained by orthonormalizing  $1, z, z^{-1}, z^2, z^{-2}, \dots$ , the linear transformation  $f(z) \rightarrow zf(z)$  (multiplication by the identity) in  $L^2(\mu)$  is represented by the matrix

$$(2.4) \quad \mathcal{C}_\mu = \begin{pmatrix} \bar{\alpha}_0 & \bar{\alpha}_1 \rho_0 & \rho_1 \rho_0 & 0 & 0 & \dots \\ \rho_0 & -\bar{\alpha}_1 \alpha_0 & -\rho_1 \alpha_0 & 0 & 0 & \dots \\ 0 & \bar{\alpha}_2 \rho_1 & -\bar{\alpha}_2 \alpha_1 & \bar{\alpha}_3 \rho_2 & \rho_3 \rho_2 & \dots \\ 0 & \rho_2 \rho_1 & -\rho_2 \alpha_1 & -\bar{\alpha}_3 \alpha_2 & -\rho_3 \alpha_2 & \dots \\ 0 & 0 & 0 & \bar{\alpha}_4 \rho_3 & -\bar{\alpha}_4 \alpha_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

with

$$(2.5) \quad |\alpha_k| < 1 \text{ and } \rho_k = \sqrt{1 - |\alpha_k|^2}$$

for every  $k \geq 0$  in the non-trivial case and for  $0 \leq k \leq n - 1$  in the trivial case, with  $|\alpha_{n-1}| = 1$  [11]. If the measure is supported by  $n$  points, then the last line is

$$\begin{cases} 0 & \dots & 0 & 0 & \bar{\alpha}_{2r}\rho_{2r-1} & -\bar{\alpha}_{2r}\alpha_{2r-1} & \text{if } n = 2r + 1, \\ 0 & \dots & 0 & \rho_{2r}\rho_{2r-1} & -\rho_{2r}\alpha_{2r-1} & -\bar{\alpha}_{2r+1}\alpha_{2r} & \text{if } n = 2r + 2, \ r \geq 0. \end{cases}$$

Actually, there is a one-to-one correspondence between such a matrix, called finite CMV matrix and a finitely supported measure. If  $\mathcal{C}$  is a such a matrix, we can take the first vector of the canonical basis as the cyclic vector  $e$ . Let  $\mu$  be the spectral measure associated to the pair  $(\mathcal{C}, e_1)$ , then  $\mathcal{C}$  represents the multiplication by  $z$  in the basis  $(\chi_k)$  of orthonormal polynomials associated to  $\mu$  and  $\mathcal{C} = \mathcal{C}_\mu$ .

More generally, if  $\mu$  is a non-trivial probability measure on  $\mathbb{T}$ , we may apply the same Gram-Schmidt process and consider the associated semi-infinite CMV matrix  $\mathcal{C}_\mu$ . Notice that now we have  $|\alpha_k| < 1$  for every  $k$ . The mapping  $\mu \mapsto \mathcal{C}_\mu$  (called here the CMV mapping) is a one to one correspondence between probability measures on  $\mathbb{T}$  having infinite support and this kind of CMV matrices. This result is sometimes called Verblunsky-Favard's theorem (see [38], p. 432).

## 2.2 The Cayley transform, random matrices and invariant models

We will switch several times between  $\mathbb{R}$  and  $\mathbb{T}$  and between distributions of unitary and Hermitian matrices. There is a natural connection between these two sets and also between these two sets of matrices. This transformation is the so-called Cayley transform or stereographical projection. We follow here partly [4] in its presentation. Let  $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$  be the compactified real line, which is topologically isomorphic to  $\mathbb{T}$ . Let  $\tau$  be the Cayley transform defined by:

$$(2.6) \quad \begin{aligned} \zeta \in \mathbb{T} \setminus \{\mathbf{1}\} &\mapsto \tau(\zeta) := i \frac{1 + \zeta}{1 - \zeta} \\ \tau(\mathbf{1}) &= \infty \end{aligned}$$

$$(2.7) \quad \begin{aligned} x \in \mathbb{R} &\mapsto \tau^{-1}(x) = \frac{x - i}{x + i} \\ \tau^{-1}(\infty) &= \mathbf{1}. \end{aligned}$$

It is clear that  $\tau^{-1}$  is a homeomorphism from  $\mathbb{R}$  onto  $\mathbb{T} \setminus \{\mathbf{1}\}$ . Let us notice the important relations

$$(2.8) \quad |\tau^{-1}(x) - \tau^{-1}(y)| = \frac{2|x - y|}{\sqrt{1 + x^2}\sqrt{1 + y^2}}, \quad |1 - \tau^{-1}(x)| = \frac{2}{\sqrt{1 + x^2}},$$

and with angular coordinates

$$\zeta = e^{i\theta} \iff x = \tau(\zeta) = -\cot \theta/2$$

and

$$(2.9) \quad d\theta = \frac{2dx}{1+x^2}.$$

At the level of measures, we will consider three spaces. First  $\mathcal{M}_1(\mathbb{R})$  and  $\mathcal{M}_1(\mathbb{T})$  are the spaces of probability measures on  $\mathbb{R}$  and  $\mathbb{T}$ , respectively, equipped with the topology of the weak convergence. Finally we need to use the set  $\mathcal{M}_{\leq 1}(\mathbb{R})$  of subprobabilities on  $\mathbb{R}$ , equipped with the topology of vague convergence. Let us define the mapping  $\hat{\tau} : \mathcal{M}_1(\mathbb{T}) \rightarrow \mathcal{M}_{\leq 1}(\mathbb{R})$  with  $\hat{\tau}(\nu)$  defined by

$$(2.10) \quad \int_{\mathbb{R}} f(x) d\hat{\tau}(\nu)(x) = \int_{\mathbb{T} \setminus \{1\}} f(\tau(\zeta)) d\nu(\zeta).$$

for any  $f \in \mathcal{C}_0(\mathbb{R})$ , continuous and vanishing at infinity. The mapping  $\hat{\tau}$  is continuous if we equip  $\mathcal{M}_{\leq 1}$  with the topology of vague convergence. Notice that  $\mathcal{M}_1(\mathbb{T})$  and  $\mathcal{M}_{\leq 1}(\mathbb{R})$  are compact sets. We endow all these sets with the corresponding Borel  $\sigma$ -algebra. The image of the uniform distribution on  $\mathbb{T}$  is the Cauchy distribution on  $\mathbb{R}$ . Let  $\mathbb{U}(n)$  be the set of unitary  $n \times n$  matrices and let  $\mathbf{I}_n$  the identity  $n \times n$  matrix. The Cayley transform induces a transformation from  $\mathbb{U}(n) \setminus \{\mathbf{I}_n\}$  onto  $\mathbb{H}(n)$ , the set of Hermitian  $n \times n$  matrices by

$$(2.11) \quad M = \tau(U) := i \frac{\mathbf{I}_n + U}{\mathbf{I}_n - U} \iff U = \tau^{-1}(M) = \frac{M - i\mathbf{I}_n}{M + i\mathbf{I}_n},$$

in the sense of functional calculus. We denote by  $\mathbb{P}^{(n)}$  the normalized Haar measure on  $\mathbb{U}(n)$ . It is classical that under  $\mathbb{P}^{(n)}$  the array of eigenvalues has a density with respect to the Lebesgue measure  $d\zeta_1 \dots d\zeta_n$  on  $\mathbb{T}^n$  which is proportional to

$$|\Delta(\zeta_1, \dots, \zeta_n)|^2,$$

where  $\Delta$  is the Vandermonde determinant. In the Haar case  $\mathcal{V} \equiv 0$  the Verblunsky coefficients are independent. More precisely, in this case the vector  $\alpha^{(n)} := (\alpha_0, \dots, \alpha_{n-1} = e^{i\phi})$  has the distribution

$$dP_0^{(n)}(\alpha_0, \dots, \alpha_{n-1}) = \left( \otimes_{j=0}^{n-2} \eta_{n-j+1}(d\alpha_j) \right) \otimes \frac{d\phi}{2\pi},$$

with  $\eta_r(d\alpha) = \frac{r-2}{\pi} (1 - |\alpha|^2)^{r-3} 1_{\mathbb{D}}(\alpha) d^2\alpha$  (see [29]), with  $\mathbb{D}$  the open unit disk.

More generally, it is usual to equip  $\mathbb{U}(n)$  with a probability measure of the form

$$(2.12) \quad d\mathbb{P}_{\mathcal{V}}^{(n)}(U) = \frac{1}{\mathcal{Z}_n^{\mathcal{V}}} e^{-n \operatorname{tr} \mathcal{V}(U)} d\mathbb{P}^{(n)}(U),$$

with  $\mathcal{Z}_n^{\mathcal{V}}$  the normalizing constant and where  $\mathcal{V}$  satisfies a convenient integrability assumption. The density of eigenvalues under  $\mathbb{P}_{\mathcal{V}}^{(n)}$  is then proportional to

$$(2.13) \quad |\Delta(\zeta_1, \dots, \zeta_n)|^2 \exp \left( -n \sum_{j=1}^n \mathcal{V}(\zeta_j) \right).$$

If  $\mathbb{Q}^{(n)}$  is the Haar measure on the additive group  $\mathbb{H}^{(n)}$  of Hermitian matrices defined by

$$d\mathbb{Q}^{(n)}(M) = \prod_{k=1}^n dM_{kk} \prod_{1 \leq k < l \leq n} d(\Re M_{kl}) \prod_{1 \leq k < l \leq n} d(\Im M_{kl}).$$

the pushforward of  $\mathbb{P}^{(n)}$  by  $\tau$  is the Cauchy ensemble whose density with respect to  $\mathbb{Q}^{(n)}$  is proportional to  $\det(\mathbf{I}_n + M^2)^{-n}$ . Let us compute the density of the (real) eigenvalues of  $M = \tau(U)$ , which are the pushforward of the eigenvalues of  $U$  by  $\tau$ , when  $\mathbb{U}(n)$  is equipped with  $\mathbb{P}_{\mathcal{V}}^{(n)}$ . From (2.8) we have, if  $\zeta_i \neq 1$  for  $i \leq n$ ,

$$|\Delta(\zeta_1, \dots, \zeta_n)| = 2^{n(n-1)/2} |\Delta(x_1, \dots, x_n)| \prod_{k=1}^n (1 + x_k^2)^{-(n-1)/2}.$$

and with (2.9) we conclude that the array of eigenvalues of  $M$  has a joint density proportional to

$$|\Delta(x_1, \dots, x_n)|^2 \exp\left(-n \sum_{j=1}^n V(x_j)\right),$$

with respect to the Lebesgue measure on  $\mathbb{R}^n$ , where the potentials  $V$  and  $\mathcal{V}$  are related by

$$(2.14) \quad V(x) = \mathcal{V}(\tau^{-1}(x)) + \log(1 + x^2).$$

The inverse relation is

$$\mathcal{V}(e^{i\theta}) = V(-\cotan \theta/2) + \log |\sin \theta/2|.$$

Of course, the same distribution of eigenvalues can be obtained by observing that the pushforward of (2.12) by  $\tau$  is

$$(2.15) \quad d\mathbb{P}_{\mathcal{V}}^{(n)}(M) = \frac{1}{Z_{\mathcal{V}}^n} e^{-n \operatorname{tr} \mathcal{V}(\tau^{-1}(M))} \det(\mathbf{I}_n + M^2)^{-n} d\mathbb{Q}^{(n)}(M).$$

Besides, it is known that in all these unitary invariant models, the matrix  $[\psi_1, \dots, \psi_n]$  of eigenvectors (defined up to multiplication of each vector by a phase) is Haar distributed on  $\mathbb{U}(n)$ . In particular, the array of weights  $(\mathbf{w}_1, \dots, \mathbf{w}_n)$  defined in (2.2) is uniformly distributed on the simplex  $\sum_1^n \mathbf{w}_k = 1$ .

Looking at the spectral measures, the above models can be generalized to log-gases. In this framework  $n$  is the number of particles (or eigenvalues), denoted by  $\lambda_1, \dots, \lambda_n$ , with the joint distributions  $\Pi_{\mathcal{V}}^{(n)}$  on  $\mathbb{T}^n$  having the density

$$(2.16) \quad \frac{d\Pi_{\mathcal{V}}^{(n)}(\lambda)}{d\lambda} = \frac{1}{Z_{\mathcal{V}}^n} e^{-n\beta' \sum_{k=1}^n \mathcal{V}(\lambda_k)} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^{\beta},$$

with respect to the Lebesgue measure  $d\lambda = d\lambda_1 \cdots d\lambda_n$ . Here  $\beta' = \beta/2$  and  $\beta > 0$  is a parameter interpreted as the inverse temperature. Then it is possible to consider the CMV matrices having these particles as eigenvalues and weights distributed according to the density proportional to

$$\prod_{k=1}^n w_k^{\beta'-1}$$

with respect to the uniform measure on the simplex (the Dirichlet distribution of parameter  $\beta'$ ). The correspondence (2.14) between the potentials is now more complicated: the pushforward by  $\tau$  gives the relation

$$(2.17) \quad V(x) = \mathcal{V}(\tau^{-1}(x)) + \left(1 - \frac{1}{n} + \frac{1}{\beta n}\right) \log(1 + x^2),$$

that is, either  $V$  or  $\mathcal{V}$  is  $n$ -dependent. It is possible to treat this general case, see Remark 3.5 in [19], but for simplicity, we consider here only the case  $\beta = 2$ .

## 2.3 Assumptions on the potentials

### 2.3.1 Real line

We will assume that the potentials  $V$  on  $\mathbb{R}$  are finite and continuous everywhere. The classical assumption on the growth of the potential is

(R1s) Strong growth:

$$(2.18) \quad \liminf_{|x| \rightarrow \infty} \frac{V(x)}{2 \log |x|} > 1.$$

Recently, Hardy [25] introduced the weaker assumption

(R1w) Weak growth:

$$(2.19) \quad \liminf_{|x| \rightarrow \infty} (V(x) - 2 \log |x|) > -\infty.$$

Under (R1w), the empirical distribution  $\mu_{\mathbf{u}}^{(n)}$  of eigenvalues  $\lambda_1, \dots, \lambda_n$  has a limit  $\mu_V$  (in probability), which is the unique minimizer of

$$(2.20) \quad \mu \mapsto \mathcal{E}_V(\mu) := \int_{\mathbb{R}} V(x) d\mu(x) - \iint_{\mathbb{R}^2} \log |x - y| d\mu(x) d\mu(y), \quad \mu \in \mathcal{M}_1(\mathbb{R}).$$

The minimal value is denoted as

$$F_V = \mathcal{E}_V(\mu_V).$$

Under (R1s), it is known that the support of  $\mu_V$  is compact. We will make in this case the additional assumption

(R2) One-cut regime: the support of  $\mu_V$  is a single compact interval  $[\alpha^-, \alpha^+]$  ( $\alpha^- < \alpha^+$ ).

The minimizer  $\mu_V$  is characterized by the Euler-Lagrange variational equations

$$(2.21) \quad \mathcal{J}_V(x) \begin{cases} = 2\xi_V & \text{if } x \in [\alpha^-, \alpha^+] \\ \geq 2\xi_V & \text{if } x \notin [\alpha^-, \alpha^+] \end{cases}$$

where  $\mathcal{J}_V$  is the effective potential

$$(2.22) \quad \mathcal{J}_V(x) := V(x) - 2 \int_{\mathbb{R}} \log |x - \xi| d\mu_V(\xi),$$

and  $\xi_V$  is the so-called modified Robin constant. We will make use of the following assumption

(R3) Control (of large deviations):

$\mathcal{J}_V$  achieves its global minimum value on the complement of  $(\alpha^-, \alpha^+)$  only on the boundary of this set.

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

(R4) Offcriticality:

$$d\mu_V(x) = \frac{1}{\pi} S(x) \sqrt{(\alpha^+ - x)(x - \alpha_-)} dx$$

where  $S > 0$  on  $[\alpha^-, \alpha^+]$ .

(R5) Analyticity:  $V$  can be extended as a holomorphic function in some open neighborhood of  $[\alpha^-, \alpha^+]$ .

We remark that for  $V$  strictly convex, the assumptions (R2), (R3) and (R4) are fulfilled (see [7] and [28]).

**Proposition 2.1** ([20] Prop. 3.2) *If the conditions (R1s), and (R2) to (R5) are satisfied, then*

$$(2.23) \quad \mathcal{J}_V(x) - \xi_V = \begin{cases} \int_{\alpha^+}^x S(t) \sqrt{(t - \alpha^-)(t - \alpha^+)} dt & \text{if } x > \alpha^+ \\ \int_x^{\alpha^-} S(t) \sqrt{(\alpha^- - t)(\alpha^+ - t)} dt & \text{if } x < \alpha^- \end{cases}$$

### 2.3.2 Unit circle

Let  $\varphi, \psi \in [0, 2\pi]$  be two angles with  $\varphi < \psi$ . We define  $\widehat{[\varphi, \psi]}$  to be the arc  $[e^{i\varphi}, e^{i\psi}] \subset \mathbb{T}$  where we go from  $e^{i\varphi}$  to  $e^{i\psi}$  in a counterclockwise direction. The potential  $\mathcal{V}$  is supposed to be continuous on  $\mathbb{T} \setminus \{1\}$ . We make the additional assumption:

(T1)  $\mathcal{V}$  is semicontinuous in  $1$ . Without loss of generality we may assume that

$$\mathcal{V}(0) = \liminf_{z \rightarrow 1} \mathcal{V}(z) \in (-\infty, \infty].$$

This implies that there is a unique minimizer  $\mu_{\mathcal{V}}$  of

$$(2.24) \quad \mu \mapsto \mathcal{E}_{\mathcal{V}}(\mu) = \int_{\mathbb{T}} \mathcal{V}(z) d\mu(z) - \iint_{\mathbb{T}^2} \log |z - \zeta| d\mu(z) d\mu(\zeta), \quad \mu \in \mathcal{M}_1(\mathbb{T}).$$

The minimal value is denoted by

$$(2.25) \quad F_{\mathcal{V}} = \mathcal{E}_{\mathcal{V}}(\mu_{\mathcal{V}}).$$

We will suppose that either the support of  $\mu_{\mathcal{V}}$  is  $\mathbb{T}$  or

(T2) One-cut regime: the support of  $\mu_{\mathcal{V}}$  is a single arc  $\widehat{[\alpha^-, \alpha^+]} \subset \widehat{(0, 2\pi)}$ .

In this case,  $\mu_{\mathcal{V}}$  is characterized by the Euler-Lagrange equations:

$$(2.26) \quad \mathcal{J}_{\mathcal{V}}(e^{i\theta}) \begin{cases} = 2\xi_{\mathcal{V}} & \text{if } \theta \in [\widehat{\alpha^-}, \widehat{\alpha^+}] \\ \geq 2\xi_{\mathcal{V}} & \text{if } \theta \notin [\widehat{\alpha^-}, \widehat{\alpha^+}], \end{cases}$$

where  $\mathcal{J}_{\mathcal{V}}$  is the effective potential

$$(2.27) \quad \mathcal{J}_{\mathcal{V}}(e^{i\theta}) := V(e^{i\theta}) - 2 \int_{\mathbb{R}} \log |e^{i\theta} - \zeta| d\mu_{\mathcal{V}}(\zeta),$$

and  $\xi_{\mathcal{V}}$  is the modified Robin constant. Like in the case of the real line, we make the assumption

(T3) Control (of large deviations):

$\mathcal{J}_{\mathcal{V}}$  achieves its global minimum value on the complement of  $[\widehat{\alpha^-}, \widehat{\alpha^+}]$  only on the boundary of this set.

When  $\theta \mapsto v(\theta) := \mathcal{V}(e^{i\theta})$  is convex, this condition is satisfied. Indeed, it is

$$(2.28) \quad \int_{\mathbb{T}} \log |e^{i\theta} - \zeta| d\mu_{\mathcal{V}}(\zeta) = \int_{e^{i\alpha^-}}^{e^{i\alpha^+}} \log \left| \sin \frac{\theta - \varphi}{2} \right| d\mu_{\mathcal{V}}(e^{i\varphi}) + \log 2$$

so that, for  $0 < \theta < \alpha^-$ , the function  $\theta \mapsto \mathcal{J}_{\mathcal{V}}(e^{i\theta})$  is strictly convex, nonnegative and vanishes for  $\theta = \alpha^-$ , hence is positive on  $[0, \theta^-)$ . An analogous argument can be made (mutatis mutandis) for  $\alpha^+ < \theta < 2\pi$ .

The additional assumptions to obtain a non-variational expression for the rate are on the unit circle:

(T4) Offcriticality:

$$d\mu_{\mathcal{V}}(z) = \frac{1}{\pi} S(e^{i\theta}) \sqrt{|(e^{i\theta} - e^{i\alpha^-})(e^{i\theta} - e^{i\alpha^+})|} d\theta$$

where  $S(e^{i\theta}) > 0$  for  $\theta \in [\alpha^-, \alpha^+]$ .

(T5) Analyticity:  $\mathcal{V}$  can be extended as a holomorphic function in some open neighbourhood in  $\mathbb{C}$  of the arc  $[\widehat{\alpha^-}, \widehat{\alpha^+}]$ .

**Remark 2.2** Assumption (T1) is equivalent via (2.6) and (2.17) to Hardy's assumption (2.19).

**Proposition 2.3** If  $\mathcal{V}$  satisfies assumptions (T1) to (T5), then

$$(2.29) \quad \mathcal{J}_{\mathcal{V}}(e^{i\theta}) - \xi_{\mathcal{V}} = \begin{cases} \int_{\theta}^{\alpha^-} S(e^{i\tau}) \sqrt{|(e^{i\tau} - e^{i\alpha^-})(e^{i\tau} - e^{i\alpha^+})|} d\tau & \text{if } \theta \in (0, \alpha^-] \\ \int_{\alpha^+}^{\theta} S(e^{i\tau}) \sqrt{|(e^{i\tau} - e^{i\alpha^-})(e^{i\tau} - e^{i\alpha^+})|} d\tau & \text{if } \theta \in [\alpha^+, 2\pi) \end{cases}$$

## 2.4 Large deviations

### 2.4.1 Introduction

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 [14] or to the Appendix D of [1].

Let  $U$  be a topological Hausdorff space with Borel  $\sigma$ -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  $\mathcal{I} : U \rightarrow [0, \infty]$  if:

- (i)  $\mathcal{I}$  is lower semicontinuous.
- (ii) For all closed sets  $F \subset U$ : 
$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \log P_n(F) \leq - \inf_{x \in F} \mathcal{I}(x)$$
- (iii) For all open sets  $O \subset U$ : 
$$\liminf_{n \rightarrow \infty} \frac{1}{a_n} \log P_n(O) \geq - \inf_{x \in O} \mathcal{I}(x)$$

The rate function  $\mathcal{I}$  is good if its level sets  $\{x \in U \mid \mathcal{I}(x) \leq a\}$  are compact for all  $a \geq 0$ . 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 tightness is fulfilled: For every  $M > 0$  there exists a compact set  $K_M \subset U$  such that

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

In our case, the measures  $P_n$  will be the distributions of the random spectral measures  $\mu_n$  and we will say that the sequence of measures  $\mu_n$  satisfies a LDP. All LDPs for spectral measures in this section are in the weak topology.

### 2.4.2 LDP for ESD

The most famous LDP in random matrix theory is for the sequence of empirical spectral measures (ESD) as defined in (2.3). The improved version (in the case  $\beta = 2$ ) is

**Theorem 2.4 (Hardy [25] Thm. 1.1)** *If the potential  $V$  in  $\mathbb{R}$  satisfies assumption (R1w), then under  $\mathbb{P}_V^{(n)}$ , the sequence of random probability measures  $(\mu_n^{(n)})$  satisfies in  $\mathcal{M}_1(\mathbb{R})$  a LDP with speed  $n^2$  and good rate function*

$$I_V(\mu) := \mathcal{E}_V(\mu) - F_V$$

where  $\mathcal{E}_V$  is defined in (2.20).

An equivalent statement may be claimed via the Cayley transform (see also Remark 2.4 in [25]).

**Theorem 2.5** *If the potential  $\mathcal{V}$  satisfies assumption (T1), then under  $\mathbb{P}_{\mathcal{V}}^{(n)}$  (see (2.12)), the sequence of random probability measures  $(\mu_{\mathfrak{a}}^{(n)})$  satisfies in  $\mathcal{M}_1(\mathbb{T})$  a LDP with speed  $n^2$  and good rate function*

$$I_{\mathcal{V}}(\mu) := \mathcal{E}_{\mathcal{V}}(\mu) - F_{\mathcal{V}}$$

where  $\mathcal{E}_{\mathcal{V}}$  is defined in (2.24).

## 3 Our two main examples of matrix ensembles

### 3.1 Gross-Witten ensemble

The following model is important in the analysis of problems involving random permutations [2]. Let us consider the Gross-Witten measure on  $\mathbb{U}(n)$ , absolutely continuous with respect to the Haar measure  $\mathbb{P}^{(n)}$  with density:

$$(3.1) \quad \frac{d\text{GW}_{\mathfrak{g}}^{(n)}}{d\mathbb{P}^{(n)}}(U) := \frac{1}{\mathcal{Z}_n(\mathfrak{g})} \exp \left[ \frac{n\mathfrak{g}}{2} \text{tr}(U + U^\dagger) \right],$$

where  $\mathfrak{g} \in \mathbb{R}$  and  $\mathcal{Z}_n(\mathfrak{g})$  is the normalizing constant and we denote by  $U^\dagger$  the adjoint of  $U$ . For details and applications of this distribution we refer to [26] p.203, [24], [42]. The potential is

$$(3.2) \quad \mathcal{V}_{\mathfrak{g}}(z) = -\mathfrak{g}\Re(z).$$

If  $|\mathfrak{g}| \leq 1$  (ungapped or strongly coupled phase), the equilibrium measure is supported by  $\mathbb{T}$  and is

$$(3.3) \quad \text{GW}_{\mathfrak{g}}(dz) = \frac{1}{2\pi} (1 + \mathfrak{g} \cos \theta) d\theta$$

for  $z = e^{i\theta}$ ,  $\theta \in [-\pi, \pi]$ . Moreover, the values for  $F_{\mathfrak{g}}^{GW} = F_{\mathcal{V}_{\mathfrak{g}}}$  and  $\xi_{\mathfrak{g}}^{GW} = \xi_{\mathcal{V}_{\mathfrak{g}}}$  are given by

$$(3.4) \quad F_{\mathfrak{g}}^{GW} = \mathfrak{g}^2/2,$$

$$(3.5) \quad \xi_{\mathfrak{g}}^{GW} = \mathfrak{g}^2/4.$$

Let us recall from Simon [39], p. 86 that the equilibrium measure has Verblunsky coefficients

$$(3.6) \quad \alpha_n(\text{GW}_{\mathfrak{g}}) = -\frac{x_+ - x_-}{x_+^{n+2} - x_-^{n+2}}$$

where  $x_{\pm} = -\mathfrak{g}^{-1} \pm \sqrt{\mathfrak{g}^{-2} - 1}$  are roots of the equation

$$x + \frac{1}{x} = -\frac{2}{\mathfrak{g}},$$

when  $|\mathbf{g}| < 1$  and  $\alpha_n(\text{GW}_{\mathbf{g}}) = (-\mathbf{g})^{n+1}/(n+2)$  if  $|\mathbf{g}| = 1$ . We remark that these distributions have only nontrivial moments of order  $\pm 1$ .

For  $|\mathbf{g}| > 1$  (gapped or weakly coupled phase), let  $\theta_{\mathbf{g}} \in [0, \pi]$  be such that

$$(3.7) \quad \sin^2\left(\frac{\theta_{\mathbf{g}}}{2}\right) = |\mathbf{g}|^{-1}.$$

For  $\mathbf{g} > 1$ , the equilibrium measure is

$$(3.8) \quad \text{GW}_{\mathbf{g}}(dz) = \frac{\mathbf{g}}{\pi} \cos\left(\frac{\theta}{2}\right) \sqrt{\sin^2\left(\frac{\theta_{\mathbf{g}}}{2}\right) - \sin^2\left(\frac{\theta}{2}\right)} 1_{[-\theta_{\mathbf{g}}, \theta_{\mathbf{g}}]} d\theta,$$

for  $z = e^{i\theta}$ ,  $\theta \in [-\pi, \pi]$ . Moreover, the free energy and the modified Robin constant are in the gapped case

$$(3.9) \quad F_{\mathbf{g}}^{GW} = -\mathbf{g} + \frac{1}{2} \log(\mathbf{g}) + \frac{3}{4},$$

$$(3.10) \quad \xi_{\mathbf{g}}^{GW} = \frac{1}{2}(\log(\mathbf{g}) - \mathbf{g} + 1).$$

The result (3.9) is known since [24] and (3.10) is formula (4.14) in [2].

When  $\mathbf{g} < -1$ , the equilibrium measure is

$$(3.11) \quad \text{GW}_{\mathbf{g}}(dz) = \frac{|\mathbf{g}|}{\pi} \sin\left(\frac{\theta}{2}\right) \sqrt{\sin^2\left(\frac{\theta}{2}\right) - \cos^2\left(\frac{\theta_{\mathbf{g}}}{2}\right)} 1_{[\pi - \theta_{\mathbf{g}}, \pi + \theta_{\mathbf{g}}]} d\theta,$$

for  $z = e^{i\theta}$ ,  $\theta \in [0, 2\pi]$  and  $\theta_{\mathbf{g}}$  is the same as before. It is the same to say that the support of  $\text{GW}_{\mathbf{g}}$  is  $[\pi - \theta_{\mathbf{g}}, \pi + \theta_{\mathbf{g}}]$ .

Let  $\widetilde{\text{GW}}_{-\mathbf{g}}^{(n)}$  be the probability measures on  $\mathbb{H}(n)$  obtained by pushing forward  $\text{GW}_{-\mathbf{g}}^{(n)}$  by  $\tau$ . We get

$$\frac{d\widetilde{\text{GW}}_{-\mathbf{g}}^{(n)}}{d\mathbb{Q}^{(n)}}(H) := \frac{1}{\widetilde{\mathcal{Z}}_n(\mathbf{g})} \exp\left[n\mathbf{g} \operatorname{tr} \frac{I - H^2}{I + H^2}\right] [\det(I + H^2)]^{-n}.$$

The potential is

$$(3.12) \quad V_{-\mathbf{g}}(x) = \mathbf{g} \frac{x^2 - 1}{x^2 + 1} + \log(1 + x^2).$$

For  $0 \leq \mathbf{g} \leq 1$  the equilibrium measure (supported by  $(-\infty, \infty)$ ) is

$$(3.13) \quad \widetilde{\text{GW}}_{-\mathbf{g}}(dx) = \frac{(1 - \mathbf{g})x^2 + 1 + \mathbf{g}}{\pi(x^2 + 1)^2} dx$$

(for  $\mathbf{g} = 0$  it is the Cauchy distribution). For  $\mathbf{g} > 1$ , the equilibrium measure has a compact support:

$$(3.14) \quad \widetilde{\text{GW}}_{-\mathbf{g}}(dx) = \frac{2\sqrt{1 + \mathbf{m}^2} \sqrt{\mathbf{m}^2 - x^2}}{\pi \mathbf{m}^2 (1 + x^2)^2} 1_{[-\mathbf{m}, \mathbf{m}]}(x) dx,$$

where  $\mathbf{m}^2 = (\mathbf{g} - 1)^{-1}$ .

### 3.2 Hua-Pickrell ensemble

The following distribution was introduced by [27] and reconsidered by [35]. We also refer to [33], who introduced a complex parameter and for further analysis to [6] and [9]. The Hua-Pickrell ensemble is defined by the density

$$(3.15) \quad \frac{d\mathbb{H}\mathbb{P}_\delta^{(n)}}{d\mathbb{P}^{(n)}}(U) := \frac{1}{\mathcal{Z}_n(\delta)} [\det(\mathbb{I}_n - U)]^{\bar{\delta}} [\det(\mathbb{I}_n - \bar{U})]^\delta$$

with respect to the Haar measure on  $\mathbb{U}^{(n)}$ , where  $\delta$  is a complex parameter such that  $\Re\delta > -1/2$ . Let  $\widetilde{\mathbb{H}\mathbb{P}}_\delta^{(n)}$  denote the probability measure on  $\mathbb{H}(n)$  obtained by pushing forward  $\mathbb{H}\mathbb{P}_\delta^{(n)}$  by  $\tau$ . We get

$$\frac{d\widetilde{\mathbb{H}\mathbb{P}}_\delta^{(n)}}{d\mathbb{Q}^{(n)}}(H) := \frac{1}{\widetilde{\mathcal{Z}}_n(\delta)} [\det(\mathbb{I}_n + H^2)]^{-n} [\det(\mathbb{I}_n + iH)]^{-\bar{\delta}} [\det(\mathbb{I}_n - iH)]^{-\delta}.$$

A particularly interesting case is the regime  $\delta = \mathfrak{d}n$ , which requires  $\Re\mathfrak{d} \geq 0$  for integrability. The case  $\mathfrak{d} = 0$  is the same as  $\mathfrak{g} = 0$  above and corresponds to the Cauchy ensemble. For simplicity of the computations we will consider here the symmetric case  $\mathfrak{d} > 0$ , although it is possible to treat the general case. In the framework laid out in Section 2.2, this corresponds to the potential

$$(3.16) \quad \mathcal{V}_\mathfrak{d}(z) = -2\mathfrak{d} \log |1 - z|.$$

The equilibrium measure is

$$(3.17) \quad \mathbb{H}\mathbb{P}_\mathfrak{d}(dz) = (1 + \mathfrak{d}) \frac{\sqrt{\sin^2(\frac{\theta}{2}) - \sin^2(\frac{\theta_\mathfrak{d}}{2})}}{2\pi \sin(\frac{\theta}{2})} 1_{(\theta_\mathfrak{d}, 2\pi - \theta_\mathfrak{d})}(\theta) d\theta$$

where  $z = e^{i\theta}$ ,  $\theta \in [0, 2\pi]$  and  $\theta_\mathfrak{d} \in (0, \pi)$  is such that

$$(3.18) \quad \sin(\frac{\theta_\mathfrak{d}}{2}) = \frac{\mathfrak{d}}{1 + \mathfrak{d}}.$$

The support of the equilibrium measure is thus the arc  $[\theta_\mathfrak{d}, \widehat{2\pi - \theta_\mathfrak{d}}]$ . Moreover, we have for the free energy and the modified Robin constant

$$(3.19) \quad F_\mathfrak{d}^{HP} = (\mathfrak{d} + 1)^2 \log(\mathfrak{d} + 1) + \mathfrak{d}^2 \log \mathfrak{d} - \frac{1}{2}(1 + 2\mathfrak{d})^2 \log(1 + 2\mathfrak{d}) + 2\mathfrak{d}^2 \log 2,$$

$$(3.20) \quad \xi_\mathfrak{d}^{HP} = (\mathfrak{d} + 1) \log(\mathfrak{d} + 1) - \frac{1 + 2\mathfrak{d}}{2} \log(1 + 2\mathfrak{d}).$$

The orthogonal polynomials are the Geronimus polynomials with constant Verblunsky coefficients

$$(3.21) \quad \alpha_k = -\frac{\mathfrak{d}}{1 + \mathfrak{d}}, \quad k \geq 0.$$

Pushing forward this measure on the set  $\mathbb{H}(n)$  of  $n \times n$  Hermitian matrices, we get the potential

$$(3.22) \quad V_{\mathbf{d}}(x) = (\mathbf{d} + 1) \log(1 + x^2)$$

This model is sometimes called the modified Cauchy ensemble, see [18], [32], [31], [34] Problem 11.4.15, or the Lorentzian ensemble [10]. The equilibrium measure on the real line is

$$(3.23) \quad \widetilde{\text{HP}}_{\mathbf{d}}(dx) = \frac{1}{\pi(\sqrt{1 + \mathbf{p}^2} - 1)} \frac{\sqrt{\mathbf{p}^2 - x^2}}{1 + x^2} 1_{[-\mathbf{p}, \mathbf{p}]}(x) dx,$$

where  $\mathbf{p}^2 = (1 + 2\mathbf{d})\mathbf{d}^{-2}$  (see [5] Prop. 11.2.2, p. 359). Moreover

$$(3.24) \quad \begin{aligned} \widetilde{F}_{\mathbf{d}}^{HP} &= (\mathbf{d} + 1)^2 \log(\mathbf{d} + 1) + \mathbf{d}^2 \log \mathbf{d} \\ &\quad - \frac{1}{2}(1 + 2\mathbf{d})^2 \log(1 + 2\mathbf{d}) + (2\mathbf{d}^2 - 1) \log 2, \end{aligned}$$

$$(3.25) \quad \widetilde{\xi}_{\mathbf{d}}^{HP} = \left( \mathbf{d} + \frac{1}{2} \right) \log(2\mathbf{d} + 1) - \mathbf{d} \log \mathbf{d} - (2\mathbf{d} + 1) \log 2.$$

**Remark 3.1** *The corresponding Jacobi coefficients of the tridiagonal representation are*

$$(3.26) \quad a_1 = \sqrt{\frac{2(1 + 2\mathbf{d})}{(1 + \mathbf{d})^3}}, \quad a_k = \frac{1 + 2\mathbf{d}}{(1 + \mathbf{d})^2} \quad (k > 1),$$

$$(3.27) \quad b_1 = -\frac{2\mathbf{d}}{1 + \mathbf{d}}, \quad b_k = -2\frac{\mathbf{d}^2}{(1 + \mathbf{d})^2}.$$

*We didn't find easily the corresponding values in the literature.*

As an application of Theorem 2.5 and Theorem 2.4, we have the following theorem, collecting all the LDP's for the empirical spectral measure as in (2.3) in our basic models.

### Theorem 3.2

1. For any  $\mathbf{g} \in \mathbb{R}$ , the sequence of distributions of  $(\mu_{\mathbf{u}}^{(n)})$  under  $\mathbb{G}\mathbb{W}_{\mathbf{g}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$ , with speed  $n^2$  and good rate function  $I_{\mathcal{V}}$  with  $\mathcal{V} = \mathcal{V}_{\mathbf{g}}$  given by (3.2).
2. For any  $\mathbf{d} > 0$ , the sequence of distributions of  $(\mu_{\mathbf{u}}^{(n)})$  under  $\mathbb{H}\mathbb{P}_{\mathbf{d}\mathbf{n}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$ , with speed  $n^2$  and good rate function  $I_{\mathcal{V}}$  with  $\mathcal{V} = \mathcal{V}_{\mathbf{d}}$  given by (3.16).
3. For any  $\mathbf{g} \in \mathbb{R}$ , the sequence of distributions of  $(\mu_{\mathbf{u}}^{(n)})$  under  $\widetilde{\mathbb{G}\mathbb{W}}_{\mathbf{g}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{R})$ , with speed  $n^2$  and good rate function  $I_{\mathcal{V}}$  with  $\mathcal{V} = \mathcal{V}_{-\mathbf{g}}$  given by (3.12).
4. For any  $\mathbf{d} > 0$ , the sequence of distributions of  $(\mu_{\mathbf{u}}^{(n)})$  under  $\widetilde{\mathbb{H}\mathbb{P}}_{\mathbf{d}\mathbf{n}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{R})$ , with speed  $n^2$  and good rate function  $I_{\mathcal{V}}$  with  $\mathcal{V} = \mathcal{V}_{\mathbf{d}}$  given by (3.22).

Point 1. is in [26] p.225 and point 2. is in [9] Theorem 5.5. The points 3. and 4. are obtained carrying the results to the real line by the Cayley transform.

## 4 LDP for spectral measures

### 4.1 Measure encoding approach

In this subsection, we state LDPs for the weighted measures given in (2.2). They are element of  $\mathcal{M}_1(\mathbb{T})$ . We first recall the main theorem of [20] on  $\mathbb{R}$ , then we state the LDP on  $\mathbb{T}$  which allows to improve the result on  $\mathbb{R}$  with weaker assumptions. First we recall the definition of the Kullback-Leibler divergence, with a slight generalization for sub-probabilities.

Let  $\mu$  be a probability measure and  $\nu$  be a non-zero sub-probability measures on some measurable space. The Kullback-Leibler divergence between  $\mu$  and  $\nu$  is given by

$$(4.1) \quad \mathcal{K}(\mu|\nu) = \int \log \left( \frac{d\mu}{d(\nu)} \right) d\mu$$

if  $\mu$  is absolutely continuous with respect to  $\nu$  and  $\mathcal{K}(\mu|\nu) = \infty$  otherwise. In our LDP, the rate function will involve the reversed Kullback-Leibler distance, where  $\mu$  will be the reference measure and  $\nu$  is the argument.

Recall the definition of the set  $\mathcal{S}_1^{\mathbb{R}}(\alpha^-, \alpha^+)$  in the introduction, consisting of probability measures

$$(4.2) \quad \mu = \mu|_I + \sum_{i=1}^{N^+} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \gamma_i^- \delta_{\lambda_i^-}.$$

In our extension of the Killip-Simon sum rule we will also consider reference measures supported by the whole real line. To keep a consistent notation, we write  $\mathcal{S}_1^{\mathbb{R}}(-\infty, \infty)$  for the set of probability measure with support  $\mathbb{R}$ . In this case,  $N^+ = N^- = 0$ . In the same vein, we define  $\mathcal{S}_{\leq 1}^{\mathbb{R}}(\alpha^-, \alpha^+)$  for subprobabilities. Notice that this last set may be seen as  $\mathcal{S}_1^{\mathbb{R}}(\alpha^-, \alpha^+)$ .

We now introduce the analogue on the circle. If  $[\alpha^-, \alpha^+]$  is an interval in  $(0, 2\pi)$ , let  $I = [\widehat{\alpha^-}, \widehat{\alpha^+}]$  and let  $\mathcal{S}_1^{\mathbb{T}} = \mathcal{S}_1^{\mathbb{T}}(\alpha^-, \alpha^+)$  be the set of all probability measures  $\mu$  on  $\mathbb{T}$  with

- (i)  $\text{supp}(\mu) = J \cup \{e^{i\theta_i^-}\}_{i=1}^{N^-} \cup \{e^{i\theta_i^+}\}_{i=1}^{N^+}$ , where  $J \subset I$ ,  $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$  and  $\theta_i^\pm \in [0, 2\pi)$ .  
Furthermore,

$$0 \leq \theta_1^- < \theta_2^- < \dots < \alpha^- \quad \text{and} \quad \theta_1^+ > \theta_2^+ > \dots > \alpha^+.$$

- (ii) If  $N^-$  (resp.  $N^+$ ) is infinite, then  $\theta_j^-$  converges towards  $\alpha^-$  (resp.  $\theta_j^+$  converges to  $\alpha^+$ ).

We will also write  $\lambda_i^\pm = e^{i\theta_i^\pm}$  as in the real case for the outlying support points. For a measure  $\mu \in \mathcal{S}_1^{\mathbb{T}}(\alpha^-, \alpha^+)$  we have then the analogous way of writing it as in (4.2). Similar to the real case, we write  $\mathcal{S}_1^{\mathbb{T}}(0, 2\pi)$  for the probability measures supported by  $\mathbb{T}$ . It should be clear that the Cayley transform carries  $\{\mu \in \mathcal{S}_1^{\mathbb{T}} | \mu(1) = 0\}$  onto  $\mathcal{S}_1^{\mathbb{R}}$  and  $\mathcal{S}_1^{\mathbb{T}}$  onto  $\mathcal{S}_{\leq 1}^{\mathbb{R}}$ . We remark that since the circle is rotationally invariant, separating between  $\theta_i^+$  and  $\theta_i^-$  at 1 is essentially arbitrary, but it is consistent with this mapping of measures. We endow the sets  $\mathcal{S}_1^{\mathbb{T}}$  and  $\mathcal{S}_1^{\mathbb{R}}$  with the weak topology and  $\mathcal{S}_{\leq 1}^{\mathbb{T}}$  and  $\mathcal{S}_{\leq 1}^{\mathbb{R}}$  with the vague topology and the corresponding Borel  $\sigma$ -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,

$$(4.3) \quad \mathcal{F}_V^+(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi \in \mathbb{R}} \mathcal{J}_V(\xi) & \text{if } x \geq \alpha^+, \\ \infty & \text{otherwise,} \end{cases}$$

$$(4.4) \quad \mathcal{F}_V^-(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi \in \mathbb{R}} \mathcal{J}_V(\xi) & \text{if } x \leq \alpha^-, \\ \infty & \text{otherwise.} \end{cases}$$

On the unit circle, we have similar notations, with  $V$  replaced by  $\mathcal{V}$ . Notice that if  $\mathcal{V}(\mathbf{1}) < \infty$ , then  $\mathcal{F}_{\mathcal{V}}(\mathbf{1}) < \infty$ . In this case let us denote

$$(4.5) \quad \kappa_{\mathcal{V}} = \mathcal{F}_{\mathcal{V}}(\mathbf{1}).$$

**Theorem 4.1 ([20] Thm. 3.1)** *Assume that the potential  $V$  satisfies the assumptions (R1s), (R2) and (R3). Then the sequence of spectral measures  $\mu^{(n)}$  under  $\mathbb{P}_V^{(n)}$  satisfies the LDP with speed  $n$  and good 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  $\mu \in \mathcal{S}_1^{\mathbb{R}}(\alpha^-, \alpha^+)$  and  $\mathcal{I}_V(\mu) = \infty$  otherwise.

On the unit circle, we claim

**Theorem 4.2**

1. *Assume that the potential satisfies (T1) and that the support of  $\mu_{\mathcal{V}}$  is  $\mathbb{T}$ . Then the sequence of spectral measures  $\mu^{(n)}$  under  $\mathbb{P}_{\mathcal{V}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and good rate function*

$$(4.6) \quad \mathcal{I}_{\mathcal{V}}(\mu) = \mathcal{K}(\mu_{\mathcal{V}} | \mu).$$

2. *Assume that the potential  $\mathcal{V}$  satisfies the assumptions (T1), (T2) and (T3). Then the sequence of spectral measures  $\mu^{(n)}$  under  $\mathbb{P}_{\mathcal{V}}^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and good rate function*

$$(4.7) \quad \mathcal{I}_{\mathcal{V}}(\mu) = \mathcal{K}(\mu_{\mathcal{V}} | \mu) + \sum_{n=1}^{N^+} \mathcal{F}_{\mathcal{V}}^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_{\mathcal{V}}^-(\lambda_n^-)$$

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

To transfer the LDP in Theorem 4.2 to the real line we use the mapping  $\hat{\tau}$  given in (2.10). We get the following proposition.

**Proposition 4.3**

1. Assume that the potential  $V$  satisfies the assumption (R1w) and that the support of  $\mu_V$  is  $\mathbb{R}$ . Then the sequence of spectral measures  $\mu^{(n)}$  under  $\mathbb{P}_V^{(n)}$  satisfies the LDP in  $\mathcal{M}_{\leq 1}(\mathbb{R})$  with speed  $n$  and rate function

$$(4.8) \quad \mathcal{I}_V(\mu) = \mathcal{K}(\mu_V | \mu)$$

2. Assume that the potential  $V$  satisfies the assumptions (R1w), (R2) and (R3). Then the sequence of spectral measures  $\mu^{(n)}$  under  $\mathbb{P}_V^{(n)}$  satisfies the LDP in  $\mathcal{M}_{\leq 1}(\mathbb{R})$  with speed  $n$  and rate function

$$\mathcal{I}_V(\mu) = \mathcal{K}(\mu_V | \mu) + \sum_{n=1}^{N^+} \mathcal{F}_V^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_V^-(\lambda_n^-) + \kappa_V \mathbf{1}_{\mu(\mathbb{R}) < 1}$$

if  $\mu \in \mathcal{S}_{\leq 1}^{\mathbb{R}}(\alpha^-, \alpha^+)$  and  $\mathcal{I}_V(\mu) = \infty$  otherwise.

**Proof:** We only prove the second point, since the other one is simpler. From the definitions,  $\mathbf{1}$  is  $\mathbb{P}_V^{(n)}$  almost surely not an eigenvalue, so that we may consider the two random measures

$$\nu^{(n)} = \sum_{k=1}^n \mathbf{w}_k \delta_{\zeta_k} \quad \text{and} \quad \mu^{(n)} = \hat{\tau}(\nu^{(n)}) = \sum_{k=1}^n \mathbf{w}_k \delta_{\tau(\zeta_k)}$$

The mapping  $\hat{\tau}$  is continuous, and  $\mathcal{I}_V$  is good. We may apply the contraction principle (Theorem 4.2.1 in [14]). We obtain a LDP in  $\mathcal{M}_{\leq 1}(\mathbb{R})$  with good rate function

$$\widehat{\mathcal{I}}(\mu) = \inf_{\nu: \hat{\tau}(\nu) = \mu} \mathcal{I}_V(\nu).$$

Actually only those  $\nu$  such that  $\mathcal{I}_V(\nu)$  is finite contribute to the infimum. But then  $\nu \in \mathcal{S}_1^{\mathbb{T}}(\alpha_V^-, \alpha_V^+)$  implies  $\hat{\tau}(\nu) \in \mathcal{S}_{\leq 1}^{\mathbb{R}}(\alpha_V^-, \alpha_V^+)$  with  $\alpha_V^{\pm} = \tau(\alpha_V^{\pm})$ . Under our assumptions,  $\mu_V$  has no atom at  $\mathbf{1}$  and  $\mu_V = \hat{\tau}(\mu_V)$ . For a  $\nu$  such as above, we have by pushforward by  $\hat{\tau}$

$$\mathcal{K}(\mu_V | \nu) = \mathcal{K}(\hat{\tau}(\mu_V) | \hat{\tau}(\nu)) = \mathcal{K}(\mu_V | \mu).$$

Moreover, the outliers of  $\nu$  different from  $\mathbf{1}$  are carried upon outliers of  $\mu$ , and  $\mathcal{F}_V(\zeta) = \mathcal{F}_V(\tau(\zeta))$ . Now, assume that  $\nu$  has an outlier at  $\mathbf{1}$ , say  $\nu = \nu_0 + a\delta_{\mathbf{1}}$  then  $\mu(\mathbb{R}) = \nu_0(\mathbb{T}) = 1 - a$ , and the contribution of  $\mathbf{1}$  in  $\mathcal{I}_V(\nu)$  hence in  $\widehat{\mathcal{I}}(\mu)$  is  $\kappa_V$ . This proves that  $\widehat{\mathcal{I}} = \mathcal{I}_V$  and ends the proof of the corollary.  $\square$

As a consequence, we have for our models the following results.

**Proposition 4.4**

1. Under  $\mathbb{HIP}_{nd}^{(n)}$ , the sequence of spectral measures  $\mu^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and rate function  $\mathcal{I}_V$  where  $\mu_V = \text{HP}_a$  is given in (3.17) and  $\mathcal{F}_V^\pm = \mathcal{F}_{HP}^\pm$ , where for  $0 < \theta \leq \theta_a$

$$(4.9) \quad \mathcal{F}_{HP}^-(e^{i\theta}) := \int_{\theta}^{\theta_a} (1 + d) \frac{\sqrt{\sin^2(\theta_a/2) - \sin^2(\varphi/2)}}{2 \sin(\varphi/2)} d\varphi$$

and for  $\theta \in (2\pi - \theta_a, 2\pi)$ ,  $\mathcal{F}_{HP}^+(e^{i\theta}) := \mathcal{F}_{HP}^-(e^{-i\theta})$ .

2. Under  $\widetilde{\mathbb{HIP}}_{nd}^{(n)}$ , the sequence of spectral measures  $\mu^{(n)}$  satisfies the LDP with speed  $n$  and rate function  $\mathcal{I}_V$ , where  $\mu_V = \widetilde{\text{HP}}_a$  is given by (3.23) and  $\mathcal{F}_V^\pm = \mathcal{F}_{HP}^\pm$ , where for  $x \geq p$

$$(4.10) \quad \mathcal{F}_{HP}^+(x) = \int_p^x \frac{2}{\sqrt{1+p^2}-1} \frac{\sqrt{\xi^2-p^2}}{1+\xi^2} d\xi,$$

and  $\mathcal{F}_{HP}^-(x) = \mathcal{F}_{HP}^+(-x)$  for  $x \leq -p$ .

### Proposition 4.5

1. Under  $\mathbb{GW}_g^{(n)}$ ,  $g \leq 1$ , the sequence of spectral measures  $\mu^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and rate function  $\mathcal{I}_V$  where  $\mu_V = \text{GW}_g$  is given in (3.3) and (3.11).  
If  $|g| \leq 1$  there are no outliers and the rate function reduces to

$$\mathcal{I}_V(\mu) = \mathcal{K}(\text{GW}_g | \mu).$$

If  $g < -1$ , we have  $\mathcal{F}_V^\pm = \mathcal{F}_{GW}^\pm$ , where for  $0 < \theta < \pi - \theta_g$

$$\mathcal{F}_{GW}^-(e^{i\theta}) = \int_{\theta}^{\pi - \theta_g} 2|g| \sin \frac{\varphi}{2} \sqrt{\cos^2 \frac{\theta_g}{2} - \sin^2 \frac{\varphi}{2}} d\varphi = 4 \int_1^{\sqrt{|g| \cos \frac{\theta}{2}}} \sqrt{u^2 - 1} du,$$

and  $\mathcal{F}_{GW}^+(e^{i\theta}) = \mathcal{F}_{GW}^-(e^{-i\theta})$  if  $\pi + \theta_g < \theta < 2\pi$ .

2. Under  $\widetilde{\mathbb{GW}}_{-g}^{(n)}$ ,  $g \geq 0$ , the sequence of spectral measures  $\mu^{(n)}$  satisfies the LDP with speed  $n$  and good rate function  $\mathcal{I}_V$  with  $\mu_V = \widetilde{\text{GW}}_{-g}$  as in (3.13) and (3.14).  
If  $0 \leq g \leq 1$ , the support of  $\mu_V$  is  $\mathbb{R}$ , the LDP is in  $\mathcal{M}_{\leq 1}(\mathbb{R})$  and the rate function is

$$\mathcal{I}_V(\mu) = \mathcal{K}(\widetilde{\text{GW}}_{-g} | \mu).$$

If  $g > 1$ , the LDP is in  $\mathcal{M}_{\leq 1}(\mathbb{R})$ . We have  $\mathcal{F}_V^\pm = \mathcal{F}_{GW}^\pm$  where for  $x > m$

$$\mathcal{F}_{GW}^+(x) = \int_m^x \frac{4\sqrt{1+m^2}}{m^2} \frac{\sqrt{\xi^2-m^2}}{(1+\xi^2)^2} d\xi = 4 \int_m^{\frac{x|g|}{\sqrt{1+x^2}}} \sqrt{u^2 - 1} du,$$

and for  $x < -m$ ,  $\mathcal{F}_{GW}^-(x) = \mathcal{F}_{GW}^+(-x)$ .

## 4.2 Verblunsky coefficient encoding approach

To begin with, let us recall the simplest example. It is the Circular Unitary ensemble where  $\mathbb{U}(n)$  is equipped with the Haar measure. Then the Verblunsky coefficients are independent. More precisely, the  $n$ -uple  $\alpha^{(n)} := (\alpha_0, \dots, \alpha_{n-1} = e^{i\phi})$  has the distribution

$$(4.11) \quad dP_0^{(n)}(\alpha_0, \dots, \alpha_{n-1}) = \left( \otimes_{j=0}^{n-2} \eta_{n-j+1}(d\alpha_j) \right) \otimes \frac{d\phi}{2\pi}$$

([29]). From that, it is deduced in [21], Section 5.2 that the family of distributions of  $\mu^{(n)}$  under  $\text{CUE}(n)$  satisfies the LDP (in  $\mathcal{M}_1(\mathbb{T})$  equipped with the weak topology) with speed  $n$  and good rate function

$$I^0(\mu) = \sum_{j=0}^{\infty} -\log(1 - |\alpha_j|^2),$$

when  $\alpha_j, j \geq 0$  are the Verblunsky coefficients of  $\mu$ . In the Hua-Pickrell case, the Verblunsky coefficients are no more independent (except when  $\mathbf{d} = 0$ ). To recover a structure of independence, it is necessary to introduce the so-called *deformed Verblunsky coefficients*. Given a measure  $\mu \in \mathcal{M}_1(\mathbb{T})$  with at least  $n$  distinct support points and monic orthogonal polynomials  $\phi_0, \dots, \phi_{n-1}$ , define

$$(4.12) \quad b_k = \frac{\phi_k(1)}{\phi_k^*(1)} \quad \text{and} \quad \gamma_k = \bar{\alpha}_k(b_k)^{-1}, \quad k = 0, \dots, n-1.$$

This is equivalent to the recursive definition

$$(4.13) \quad \gamma_0 = \bar{\alpha}_0, \quad \gamma_k = \bar{\alpha}_k \prod_{j=0}^{k-1} \frac{1 - \bar{\gamma}_j}{1 - \gamma_j}, \quad k = 1, \dots, n-1.$$

For a more detailed description and meaning of these quantities we refer to [9], Section 2.2. In Theorem 3.2 therein, it is proved that under  $\mathbb{HP}_\delta^{(n)}$ , the random variables  $\gamma_0^{(n)}, \dots, \gamma_{n-1}^{(n)}$  are independent and for  $k = 0, \dots, n-2$ , the density of  $\gamma_k^{(n)}$  on  $\mathbb{D}$  is

$$(4.14) \quad \frac{\Gamma(n-k+\delta)\Gamma(n-k+\bar{\delta})}{\pi\Gamma(n-k-1)\Gamma(n-k+\delta+\bar{\delta})} (1-|z|^2)^{n-k-2} (1-z)^{\bar{\delta}} (1-\bar{z})^\delta,$$

and  $\gamma_{n-1}^{(n)} \in \mathbb{T}$  has the density

$$(4.15) \quad \frac{\Gamma(1+\delta)\Gamma(1+\bar{\delta})}{\Gamma(1+\delta+\bar{\delta})} (1-\zeta)^{\bar{\delta}} (1-\bar{\zeta})^\delta$$

with respect to the Haar measure on  $\mathbb{T}$ .

When  $\delta = n\mathbf{d}$ ,  $\mathbf{d} \geq 0$ , a straightforward study of the density (4.14) leads to a LDP for  $\gamma_j^{(n)}$  for  $j$  fixed. It is a particular case of the matricial result, Proposition 6.6, whose proof is in Section 7.3.3.

**Lemma 4.6** For fixed  $k$ ,  $(\gamma_0^{(n)}, \gamma_1^{(n)}, \dots, \gamma_k^{(n)})_{n \geq k}$  satisfies under  $\mathbb{HIP}_{nd}^{(n)}$  the LDP in  $\bar{\mathbb{D}}^k$  with speed  $n$  and good rate function

$$I_k(\gamma_0, \dots, \gamma_k) = \sum_{j=0}^k H_d(\gamma_j),$$

where

$$(4.16) \quad H_d(\gamma) = -\log(1 - |\gamma|^2) - 2d \log |1 - \gamma| + H_d(0),$$

$$(4.17) \quad H_d(0) = (1 + 2d) \log(1 + 2d) - 2(1 + d) \log(1 + d).$$

Using the classical method of projective limits (see the proof in Section 8), this allows to claim

**Theorem 4.7** Under  $\mathbb{HIP}_{nd}^{(n)}$ , the sequence of measures  $\mu^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and good rate function

$$J_d^{HP}(\mu) = \sum_{k=0}^{\infty} H_d(\gamma_k).$$

In the Gross-Witten case, the Verblunsky coefficients are not independent (except when  $\mathbf{g} = 0$ ). More precisely, the joint distribution is given by the following lemma.

**Lemma 4.8** The law of  $(\alpha_0^{(n)}, \dots, \alpha_{N-1}^{(n)})$  under  $\mathbb{GW}_{\mathbf{g}}^{(n)}$  is given by

$$(4.18) \quad dP_{\mathbf{g}}^{(n)}(\alpha_0, \dots, \alpha_{n-1} = e^{i\phi}) = \mathcal{Z}_n(\mathbf{g})^{-1} \exp \left[ n\mathbf{g}\Re \left( \alpha_0 - \sum_{k=0}^{n-1} \alpha_k \bar{\alpha}_{k-1} \right) \right] \left( \otimes_{j=0}^{n-2} \eta_{n-j+1}(d\alpha_j) \right) \otimes \frac{d\phi}{2\pi}.$$

**Proof:** By definition, we have

$$\frac{d\mathbb{P}_{\mathbf{g}}}{d\mathbb{P}_0}(U) = \exp \left( \frac{n\mathbf{g}}{2} \text{tr}(U + U^\dagger) \right) = \exp(n\mathbf{g}\Re(\text{tr } U)).$$

Now, from the CMV representation (2.4), we get

$$(4.19) \quad \text{tr } U = -\alpha_0 + \sum_{k=0}^{N-1} \alpha_k \bar{\alpha}_{k-1},$$

(see also Simon [39] p. 273). It remains to use (4.11).  $\square$

Given the explicit density in Lemma 4.8, we may conjecture a LDP for the spectral measure in terms of its Verblunsky coefficients.

**Conjecture 4.9** Under  $\mathbb{GW}_{\mathbf{g}}^{(n)}$ , the sequence of measures  $\mu^{(n)}$  satisfies the LDP in  $\mathcal{M}_1(\mathbb{T})$  with speed  $n$  and rate function

$$J_{\mathbf{g}}^{GW}(\mu) = H(\mathbf{g}) - \mathbf{g}\Re \left( \alpha_0 - \sum_{k=1}^{\infty} \alpha_k \bar{\alpha}_{k-1} \right) - \sum_{k=0}^{\infty} \log(1 - |\alpha_k|^2).$$

## 5 Sum rules from large deviations

### 5.1 Hua-Pickrell case

Our new sum rule is a straightforward consequence of Theorem 4.7 and Theorem 4.2.

**Theorem 5.1** *Let  $\mu \in \mathcal{M}_1(\mathbb{T})$  with infinite support and let  $(\gamma_k)_{k \geq 0} \in \mathbb{D}^{\mathbb{N}}$  be the sequence of its deformed Verblunsky coefficients. Then for any  $\mathbf{d} \geq 0$ , we have  $\sum_{k=0}^{\infty} H_{\mathbf{d}}(\gamma_k) = \infty$  if  $\mu \notin \mathcal{S}_1^{\mathbb{T}}(\theta_{\mathbf{d}}, 2\pi - \theta_{\mathbf{d}})$ . If  $\mu \in \mathcal{S}_1^{\mathbb{T}}(\theta_{\mathbf{d}}, 2\pi - \theta_{\mathbf{d}})$ , we have*

$$(5.1) \quad \mathcal{K}(\text{HP}_{\mathbf{d}} | \mu) + \sum_{n=1}^{N^+} \mathcal{F}_{HP}^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_{HP}^-(\lambda_n^-) = \sum_{k=0}^{\infty} H_{\mathbf{d}}(\gamma_k),$$

where both sides may be infinite simultaneously.

### 5.2 Gross-Witten case

As we saw above, we do not have an easy structure of the density of Verblunsky coefficients and could not succeed in finding a LDP in this encoding. Nevertheless, Simon found a sum rule. Here the reference measure is supported by the full circle  $\mathbb{T}$  and there is no contribution of outliers.

**Theorem 5.2 ([39] Thm. 2.8.1)** *Let  $\mu$  be a probability measure on  $\mathbb{T}$  with Verblunsky coefficients  $(\alpha_k)_{k \geq 0} \in \mathbb{D}^{\mathbb{N}}$ . Then*

$$(5.2) \quad \mathcal{K}(\text{GW}_{-1} | \mu) = 1 - \log 2 + \Re(\alpha_0) + \frac{|\alpha_0|^2}{2} + \frac{1}{2} \sum_{k=1}^{\infty} |\alpha_k - \alpha_{k-1}|^2 + \sum_{k=0}^{\infty} h(\alpha_k),$$

where

$$h(\alpha) = -\log(1 - |\alpha|^2) - |\alpha|^2.$$

In particular,

$$(5.3) \quad \mathcal{K}(\text{GW}_{-1} | \mu) < \infty \iff \sum_{k=0}^{\infty} |\alpha_{k+1} - \alpha_k|^2 + |\alpha_k|^4 < \infty.$$

As an easy corollary, we have

**Corollary 5.3** *Let  $\mu$  be a probability measure on  $\mathbb{T}$  with Verblunsky coefficients  $(\alpha_k)_{k \geq 0} \in \mathbb{D}^{\mathbb{N}}$ . Then for  $0 \leq \mathbf{g} < 1$ , we have*

$$(5.4) \quad \mathcal{K}(\text{GW}_{-\mathbf{g}} | \mu) = H(\mathbf{g}) + \mathbf{g} \left( \Re(\alpha_0) + \frac{|\alpha_0|^2}{2} + \frac{1}{2} \sum_{k=1}^{\infty} |\alpha_k - \alpha_{k-1}|^2 \right) + \sum_{k=0}^{\infty} -\log(1 - |\alpha_k|^2) - \mathbf{g}|\alpha_k|^2,$$

where

$$H(\mathbf{g}) := 1 - \sqrt{1 - \mathbf{g}^2} + \log \frac{1 + \sqrt{1 - \mathbf{g}^2}}{2}.$$

In particular, we have

$$(5.5) \quad \mathcal{K}(\text{GW}_{-\mathbf{g}}|\mu) < \infty \iff \sum_{k=0}^{\infty} |\alpha_k|^2 < \infty.$$

**Remark 5.4** In the way to prove (5.2), Simon arrived at the equivalent expression:

$$(5.6) \quad \mathcal{K}(\text{GW}_{-1}|\mu) = 1 - \log 2 + \Re \left( \alpha_0 - \sum_{k=1}^{\infty} \alpha_k \bar{\alpha}_{k-1} \right) + \sum_{k=0}^{\infty} -\log(1 - |\alpha_k|^2).$$

In the same vein, (5.4) is equivalent to

$$(5.7) \quad \mathcal{K}(\text{GW}_{-\mathbf{g}}|\mu) = H(\mathbf{g}) + \mathbf{g} \Re \left( \alpha_0 - \sum_{k=1}^{\infty} \alpha_k \bar{\alpha}_{k-1} \right) + \sum_{k=0}^{\infty} -\log(1 - |\alpha_k|^2).$$

For  $|\mathbf{g}| > 1$ , we may still propose a sum rule. The left hand side of such an identity would be given by the rate function of the LDP for the spectral measure encoded by the eigenvalues and the weights. It will be a particular case of our general Theorem 4.2. The right hand side can be only conjectured. It is natural to state the following conjecture.

**Conjecture 5.5** Let  $\mu$  be a probability measure on  $\mathbb{T}$  with Verblunsky coefficients  $(\alpha_k)_{k \geq 0} \in \mathbb{D}^{\mathbb{N}}$ . Then for any  $\mathbf{g} < -1$  and  $\mu \in \mathcal{S}_1^{\mathbb{T}}(\pi - \theta_{\mathbf{g}}, \pi + \theta_{\mathbf{g}})$ ,

$$(5.8) \quad \mathcal{K}(\text{GW}_{\mathbf{g}}|\mu) + \sum_{n=1}^{N^-} \mathcal{F}_{\mathbf{g}}^+(\lambda_n^+) + \sum_{n=1}^{N^+} \mathcal{F}_{\mathbf{g}}^-(\lambda_n^-) = H(\mathbf{g}) - \mathbf{g} \Re \left( \alpha_0 - \sum_{k=1}^{\infty} \alpha_k \bar{\alpha}_{k-1} \right) - \sum_{k=0}^{\infty} \log(1 - |\alpha_k|^2).$$

If  $\mu \notin \mathcal{S}_1^{\mathbb{T}}(\pi - \theta_{\mathbf{g}}, \pi + \theta_{\mathbf{g}})$ , the right hand side equals  $+\infty$ .

This statement would be a direct consequence of Theorem 4.2 and of Conjecture 4.9, as soon as the latter is true.

## 6 Matrix extensions

In this section we show how several results can be extended to the case of operator values measures. Since the arguments necessary for the proofs are mostly identical to the scalar case or can be found in the companion paper [19], we omit most of the proofs.

## 6.1 Matrix spectral measure

A matrix measure  $\Sigma = (\Sigma_{i,j})_{i,j}$  of size  $p \times p$  on  $\mathbb{T}$  is a matrix of signed complex measures, such that  $\Sigma(A) = (\Sigma_{i,j}(A))_{i,j} \in \mathbb{H}(p)$  is Hermitian and non-negative definite for any Borel set  $A \subset \mathbb{T}$ . A matrix measure on  $\mathbb{T}$  is normalized, if  $\Sigma(\mathbb{T}) = I_n$ . We denote by  $\mathcal{M}_{p,1}(T)$  the set of normalized  $p \times p$  matrix measures with support in  $T \subset \mathbb{T}$ . Given a unitary operator  $U$  and a collection of vectors  $e_1, \dots, e_p$  cyclic for  $U$ , one can define the spectral matrix measure  $\Sigma$  of  $(U, e_1, \dots, e_p)$  similar to 2.1 by the relation

$$(6.1) \quad \langle e_i, U^k e_j \rangle = \int_{\mathbb{T}} z^k d(e_i^\dagger \Sigma e_j)(z), \quad k \in \mathbb{Z}.$$

In fact, if  $U \in \mathbb{U}(n)$  with eigenvalues  $\lambda_1 = e^{i\theta_1}, \dots, \lambda_n = e^{i\theta_n}$  and  $\psi_1, \dots, \psi_n$  a corresponding system of orthonormal eigenvectors, the spectral matrix measure is given by

$$(6.2) \quad \Sigma^{(n)} = \sum_{k=1}^n W_k \delta_{\lambda_k},$$

where  $W_k = \psi_k^{(p)} (\psi_k^{(p)})^\dagger$  and  $\psi_k^{(p)}$  is the projection of  $\psi_k$  onto the first  $p$  coordinates.

If  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$  is a *quasi scalar* measure, that is if  $\Sigma = I_p \sigma$  with  $\sigma \in \mathcal{M}_1(\mathbb{T})$  a scalar measure and if  $\Pi$  is a normalized matrix measure with Lebesgue decomposition

$$\Pi(dz) = h(z) \sigma(dz) + \Pi^s(dz),$$

we define

$$(6.3) \quad \mathcal{K}(\Sigma|\Pi) := - \int \log \det h(z) \sigma(dz).$$

Note that if we define a density

$$\frac{d\Sigma}{d\Pi} = \left( \frac{d\sigma}{d\Pi_{i,j}} \right)_{i,j}$$

componentwise, then it is possible to rewrite the above quantity in the flavour of the Kullback-Leibler information (or relative entropy)

$$\mathcal{K}(\Sigma|\Pi) = \int \log \det \frac{d\Sigma}{d\Pi}(z) d\sigma(z),$$

if the density  $\frac{d\Sigma}{d\Pi}$  exists and infinity otherwise (see [30] or [37]).

As in the scalar case, we can define matrix versions of Verblunsky coefficients, where now the correspondence is via matrix orthogonal polynomials on the unit circle (MOPUC).

## 6.2 MOPUC

We follow the notation of [23] and [12]. A  $p \times p$  matrix polynomials  $\mathbf{F}$  is a polynomial with coefficients in  $\mathbb{C}^{p \times p}$ . Given a measure  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$ , we define two inner products on the space of  $p \times p$  matrix polynomials by setting

$$\begin{aligned}\langle\langle \mathbf{F}, \mathbf{G} \rangle\rangle_R &= \int_{\mathbb{T}} \mathbf{F}(z)^\dagger d\Sigma(z) \mathbf{G}(z) \in \mathbb{C}^{p \times p}, \\ \langle\langle \mathbf{F}, \mathbf{G} \rangle\rangle_L &= \int_{\mathbb{T}} \mathbf{G}(z) d\Sigma(z) \mathbf{F}(z)^\dagger \in \mathbb{C}^{p \times p}.\end{aligned}$$

A sequence of matrix polynomials  $(\varphi_j)$  is called right-orthonormal if and only if

$$\langle\langle \varphi_i, \varphi_j \rangle\rangle_R = \delta_{ij} \mathbf{I}_p.$$

Analogous to the scalar case, we can construct orthonormal polynomials satisfying a recursion and the matrices appearing in this recursion are the so-called matrix Verblunsky coefficients (see [12] and an historical introduction therein). For the sake of completeness, we give some more details. First, assume that the support of  $\Sigma$  is infinite. We define the right monic matrix orthogonal polynomials  $\Phi_n^R$  by applying the block Gram-Schmidt algorithm to  $\{\mathbf{I}_p, z\mathbf{I}_p, z^2\mathbf{I}_p, \dots\}$ . In other words,  $\Phi_n^R$  is the unique matrix polynomial  $\Phi_n^R(z) = z^n \mathbf{I}_p + \text{lower order terms}$ , such that  $\langle\langle z^k \mathbf{I}_p, \Phi_n^R \rangle\rangle_R = 0$  for  $k = 0, \dots, n-1$ . The normalized orthogonal polynomials are defined by

$$\varphi_0 = \mathbf{I}_p, \quad \varphi_n^R = \Phi_n^R \kappa_n^R$$

where the sequence of  $p \times p$  matrices  $\kappa_n^R$  satisfies, for all  $n$  the condition  $(\kappa_n^R)^{-1} \kappa_{n+1}^R > 0_p$  and is such that the set  $\{\varphi_n^R\}$  is orthonormal. We define the sequence of left-orthonormal polynomials  $\{\varphi_n^L\}$  in the same way except that the above condition is replaced by  $\kappa_{n+1}^L (\kappa_n^L)^{-1} > 0$ . The matrix Szegő recursion is then

$$(6.4) \quad z\varphi_n^L - \rho_n^L \varphi_{n+1}^L = \alpha_n^\dagger (\varphi_n^R)^*$$

$$(6.5) \quad z\varphi_n^R - \varphi_{n+1}^R \rho_n^R = (\varphi_n^L)^* \alpha_n^\dagger,$$

where for all  $n \in \mathbb{N}$ ,

- $\alpha_n$  belongs to  $\mathbb{B}_p$ , the closed unit ball of  $\mathbb{C}^{p \times p}$  defined by

$$(6.6) \quad \mathbb{B}_p := \{M \in \mathbb{C}^{p \times p} : MM^\dagger \leq \mathbf{I}_p\},$$

- $\rho_n$  is the so-called defect matrix defined by

$$(6.7) \quad \rho_n^R := (\mathbf{I}_p - \alpha_n \alpha_n^\dagger)^{1/2}, \quad \rho_n^L = (\mathbf{I}_p - \alpha_n^\dagger \alpha_n)^{1/2},$$

- for a matrix polynomial  $\mathbf{P}$  with degree  $n$ , the reversed polynomial  $\mathbf{P}^*$  is defined by

$$\mathbf{P}^*(z) := z^n \mathbf{P}(1/\bar{z})^\dagger.$$

Notice that the construction of the recursion coefficients uses only the matrix moments. Verblunski's theorem (the analogue of Favard's theorem for matrix orthogonal polynomials on the unit circle) establishes a one-to-one correspondance between matrix measures on  $\mathbb{T}$  with infinite support and sequences of elements of the interior of  $\mathbb{B}_p$  (Theorem 3.12 in [12]).

Now, for a matrix measure having a finite support, the construction of the Verblunsky coefficients is not obvious. In [17] Theorem 2.1, a sufficient condition on the moments for such a construction is provided. It is related to the positivity of a block-Toeplitz matrix, as it is also mentioned in [39] at the top of p. 208.

In the basis  $(\chi_k)$ , the matrix of  $U$  is (see formula (3.70) in [12])

$$(6.8) \quad \mathcal{C}_\Sigma = \begin{pmatrix} \alpha_0^\dagger & \rho_0^L \alpha_1^\dagger & \rho_0^L \rho_1^L & 0 & 0 & \dots \\ \rho_0^R & -\alpha_0 \alpha_1^\dagger & -\alpha_0 \rho_1^L & 0 & 0 & \dots \\ 0 & \alpha_2^\dagger \rho_1^R & -\alpha_2^\dagger \alpha_1 & \rho_2^L \alpha_3^\dagger & \rho_2^L \rho_3^L & \dots \\ 0 & \rho_2^R \rho_1^R & -\rho_2^R \alpha_1 & -\alpha_2 \alpha_3^\dagger & -\alpha_2 \rho_3^L & \dots \\ 0 & 0 & 0 & \alpha_4^\dagger \rho_3^R & -\alpha_4^\dagger \alpha_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

with

$$(6.9) \quad 0_p \leq \alpha_k \alpha_k^\dagger < I_p$$

for every  $k \geq 0$  in the non-trivial case. If the measure  $\Sigma$  is supported by  $N = np$  points, then  $\Sigma$  is non non-trivial and  $\alpha_{n-1} \in \mathbb{U}(p)$ . In this case, the last line is

$$\begin{array}{llllll} 0 & \dots & 0 & 0 & \alpha_{2r}^\dagger \rho_{2r-1}^R & -\alpha_{2r}^\dagger \alpha_{2r-1} & \text{if } n = 2r + 1, \\ 0 & \dots & 0 & \rho_{2r}^R \rho_{2r-1}^R & -\rho_{2r}^R \alpha_{2r-1} & -\alpha_{2r} \alpha_{2r+1}^\dagger & \text{if } n = 2r + 2, r \geq 0. \end{array}$$

### 6.3 Deformed Verblunsky coefficients

This section is devoted to a detailed study of the deformed Verblunsky coefficients in the matrix setting, consisting in identification of their different definitions and properties. To make the reading easier, we recall the essential results of the scalar case proved in [9].

#### 6.3.1 Scalar case

Motivated by the study of the (scalar) Hua-Pickrell ensemble, Bourgade et al. [9] introduced the so-called *deformed Verblunsky coefficients*. They could be defined in various ways.

**OPUC recursion and the Schur machinery** Let us assume that  $\mu \in \mathcal{M}_1(\mathbb{T})$  has either a finite support consisting in  $n$  points, or infinite support and we will say  $n = \infty$  and  $k \leq n - 1$  will mean  $k \geq 0$ . Then, starting with the orthogonal polynomials  $\Phi_k$  in  $L^2(\mu)$  we define for  $k \leq n - 1$

the functions

$$(6.10) \quad b_k(z) := \frac{\Phi_k(z)}{\Phi_k^*(z)}$$

$$(6.11) \quad \gamma_k(z) := z - \frac{\Phi_{k+1}(z)}{\Phi_k(z)}.$$

From the Szegő recursion (1.1), we have

$$\gamma_k(z) = \frac{\bar{\alpha}_k}{b_k(z)}$$

and successively

$$(6.12) \quad \Phi_k(z) = \prod_0^{k-1} (z - \gamma_j(z)),$$

$$(6.13) \quad \gamma_k(z) = \bar{\alpha}_k \prod_0^{k-1} \frac{1 - z\tilde{\gamma}_j(z)}{z - \gamma_j(z)}, \quad \text{with } \tilde{\gamma}_j(z) = \overline{\gamma_j(\bar{z}^{-1})}.$$

The deformed Verblunsky coefficients are by definition

$$\gamma_k := \gamma_k(1)$$

and may be computed recursively as

$$\gamma_0 = \bar{\alpha}_0, \quad \gamma_k = \bar{\alpha}_k \prod_{j=0}^{k-1} \frac{1 - \bar{\gamma}_j}{1 - \gamma_j}.$$

If  $n$  is finite, the obvious relation  $\Phi_n(1) = \prod_0^{n-1} (1 - \gamma_k)$  may be lifted up, when  $(U, e)$  is given, as

$$\det(I - U) = \Phi_n(1) = \prod_{j=0}^{n-1} (1 - \gamma_j).$$

To explain the connection with Schur parameters, let us recall that the Carathéodory function of a measure  $\mu \in \mathcal{M}_1(\mathbb{T})$  is defined as

$$(6.14) \quad F(z) = \int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(e^{i\theta})$$

and its Schur function  $f : \mathbb{D} \rightarrow \mathbb{D}$  is defined by means of  $F$  by:

$$f(z) = \frac{1 F(z) - 1}{z F(z) - 1}.$$

The Schur algorithm allows to parametrize the Schur function  $f$  by a sequence of so-called Schur parameters. For  $\alpha \in \mathbb{D}$ , let

$$(6.15) \quad T_\alpha : \zeta \mapsto (\zeta - \alpha)(1 - \bar{\alpha}\zeta)^{-1},$$

The reverse mapping is  $T_{-\alpha}$ . Let us define the sequence

$$(6.16) \quad f_0(z) = f(z), \quad f_{j+1}(z) = z^{-1}T_{\alpha_j}(f_j(z)), \quad \alpha_j = f_j(0).$$

We say that  $f$  is the Schur function associated with the sequence  $(\alpha_0, \alpha_1, \dots)$ .

The Geronimus theorem states that these are exactly the Verblunsky coefficients.

From the basic recursion and its *star* version

$$(6.17) \quad \begin{aligned} \Phi_{n+1}(z) &= z\Phi_n(z) - \bar{\alpha}_n\Phi_n^*(z) \\ \Phi_{n+1}^*(z) &= \Phi_n^*(z) - \alpha_n z\Phi_n(z), \end{aligned}$$

we deduce that the sequence of quotients  $b_k(z)$  defined in (6.10) satisfies the recursion

$$(6.18) \quad b_k(z) = \frac{zb_{k-1}(z) - \bar{\alpha}_{k-1}}{1 - z\alpha_{k-1}b_{k-1}(z)},$$

i.e.

$$b_{k-1}(z) = z^{-1}T_{-\bar{\alpha}_{k-1}}(b_k(z)).$$

In other words,  $b_k$  is the Schur function corresponding to the reversed sequence  $(-\bar{\alpha}_{k-1}, \dots, -\bar{\alpha}_0, 1)$  (see [40] Prop. 9.2.3), we say that the sequence  $(b_k)_k$  is the sequence of inverse Schur iterates.

**Decomposition by reflections** Moreover, when  $n$  is finite, a geometrical interpretation is given, with a decomposition of  $U$  into a product of complex reflections parametrized by the coefficients  $\gamma_k, k = 0, \dots, n-1$ .

A  $n$ -(complex) reflection is an element of  $\mathbb{U}(n)$  such that  $r - I_n$  has rank 0 or 1. If  $e$  and  $m \neq e$  are unit vectors of  $\mathbb{C}^n$ , there is a unique reflection  $r$  such that  $r(e) = m$ , and it is

$$(6.19) \quad r = I_n - \frac{1}{1 - \langle m, e \rangle} (m - e) \langle (m - e) | \cdot$$

If  $F := \text{span}\{e, m\}$ , then  $r$  leaves  $F^\perp$  invariant. Now setting

$$\gamma = \langle e, m \rangle, \quad \rho = \sqrt{1 - |\gamma|^2}, \quad e^{i\varphi} = \frac{1 - \gamma}{1 - \bar{\gamma}},$$

then, in the basis  $(e, g)$  of  $F$  obtained by the Gram-Schmidt procedure, the restriction of  $r$  to  $F$  has the matrix

$$\Xi(\gamma) = \begin{pmatrix} \gamma & \rho e^{i\varphi} \\ \rho & -\bar{\gamma} e^{i\varphi} \end{pmatrix}.$$

Let  $U \in \mathbb{U}(n)$ ,  $e$  cyclic for  $U$  and let  $(\varepsilon_1, \dots, \varepsilon_n)$  be the orthonormal basis obtained from the Gram-Schmidt procedure applied to  $(e, Ue, \dots, U^{n-1}e)$ . We define recursively  $n$  reflections as follows.  $r_1$  is the reflection mapping  $e = \varepsilon_1$  onto  $Ue = U\varepsilon_1$  and for  $k \geq 2$   $r_k$  is the reflection mapping  $\varepsilon_k$  onto  $r_{k-1}^{-1}r_{k-2}^{-1} \dots r_1^{-1}U\varepsilon_k$ . Then  $U = r_1 \dots r_n$  and

$$\langle \varepsilon_k, r_k \varepsilon_k \rangle = \gamma_k.$$

### 6.3.2 The matrix case

By convention,  $\mathbf{1} = I_p$  in the following.

**MOPUC recursion and the Schur machinery** Let us define, for  $k = 0, \dots, n-1$ ,

$$(6.20) \quad \mathbf{b}_0(z) = \mathbf{1}, \quad \mathbf{b}_k(z) = \varphi_k^L(1) \left( \varphi_k^{R,*}(1) \right)^{-1},$$

which is consistent with the definition, given in (6.31) of  $\mathbf{b}_k = \mathbf{b}_k(1)$ :

$$(6.21) \quad \mathbf{b}_k = \varphi_k^L(1) \left( \varphi_k^{R,*}(1) \right)^{-1} \quad \text{and} \quad \gamma_k = \mathbf{b}_k^{-1} \alpha_k^\dagger.$$

As in the scalar case, we can make the connection with the inverse Schur iterates.

The Carathéodory function  $\mathbf{F}$  is now matrix-valued, defined again by (6.14), and the Schur function is ([12] Prop. 3.15)

$$\mathbf{f}(z) = z^{-1}(\mathbf{F}(z) - \mathbf{1})(\mathbf{F}(z) + \mathbf{1})^{-1}.$$

To define the Schur algorithm, we set for  $\alpha \in \mathbb{B}_p$  with  $\alpha \alpha^\dagger < \mathbf{1}$ ,

$$\mathbf{T}_\alpha(\zeta) = (\rho^R)^{-1}(\zeta - \alpha)(\mathbf{1} - \alpha^\dagger)^{-1} \rho^L$$

The reverse mapping is  $\mathbf{T}_{-\alpha}$ , and we notice that

$$(6.22) \quad (\mathbf{T}_\alpha(\zeta))^{-1} = \mathbf{T}_{\alpha^\dagger}(\zeta^{-1}).$$

**Theorem 6.1** ([12] Th. 3.19) *For the Schur functions  $\mathbf{f}_0, \mathbf{f}_1, \dots$  associated with Verblunsky coefficients  $\alpha_0, \alpha_1, \dots$ , the following relations hold:*

$$(6.23) \quad \mathbf{f}_{j+1}(z) = z^{-1} \mathbf{T}_{\alpha_j}(\mathbf{f}_j(z))$$

$$(6.24) \quad \alpha_j = \mathbf{f}_j(0)$$

The connection with (6.20) is the following.

**Proposition 6.2** ([12] Prop. 3.26) *For  $k \geq 1$ ,  $\mathbf{b}_k(z)$  is the Schur function associated with the reversed sequence  $(-\alpha_{k-1}^\dagger, \dots, -\alpha_0^\dagger, \mathbf{1})$ .*

**Decomposition by reflections** Let us first fix the notations. Let  $\mathbf{e} = [e_1, \dots, e_p]$  a  $N \times p$  matrix consisting in  $p$  column vectors of dimension  $N = np$ . If  $U \in \mathbb{U}(N)$ , we denote by  $U\mathbf{e}$  the  $N \times p$  matrix  $U\mathbf{e} := [Ue_1, \dots, Ue_p]$ . The pseudo-scalar product of  $\mathbf{e}$  with  $\mathbf{f} = [f_1 \cdots f_p]$  is a matrix  $p \times p$  denoted by  $\ll \mathbf{f}, \mathbf{e} \gg$  and defined by

$$\ll \mathbf{f}, \mathbf{e} \gg_{i,j} = \langle f_i, e_j \rangle \quad i, j = 1, \dots, p.$$

Assume that  $\mathbf{e}$  is cyclic for  $U$  (see definition 2.3 in [23]). If  $N = np$  with  $n \geq 2$ , let  $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_Q)$  be the orthonormal basis obtained from the Gram-Schmidt procedure applied to  $(\mathbf{e}, U\mathbf{e}, \dots, U^{Q-1}\mathbf{e})$ . Neretin [33] defined a sequence of operations on unitary matrices of decreasing dimensions. Up to a slight change of notations, it is the following. For  $m < n$  we decompose a unitary matrix  $U \in \mathbb{U}(n)$  into four blocks

$$U = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with  $A$  a  $m \times m$  matrix, and then define

$$\Xi_n^m(U) = D + C(\mathbf{I}_m - A)^{-1}B \in \mathbb{U}(n - m).$$

Actually,  $\mathbf{I}_{n-m} - \Xi_n^m(U) = (\mathbf{I}_n - U) \diagdown (\mathbf{I}_m - A)$  where  $M \diagdown N$  is the Schur complement of  $M$  with respect to its upper left block (submatrix)  $N$ . This doubly indexed sequence of transformation enjoys the projective property:

$$(6.25) \quad \Xi_{n-m}^r \circ \Xi_n^m = \Xi_n^{r+m},$$

as soon as  $r + m < n$  ([33] Prop. 0.1). In the sequel, for  $q > p$ , we denote by  $[M]_p$  the upper left block of the  $q \times q$  matrix  $M$ .

We define the successive iterations

$$(6.26) \quad \mathbf{c}_0(U) := [U]_p, \quad \mathbf{c}_r(U) := [\Xi_N^{rp}(U)]_p, \quad 1 \leq r \leq n - 1.$$

Then Neretin proved ([33] Sect. 1.5)

$$(6.27) \quad \det(\mathbf{I}_N - U) = \prod_{r=0}^{n-1} \det(\mathbf{1} - \mathbf{c}_r(U)).$$

These operators  $\Xi$  are used to define the successive reflections. More precisely, we define

$$\hat{\pi}_0(U) = U, \quad \hat{\pi}_k(U) = \mathbf{I}_{kp} \oplus \Xi_N^{kp}, \quad 1 \leq k \leq n - 1, \quad \hat{\pi}_Q(U) = \mathbf{I}_N$$

$$R_j(U) = \hat{\pi}_{j-1}(U) \hat{\pi}_j(U)^\dagger, \quad 1 \leq j \leq n.$$

Then, when  $U$  is written in an orthonormal basis  $(\mathbf{e}, \mathbf{e}_2, \dots, \mathbf{e}_n)$ , then  $R_1$  maps  $\mathbf{e}$  onto  $U\mathbf{e}$  and is a reflection since the rank of  $R_1 - \mathbf{I}_N$  is the same as the rank of  $U - (\mathbf{1} \oplus \Xi_N^p)$  which is at most  $p$  (see Prop. 2.5 in [8]).

More generally, for  $k \geq 2$ ,  $R_k$  is a reflection mapping  $\mathbf{e}_k$  onto  $R_{k-1}^\dagger R_{k-2}^\dagger \dots R_1^\dagger U \mathbf{e}_k$  and

$$U = R_1 \dots R_Q$$

In particular, let  $\mathcal{G}$  be the matrix of a unitary operator  $U$  written in the basis  $(\boldsymbol{\varepsilon}_k)_k$  obtained by orthonormalizing the sequence  $\mathbf{e}, U\mathbf{e}, \dots, U^{n-1}\mathbf{e}$ . Usually  $\mathcal{G}$  is called the block GGT matrix :

$$\mathcal{G} := \mathcal{G}^R(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \dots) = \begin{pmatrix} \boldsymbol{\alpha}_0^\dagger & \boldsymbol{\rho}_0^L \boldsymbol{\alpha}_1^\dagger & \boldsymbol{\rho}_0^L \boldsymbol{\rho}_1^L \boldsymbol{\alpha}_2^\dagger & \boldsymbol{\rho}_0^L \boldsymbol{\rho}_1^L \boldsymbol{\rho}_2^L \boldsymbol{\alpha}_3^\dagger & \dots \\ \boldsymbol{\rho}_0^R & -\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_1^\dagger & -\boldsymbol{\alpha}_0 \boldsymbol{\rho}_1^L \boldsymbol{\alpha}_2^\dagger & -\boldsymbol{\alpha}_0 \boldsymbol{\rho}_1^L \boldsymbol{\rho}_2^L \boldsymbol{\alpha}_3^\dagger & \dots \\ 0 & \boldsymbol{\rho}_1^R & -\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2^\dagger & -\boldsymbol{\alpha}_1 \boldsymbol{\rho}_2^L \boldsymbol{\alpha}_3^\dagger & \dots \\ 0 & 0 & \boldsymbol{\rho}_2^R & -\boldsymbol{\alpha}_2 \boldsymbol{\alpha}_3^\dagger & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Write  $\mathcal{R}_j$  for  $R_j(\mathcal{G})$ . Then  $\mathcal{R}_1$  maps  $\varepsilon_1$  onto  $\mathcal{G}\varepsilon_1$  and is a reflection since the rank of  $\mathcal{R}_1 - I_N$  is the same as the rank of  $\mathcal{G} - (1 \oplus \Xi_N^p)$  which is at most  $p$  (see Prop. 2.5 in [8]).

More generally, for  $k \geq 2$ ,  $\mathcal{R}_k$  is a reflection mapping  $\varepsilon_k$  onto  $\mathcal{R}_{k-1}^\dagger \mathcal{R}_{k-2}^\dagger \dots \mathcal{R}_1^\dagger \mathcal{G}\varepsilon_k$  and

$$\mathcal{G} = \mathcal{R}_1 \dots \mathcal{R}_Q$$

Of course, we have

$$(6.28) \quad \det(\mathbf{1} - U) = \det(I_N - \mathcal{G}) = \prod_{r=0}^{n-1} \det(\mathbf{1} - \mathbf{c}_r(\mathcal{G}))$$

We have the following identification.

**Proposition 6.3** *Let  $(U, \mathbf{e})$  be given and call  $\mathcal{G}$  the matrix of  $U$  written in the  $\varepsilon$  basis. Then for  $k = 1, \dots, n$*

$$(6.29) \quad \mathbf{c}_{k-1}(\mathcal{G}) = \ll \varepsilon_k, \mathcal{R}_k \varepsilon_k \gg = \gamma_{k-1}.$$

## 6.4 LDP for matrix Verblunsky coefficients

In a previous work ([23]), the first and last author studied the CUE case. If  $N = np$ , the matrix Verblunsky coefficients  $\alpha_0, \dots, \alpha_{n-1}$  are independent, and for  $j \leq n-2$ ,  $\alpha_j$  has the density in  $\mathbb{B}_p$

$$(6.30) \quad K_{p,(n-k)p} \det(I_p - \alpha \alpha^\dagger)^{(n-2-k)p},$$

where  $K_{p,(n-k)p}$  is some explicit constant. Note that all densities in this section are with respect to

$$dM = \prod_{1 \leq k, l \leq n} d(\Re M_{kl}) \prod_{1 \leq k, l \leq n} d(\Im M_{kl}).$$

From this density, we deduced the LDP:

**Proposition 6.4 ([23] Theorem 3.6)** *Let  $U \in \mathbb{U}(N)$  with  $N = np$  be chosen according to the Haar measure  $\mathbb{P}^{(N)}$  and let  $(\alpha_j^{(n)})_{0 \leq j \leq n-1}$  be the matrix Verblunsky coefficients of the spectral matrix measure of  $(U, e_1, \dots, e_p)$ . Then, for any fixed  $k \geq 1$ ,  $(\alpha_0^{(n)}, \alpha_1^{(n)}, \dots, \alpha_k^{(n)})_{n \geq k}$  satisfies the LDP in  $(\mathbb{B}_p)^k$  with speed  $N$  and good rate function*

$$I_k(\alpha_0, \dots, \alpha_k) = \sum_{j=0}^k -\log \det(I_p - \alpha_j \alpha_j^\dagger).$$

To study the Hua-Pickrell case, we define the deformed matrix Verblunsky coefficients as

$$(6.31) \quad \mathbf{b}_k = \varphi_k^L(1) \left( \varphi_k^{R,*}(1) \right)^{-1} \quad \text{and} \quad \gamma_k = \mathbf{b}_k^{-1} \alpha_k^\dagger.$$

Their interest is broader than the simple study in the Hua-Pickrell case. We have the following result for the distribution of the deformed matrix Verblunsky coefficients, the proof is in Section 6.3.

**Theorem 6.5** *Let  $U$  be chosen at random in  $\mathbb{U}(N)$ ,  $N = np$  with  $n > 2$ , according to the Hua-Pickrell measure with parameter  $\delta$ . Let  $(\gamma_j^{(n)})_{0 \leq j \leq n-1}$  be the deformed matrix Verblunsky coefficients of the spectral matrix measure of  $(U, e_1, \dots, e_p)$ . Then,  $\gamma_1^{(n)}, \dots, \gamma_{n-1}^{(n)}$  are independent. Moreover, for  $j \leq n-2$ ,  $\gamma_j^{(n)}$  has in  $\mathbb{B}_p$  the density*

$$(6.32) \quad K_{n,j}^{(\delta)} \det(\mathbb{I}_p - \gamma)^{\bar{\delta}} \det(\mathbb{I}_p - \gamma^\dagger)^\delta \det(\mathbb{I}_p - \gamma\gamma^\dagger)^{(n-2-j)p}$$

where

$$(6.33) \quad K_{n,j}^{(\delta)} = \pi^{-p^2} \prod_{k=1}^p \frac{\Gamma(N - (j+1)p + k + \delta) \Gamma(N - (j+1)p + k + \bar{\delta})}{\Gamma(N - (j+2)p + k) \Gamma(N - (j+1)p + k + \delta + \bar{\delta})}$$

and  $\gamma_{n-1}$  follows the Hua-Pickrell distribution on  $\mathbb{U}(p)$  with parameter  $\delta$ .

If  $\delta = Nd$ , we get the following LDP for the deformed coefficients. We remark that if  $d = 0$ , the rate reduces to the rate in Proposition 6.4, since the matrices  $\mathbf{b}_k$  are unitary by Theorem 3.9 in [12].

**Proposition 6.6** *Let  $U$  be chosen at random in  $\mathbb{U}(N)$ ,  $N = np$ , according to the Hua-Pickrell measure  $\mathbb{HIP}_{Nd}^{(N)}$  with  $d \geq 0$ . Let  $(\gamma_j^{(n)})_{0 \leq j \leq n-1}$  be the deformed matrix Verblunsky coefficients of the spectral matrix measure of  $(U, e_1, \dots, e_p)$ . Then, for any fixed  $k$ ,  $(\gamma_0^{(n)}, \gamma_1^{(n)}, \dots, \gamma_k^{(n)})_{n \geq k}$  satisfies the LDP in  $(\mathbb{B}_p)^k$  with speed  $N$  and good rate function*

$$I_k(\gamma_0, \dots, \gamma_k) = \sum_{j=0}^k H_{d,p}(\gamma_j),$$

with

$$(6.34) \quad H_{d,p}(\gamma) = -\log \det(\mathbb{I}_p - \gamma\gamma^\dagger) - d \log \det((\mathbb{I}_p - \gamma)(\mathbb{I}_p - \gamma^\dagger)) + pH_d(0)$$

where  $H_d$  is as in (4.17).

## 6.5 LDP for matrix spectral measures

Our next LDP holds for matrix spectral measures of  $(U, e_1, \dots, e_p)$ , when  $U$  is chosen randomly according to a general measure  $\mathbb{P}_\gamma^{(N)}$  on  $\mathbb{U}(N)$  as defined in (2.12). In this case, the eigenvector matrix is again Haar distributed, the weights  $(W_1, \dots, W_N)$  are independent of the eigenvalues and follow a distribution, which is a matrix analogue of the Dirichlet distribution. For the precise statement, we refer to Proposition 3.1 in [23].

Let us introduce the matrix analogue to the set  $\mathcal{S}_1^\mathbb{T} = \mathcal{S}_1^\mathbb{T}(\alpha^-, \alpha^+)$ . For  $[\alpha^-, \alpha^+]$  an interval in  $(0, 2\pi)$ , let  $I = [\widehat{\alpha^-, \alpha^+}]$  and let  $\mathcal{S}_{p,1}^\mathbb{T} = \mathcal{S}_{p,1}^\mathbb{T}(\alpha^-, \alpha^+)$  be the set of all normalized measures  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$  with

- (i)  $\text{supp}(\Sigma) = J \cup \{e^{i\theta_i^-}\}_{i=1}^{N^-} \cup \{e^{i\theta_i^+}\}_{i=1}^{N^+}$ , where  $J \subset I$ ,  $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$  and  $\theta_i^\pm \in [0, 2\pi)$ . Furthermore,

$$0 \leq \theta_1^- < \theta_2^- < \dots < \alpha^- \quad \text{and} \quad \theta_1^+ > \theta_2^+ > \dots > \alpha^+.$$

- (ii) If  $N^-$  (resp.  $N^+$ ) is infinite, then  $\theta_j^-$  converges towards  $\alpha^-$  (resp.  $\theta_j^+$  converges to  $\alpha^+$ ).

We can write such a measure  $\Sigma$  as

$$(6.35) \quad \Sigma = \Sigma|_I + \sum_{i=1}^{N^+} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \Gamma_i^- \delta_{\lambda_i^-},$$

for some nonnegative Hermitian matrices  $\Gamma_1^+, \dots, \Gamma_{N^+}^+, \Gamma_1^-, \dots, \Gamma_{N^-}^-$  and  $\lambda_i^\pm = e^{i\theta_i^\pm}$ . As before,  $\mathcal{S}_{p,1}^{\mathbb{T}}(0, 2\pi)$  is the set of measures supported by  $\mathbb{T}$ . Then we have the following result. We omit the proof, the necessary steps to extend the scalar case in Theorem 4.2 to the matrix case are exactly the same as in [19], where the matrix LDP for measures on the real line is established.

**Theorem 6.7** *Assume that  $U$  is distributed according to  $\mathbb{P}_{\mathcal{V}}^{(N)}$ ,  $N = np$ , and the potential  $\mathcal{V}$  satisfies assumptions (T1), (T2), (T3). Then the sequence of matrix spectral measures  $(\Sigma_p^{(N)})_n$  of  $(U, e_1, \dots, e_p)$  satisfies the LDP in  $\mathcal{M}_{p,1}(\mathbb{T})$  equipped with the weak topology, with speed  $N$  and rate function*

$$(6.36) \quad \mathcal{I}_{\mathcal{V}}^p(\Sigma) = \mathcal{K}(\mathbb{I}_p \cdot \mu_{\mathcal{V}} | \Sigma) + \sum_{n=1}^{N^+} \mathcal{F}_{\mathcal{V}}^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_{\mathcal{V}}^-(\lambda_n^-).$$

if  $\Sigma \in \mathcal{S}_{p,1}^{\mathbb{T}}(\alpha^-, \alpha^+)$  and  $\mathcal{I}_{\mathcal{V}}^p(\Sigma) = +\infty$  otherwise. Here,  $\mu_{\mathcal{V}}$  is the scalar measure as in assumption (T2).

## 6.6 Sum rules

The matrix version of Szegő's formula was established in [13] (see more recently [15]) and a probabilistic proof is in [23].

**Theorem 6.8** *Let  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$  with infinite support and let  $(\alpha_k)_{k \geq 0} \in (\mathbb{B}_p)^{\mathbb{N}}$  be the sequence of its Verblunsky coefficients. Then*

$$(6.37) \quad \mathcal{K}(\mathbb{I}_p \cdot \text{UNIF} | \Sigma) = \sum_{j=0}^{\infty} -\log \det(\mathbb{I}_p - \alpha_j \alpha_j^\dagger).$$

The matrix version of Theorem 5.1 is the following result. It is a combination of Proposition 6.6 and Theorem 6.7, when  $\mathcal{V}$  is the potential of the Hua-Pickrell ensemble. The proof is as in the scalar case: Proposition 6.6 yields by the projective method a complimentary LDP for a measure distributed according to the Hua-Pickrell ensemble, the statement follows then from the uniqueness of rate functions.

**Theorem 6.9** Let  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$  with infinite support and let  $(\gamma_k)_{k \geq 0} \in (\mathbb{B}_p)^\mathbb{N}$  be the sequence of its deformed matrix Verblunsky coefficients. Then for any  $\mathbf{d} \geq 0$ ,  $\sum_{k=0}^{\infty} H_{\mathbf{d},p}(\gamma_k) = \infty$  if  $\Sigma \notin \mathcal{S}_{1,p}^\mathbb{T}(\theta_{\mathbf{d}}, 2\pi - \theta_{\mathbf{d}})$ . For  $\Sigma \in \mathcal{S}_{p,1}^\mathbb{T}(\theta_{\mathbf{d}}, 2\pi - \theta_{\mathbf{d}})$ ,

$$(6.38) \quad \mathcal{K}(\mathbb{I}_p \cdot \text{HP}_{\mathbf{d}}|\Sigma) + \sum_{n=1}^{N^+} \mathcal{F}_{HP}^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_{HP}^-(\lambda_n^-) = \sum_{j=0}^{\infty} H_{\mathbf{d},p}(\gamma_j),$$

where both sides may be infinite simultaneously, and  $H_{\mathbf{d},p}$  is defined in (6.34).

For the Gross-Witten ensemble, it seems difficult (at least at a first attempt) to adapt Simon's proof to the matrix setup. Nevertheless, since the density of  $\mathbb{GW}_{\mathbf{g}}^{(N)}$  with respect to the Haar measure is proportional to

$$\exp(N\mathbf{g} \Re \text{tr} U)$$

and  $\text{tr} U = \text{tr} \mathcal{C}_\Sigma$  which can be computed in matrix terms, taking into account the CMV form of  $U$  (6.8):

$$\text{tr} \mathcal{C}_\Sigma = \text{tr} T_n(\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_{n-1})$$

where

$$T_n(\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_{n-1}) = \boldsymbol{\alpha}_0^\dagger - \boldsymbol{\alpha}_0 \boldsymbol{\alpha}_1^\dagger - \dots \begin{cases} -\boldsymbol{\alpha}_{2r}^\dagger \boldsymbol{\alpha}_{2r-1} & \text{if } n = 2r + 1, \\ -\boldsymbol{\alpha}_{2r} \boldsymbol{\alpha}_{2r+1}^\dagger & \text{if } n = 2r + 2, r \geq 0. \end{cases}$$

So with

$$(6.39) \quad T(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \dots) = \boldsymbol{\alpha}_0^\dagger - \boldsymbol{\alpha}_0 \boldsymbol{\alpha}_1^\dagger - \sum_{j=1}^{\infty} (\boldsymbol{\alpha}_{2k} \boldsymbol{\alpha}_{2k+1}^\dagger + \boldsymbol{\alpha}_{2k}^\dagger \boldsymbol{\alpha}_{2k-1}),$$

we are in the position to propose the following conjecture:

**Conjecture 6.10** Let  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{T})$  with infinite support and let  $(\boldsymbol{\alpha}_k)_{k \geq 0} \in (\mathbb{B}_p)^\mathbb{N}$  be the sequence of its matrix Verblunsky coefficients.

1. If  $|\mathbf{g}| \leq 1$ , then

$$(6.40) \quad \mathcal{K}(\mathbb{I}_p \cdot \text{GW}_{-\mathbf{g}}|\Sigma) = \sum_{j=0}^{\infty} -\log \det(I_p - \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^\dagger) + \mathbf{g} \Re \text{tr} T(\boldsymbol{\alpha}_0, \dots) + H_p(\mathbf{g}),$$

where  $T(\boldsymbol{\alpha}_0, \dots)$  is given by (6.39) and  $H_p(\mathbf{g})$  is some constant.

2. If  $|\mathbf{g}| > 1$ , then a similar identity holds, with an additional term on the left hand side which is

$$\sum_{n=1}^{N^-} \mathcal{F}_{-\mathbf{g}}^+(\lambda_i^-) + \sum_{n=1}^{N^+} \mathcal{F}_{-\mathbf{g}}^-(\lambda_i^+)$$

as in Conjecture 5.8.

## 7 Proofs

### 7.1 Proofs of Section 2.4

#### 7.1.1 Proof of Theorem 4.2

We can follow verbatim the proof of the corresponding theorem in the real case. The main idea is to apply the projective method (the Dawson-Gärtner Theorem, see [14]) to a non-normalized version of the spectral measure. In a first step, we consider instead of  $\mu^{(n)}$  the measure

$$\tilde{\mu}^{(n)} = \sum_{i=1}^n \gamma_i \delta_{\lambda_i},$$

with  $\gamma_1, \dots, \gamma_n$  independent  $\text{Gamma}(1, n^{-1})$  distributed random variables with mean  $n^{-1}$ . Since the self-normalized vector of these gamma-distributed random variables has a uniform distribution on the simplex,  $\tilde{\mu}^{(n)}/\tilde{\mu}^{(n)}(\mathbb{T})$  recovers the original distribution of  $\mu^{(n)}$ . Then, we consider the measure

$$(7.1) \quad \pi_j(\tilde{\mu}^{(n)}) = \tilde{\mu}_{|I}^{(n)} + \sum_{i=1}^{N^+ \wedge j} \gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^- \wedge j} \gamma_i^- \delta_{\lambda_i^-}$$

as in (4.2). Note that this projection is not continuous in the weak topology, and in [20] we introduce a new topology generated by  $\tilde{\mu}_{|I}^{(n)}$  and the vector of outliers. On the set of normalized measures, this topology is stronger than the weak topology and we can claim the LDP in the latter topology. Ultimately, this also explains why our arbitrary distinction between  $\lambda_i^+$  and  $\lambda_i^-$  creates no problems: the transition of an eigenvalue from  $e^{i\theta_1^-}$  to  $e^{i\theta_1^+}$  is continuous in the weak topology, but not in our new one.

A crucial ingredient in the LDP for  $\pi_j(\tilde{\mu}^{(n)})$  is the LDP for a finite collection of extreme eigenvalues. Of course, for the first statement of our theorem, this can be omitted. To simplify notation, let  $A^{\uparrow j}$  (or  $A^{\downarrow j}$ ) denote subset of  $A^j$  with non-decreasing (resp. non-increasing) coordinates.

**Proposition 7.1** *Let  $j$  and  $\ell$  be fixed integers. Assume that  $\mathcal{V}$  satisfies (T1), (T2) and the control condition (T3). If  $0 < \alpha^+$  and  $\alpha^- < 2\pi$ , then the law of  $(\theta_1^+, \dots, \theta_j^+, \theta_1^-, \dots, \theta_\ell^-)$  under  $\mathbb{P}_{\mathcal{V}}^{(n)}$  satisfies the LDP in  $\mathbb{R}^{j+\ell}$  with speed  $n$  and rate function*

$$\mathcal{I}_{\theta^\pm}(\theta^+, \theta^-) = \sum_{k=1}^j \mathcal{F}_{\mathcal{V}}^+(e^{i\theta_k^+}) + \sum_{k=1}^{\ell} \mathcal{F}_{\mathcal{V}}^-(e^{i\theta_k^-})$$

if  $\theta^+ = (\theta_1^+, \dots, \theta_j^+) \in [\alpha^+, 2\pi]^{\downarrow j}$  and  $\theta^- = (\theta_1^-, \dots, \theta_\ell^-) \in [0, \alpha^-]^{\uparrow \ell}$  and  $\mathcal{I}_{\theta^\pm}(\theta^+, \theta^-) = \infty$  otherwise.

**Proof:** We first mention the main points in the proof of the large deviation upper bound. Let us stress that exponential tightness is inherent on the circle. The proof follows the same lines as in [20] and makes use of the following lemmas.

**Lemma 7.2** *Let  $\mathcal{V}$  be a continuous potential on  $\mathbb{T} \setminus \{1\}$  satisfying (T1). and let  $r$  be a fixed integer. If  $\mathbb{P}_{\mathcal{V}_n}^{(n)}$  is the probability measure associated to the potential  $\mathcal{V}_n = \frac{n+r}{n}\mathcal{V}$ , then the law of  $\mu_{\mathbb{U}}^{(n)}$  under  $\mathbb{P}_{\mathcal{V}_n}^{(n)}$  satisfies the LDP with speed  $n^2$  with good rate function*

$$(7.2) \quad \mu \mapsto \mathcal{E}(\mu) - \inf_{\nu} \mathcal{E}(\nu)$$

where  $\mathcal{E}$  is defined in (2.20).

**Lemma 7.3** *If the potential  $\mathcal{V}$  is continuous on  $\mathbb{T} \setminus \{1\}$  and satisfies (T1), we have for every  $p \geq 1$*

$$(7.3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{Z_{\mathcal{V}}^{(n)}}{Z_{\frac{n-p}{n}\mathcal{V}}^{(n-p)}} = - \inf_{z_1, \dots, z_k} \sum_{k=1}^p \mathcal{J}_{\mathcal{V}}(z_k) = -p \inf_z \mathcal{J}_{\mathcal{V}}(z).$$

Actually, to prove Lemma 7.2, it is enough to notice that there exists two constants  $c_1 < c_2$  such that

$$c_1 \leq \frac{1}{n} \log \frac{d\mathbb{P}_{\mathcal{V}_n}^{(n)}}{d\mathbb{P}_{\mathcal{V}}^{(n)}} \leq c_2.$$

Since the arguments for the two lemmas is the same as in the real case, we omit the proofs.

For the proof of the large deviation lower bound, we may make the same remark as above, we don't need to show exponential tightness anymore. Besides we need the fact that under  $\mathbb{P}_{\frac{n+r}{n}\mathcal{V}}^{(n)}$ , the extremal eigenvalues converges to the endpoints of the support of  $\mu_{\mathcal{V}}$  if its support is a proper arc. It was a separate lemma in [20], but it is a direct consequence of the upper bound and assumption (T3).

**Lemma 7.4** *Under Assumption (T1) and (T3), the distance of  $\theta_i^+$  and  $\theta_i^-$  to  $\{\alpha^-, \alpha^+\}$  converges in probability to 0 for all  $i \geq 1$ .*

**Proof:** We may utilize the large deviation upper bound of Proposition 7.1. The upper bound involves the rate function  $\mathcal{J}_{\mathcal{V}} - \inf_x \mathcal{J}_{\mathcal{V}}(x)$ . So, if this rate function which vanishes on the support of  $\mu_{\mathcal{V}}$  does not vanish outside by assumption (T3), that means that the probability that the distance to  $\{\alpha^-, \alpha^+\}$  is greater than  $\varepsilon$  is exponentially small.  $\square$

The next step in the proof of Theorem 4.2 is a joint LDP for the measure  $\tilde{\mu}_{|I(j)}^{(n)}$  restricted to  $I(j) = I \setminus \{\lambda^+, \lambda^-\}$  and the extremal eigenvalues. The crucial ingredients are the independence of eigenvalues and weights and the LDP for  $\mu_{\mathbb{U}}^{(n)}$  on the faster scale  $n^2$ . The following result is a straightforward adaption of Theorem 4.2 in [20].

### Proposition 7.5

1. *Assume that the potential satisfies (T1) and that the support of  $\mu_{\mathcal{V}}$  is  $\mathbb{T}$ . Then the sequence of measures  $\tilde{\mu}^{(n)}$  satisfies the LDP with speed  $n$  and good rate function*

$$\mathcal{I}(\mu) = \mathcal{K}(\mu_{\mathcal{V}} | \mu) + \mu(\mathbb{T}) - 1.$$

2. Assume that the potential  $\mathcal{V}$  satisfies the assumptions (T1), (T2) and (T3). Then the sequence

$$(\tilde{\mu}_{|I(j)}^{(n)}, \theta^+, \theta^-)$$

with  $\theta^\pm = (\theta_1^\pm, \dots, \theta_j^\pm)$  satisfies the LDP with speed  $n$  and good rate function

$$\mathcal{I}(\mu, x^+, x^-) = \mathcal{K}(\mu_{\mathcal{V}} | \mu) + \mu(I) - 1 + \mathcal{I}_{\theta^\pm}(\theta^+, \theta^-).$$

The weights  $(\gamma_1^+, \dots, \gamma_j^+, \gamma_1^-, \dots, \gamma_j^-)$  associated with the extremal eigenvalues satisfy a LDP in  $\mathbb{R}^{2j}$  with speed  $n$  and good rate

$$\mathcal{I}_\gamma(y) = \sum_{i=1}^{2j} y_i$$

if all  $y_i$  are nonnegative and  $\mathcal{I}_\gamma(y) = \infty$  otherwise. Using again the independence, we obtain from Proposition 7.5 the joint LDP for

$$(\tilde{\mu}_{|I(j)}^{(n)}, \theta^+, \theta^-, \gamma^+, \gamma^-)$$

(omitting the extremal eigenvalues if the support of  $\mu_{\mathcal{V}}$  is  $\mathbb{T}$ ) with rate the sum of the rate in Proposition 7.5 and  $\mathcal{I}_\gamma$ . From this collection, we may now conclude the LDP for the projected measure  $\pi_j(\tilde{\mu}^{(n)})$ , by mapping continuously

$$(\tilde{\mu}_{|I(j)}^{(n)}, \theta^+, \theta^-, \gamma^+, \gamma^-) \mapsto \tilde{\mu}_{|I(j)}^{(n)} + \sum_{i=1}^j \left( \gamma_i^+ \delta_{e^{i\theta_i^+}} + \gamma_i^- \delta_{e^{i\theta_i^-}} \right),$$

which yields by the contraction principle the LDP for  $\pi_j(\tilde{\mu}^{(n)})$  with good rate

$$\mathcal{I}_j(\mu) = \mathcal{K}(\mu_{\mathcal{V}} | \mu) + \mu(\mathbb{T}) - 1 + \sum_{i=1}^{N^+ \wedge j} \mathcal{F}_{\mathcal{V}}^+(e^{i\theta_i^+}) + \sum_{i=1}^{N^- \wedge j} \mathcal{F}_{\mathcal{V}}^+(e^{i\theta_i^-}).$$

Finally, the LDP for  $\mu^{(n)}$  follows by taking the projective limit and normalizing. The arguments are as in Section 4.4 in [20]. This concludes the proof of Theorem 4.2.

### 7.1.2 Proof of Theorem 4.7

We mimick the proof of Theorem 4.3 and 4.4 in [22]. The weak convergence topology on  $\mathcal{M}_1(\mathbb{T})$  is equivalent to the topology of convergence of moments on  $\bar{\mathbb{D}}^{\mathbb{N}}$ , which is equivalent to the convergence of deformed Verblunsky coefficients. Let  $\Gamma : \mathcal{M}_1(\mathbb{T}) \rightarrow \bar{\mathbb{D}}^{\mathbb{N}}$  denote the mapping which associates to each measure  $\mu$  the sequence of deformed Verblunsky coefficients

$$\Gamma(\mu) = (\gamma_0(\mu), \gamma_1(\mu), \dots),$$

where we set  $\gamma_{n+1}(\mu) = \gamma_{n+2}(\mu) = \dots = 1$  if  $\gamma_n(\mu) \in \mathbb{T}$ . Let  $p_k : \bar{\mathbb{D}}^{\mathbb{N}} \rightarrow \bar{\mathbb{D}}^k$  be the projection onto the first  $k$  coordinates. By Lemma 4.6,  $p_k(\Gamma(\mu^{(n)}))$  satisfies under  $\mathbb{HPP}_{nd}^{(n)}$  the LDP in  $\bar{\mathbb{D}}^k$  with speed  $n$  and good rate  $I_k$ . We now apply the Dawson-Gärtner theorem. When equipped with the product topology,  $\bar{\mathbb{D}}^{\mathbb{N}}$  can be viewed as the projective limit

$$\bar{\mathbb{D}}^{\mathbb{N}} = \lim_{\leftarrow} \bar{\mathbb{D}}^k,$$

so that under  $\mathbb{HPP}_{nd}^{(n)}$ , the sequence  $\Gamma(\mu^{(n)})$  satisfies the LDP in  $\bar{\mathbb{D}}^{\mathbb{N}}$  with speed  $n$  and good rate function

$$I_{\infty}(\gamma_0, \gamma_1, \dots) = \sup_k I_k(p_k(\gamma_0, \dots)) = \sup_k \sum_{j=0}^k H_d(\gamma_j) = \sum_{j=0}^{\infty} H_d(\gamma_j).$$

Actually, this LDP holds in  $\Gamma(\mathcal{M}_1(\mathbb{T}))$ , as on the complement the rate function equals  $+\infty$ . Finally, it remains to apply the continuous mapping  $\Gamma^{-1}$  to obtain the LDP for  $\mu^{(n)}$  with the good claimed rate function  $J_d^{HP}$ .  $\square$

## 7.2 Proofs of Section 5

### 7.2.1 Proof of Corollary 5.3

The elementary decomposition

$$(1 - g \cos \theta) = g(1 - \cos \theta) + (1 - g)$$

and the definition of  $\text{GW}_{-g}$  give

$$\mathcal{K}(\text{GW}_{-g} | \mu) = G(g) + g\mathcal{K}(\text{GW}_{-1} | \mu) + (1 - g)\mathcal{K}(\text{GW}_0 | \mu) - gG(1)$$

where

$$G(a) := \int_0^{2\pi} (1 - a \cos \theta) \log(1 - a \cos \theta) \frac{d\theta}{2\pi}.$$

Now a routine computation gives:

$$(7.4) \quad G(a) = 1 - \sqrt{1 - a^2} + \log \frac{1 + \sqrt{1 - a^2}}{2}.$$

**Remark 7.6** *The minimum in formula (5.4) is 0. It is reached uniquely in  $\mu = \text{GW}_{-g}$ , which corresponds to the Verblunsky coefficients given in (3.6).*

## 7.3 Proofs of Section 6

### 7.3.1 Proposition 6.3

First, we have

$$\begin{aligned} \ll \varepsilon_k, \mathcal{R}_k \varepsilon_k \gg &= \ll \varepsilon_k, \mathcal{R}_{k-1}^* \mathcal{R}_{k-2}^* \dots \mathcal{R}_1^* \mathcal{G} \varepsilon_k \gg = \ll \mathcal{R}_1 \dots \mathcal{R}_{k-1} \varepsilon_k, \mathcal{G} \varepsilon_k \gg \\ &= \ll \mathcal{G} \hat{\pi}_k(\mathcal{G})^* \varepsilon_k, \mathcal{G} \varepsilon_k \gg = \ll \hat{\pi}_k(U)^* \varepsilon_k, \varepsilon_k \gg = \ll \varepsilon_k, \hat{\pi}_k(\mathcal{G}) \varepsilon_k \gg = \mathbf{c}_{k-1}(\mathcal{G}). \end{aligned}$$

To compute  $\mathbf{c}_k(\mathcal{G})$  we start from the definitions of  $\mathcal{G}$  and  $\Xi_N^p$ , which yield

$$(7.5) \quad \Xi_N^p(\mathcal{G}^R) = \hat{\Theta}(u_0)\mathcal{G}^R(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots)$$

where, if  $u \in \mathbb{U}(p)$

$$\hat{\Theta}(u) = \begin{pmatrix} u & 0_{p, N-2p} \\ 0_{N-2p, p} & I_{N-2p} \end{pmatrix}$$

and

$$u_0 = -\boldsymbol{\alpha}_0 + \boldsymbol{\rho}_1^R(\mathbf{1} - \boldsymbol{\alpha}_0^\dagger)^{-1}\boldsymbol{\rho}_0^L,$$

so that

$$\mathbf{c}_1(\mathcal{G}^R) = [\Xi_N^p(\mathcal{G}^R)]_p = u_0\boldsymbol{\alpha}_1^\dagger.$$

More generally, looking for a recursion - thinking of (6.25) - , we notice that

$$(7.6) \quad \Xi_n^p\left(\hat{\Theta}(u)\mathcal{G}^R(\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_{n-1})\right) = \hat{\Theta}(v)\mathcal{G}^R(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n-1})$$

where

$$(7.7) \quad v = v(u, \boldsymbol{\alpha}_0) = -\boldsymbol{\alpha}_0 + \boldsymbol{\rho}_0^R(\mathbf{1} - u\boldsymbol{\alpha}_0^\dagger)^{-1}u\boldsymbol{\rho}_0^L.$$

We need the following observation.

**Lemma 7.7** *If  $\boldsymbol{\alpha} \in \mathbb{B}(p)$  and  $u \in \mathbb{U}(p)$  then*

$$(7.8) \quad -\boldsymbol{\alpha} + \boldsymbol{\rho}^R(\mathbf{1} - u\boldsymbol{\alpha}^\dagger)^{-1}u\boldsymbol{\rho}^L = (\boldsymbol{\rho}^R)^{-1}(u - \boldsymbol{\alpha})(\mathbf{1} - \boldsymbol{\alpha}^\dagger u)^{-1}\boldsymbol{\rho}^L = \mathbf{T}_{\boldsymbol{\alpha}}(u).$$

Let us assume that

$$(7.9) \quad \Xi_n^{jp} = \hat{\Theta}(u_j)\mathcal{G}^R(\boldsymbol{\alpha}_j, \dots, \boldsymbol{\alpha}_{n-1}),$$

where  $u_j$  depends on  $\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_{j-1}$ .

Applying (6.25), (7.6) and Lemma 7.7 we get

$$\begin{aligned} \Xi_N^{(j+1)p}(\mathcal{G}^R(\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_{n-1})) &= \Xi_{N-jp}^p\left(\hat{\Theta}(u_p)\mathcal{G}^R(\boldsymbol{\alpha}_j, \dots, \boldsymbol{\alpha}_{n-1})\right) \\ &= \hat{\Theta}(\mathbf{T}_{\boldsymbol{\alpha}_j}(u_j))\mathcal{G}^R(\boldsymbol{\alpha}_{j+1}, \dots, \boldsymbol{\alpha}_{n-1}), \end{aligned}$$

and the assumption (7.9) is satisfied at rank  $j + 1$  with

$$(7.10) \quad u_{j+1} = \mathbf{T}_{\boldsymbol{\alpha}_j}(u_j).$$

Passing to the upperleft block, we obtain easily, for every  $j \leq n - 1$

$$\mathbf{c}_j = u_j\boldsymbol{\alpha}_j^\dagger.$$

Now, using (6.22), we see that  $\boldsymbol{\beta}_j := u_j^{-1}$  satisfies the recursion

$$\boldsymbol{\beta}_{j+1} = \mathbf{T}_{\boldsymbol{\alpha}_j^\dagger}(\boldsymbol{\beta}_j)$$

or, reversing

$$\boldsymbol{\beta}_j = \mathbf{T}_{-\boldsymbol{\alpha}_j^\dagger}(\boldsymbol{\beta}_{j+1})$$

which allows to conclude that  $\boldsymbol{\beta}_j = \mathbf{b}_j$  and ends the proof of Proposition 6.3.

**Proof of Lemma 7.7 :** We have to prove  $E = (\rho^R)^{-1}F\rho^L$  (say). Actually, since  $\rho^R\alpha = \alpha\rho^L$  and  $(\rho^R)^2 = \mathbf{1} - \alpha\alpha^\dagger$ , we have

$$\rho^R E (\rho^L)^{-1} = -\alpha + (\mathbf{1} - \alpha\alpha^\dagger)(\mathbf{1} - u\alpha^\dagger)^{-1}u.$$

Now,  $(\mathbf{1} - u\alpha^\dagger)u = u(\mathbf{1} - \alpha^\dagger u)$  and then

$$\rho^R E (\rho^L)^{-1} = -\alpha (\mathbf{1} + \alpha^\dagger u (\mathbf{1} - \alpha^\dagger u)^{-1}) + u(\mathbf{1} - \alpha^\dagger u)^{-1},$$

which is exactly  $F = (u - \alpha)(\mathbf{1} - \alpha^\dagger u)^{-1}$ . ■

### 7.3.2 Proof of Theorem 6.5

In [23] it is proved that if  $\mathbb{U}(N)$  is equipped with the Haar measure, the distribution of  $(\alpha_0, \dots, \alpha_{n-1})$  is

$$(7.11) \quad \text{const} \cdot \prod_{r=0}^{n-2} \det(\mathbf{1} - \alpha_r^\dagger \alpha_r)^{N-(r+2)p} \prod_{r=0}^{n-1} d\alpha_r,$$

where, for  $r = 0, \dots, n-2$ ,  $d\alpha_r$  is the Lebesgue measure on  $\mathbb{B}_p$ , and  $d\gamma_{n-1}$  is the Haar measure on  $\mathbb{U}(p)$ .

Since  $\gamma_r$  is  $\alpha_r^\dagger$  up to multiplication by a unitary matrix depending only on  $(\alpha_0, \dots, \alpha_{r-1})$ , we deduce that, the pushforward of the Haar measure on  $\mathbb{U}(N)$  by  $(\gamma_0, \gamma_1, \dots, \gamma_{n-1})$  has the distribution

$$(7.12) \quad \text{const} \cdot \prod_{r=0}^{n-2} \det(\mathbf{1} - \gamma_r^\dagger \gamma_r)^{N-(r+2)p} \prod_{r=0}^{n-1} d\gamma_r,$$

Now, by definition

$$\frac{d\mathbb{HIP}_\delta^{(N)}}{d\mathbb{P}_0^{(N)}}(U) = \text{const} \cdot \det(\mathbf{I}_N - U)^\delta \det(\mathbf{I}_N - U^\dagger)^\delta$$

It remains to apply (6.28) and (6.3) to conclude that under the  $\mathbb{HIP}_\delta^{(N)}$  probability, the variables  $(\gamma_0, \gamma_1, \dots, \gamma_{n-1})$  are independent and for  $0 \leq r \leq n-2$  the density of  $\gamma_r^p$  in  $\mathbb{B}_p$  is then

$$\text{const} \cdot (\det(\mathbf{1} - \gamma))^\delta (\det(\mathbf{1} - \gamma^\dagger))^\delta \det(\mathbf{1} - \gamma^\dagger \gamma)^{N-(r+2)p},$$

and the variable  $\gamma_{n-1}$  is  $\mathbb{HIP}_\delta^{(p)}$  distributed on  $\mathbb{U}(p)$ . The value of the normalizing constant (6.33) is then taken from formulas (2.9) in [33].

**Remark 7.8** *Theorem 1.3 of Neretin [33] says that if  $\mathbb{U}(N)$  is equipped with the Haar measure, then the distribution of  $(\mathbf{c}_0(U), \dots, \mathbf{c}_{n-1}(U))$  is also (7.12). From (6.27) we deduce that, under  $\mathbb{HIP}_\delta^{(n)}$ ,  $(\mathbf{c}_0(U), \dots, \mathbf{c}_{n-1}(U))$  and  $(\gamma_0, \gamma_1, \dots, \gamma_{n-1})$  have the same distribution. The difference is that the second array depends only on the spectral measure, and the first one depends more deeply on  $U$ . In particular, we do not know the connection between these coefficients  $\mathbf{c}(U)$  and  $\alpha$ .*

### 7.3.3 Proof of Proposition 6.6

By independence, it suffices to prove the LDP for a single  $\gamma_j^{(n)}$  with rate  $H_{a,p}$ . Since the LDP is a standard consequence of the explicit density in (6.32), we only give a sketch of the proof. First, we get from the explicit expression of the constant in (6.33)

$$\lim_{n \rightarrow \infty} \frac{1}{np} \log K_{n,j}^{(nd)} = pH_a(0).$$

Then, note that on the set  $\{M \in \mathbb{C}^{p \times p} | MM^\dagger < I_p\}$  the rate function is finite and continuous, since if  $\gamma \in \mathbb{C}^{p \times p}$  is a matrix with singular values smaller than 1,  $I_p - \gamma$  is non-singular. On the other hand, if  $\gamma \in \mathbb{B}_p \setminus \{M \in \mathbb{C}^{p \times p} | MM^\dagger < I_p\}$ , we have  $H_{a,p}(\gamma) = \infty$ . This implies for any  $\gamma \in \mathbb{B}_p$ , with  $B_\varepsilon(\gamma)$  the open ball centered at  $\gamma$  with radius  $\varepsilon$  in the Frobenius norm,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{np} \log \mathbb{HPP}_{Nd}^{(N)}(\gamma_j^{(n)} \in B_\varepsilon(\gamma)) &= H_{a,p}(\gamma), \\ \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{np} \log \mathbb{HPP}_{Nd}^{(N)}(\gamma_j^{(n)} \in B_\varepsilon(\gamma)) &= H_{a,p}(\gamma). \end{aligned}$$

From these limits, we get that  $(\gamma_j^{(n)})_n$  satisfies the weak LDP with speed  $N = np$  and good rate  $H_{a,p}$ . Necessarily, this sequence is exponentially tight, since it lives on the compact set  $\mathbb{B}_p$ , such that the full LDP follows.  $\square$

## References

- [1] G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, Cambridge, 2010.
- [2] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.
- [3] E.L. Basor and Y. Chen. Toeplitz determinants from compatibility conditions. *Ramanujan J.*, 16(1):25–40, 2008.
- [4] D. Benko, S.B. Damelin, and P.D. Dragnev. On the support of the equilibrium measure for arcs of the unit circle and for real intervals. *Electron. Trans. Numer. Anal.*, 25:27–40, 2006.
- [5] G. Blower. *Random matrices: high dimensional phenomena*, volume 367 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2009.
- [6] A. Borodin and G. Olshanski. Infinite random matrices and ergodic measures. *Comm. Math. Phys.*, 223(1):87–123, 2001.
- [7] G. Borot and A. Guionnet. Asymptotic expansion of  $\beta$  matrix models in the one-cut regime. *Comm. Math. Phys.*, 317(2):447–483, 2013.

- [8] P. Bourgade, J. Najnudel, and A. Nikeghbali. A unitary extension of virtual permutations. *Int. Math. Res. Not.*, 18:4101–4134, 2013.
- [9] P. Bourgade, A. Nikeghbali, and A. Rouault. Circular Jacobi ensembles and deformed Verblunsky coefficients. *Int. Math. Res. Not.*, (23):4357–4394, 2009.
- [10] PW Brouwer. Generalized circular ensemble of scattering matrices for a chaotic cavity with nonideal leads. *Physical Review B*, 51(23):16878, 1995.
- [11] M. J. Cantero, L. Moral, and L. Velázquez. Five-diagonal matrices and zeros of orthogonal polynomials on the unit circle. *Linear Algebra Appl.*, 362:29–56, 2003.
- [12] D. Damanik, A. Pushnitski, and B. Simon. The analytic theory of matrix orthogonal polynomials. *Surv. Approx. Theory*, 4:1–85, 2008.
- [13] P. Delsarte, Y.V. Genin, and Y.G. Kamp. Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits and Systems*, pages 149–160, 1978.
- [14] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.
- [15] M. Derevyagin, O. Holtz, S. Khrushchev, and M. Tyaglov. Szegő’s theorem for matrix orthogonal polynomials. *J. Approx. Theory*, 164(9):1238–1261, 2012.
- [16] H. Dette and W. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics, 1997.
- [17] H. Dette and J. Wagoner. Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the Geronimus relations. *Lin. Alg. and its Appl.*, 432:1609–1626, 2010.
- [18] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.
- [19] F. Gamboa, J. Nagel, and A Rouault. Sum rules and large deviations for spectral matrix measures. preprint hal-01264137, version 1.
- [20] F. Gamboa, J. Nagel, and A Rouault. Sum rules via large deviations. *J. Funct. Anal.*, (270):509–559, 2016.
- [21] F. Gamboa and A. Rouault. Canonical moments and random spectral measures. *J. Theoret. Probab.*, 23:1015–1038, 2010.
- [22] F. Gamboa and A. Rouault. Large deviations for random spectral measures and sum rules. *Appl. Math. Res. Express AMRX*, (2):281–307, 2011.
- [23] F. Gamboa and A. Rouault. Operator-valued spectral measures and large deviations. *J. Statist. Plann. Inference*, (154):72–86, 2014.
- [24] D.J. Gross and E. Witten. Possible third-order phase transition in the large-N lattice gauge theory. *Phys. Rev. D*, 21(2):446–453, 1980.

- [25] A. Hardy. A note on large deviations for 2D Coulomb gas with weakly confining potential. *Electron. Commun. Probab*, 17(19):1–12, 2012.
- [26] F. Hiai and D. Petz. *The Semicircle Law, Free Random Variables and Entropy*, volume 77 of *Mathematical Surveys and Monographs*. Amer. Math. Soc., Providence, 2000.
- [27] Luogeng Hua. *Harmonic analysis of functions of several complex variables in the classical domains*. Number 6. American Mathematical Soc., 1963.
- [28] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.
- [29] R. Killip and I. Nenciu. Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701, 2004.
- [30] V. Mandrekar and H. Salehi. On singularity and Lebesgue type decomposition for operator-valued measures. *Journal of Multivariate Analysis*, 1(2):167–185, 1971.
- [31] S. Mizoguchi. On unitary/Hermitian duality in matrix models. *Nuclear Phys. B*, 716(3):462–486, 2005.
- [32] J. Najnudel, A. Nikeghbali, and F. Rubin. Scaled limit and rate of convergence for the largest eigenvalue from the generalized Cauchy random matrix ensemble. *J. Stat. Phys.*, 137(2):373–406, 2009.
- [33] Y. A. Neretin. Hua-type integrals over unitary groups and over projective limits of unitary groups. *Duke Math. J.*, 114(2), 2002.
- [34] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171. American Mathematical Society Providence, RI, 2011.
- [35] Doug Pickrell. Measures on infinite dimensional grassmann manifolds. *Journal of Functional Analysis*, 70(2):323–356, 1987.
- [36] I.E. Pritsker. Weighted energy problem on the unit circle. *Constr. Approx.*, 23(1):103–120, 2005.
- [37] J.B. Robertson and M. Rosenberg. The decomposition of matrix-valued measures. *Michigan Math. J*, 15:353–368, 1968.
- [38] B. Simon. OPUC on one foot. *Bull. Amer. Math. Soc. (N.S.)*, 42(4):431–460, 2005. MR2163705 (2006e:42039).
- [39] B. Simon. *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. Colloquium Publications. American Mathematical Society 54, Part 1. Providence, RI: American Mathematical Society (AMS), 2005.

- [40] B. Simon. *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*. Colloquium Publications. American Mathematical Society 51, Part 2. Providence, RI: American Mathematical Society, 2005.
- [41] B. Simon. *Szegő's theorem and its descendants*. M. B. Porter Lectures. Princeton University Press, Princeton, NJ, 2011.
- [42] S. Wadia. A study of  $U(N)$  lattice gauge theory in 2-dimensions. *arXiv preprint arXiv:1212.2906*, 2012.