



**HAL**  
open science

# Sum rules and large deviations for spectral matrix measures in the Jacobi ensemble

Fabrice Gamboa, Jan Nagel, Alain Rouault

► **To cite this version:**

Fabrice Gamboa, Jan Nagel, Alain Rouault. Sum rules and large deviations for spectral matrix measures in the Jacobi ensemble. 2018. hal-01925648

**HAL Id: hal-01925648**

**<https://hal.science/hal-01925648v1>**

Preprint submitted on 16 Nov 2018

**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 matrix measures in the Jacobi ensemble

Fabrice Gamboa<sup>\*</sup>

Jan Nagel<sup>†</sup>

Alain Rouault<sup>‡</sup>

November 15, 2018

## Abstract

We continue to explore the connections between large deviations for objects coming from random matrix theory and sum rules. This connection was established in [18] for spectral measures of classical ensembles (Gauss-Hermite, Laguerre, Jacobi) and it was extended to spectral matrix measures of the Hermite and Laguerre ensemble in [21]. In this paper, we consider the remaining case of spectral matrix measures of the Jacobi ensemble. Our main results are a large deviation principle for such measures and a sum rule for matrix measures with reference measure the Kesten-McKay law. As an important intermediate step, we derive the distribution of canonical moments of the matrix Jacobi ensemble.

## 1 Introduction

A probability measure on a compact subset of  $\mathbb{R}$  or on the unit circle may be encoded by the sequence of its moments or by the coefficients of the recursion satisfied by the corresponding orthogonal polynomials. It is however not easy to relate information on the measure, (for example on its support), with information on the recursion coefficients. Sum rules give a way to translate between these two languages. Indeed, a sum rule is an identity relating a functional of the

---

<sup>\*</sup>Université Paul Sabatier, Institut de Mathématiques de Toulouse, 31062-Toulouse Cedex 9, France, [fabrice.gamboa@math.univ-toulouse.fr](mailto:fabrice.gamboa@math.univ-toulouse.fr)

<sup>†</sup>Eindhoven University of Technology, Department of Technology and Computer Science, 5600 MB Eindhoven, Netherlands, e-mail: [j.h.nagel@tue.nl](mailto:j.h.nagel@tue.nl)

<sup>‡</sup>Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035-Versailles Cedex France, e-mail: [alain.rouault@uvsq.fr](mailto:alain.rouault@uvsq.fr)

probability measure, usually in the form of a relative entropy, and a functional of its recursion coefficients. The "*measure side*" of the identity gives the discrepancy between the measure and a reference measure and the "*coefficient side*" gives the discrepancy between the corresponding series of recursion coefficients.

One of the most classical example of such a sum rule is the Szegő-Verblunsky theorem for measures on the unit circle  $\mathbb{T}$ , see Chapter 1 of [33]. Here, the reference measure is the uniform measure on  $\mathbb{T}$  and the coefficient side involves a sum of functions of the Verblunsky coefficients. The most famous sum rule for measures on the real line is the Killip-Simon sum rule [27] (see also [33] Section 3.5). In this case, the reference measure is the semicircle distribution. In [18], we gave a probabilistic interpretation of the Killip-Simon sum rule (KS-SR) and a general strategy to construct and prove new sum rules. The starting point is a  $N \times N$  random matrix  $X_N$  chosen according to the Gaussian unitarily invariant ensemble. The random spectral measure  $\mu_N$  of this random matrix is then defined through its moments, by the relation

$$\int x^k d\mu_N = (X_N^k)_{1,1}.$$

It was shown in [22], that as  $N$  tends to infinity, the sequence  $(\mu_N)_N$  satisfies a large deviation principle (LDP). The rate function  $\mathcal{I}_c$  is a functional of the recursion coefficients. Surprisingly, this functional is exactly the coefficient side of the KS-SR. Later, in [18], we gave an alternative proof of this LDP, with a rate function  $\mathcal{I}_m$  that is exactly the measure side of KS-SR. Since a large deviation rate function is unique, this implies the sum rule identity  $\mathcal{I}_c = \mathcal{I}_m$ . Working with a random matrix of one of the other two classical ensembles, the Laguerre and Jacobi ensemble, this method leads to new sum rules. Here the reference measures are the Marchenko-Pastur law and the Kesten-McKay law, respectively [18]. We also refer to recent interesting developments of the method explored in [2] and [3].

One of the ingredient to prove the LDP in terms of the coefficients is the fact that these coefficients are independent and have explicit distributions. To be more precise, it has been shown in [14], that in the Gaussian case the coefficients are independent random variables with normal or gamma distributions. The Laguerre case has also been considered in [14]. In this last frame, the convenient encoding is not directly by the recursion coefficients, but by decomposition of them into independent variables. In [26], a further decomposition is shown for the Jacobi ensemble. Actually these variables are the Verblunsky coefficient of the measure lifted to the unit circle, which are sometimes also called canonical moments, see the monograph [10].

A natural extension of scalar measures are measures with values in the space of Hermitian non-negative definite matrices. There is a rich theory of polynomials orthogonal with respect to such

a matrix measure, and we refer the interested reader to [34], [15], [16] or [5] and references therein. Surprisingly, sum rule identities also hold in the matrix frame. In [4], a matricial version of KS-SR is proved (see also Section 4.6 of [33]). In [21], we have extended our probabilistic method to the matrix case as well, and have proved an LDP for random matrix valued spectral measures. This  $p \times p$  measure  $\Sigma_N$  is now defined by its matrix moments

$$(1.1) \quad \int x^k d\Sigma_N(x) = (X_N^k)_{i,j=1,\dots,p}, \quad k \geq 1,$$

where  $X_N$  is as before a random  $N \times N$  matrix and  $N \geq p$ . Using the explicit construction of random matrices of the Gaussian and Laguerre ensemble, it is possible to derive the distribution of the recursion coefficients of  $\Sigma_N$ , which are now  $p \times p$  matrices, and prove an LDP for them, generalizing the results of [14] and [22]. Collecting these two LDPs and different representations of the rate function, we obtain the matrix sum rule both for Gaussian and Laguerre cases. A large deviation principle for the coefficients in the matricial Jacobi case, and consequently a new sum rule, has been open so far.

In this paper, we complete the trio of matrix measures of classical ensembles by addressing the Jacobi case. We prove an LDP for the spectral matrix measure in Theorem 5.1, which then implies the new matrix sum rule stated in Theorem 3.1. A crucial ingredient for the proof of Theorem 5.1 is Theorem 4.2, where we derive the distribution of matricial canonical moments of  $\Sigma_N$ . Up to our knowledge, this result is new. Actually, we have to consider for our probabilistic approach certain Hermitian versions of the canonical moments and we show that these versions are independent and each distributed as  $p \times p$  matrices of the Jacobi ensemble, thereby generalizing the results of [26] to the matrix case. An additional difficulty is that the measures we need to consider are finitely supported and then are not nontrivial. In this case, many arguments used in the scalar case cannot be extended directly. The fact that there is still a one-to-one correspondence between the spectral measure  $\Sigma_N$  and its canonical moments might therefore be of independent interest.

Let us explain the main obstacle that so far impeded a large deviation analysis of the coefficients in the matrix Jacobi case. For the Gaussian or Laguerre ensemble, the distribution of recursion coefficients can be derived through repeated Householder reflections applied to the full matrix  $X_N$ . In the Jacobi case, it seems impossible to control the effect of these transformation on the different subblocks of  $X_N$ . Instead, looking at the scalar case, there are two potential strategies. First, by identifying the canonical moments as variables appearing in the CS-decomposition of  $X_N$ . In the scalar case, this goes back to [35] and [17]. Any effort to generalize this to a block-CS-decomposition seems to fail due to non-commutativity of the blocks. The other possible strategy is to follow the path of [26] applying the (inverse) Szegő mapping. This yields a symmetric measure on the unit circle  $\mathbb{T}$ . Then apply the Householder algorithm to the corresponding unitary matrix.

Unfortunately, in the matrix case, the Szegő mapping does not give a good symmetric measure on  $\mathbb{T}$  in the matrix case. We refer to Section 4.2 for a discussion of this difficulty. In the present paper, we obtain the distribution of canonical moments by directly computing the Jacobian of a compound map. The first application maps support points and weights of  $\Sigma_N$  to the recursion coefficients. Then the recursion coefficients are mapped to a suitable Hermitian version of the canonical moments. We give two different ways to compute this distribution. One proof follows by direct calculation. The other one is more subtle. It uses the relation between the canonical coefficients and the matrix Verblunsky coefficients.

This paper is organized as follows. In Section 2, we first give notations and explain the different representations for the matrix measures. We also discuss finitely supported matrix measures. In Section 3, we give our new sum rule. Section 4 is devoted to the set up of probability distributions of the matrix models and of the canonical moments. This leads in Section 5 to an LDP for the coefficient side. Section 6 contains the proof of our three main results, subject to technical lemmas, whose proofs are postponed to Section 7.

## 2 Matrix measures and representation of coefficients

All along this paper,  $p$  will be a fixed integer. A  $p \times p$  matrix measure  $\Sigma$  on  $\mathbb{R}$  is a matrix of complex valued Borel measures on  $\mathbb{R}$  such that for every Borel set  $A \subset \mathbb{R}$  the matrix  $\Sigma(A)$  is nonnegative definite, i.e.  $\Sigma(A) \geq 0$ . When its  $k$ -th moment is finite, it is denoted by

$$M_k(\Sigma) = \int x^k d\Sigma(x), \quad k \geq 1,$$

writing  $M_k$  for  $M_k(\Sigma)$  if the measure is clear from context. We keep, as much as possible, the notations close to those of [21]. All matrix measures in this paper will be of size  $p \times p$ . Let  $\mathbf{1}$  be the  $p \times p$  identity matrix and  $\mathbf{0}$  be the  $p \times p$  null matrix. For every integer  $n$ ,  $I_n$  denotes the  $np \times np$  identity matrix. The set of all matrix measures with support in some set  $A$  is denoted by  $\mathcal{M}_p(A)$ , and we let  $\mathcal{M}_{p,1}(A) := \{\Sigma \in \mathcal{M}_p(A) : \Sigma(A) = \mathbf{1}\}$  denoting the set of normalized measures.

For the remainder of this section, let  $\Sigma \in \mathcal{M}_{p,1}(\mathbb{R})$  have compact support. Such a measure  $\Sigma$  can be uniquely described by its sequence of moments  $(M_1(\Sigma), M_2(\Sigma), \dots)$ . Another particular convenient set of parameters characterizing the measure is given by the coefficients in the recursion of orthogonal matrix polynomials, introduced in the following subsection. We will follow largely the exposition developed in [5]. For matrix measures supported by  $[0, 1]$ , there exists, just as in the scalar case, a remarkable decomposition of the recursion coefficients into a set of so-called

canonical moments. The parametrization of  $\Sigma$  in terms of these canonical moments is one of the main tools for our probabilistic results.

## 2.1 Orthogonal matrix polynomials

The (right) inner product of two matrix polynomials  $\mathbf{f}, \mathbf{g}$ , i.e., polynomials whose coefficients are complex  $p \times p$  matrices, is defined by

$$\langle\langle \mathbf{f}, \mathbf{g} \rangle\rangle = \int \mathbf{f}(x)^\dagger d\Sigma(x) \mathbf{g}(x).$$

A matrix measure is called nontrivial, if for any non zero polynomial  $P$  we have

$$(2.1) \quad \text{tr}\langle\langle P, P \rangle\rangle > 0,$$

see Lemma 2.1 of [5] for equivalent characterizations of nontriviality. Let us first suppose that  $\Sigma$  is nontrivial. Lemma 2.3 of [5] shows that then  $\langle\langle Q, Q \rangle\rangle$  is positive definite for any monic polynomial  $Q$  (with leading coefficient  $\mathbf{1}$ ). We may then apply the Gram-Schmidt procedure to  $\{\mathbf{1}, x\mathbf{1}, \dots\}$  and obtain a sequence of monic matrix polynomials  $P_n, n \geq 0$ , where  $P_n$  has degree  $n$  and which are orthogonal with respect to  $\Sigma$ , that is,  $\langle\langle P_n, P_m \rangle\rangle = \mathbf{0}$  if  $m \neq n$ . The polynomials satisfy the recurrence

$$(2.2) \quad xP_n = P_{n+1} + P_n u_n + P_{n-1} v_n, \quad n \geq 0,$$

where, setting

$$(2.3) \quad \gamma_n := \langle\langle P_n, P_n \rangle\rangle,$$

$\gamma_n$  is Hermitian and positive definite, and for  $n \geq 1$

$$(2.4) \quad u_n = \gamma_n^{-1} \langle\langle P_n, xP_n \rangle\rangle, \quad v_n = \gamma_{n-1}^{-1} \gamma_n,$$

with  $v_0 = \mathbf{0}$ . This defines a one-to-one correspondence between the sequence  $(u_0, v_1, u_2, \dots)$  and the measure  $\Sigma$ .

From the matrix coefficients  $u_n, v_n$ , we can then define a sequence of very useful Hermitian matrices. We first define matrices related to orthonormal polynomials recursion. Let for  $n \geq 0$

$$(2.5) \quad \tilde{\mathcal{A}}_{n+1} := \gamma_n^{1/2} v_{n+1} \gamma_{n+1}^{-1/2} = \gamma_n^{-1/2} \gamma_{n+1}^{1/2},$$

$$(2.6) \quad \mathcal{B}_{n+1} := \gamma_n^{1/2} u_n \gamma_n^{-1/2} = \gamma_n^{-1/2} \langle\langle P_n, xP_n \rangle\rangle \gamma_n^{-1/2}.$$

Obviously, setting

$$\mathbf{p}_n = P_n \gamma_n^{-1/2}, \quad n \geq 0,$$

defines a sequence of matrix orthonormal polynomials. These polynomials satisfy the recursion

$$(2.7) \quad x\mathbf{p}_n = \mathbf{p}_{n+1}\tilde{\mathcal{A}}_{n+1}^\dagger + \mathbf{p}_n\mathcal{B}_{n+1} + \mathbf{p}_{n-1}\tilde{\mathcal{A}}_n, \quad n \geq 0),$$

taking  $\mathbf{p}_{-1} = \mathbf{0}$ . The matrices  $\tilde{\mathcal{A}}_n$  and  $\mathcal{B}_n$  play the role of matrix Jacobi coefficients in the following sense. Define the infinite block-tridiagonal matrix

$$(2.8) \quad J = \begin{pmatrix} \mathcal{B}_1 & \tilde{\mathcal{A}}_1 & & \\ \tilde{\mathcal{A}}_1^\dagger & \mathcal{B}_2 & \ddots & \\ & \ddots & \ddots & \end{pmatrix}.$$

On the space of matrix polynomials, the map  $f \mapsto (x \mapsto xf(x))$  is a right homomorphism, represented in the (right-module) basis  $\mathbf{p}_0, \mathbf{p}_1, \dots$  by the matrix  $J$ . Moreover, the measure  $\Sigma$  is nothing more than the spectral measure of the matrix  $J$  defined through its moments by

$$e_i^* \int x^k d\Sigma(x) e_j = e_i^* J e_j, \quad i, j = 1, \dots, p.$$

(See for example Theorem 2.11 of [5]).

The matrix  $\mathcal{B}_n$  is Hermitian and we define the Hermitian square of  $\tilde{\mathcal{A}}_n$  by

$$(2.9) \quad \mathcal{A}_n = \tilde{\mathcal{A}}_n \tilde{\mathcal{A}}_n^\dagger = \gamma_{n-1}^{-1/2} \gamma_n \gamma_{n-1}^{-1/2}.$$

Note that  $\mathcal{A}_n$  is Hermitian positive definite.

## 2.2 Measures on $[0, 1]$

Now suppose that  $\Sigma$  is a nontrivial matrix measure supported by a subset of  $[0, 1]$ . We present two (equivalent) ways to parametrize  $\Sigma$ , extending the corresponding parametrization of the scalar case. The first one uses the canonical moments, the second one uses the Szegő mapping and Verblunsky coefficients.

### 2.2.1 Encoding via canonical coefficients

Detle and Studden [11] proved the following matrix version of Favard's Theorem for measures on  $[0, 1]$ : If  $\Sigma$  has support in  $[0, 1]$ , there exist matrices  $U_n$ ,  $n \geq 1$ , such that the recursion coefficients defined in (2.2) can may be decomposed as

$$(2.10) \quad u_n = \zeta_{2n+1} + \zeta_{2n}, \quad v_n = \zeta_{2n-1} \zeta_{2n}, \quad n \geq 1,$$

where  $\zeta_0 = \mathbf{0}$ ,  $\zeta_1 = U_1$  and for  $n > 1$

$$(2.11) \quad \zeta_n = (\mathbf{1} - U_{n-1})U_n.$$

Moreover,  $U_n$  has the following geometric interpretation. Suppose  $M_1, \dots, M_{n-1}$  are the first  $n-1$  matrix moments of some nontrivial matrix probability measure on  $[0, 1]$ . Then there exist Hermitian matrices  $M_n^-, M_n^+$ , which are upper and lower bounds for the  $n$ -th matrix moment. More precisely,  $M_1, \dots, M_n$  are the first  $n$  moments of some nontrivial measure with support in  $[0, 1]$ , if and only if

$$(2.12) \quad M_n^- < M_n < M_n^+.$$

Here we use the partial Loewner ordering, that is,  $A > B$  ( $A \geq B$ ) for Hermitian matrices  $A, B$ , if and only if  $A - B$  is positive (non-negative) definite. Then, if  $M_n$  are the moments of a nontrivial measure, the following representation holds:

$$(2.13) \quad U_n = (M_n^+ - M_n^-)^{-1}(M_n - M_n^-).$$

So that,  $U_n$  is the relative position of  $M_n$  within the set of all possible  $n$ -th matrix moments, given the matrix moments of lower order. For this reason,  $U_n$  is also called *canonical moment*. Let us define

$$(2.14) \quad R_n = M_n^+ - M_n^-, \quad H_n = M_n - M_n^-,$$

so that  $U_n = R_n^{-1}H_n$ . A Hermitian version of the canonical moments can be defined by

$$(2.15) \quad \mathcal{U}_n = R_n^{1/2}U_nR_n^{-1/2} = R_n^{-1/2}H_nR_n^{-1/2}.$$

The matrices  $\mathcal{U}_n$  have been considered previously in [8], to study asymptotics in the random matrix moment problem. Note that  $U_n$  and  $\mathcal{U}_n$  are similar and

$$\mathbf{0} < \mathcal{U}_n < \mathbf{1}.$$

Finally, we remark that  $M_n^-, M_n^+$  are continuous functions of  $M_1, \dots, M_{n-1}$ , and that

$$(2.16) \quad H_{2n} = \gamma_n.$$

## 2.2.2 Encoding via Szegő mapping

The Szegő mapping is two-one from  $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$  to  $[-2, 2]$  defined by

$$(2.17) \quad z \in \mathbb{T} \mapsto z + z^{-1} \in [-2, 2].$$



This induces a bijection  $\Sigma \mapsto \text{Sz}(\Sigma)$  between matrix probability measures on  $\mathbb{T}$  invariant by  $z \mapsto \bar{z}$  and matrix probability measures on  $[-2, 2]$ . On  $\mathbb{T}$ , a matrix measure is characterized by the system of its matricial Verblunsky coefficients, ruling the recursion of (right) orthogonal polynomials. When the measure is invariant, the Verblunsky coefficients  $(\alpha_n)_{n \geq 0}$  are Hermitian ([5] Lemma 4.1) and satisfy  $\mathbf{0} < \alpha_n^2 < \mathbf{1}$  for every  $n$ .

The Verblunsky coefficients of such a matrix probability measure on  $\mathbb{T}$  and the Jacobi coefficients of the corresponding matrix measure on  $[-2, 2]$  are connected by the Geronimus relations ([5] Theorem 4.2). It is more convenient here to consider the matrix measure on  $[0, 1]$  denoted by  $\tilde{\text{Sz}}(\Sigma)$ , obtained by pushing forward  $\text{Sz}(\Sigma)$  by the affine mapping  $x \mapsto (2 - x)/4$ .

For  $n \geq 0$ , let  $\alpha_n$  be the Verblunsky coefficient of  $\Sigma$  and  $\mathcal{U}_{n+1}$  the Hermitian canonical moment of  $\tilde{\text{Sz}}(\Sigma)$ . Then, the following equality holds:

$$(2.18) \quad \alpha_n = 2\mathcal{U}_{n+1} - 1.$$

The correspondance between the two above encodings is proven in [13], Theorem 4.3, for real-valued matrix measure. The general complex case is considered in [36].

**Remark 2.1** *In the scalar case, the canonical parameters  $U_n$  can be identified in the CS decomposition (see Edelman-Sutton [17]). In the matrix case, this approach does not seem to work, due to the lack of commutativity.*

### 2.3 Finitely supported measures

When the support of  $\Sigma$  consists of  $N = np$  distinct points, then (7.14) cannot be satisfied for all non zero polynomials and  $\Sigma$  is not nontrivial. However, if (7.14) is satisfied for all non zero polynomials of degree at most  $n - 1$ , then actually  $\langle\langle Q, Q \rangle\rangle$  is positive definite for all monic polynomials of degree at most  $n - 1$ , see Lemma 2.3 in [5]. This implies that we can use the Gram-Schmidt method to define monic orthogonal polynomials up to degree  $n$ . Further,  $\gamma_k = \langle\langle P_k, P_k \rangle\rangle$  is positive definite for  $k \leq n - 1$ . Therefore, the orthogonal polynomials allow also to define the recursion coefficients  $u_0, \dots, u_{n-1}; v_1, \dots, v_{n-1}$ . So that, we can construct  $\tilde{\mathcal{A}}_1, \dots, \tilde{\mathcal{A}}_{n-1}; \mathcal{B}_1, \dots, \mathcal{B}_n$  as well, with  $\tilde{\mathcal{A}}_k$  nonsingular for  $k = 1, \dots, n - 1$ . Let us denote by  $J_n$  the  $np \times np$  Hermitian block matrix of Jacobi coefficients

$$(2.19) \quad J_n = \begin{pmatrix} \mathcal{B}_1 & \tilde{\mathcal{A}}_1 & & & \\ \tilde{\mathcal{A}}_1^\dagger & \mathcal{B}_2 & \ddots & & \\ & \ddots & \ddots & \tilde{\mathcal{A}}_{n-1} & \\ & & \tilde{\mathcal{A}}_{n-1}^\dagger & \mathcal{B}_n & \end{pmatrix}.$$

Let  $\Sigma^{J_n}$  denote the spectral measure of  $J_n$ , as defined by (1.1). The same calculation as in the scalar case shows that the first  $2n - 1$  moments of  $\Sigma^{J_n}$  coincide with those of  $\Sigma$ . Since these matrix moments determine uniquely the recursion coefficients of monic orthogonal polynomials, the entries of the matrix (2.19) are then also the recursion coefficients of orthonormal polynomials for  $\Sigma^{J_n}$ .

Now, suppose that the support points of  $\Sigma$  lie in  $[0, 1]$ . The existence of the canonical moments is tackled in the following lemma, proved in Section 7. It requires some additional assumption and is not so obvious.

**Lemma 2.2** *Suppose  $\Sigma \in \mathcal{M}_{p,1}([0, 1])$  is such that  $\text{tr}\langle\langle P, P \rangle\rangle > 0$  for all non zero polynomials of degree at most  $n - 1$ . Suppose further  $\Sigma(\{0\}) = \Sigma(\{1\}) = \mathbf{0}$ . Then, the matrices  $M_k^-, M_k^+$  for  $k \leq 2n - 1$  still exist and they satisfy  $M_k^- < M_k < M_k^+$  for  $k \leq 2n - 1$ . Moreover, the matrices*

$$(2.20) \quad U_k = (M_k^- - M_k^+)^{-1}(M_k - M_k^-), \quad 1 \leq k \leq 2n - 1,$$

*are related to the recursion coefficients  $u_0, \dots, u_{n-1}; v_1, \dots, v_{n-1}$  of  $\Sigma$  as in (2.10) and (2.11).*

Lemma 2.2 implies that we may still define the Hermitian variables  $\mathcal{U}_1, \dots, \mathcal{U}_{2n-1}$ , if the measure  $\Sigma$  is sufficiently nontrivial. In conclusion, for any measure satisfying the assumptions of Lemma 4.1, we have a one-to-one correspondence between:

- matrix moments  $M_1, \dots, M_{2n-1}$ , with  $M_k^- < M_k < M_k^+$  for  $k = 1, \dots, 2n - 1$ ,
- recursion coefficients  $\mathcal{B}_1, \dots, \mathcal{B}_n$  as in (2.6) and positive definite  $\mathcal{A}_1, \dots, \mathcal{A}_{n-1}$  as in (2.9),
- canonical moments  $\mathcal{U}_1, \dots, \mathcal{U}_{2n-1}$  as in (2.15), with  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$  for  $k \leq 2n - 1$ .

### 3 The Jacobi sum rule

The reference measure for the sum rule in the Jacobi case is the matricial version of the Kesten-McKay law. In the scalar case, this measure is defined for parameters  $\kappa_1, \kappa_2 \geq 0$  by

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

where

$$(3.1) \quad u^\pm := \frac{1}{2} + \frac{\kappa_1^2 - \kappa_2^2 \pm 4\sqrt{(1 + \kappa_1)(1 + \kappa_2)(1 + \kappa_1 + \kappa_2)}}{2(2 + \kappa_1 + \kappa_2)^2}.$$

It appears (sometimes in other parametrizations) as a limit law for spectral measures of regular graphs (see [30]), as the asymptotic eigenvalue distribution of the Jacobi ensemble (see [7]), or in the study of random moment problems (see [12]). For  $\kappa_1 = \kappa_2 = 0$ , it reduces to the arcsine law. The matrix version is then denoted by

$$(3.2) \quad \Sigma_{\text{KMK}(\kappa_1, \kappa_2)} := \text{KMK}(\kappa_1, \kappa_2) \cdot \mathbf{1}.$$

The canonical moments of  $\Sigma_{\text{KMK}(\kappa_1, \kappa_2)}$  of even/odd order are given by

$$(3.3) \quad U_{2k} = U_e := \frac{1}{2 + \kappa_1 + \kappa_2} \cdot \mathbf{1}, \quad U_{2k-1} = U_o := \frac{1 + \kappa_1}{2 + \kappa_1 + \kappa_2} \cdot \mathbf{1}.$$

(See [23] Sect. 6 for the scalar case, which can obviously be extended to the matrix case.)

Both sides of our sum rule (Theorem 3.1) will only be finite for measures satisfying a certain condition on their support, related to the Kesten-McKay law. Let  $I = [u^-, u^+]$ . We define  $\mathcal{S}_p = \mathcal{S}_p(u^-, u^+)$  as the set of all bounded nonnegative matrix measures  $\Sigma \in \mathcal{M}_p(\mathbb{R})$  that can be written as

$$(3.4) \quad \Sigma = \Sigma_I + \sum_{i=1}^{N^+} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \Gamma_i^- \delta_{\lambda_i^-},$$

where  $\text{supp}(\Sigma_I) \subset I$ ,  $N^-, N^+ \in \mathbb{N}_0 \cup \{\infty\}$ ,  $\Gamma_i^\pm$  are rank 1 Hermitian matrices and

$$0 \leq \lambda_1^- \leq \lambda_2^- \leq \dots < u^- \quad \text{and} \quad 1 \geq \lambda_1^+ \geq \lambda_2^+ \geq \dots > u^+.$$

We assume that  $\lambda_j^-$  converges towards  $u^-$  (resp.  $\lambda_j^+$  converges to  $u^+$ ) whenever  $N^-$  (resp.  $N^+$ ) is not finite. An atom outside  $[\alpha^-, \alpha^+]$  may appear several times in the decomposition. Its multiplicity is the rank of the total matrix weight that is decomposed in a sum of rank 1 matrices. We also define

$$\mathcal{S}_{p,1} = \mathcal{S}_{p,1}(u^-, u^+) := \{\Sigma \in \mathcal{S}_p(u^-, u^+) \mid \Sigma(\mathbb{R}) = \mathbf{1}\}.$$

Furthermore, the spectral side of the sum rule of Theorem 3.1 involves the relative entropy with respect to the central measure. If  $\Sigma$  has the Lebesgue decomposition

$$(3.5) \quad \Sigma(dx) = h(x)\Sigma_{\text{KMK}}(dx) + \Sigma^s(dx),$$

with  $h$  positive  $p \times p$  Hermitian and  $\Sigma^s$  singular with respect to  $\Sigma_{\text{KMK}}$ , then we define the Kullback-Leibler distance of  $\Sigma_{\text{KMK}}$  with respect to  $\Sigma$  as

$$\mathcal{K}(\Sigma_{\text{KMK}} \mid \Sigma) = - \int \log \det h(x)\Sigma_{\text{KMK}}(dx).$$

Let us remark that if  $K(\Sigma_{\text{KMK}} | \Sigma)$  is finite, then  $h$  is positive definite almost everywhere on  $I$ , which implies that  $\Sigma$  is nontrivial. Conversely, if  $\Sigma$  is trivial, then  $K(\Sigma_{\text{KMK}} | \Sigma)$  is infinite.

Finally, for the contribution of the outlying support points, we define two functionals

$$(3.6) \quad \mathcal{F}_J^+(x) = \begin{cases} \int_{u^+}^x \frac{\sqrt{(t-u^+)(t-u^-)}}{t(1-t)} dt & \text{if } u^+ \leq x \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Similarly, let

$$(3.7) \quad \mathcal{F}_J^-(x) = \begin{cases} \int_x^{u^-} \frac{\sqrt{(u^- - t)(u^+ - t)}}{t(1-t)} dt & \text{if } 0 \leq x \leq u^-, \\ \infty & \text{otherwise.} \end{cases}$$

We are now able to formulate our main result consisting in a sum rule for the matrix Jacobi case.

**Theorem 3.1** *For  $\Sigma \in \mathcal{S}_{p,1}(u^-, u^+)$  a nontrivial measure with canonical moments  $(U_k)_{k \geq 1}$ , we have*

$$(3.8) \quad \mathcal{K}(\Sigma_{\text{KMK}} | \Sigma) + \sum_{i=1}^{N^+} \mathcal{F}_J^+(\lambda_i^+) + \sum_{i=1}^{N^+} \mathcal{F}_J^-(\lambda_i^-) = \sum_{k=1}^{\infty} \mathcal{H}_o(U_{2k+1}) + \mathcal{H}_e(U_{2k})$$

where, for a matrix  $U$  satisfying  $\mathbf{0} \leq U \leq \mathbf{1}$ ,

$$(3.9) \quad \begin{aligned} \mathcal{H}_e(U) &:= -(\log \det U - \log \det U_e) - (1 + \kappa_1 + \kappa_2) (\log \det(\mathbf{1} - U) - \log \det(\mathbf{1} - U_e)), \\ \mathcal{H}_o(U) &:= -(1 + \kappa_1) (\log \det U - \log \det U_o) - (1 + \kappa_2) (\log \det(\mathbf{1} - U) - \log \det(\mathbf{1} - U_o)), \end{aligned}$$

and where both sides may be infinite simultaneously. If  $\Sigma \notin \mathcal{S}_{p,1}(u^-, u^+)$ , the right hand side equals  $+\infty$ .

**Remark 3.2** *The arguments on the right hand side of the sum rule are the canonical moments as they appear in the decomposition of recursion coefficients in (2.10) and (2.11). For some applications, it might be more convenient to work with the Hermitian version as defined in (2.15). Indeed, since  $\mathcal{H}_e, \mathcal{H}_o$  are invariant under similarity transforms, the value of the right hand side does not change when the Hermitian canonical moments  $\mathcal{U}_k$  are considered.*

*We also point out that for trivial measures,  $U_k$  or  $1 - U_k$  will be singular for some  $k$  and then the right hand side equals  $+\infty$  (see also Theorem 5.1). Since in this case the Kullback-Leibler divergence equals  $+\infty$  as well, the equality in Theorem 3.1 is also true for trivial matrix measures.*

As in previous papers, an important consequence of this sum rule is a system of equivalent conditions for finiteness of both sides. It is a *gem*, as defined by Simon in [33] p.19. The following statement is the *gem* implied by Theorem 3.1. We give equivalent conditions on the matrices  $\mathcal{U}_k$  and the spectral measure, which characterize the finiteness of either side in the sum rule identity. The following corollary is the matrix counterpart of Corollary 2.6 in [18]. It follows immediately from Theorem 3.1, since

$$\mathcal{F}_J^\pm(u^\pm \pm h) = \frac{2\sqrt{u^+ - u^-}}{3u^\pm(1 - u^\pm)}h^{3/2} + o(h^{3/2}) \quad (h \rightarrow 0^+)$$

and, for  $H$  similar to a Hermitian matrix,

$$\begin{aligned} \mathcal{H}_e(U_e + H) &= \frac{(2 + \kappa_1 + \kappa_2)^2(\kappa_1 + \kappa_2)}{2(1 + \kappa_1 + \kappa_2)}\text{tr}H^2 + o(\|H\|^2), \\ \mathcal{H}_o(U_o + H) &= \frac{(2 + \kappa_1 + \kappa_2)^2(\kappa_2 - \kappa_1)}{2(1 + \kappa_1)(1 + \kappa_2)}\text{tr}H^2 + o(\|H\|^2), \end{aligned}$$

as  $\|H\| \rightarrow 0$ , where  $\|\cdot\|$  is any matrix norm.

**Corollary 3.3** *Let  $\Sigma$  be a nontrivial matrix probability measure on  $[0, 1]$  with canonical moments  $(U_k)_{k \geq 1}$ . Then for any  $\kappa_1, \kappa_2 \geq 0$ ,*

$$(3.10) \quad \sum_{k=1}^{\infty} [\text{tr}(U_{2k-1} - U_o)^2 + \text{tr}(U_{2k} - U_e)^2] < \infty$$

*if and only if the three following conditions hold:*

1.  $\Sigma \in \mathcal{S}_{p,1}(u^-, u^+)$
2.  $\sum_{i=1}^{N^+} (\lambda_i^+ - u^+)^{3/2} + \sum_{i=1}^{N^-} (u^- - \lambda_i^-)^{3/2} < \infty$  and additionally, if  $N^- > 0$ , then  $\lambda_1^- > 0$  and if  $N^+ > 0$ , then  $\lambda_1^+ < 1$ .
3. Writing the Lebesgue decomposition of  $\Sigma$  as in (3.5), then

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

## 4 Randomization: Classical random matrix ensembles and their spectral measures

To prove the sum rule of Theorem 3.1 by our probabilistic method, we start from some random Hermitian matrix  $X_N$  of size  $N = np$ . The random spectral measure  $\Sigma_N$  associated with

$(X_N; e_1, \dots, e_p)$ , is defined through its matrix moments:

$$(4.1) \quad M_k(\Sigma_n)_{i,j} = e_i^\dagger X_N^k e_j, \quad k \geq 0, \quad 1 \leq i, j \leq p,$$

where  $e_1, \dots, e_N$  is the canonical basis of  $\mathbb{C}^N$ . From the spectral decomposition of  $X_N$ , we see that the matrix measure  $\Sigma_N$  is

$$(4.2) \quad \Sigma_N = \sum_{j=1}^N \mathbf{v}_j \mathbf{v}_j^\dagger \delta_{\lambda_j},$$

where the support is given by the eigenvalues of  $X_N$  and  $\mathbf{v}_j$  is the projection of a unitary eigenvector corresponding to the eigenvalue  $\lambda_j$  on the subspace generated by  $e_1, \dots, e_p$ . A sum rule is then a consequence of two LDPs for the sequence  $(\Sigma_N)_n$ , the first one when the measure is encoded by its support and the weight, as in (4.2), and the second one when the measure is encoded by its recursion coefficients. The two following questions are therefore crucial:

- What is the joint distribution of  $(\lambda_1, \dots, \lambda_N; \mathbf{v}_1, \dots, \mathbf{v}_N)$ ?
- What is the distribution of the matricial recursion or canonical coefficients?

The answer to the first question is now classical (see [31] or [1]), when  $X_N$  is chosen according to a density (the joint density of all real entries, up to symmetry constraint) proportional to

$$(4.3) \quad \exp(-N \text{tr} V(X)),$$

for some potential  $V$ . In this case, the eigenvalues follow a log-gas distribution and independently, the eigenvector matrix is Haar distributed on the unitary group. In [21], the authors considered such general potentials and proved an LDP using the encoding by eigenvalues and weights. For  $X_N$  distributed according to the Hermite and Laguerre ensemble, it is also possible to answer the second question and derive the LDPs in both encodings. Remarkably, the recursion coefficients in the Hermite case are independent and are  $p \times p$  matrices of the Hermite and Laguerre ensemble. In the Laguerre case, Hermitian version of the matrices  $\zeta_k$  as in (2.10) are Laguerre-distributed. In this section, we give the answer to the second question, when  $X_N$  is a matrix of the Jacobi ensemble. We first introduce all classical ensembles.

## 4.1 The classical ensembles: GUE, LUE, JUE

We denote by  $\mathcal{N}(0, \sigma^2)$  the centered Gaussian distribution with variance  $\sigma^2 > 0$ . A random variable  $X$  taking values in  $\mathcal{H}_N$ , the set of all Hermitian  $N \times N$  matrices, is distributed according

to the Gaussian unitary ensemble  $\text{GUE}_N$ , if all real diagonal entries are distributed as  $\mathcal{N}(0, 1)$  and the real and imaginary parts of off-diagonal variables are independent and  $\mathcal{N}(0, 1/2)$  distributed (also called complex standard normal distribution). All entries are assumed to be independent up to symmetry and conjugation. The random matrix  $X$  has then a density as in (4.3) with  $V(x) = \frac{1}{2}x^2$ . The joint density of the (real) eigenvalues  $\lambda = (\lambda_1, \dots, \lambda_N)$  of  $X$  is

$$(4.4) \quad g_G(\lambda) = c_r^H \Delta(\lambda)^2 \prod_{i=1}^N e^{-\lambda_i^2/2}.$$

where

$$\Delta(\lambda) = \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|$$

is the Vandermonde determinant.

By analogy with the scalar  $\chi^2$  distribution, the Laguerre ensemble is the distribution of the "square" of Gaussian matrices. More precisely, if  $a$  is a nonnegative integer and if  $G$  denotes a  $N \times (N + a)$  matrix with independent complex standard normal entries, then  $X = GG^\dagger$  is said to be distributed according to the Laguerre ensemble  $\text{LUE}_N(N + a)$ . Its density (on the set  $\mathcal{H}_N^+$  of positive definite Hermitian matrices) is proportional to

$$(\det X)^a \exp\left(-\frac{1}{2}\text{tr} X\right).$$

The eigenvalues density in this case is

$$(4.5) \quad g_L(\lambda) = c_{N,a}^L \Delta(\lambda)^2 \prod_{i=1}^N \lambda_i^a e^{-\lambda_i} \mathbb{1}_{\{\lambda_i > 0\}}.$$

For  $a, b$  nonnegative integers, let  $L_1$  and  $L_2$  be independent matrices distributed according to  $\text{LUE}_N(N + a)$  and  $\text{LUE}_N(N + b)$ , respectively. Then the Jacobi ensemble  $\text{JUE}_N(a, b)$  is the distribution of

$$(4.6) \quad X = (L_1 + L_2)^{-1/2} L_1 (L_1 + L_2)^{-1/2}.$$

Its density on the set of Hermitian  $N \times N$  matrices satisfying  $0 < X < I_N$  is proportional to

$$(4.7) \quad \det X^a \det(I_N - X)^b.$$

The density of the eigenvalues  $(\lambda_1, \dots, \lambda_N)$  is then given by

$$(4.8) \quad g_J(\lambda) = c_{N,a,b}^J |\Delta(\lambda)|^2 \prod_{i=1}^N \lambda_i^a (1 - \lambda_i)^b \mathbb{1}_{\{0 < \lambda_i < 1\}}.$$

By extension we will say that  $X$  is distributed according to  $\text{JUE}_N(a, b)$ , if it has density (4.7), for general real parameters  $a, b \geq 0$ .

As mentioned above, in all three cases the eigenvector matrix is independent of the eigenvalues and Haar distributed on the group of unitary matrices. As a consequence, the matrix weights in the spectral measure (see (4.2)) have a distribution which is a matricial generalization of the Dirichlet distribution. Let us denote the distribution of  $(\mathbf{v}_1 \mathbf{v}_1^\dagger, \dots, \mathbf{v}_N \mathbf{v}_N^\dagger)$  by  $\mathbb{D}_{N,p}$ . It was shown in [24], that this distribution may be obtained as follows: Let  $z_1, \dots, z_N$  be random vectors in  $\mathbb{C}^p$ , with all coordinates independent complex standard normal distributed and set  $H = z_1 z_1^\dagger + \dots + z_N z_N^\dagger$ . Then we have the equality in distribution

$$(4.9) \quad (\mathbf{v}_1 \mathbf{v}_1^\dagger, \dots, \mathbf{v}_N \mathbf{v}_N^\dagger) \stackrel{d}{=} (H^{-1/2} z_1 z_1^\dagger H^{-1/2}, \dots, H^{-1/2} z_N z_N^\dagger H^{-1/2}).$$

Using this representation, we can prove the following useful lemma, which shows that although our random spectral measures are finitely supported and thus not nontrivial, it is still possible to define the first recursion coefficients or canonical moments.

**Lemma 4.1** *Let  $N = np$  and  $\Sigma_N$  be a random spectral measure as in (4.2). We assume that there are almost surely  $N$  distinct support points and that the weights are  $\mathbb{D}_{N,p}$  distributed and independent of the support points. Then, with probability one, for all nonzero matrix polynomials  $P$  of degree at most  $n - 1$ ,*

$$\text{tr}\langle\langle P, P \rangle\rangle > 0.$$

## 4.2 Distribution of coefficients

In the following, let  $N = np$ . If  $\Sigma_N$  is a spectral matrix measure of a matrix  $X_N \sim \text{GUE}_N$ , then, almost surely, the  $N$  support points of  $\Sigma_N$  are distinct and none of them equal 0 or 1. By Lemma 4.1 and the discussion in Section 2.3,  $\Sigma_N$  may be encoded by its first  $2n - 1$  coefficients in the polynomial recursion. It is known that then the random matrices  $\mathcal{B}_1, \dots, \mathcal{B}_n, \mathcal{A}_1, \dots, \mathcal{A}_{n-1}$  are independent and

$$\mathcal{A}_k \sim \text{LUE}_p((N - k)p), \quad \mathcal{B}_k \sim \text{GUE}_p.$$

For the Laguerre ensemble, the spectral measure is supported by  $[0, \infty)$  and then a decomposition as in (2.10) still holds, where now Hermitian versions of  $\zeta_1, \dots, \zeta_{2n-1}$  are distributed according to the Laguerre ensemble of dimension  $p$  with appropriate parameter. These results may be seen in [21], Lemma 6.1 and 6.2. They are extensions of the scalar results of Dumitriu-Edelman [14] and their proofs are in [19]. Since therein they are formulated in a slightly different way, we clarify



the arguments in the Hermite case when we prove Theorem 4.2 below. It is one of our main results, and shows that in the Jacobi case, the matricial canonical moments are independent and again distributed as matrices of the Jacobi ensemble.

**Theorem 4.2** *Let  $\Sigma_N$  be the random spectral matrix measure associated with the  $\text{JUE}_N(a, b)$  distribution. Then, the Hermitian canonical moments  $\mathcal{U}_1, \dots, \mathcal{U}_{2n-1}$  are independent and for  $k = 1, 2, \dots, n-1$ ,*

(4.10)

$$\mathcal{U}_{2k-1} \sim \text{JUE}_p(p(n-k) + a, p(n-k) + b), \quad \mathcal{U}_{2k} \sim \text{JUE}_p(p(n-k-1), p(n-k) + a + b)$$

and  $\mathcal{U}_{2n-1} \sim \text{JUE}_p(a, b)$ .

The Jacobi scalar case was solved by Killip and Nenciu [26]. They used the inverse Szegő mapping and actually considered the symmetric random measure on  $\mathbb{T}$  as the spectral measure of  $(U; e_1)$  where  $U$  is an element of  $\mathbb{S}\mathbb{O}(2N)$  and  $e_1$  is the first vector of the canonical basis. This measure may be written as

$$\mu = \sum_{k=1}^N \bar{\mathbf{w}}_k (\delta_{e^{i\theta_k}} + \delta_{e^{-i\theta_k}}).$$

Under the Haar measure, the support points (or eigenvalues) have the joint density proportional to

$$\Delta(\cos \theta_1, \dots, \cos \theta_N)^2$$

and the weights are Dirichlet distributed. This induces for the pushed forward eigenvalues a density proportional to

$$\Delta(\lambda)^2 \prod_{i=1}^N \lambda_i^{-1/2} (1 - \lambda_i)^{-1/2}$$

Then they used a "magic relation" to get rid of the factor  $\prod \lambda_i^{a-1/2} (1 - \lambda_i)^{b-1/2}$ .

If we consider the matricial case, i.e. if we sample  $U$  according to the Haar measure on  $\mathbb{S}\mathbb{O}(2Np)$  with  $(p \geq 2)$ , the matrix spectral measure of  $(U; e_1, \dots, e_p)$  is now

$$\Sigma = \sum_{k=1}^N (\bar{\mathbf{w}}_k \delta_{e^{i\theta_k}} + \bar{\mathbf{w}}_k \delta_{e^{-i\theta_k}}),$$

the eigenvectors of conjugate eigenvalues being conjugate of each other. Unfortunately, this measure is symmetric (i.e. invariant by  $z \mapsto \bar{z}$ ) only in the scalar case  $p = 1$ , which prohibits the use of the Szegő mapping. To find the distribution of the canonical moments, we have to follow

another strategy. First, we will use the explicit relation between the distribution of eigenvalues and weights and the distribution of the recursion coefficients, when sampling the matrix in the Gaussian ensemble. Then we will compute the Jacobian of the mapping from recursion coefficients to canonical moments using the representation in terms of moments as in (2.15).

## 5 Large deviations

In order to be self-contained, let us recall the definition of a large deviation principle. For a general reference on large deviation statements we refer to the book [6] or to the Appendix D of [1].

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

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

The rate function  $\mathcal{I}$  is good if its level sets  $\{x \in E \mid \mathcal{I}(x) \leq a\}$  are compact for all  $a \geq 0$ . We say that a sequence of  $E$ -valued random variables satisfies an LDP if their distributions satisfy an LDP.

It was shown in Theorem 3.2 of [21], that the sequence of matrix spectral measures  $\Sigma_N$  of the Jacobi ensemble  $\text{JUE}_N(\kappa_1 N, \kappa_2 N)$  satisfies an LDP with speed  $N$  and good rate function equal to the left hand side of the sum rule in Theorem 3.1. The LDP for the coefficient side is given in the following theorem. Its proof is independent of the one given in [21].

**Theorem 5.1** *Let  $\Sigma_N$  be a random spectral matrix measure of the Jacobi ensemble  $\text{JUE}_N(\kappa_1 N, \kappa_2 N)$ , with  $\kappa_1, \kappa_2 \geq 0$  and  $N = pn$ . Then the sequence  $(\Sigma_N)_N$  satisfies the LDP in  $\mathcal{M}_{p,1}([0, 1])$ , with speed  $N$  and good rate function*

$$(5.1) \quad \mathcal{I}_J(\Sigma) = \sum_{k=1}^{\infty} \mathcal{H}_o(U_{2k-1}) + \mathcal{H}_e(U_{2k})$$

for nontrivial  $\Sigma$ , where  $\mathcal{H}_o$  and  $\mathcal{H}_e$  are defined in (3.9) and  $U_k, k \geq 1$  are the canonical moments of  $\Sigma$ . If  $\Sigma$  is trivial, then  $\mathcal{I}_J(\Sigma) = +\infty$ .

The following lemma shows an LDP for the Jacobi ensemble of fixed size. It is crucial in proving the LDP for the canonical moments and consequently Theorem 5.1.

**Lemma 5.2** *For  $\alpha, \alpha' > 0$  suppose that  $X_n \sim \text{JUE}_p(\alpha n + a, \alpha' n + b)$ . Then  $(X_n)_n$  satisfies the LDP in the set of Hermitian  $p \times p$  matrices, with speed  $n$  and good rate function  $I_{\alpha, \alpha'}$  where*

$$(5.2) \quad I_{\alpha, \alpha'}(X) = -\alpha \log \det X - \alpha' \log \det(\mathbf{1} - X) + p\alpha \log \frac{\alpha}{\alpha + \alpha'} + p\alpha' \log \frac{\alpha'}{\alpha + \alpha'}$$

for  $\mathbf{0} < X < \mathbf{1}$  and  $I_{\alpha, \alpha'}(X) = \infty$  otherwise.

The proof of Lemma 5.2 makes use of the explicit density and follows as Proposition 6.6 in [20].

## 6 Proof of the main results

In this section we prove our three main results in the order of their dependence. First, Theorem 4.2 provides the distribution of the canonical moments for the Jacobi ensemble, then Theorem 5.1 shows the LDP for the spectral measure of the Jacobi ensemble, and finally Theorem 3.1 establishes the sum rule for the Jacobi case. For these three proofs, we use the result of all our technical lemmas, whose proofs are postponed to Section 7.

### 6.1 Proof of Theorem 4.2

The starting point is the spectral measure

$$(6.1) \quad \Sigma_N = \sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^\dagger \delta_{\lambda_i},$$

when the distribution of  $(\lambda, \mathbf{v}) = (\lambda_1, \dots, \lambda_{np}, \mathbf{v}_1, \dots, \mathbf{v}_{np})$  is the probability measure proportional to

$$(6.2) \quad \left( \Delta(\lambda)^2 \prod_{i=1}^N \lambda_i^a (1 - \lambda_i)^b \mathbb{1}_{\{0 < \lambda_i < 1\}} d\lambda_j \right) d\mathbb{D}_{N,p}(\mathbf{v}).$$

We need to calculate the pushforward of this measure under the mapping  $(\lambda, \mathbf{v}) \mapsto \mathcal{U} = (\mathcal{U}_1, \dots, \mathcal{U}_{2n-1})$  to the Hermitian canonical moments. By Lemma 4.1 and Lemma 2.2 this is well-defined and the canonical moments satisfy  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$ . The first step will be the computation of the pushforward under the mapping  $(\lambda, \mathbf{v}) \mapsto (\mathcal{A}, \mathcal{B})$ , when  $(\mathcal{A}, \mathcal{B}) := (\mathcal{A}_1, \dots, \mathcal{A}_{n-1}, \mathcal{B}_1, \dots, \mathcal{B}_n)$

are the Hermitian recursion coefficients as defined in (2.6) and (2.9). This can be done by considering the corresponding change of measure in the Gaussian case, that is, when  $\Sigma_N$  is the spectral measure of a  $\text{GUE}_N$ -distributed matrix with distribution proportional to

$$(6.3) \quad \left( \Delta(\lambda)^2 \prod_{i=1}^N e^{-\frac{1}{2}\lambda_i^2} d\lambda_j \right) d\mathbb{D}_{N,p}(\mathbf{v}).$$

As mentioned in Section 4, the correspondence in the Gaussian case was investigated in [21]. Lemma 6.1 therein shows that the spectral matrix measure  $\Sigma_N$  is also the spectral matrix measure of the block-tridiagonal matrix

$$(6.4) \quad \hat{J}_n = \begin{pmatrix} D_1 & C_1 & & & \\ C_1 & D_2 & \ddots & & \\ & \ddots & \ddots & C_{n-1} & \\ & & C_{n-1} & D_n & \end{pmatrix},$$

where  $C_k, D_k$  are Hermitian and independent, with  $D_k \sim \text{GUE}_p$  and  $C_k$  is positive definite with  $C_k^2 \sim \text{LUE}_p(p(n-k))$ . This implies that the Hermitian recursion coefficients  $\mathcal{B}_k$  and  $\mathcal{A}_k$  are given by  $\mathcal{B}_k = D_k$  and  $\mathcal{A}_k = C_k^2$ , respectively. That is, the pushforward of the measure (6.3) under the mapping  $(\lambda, \mathbf{v}) \mapsto (\mathcal{A}, \mathcal{B})$  is the measure proportional to

$$(6.5) \quad \left( \prod_{k=1}^n \exp\left(-\frac{1}{2}\text{tr}\mathcal{B}_k^2\right) d\mathcal{B}_k \right) \left( \prod_{k=1}^{n-1} (\det \mathcal{A}_k)^{p(n-k-1)} \exp\left(-\frac{1}{2}\text{tr}\mathcal{A}_k\right) d\mathcal{A}_k \right).$$

Here and in the following,  $dM$  denotes the Lebesgue measure in each of the functionally independent real entries of a Hermitian matrix  $M$ . Since

$$\text{tr}\hat{J}_n^2 = \sum_{k=1}^n \text{tr}\mathcal{B}_k^2 + \sum_{k=1}^{n-1} \text{tr}\mathcal{A}_k = \sum_{j=1}^N \lambda_j^2,$$

we conclude that the pushforward of the measure

$$(6.6) \quad \left( \Delta(\lambda)^2 \prod_{i=1}^N \mathbb{1}_{\{0 < \lambda_i < 1\}} d\lambda_i \right) d\mathbb{D}_{N,p}(\mathbf{w})$$

by the mapping  $(\lambda, \mathbf{w}) \mapsto (\mathcal{A}, \mathcal{B})$  is, up to a multiplicative constant, the measure

$$(6.7) \quad \left( \prod_{k=1}^{n-1} (\det \mathcal{A}_k)^{p(n-k-1)} d\mathcal{A}_k \right) \prod_{k=1}^n d\mathcal{B}_k.$$

Note that an indicator function is omitted in (6.7), (ensuring that the spectral measure is supported by  $(0, 1)$ ). This indicator function will appear in the condition  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$ , but it does not play a role in the following arguments.

Now two steps are remaining. First, we need to compute the pushforward of (6.7) under the mapping  $(\mathcal{A}, \mathcal{B}) \mapsto \mathcal{U} := (\mathcal{U}_1, \dots, \mathcal{U}_{2n-1})$ . Second, to express the prefactor  $\prod_{i=1}^{np} \lambda_i^a (1 - \lambda_i)^b$  in (6.2) in terms of  $\mathcal{U}$ . This is summarized in the two following technical lemmas, whose proofs are in Section 7.3 and 7.4, respectively.

**Lemma 6.1** *The pushforward of the measure (6.7) by the mapping  $(\mathcal{A}, \mathcal{B}) \mapsto \mathcal{U}$  is, up to a multiplicative constant, the measure*

$$(6.8) \quad \left( \prod_{k=1}^{n-1} \det((\mathbf{1} - \mathcal{U}_{2k-1})\mathcal{U}_{2k-1})^{p(n-k)} d\mathcal{U}_{2k-1} \right) \left( \prod_{k=1}^{n-1} \det(\mathbf{1} - \mathcal{U}_{2k})^{p(n-k)} \det(\mathcal{U}_{2k})^{p(n-k)} d\mathcal{U}_{2k} \right)$$

**Lemma 6.2**

$$(6.9) \quad \prod_{i=1}^{np} (1 - \lambda_i) = \prod_{k=1}^{2n-1} \det(\mathbf{1} - \mathcal{U}_k), \quad \prod_{i=1}^{np} \lambda_i = \prod_{k=1}^n \det \mathcal{U}_{2k-1} \prod_{k=1}^{n-1} \det(\mathbf{1} - \mathcal{U}_{2k}).$$

Gathering these results we see that the pushforward of the measure (6.2) by the mapping  $(\lambda, \mathbf{w}) \mapsto \mathcal{U}$  is, again up to a multiplicative constant,

$$(6.10) \quad \prod_{k=1}^n \det(\mathcal{U}_{2k-1})^{p(n-k)+a} \det(\mathbf{1} - \mathcal{U}_{2k-1})^{p(n-k)+b} \prod_{k=1}^{n-1} \det(\mathcal{U}_{2k})^{p(n-k-1)} \det(\mathbf{1} - \mathcal{U}_{2k})^{p(n-k)+a+b} \prod_{k=1}^{2n-1} d\mathcal{U}_k.$$

That is, the canonical moments are independent and

$$\mathcal{U}_{2k-1} \sim \text{JUE}_p(p(n-k) + a, p(n-k) + b), \quad \mathcal{U}_{2k} \sim \text{JUE}_p(p(n-k-1), p(n-k) + a + b).$$

This ends the proof.  $\square$

## 6.2 Proof of Theorem 5.1

Let  $\Sigma_N$  be the spectral measure of a  $\text{JUE}_N(\kappa_1 N, \kappa_2, N)$  distributed matrix, with  $N = np$  and  $\kappa_1, \kappa_2 \geq 0$ . By Lemma 4.1 and Lemma 2.2, the first  $2n-1$  canonical moments  $U_k^{(N)}$   $1 \leq k \leq 2n-1$  and their Hermitian versions  $\mathcal{U}_k^{(N)}$ ,  $1 \leq k \leq 2n-1$  are well-defined. They are elements of the space

$$(6.11) \quad \mathcal{Q}_j = \{(H_1, \dots, H_{2j-1}) \mid H_j \in \mathcal{H}_p \text{ and } \mathbf{0} \leq H_j \leq \mathbf{1} \text{ for all } j\}.$$

Let us define the sequence

$$(6.12) \quad \mathcal{U}^{(N)} = (\mathcal{U}_1^{(N)}, \dots, \mathcal{U}_{2n-1}^{(N)}, \mathbf{0}, \dots),$$

as a random element of

$$(6.13) \quad \mathcal{Q}_\infty = \{(H_1, H_2, \dots) \mid H_j \in \mathcal{H}_p \text{ and } \mathbf{0} \leq H_j \leq \mathbf{1} \text{ for all } j\},$$

which we endow with the product topology. By Theorem 4.2,

$$\mathcal{U}_{2k-1}^{(N)} \sim \text{JUE}_p(p(n-k) + \kappa_1 np, p(n-k) + \kappa_2 np)$$

for  $1 \leq k \leq n$ , and then we apply Lemma 5.2, to conclude that the sequence  $(\mathcal{U}_{2k-1}^{(N)})_n$  satisfies the LDP in  $\mathcal{Q}_1$  with speed  $n$  and good rate function  $I_{p+p\kappa_1, p+p\kappa_2}$ . If we instead consider the LDP at speed  $N$ , the rate function becomes

$$\begin{aligned} p^{-1}I_{p+p\kappa_1, p+p\kappa_2}(\mathcal{U}) &= -(1 + \kappa_1) \log \det(\mathcal{U}) - (1 + \kappa_2) \log \det(\mathbf{1} - \mathcal{U}) \\ &\quad + p(1 + \kappa_1) \log \frac{1 + \kappa_1}{2 + \kappa_1 + \kappa_2} + p(1 + \kappa_2) \log \frac{1 + \kappa_2}{2 + \kappa_1 + \kappa_2}, \end{aligned}$$

where the right hand side is interpreted as  $+\infty$ , if we do not have  $\mathbf{0} < \mathcal{U} < \mathbf{1}$ . Recalling (3.3) and (3.9), we see that  $p^{-1}I_{p+p\kappa_1, p+p\kappa_2} = \mathcal{H}_o$ . Turning to the canonical moments of even index, Theorem 4.2 gives,

$$\mathcal{U}_{2k}^{(N)} \sim \text{JUE}_p(p(n-k-1), p(n-k) + \kappa_1 np + \kappa_2 np)$$

for  $1 \leq k \leq n-1$ . Then Lemma 5.2 yields the LDP for  $(\mathcal{U}_{2k}^{(N)})_n$  in  $\mathcal{Q}_1$  with speed  $N$  and good rate function  $p^{-1}I_{p, p+p\kappa_1+p\kappa_2}$ , satisfying

$$\begin{aligned} p^{-1}I_{p, p+p\kappa_1+p\kappa_2}(\mathcal{U}) &= -\log \det(\mathcal{U}) - (1 + \kappa_1 + \kappa_2) \log \det(\mathbf{1} - \mathcal{U}) \\ &\quad + p \log \frac{1}{2 + \kappa_1 + \kappa_2} + p(1 + \kappa_1 + \kappa_2) \log \frac{1 + \kappa_1 + \kappa_2}{2 + \kappa_1 + \kappa_2} \\ &= \mathcal{H}_e(\mathcal{U}). \end{aligned}$$

Since the canonical moments are independent, we get for any  $j \geq 1$ , that  $(\mathcal{U}_1^{(N)}, \dots, \mathcal{U}_{2j-1}^{(N)})_{n \geq j}$  satisfies the LDP in  $\mathcal{Q}_j$  with speed  $N$  and good rate function

$$\mathcal{I}^{(j)}(\mathcal{U}_1, \dots, \mathcal{U}_{2j-1}) = \mathcal{H}_o(\mathcal{U}_1) + \mathcal{H}_e(\mathcal{U}_2) + \dots + \mathcal{H}_o(\mathcal{U}_{2j-1}).$$

We can now apply the projective method of the Dawson-Gärtner Theorem (see Theorem 4.6.1 in [6]). It yields the LDP for the full sequence  $\mathcal{U}^{(N)}$  in  $\mathcal{Q}_\infty$ , with speed  $N$  and good rate function

$$(6.14) \quad \mathcal{I}_\infty(\mathcal{U}_1, \mathcal{U}_2, \dots) = \sup_{j \geq 1} \mathcal{I}^{(j)}(\mathcal{U}_1, \dots, \mathcal{U}_{2j-1}) = \sum_{k=1}^{\infty} \mathcal{H}_o(\mathcal{U}_{2k-1}) + \mathcal{H}_e(\mathcal{U}_{2k}).$$

This rate function is finite only if  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$  for all  $k$ . In particular, the set where it is finite is a subset of the space

$$(6.15) \quad \widehat{\mathcal{Q}}_\infty = \{H \mid \mathbf{0} < H < \mathbf{1}\}^{\mathbb{N}} \cup \bigcup_{j=1}^{\infty} \left( \{H \mid \mathbf{0} < H < \mathbf{1}\}^{2j-1} \times \{\mathbf{0}\}^{\mathbb{N}} \right).$$

We also have  $\mathcal{U}^{(N)} \in \widehat{\mathcal{Q}}_\infty$  for all  $n$ , see (6.12). It follows from Lemma 4.1.5 in [6], that  $\mathcal{U}^{(N)}$  also satisfies the LDP in  $\widehat{\mathcal{Q}}_\infty$ , with speed  $N$  and good rate function the restriction of  $\mathcal{I}_\infty$  to this space. Then, we define the mapping  $\psi : \widehat{\mathcal{Q}}_\infty \rightarrow \mathcal{M}_{p,1}([0,1])$  as follows. If  $\mathcal{U} \in \widehat{\mathcal{Q}}_\infty$  is such that  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$  for all  $k$ , there is a unique nontrivial  $\Sigma \in \mathcal{M}_{p,1}([0,1])$ , such that  $\Sigma$  has Hermitian canonical moments  $\mathcal{U}$ , and we define  $\psi(\mathcal{U}) = \Sigma$ . If  $\mathcal{U}$  is such that  $\mathbf{0} < \mathcal{U}_{2j-1} < \mathbf{1}$ , but  $\mathcal{U}_k = \mathbf{0}$  for  $k > 2j - 1$ , we use the correspondence from Section 2.3: then there are moments  $M_1, \dots, M_{2j-1}$  with  $M_k^- < M_k^+$  for  $k \leq 2j - 1$ , and we define  $\psi(\mathcal{U})$  as the spectral measure of the block Jacobi matrix  $J_j$  as in (2.19), constructed with these moments. That is,  $\psi(\mathcal{U})$  is the unique spectral measure of such a Jacobi matrix with first canonical moments  $\mathcal{U}_1, \dots, \mathcal{U}_{2j-1}$ . Then  $\mathcal{U}_n \rightarrow \mathcal{U}$  implies that the block-Jacobi matrix of  $\psi(\mathcal{U}_n)$  converges entrywise to the block-Jacobi matrix of  $\psi(\mathcal{U})$ , where the latter one is extended by zeros if  $\mathcal{U}$  has less nonzero matricial entries than  $\mathcal{U}_n$ . This implies that the moments of  $\psi(\mathcal{U}_n)$  converge to the moments of  $\psi(\mathcal{U})$ . Since the convergence of moments of matrix measures on the compact set  $[0,1]$  implies weak convergence, the mapping  $\psi$  is continuous.

To end the proof, we now apply the contraction principle (Theorem 4.2.1 in [6]). We have  $\psi(\mathcal{U}^{(N)}) = \Sigma_N$ , and as  $\psi$  is continuous, the sequence  $(\Sigma_N)_n$  satisfies the LDP in  $\mathcal{M}_{p,1}([0,1])$  with speed  $N$  and good rate function

$$(6.16) \quad \mathcal{I}_J(\Sigma) = \inf_{\mathcal{U}: \psi(\mathcal{U}) = \Sigma} \mathcal{I}_\infty(\mathcal{U}).$$

This infimum is infinite, unless  $\Sigma$  is nontrivial, and in this case it is given by  $\mathcal{I}_\infty$  evaluated at the unique sequence of canonical moments of  $\Sigma$ .  $\square$

### 6.3 Proof of Theorem 3.1

Let  $\Sigma_N$  be the random spectral matrix measure of a matrix with distribution  $\text{JUE}_N(\kappa_1 N, \kappa_2 N)$ , with  $\kappa_1, \kappa_2 \geq 0$ , and suppose  $N = np$ . This distribution corresponds to a random matrix with potential

$$(6.17) \quad V(x) = -\kappa_1 \log(x) - \kappa_2 \log(1-x),$$

see (4.3). In the scalar case  $p = 1$ , the equilibrium measure (the minimizer of the Voiculescu entropy or the limit of  $\Sigma_N$ ) is given by  $\text{KMK}(\kappa_1, \kappa_2)$ , see [18], p. 515. For this potential, the assumptions (A1), (A2) and (A3) in [21] are satisfied, with matrix equilibrium measure  $\Sigma_V = \Sigma_{\text{KMK}(\kappa_1, \kappa_2)}$  and then by Theorem 3.2 of that paper, the sequence  $(\Sigma_N)_n$  satisfies the LDP in  $\mathcal{M}_{p,1}(\mathbb{R})$  with speed  $N$  and good rate function

$$(6.18) \quad \mathcal{I}_V(\Sigma) = \mathcal{K}(\Sigma_{\text{KMK}(\kappa_1, \kappa_2)} | \Sigma) + \sum_{i=1}^{N^+} \mathcal{F}_V^+(\lambda_i^+) + \sum_{i=1}^{N^+} \mathcal{F}_V^-(\lambda_i^-)$$

for  $\Sigma \in \mathcal{S}_{p,1}(u^-, u^+)$ , and  $\mathcal{I}_V(\Sigma) = +\infty$  otherwise. Here, the functions  $\mathcal{F}_V^\pm$  are given by

$$(6.19) \quad \mathcal{F}_V^+(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi \in \mathbb{R}} \mathcal{J}_V(\xi) & \text{if } u^+ \leq x \leq 1, \\ \infty & \text{otherwise,} \end{cases}$$

$$(6.20) \quad \mathcal{F}_V^-(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi \in \mathbb{R}} \mathcal{J}_V(\xi) & \text{if } 1 \leq x \leq u^-, \\ \infty & \text{otherwise,} \end{cases}$$

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

$$V(x) - 2 \int \log |x - \xi| d\text{KMK}(\kappa_1, \kappa_2)(\xi).$$

On the one hand, as discussed in Proposition 3.2 of [18] (see also the references therein), for  $V$  in (6.17), we have  $\mathcal{F}_V^\pm = \mathcal{F}_J^\pm$ , (see (3.7) and (3.6)). That is, the rate function  $\mathcal{I}_V$  is precisely the left hand side of the sum rule in Theorem 3.1.

On the other hand, as shown in Theorem 5.1, the sequence  $(\Sigma_N)_n$  satisfies the LDP with speed  $N$  and good rate function  $\mathcal{I}_J$ . Since a large deviation rate function is unique, we get for any  $\Sigma \in \mathcal{M}_{p,1}([0, 1])$  the identity

$$\mathcal{I}_V(\Sigma) = \mathcal{I}_J(\Sigma),$$

which is the sum rule of Theorem (3.1). □

## 7 Proof of the technical lemmas

### 7.1 Proof of Lemma 2.2

The following statements are true for general nonnegative matrix measures  $\Sigma \in \mathcal{M}_p([0, 1])$  that are not necessarily normalized. Let us denote the  $n$ -th moment space of nonnegative matrix



measures on  $[0, 1]$  by

$$(7.1) \quad \mathfrak{M}_{p,n} = \{(M_0(\Sigma), \dots, M_n(\Sigma)) \mid \Sigma \in \mathcal{M}_p([0, 1])\} \subset \mathcal{H}_p^{n+1}.$$

A comprehensive study of this matrix moment space and the relation between canonical moments and recursion coefficients has been addressed in [11]. Indeed, Theorem 2.7 therein shows that if  $(M_0, \dots, M_{2n-1})$  lies in the interior of  $\mathfrak{M}_{p,2n-1}$ , then the upper and lower bound for  $M_k$  satisfies  $M_k^- < M_k < M_k^+$  for  $1 \leq k \leq 2n - 1$ , and then the canonical moments

$$(7.2) \quad U_k = (M_k^- - M_k^+)^{-1}(M_k - M_k^-), \quad 1 \leq k \leq 2n - 1$$

are well defined. Theorem 4.1 of [11] shows that the recursion coefficients  $u_0, \dots, u_{n-1}; v_1, \dots, v_{n-1}$  of  $\Sigma$  satisfy the decomposition as in (2.10) and (2.11). Therefore, the statement of Lemma 2.2 follows once we show that for a measure  $\Sigma$  satisfying the assumption of the lemma,  $(M_0, \dots, M_{2n-1})$  is in the interior of the moment space  $\mathfrak{M}_{p,2n-1}$ . Since this result may be of independent interest, we formulate it as a lemma.

**Lemma 7.1** *Let  $\Sigma \in \mathcal{M}_p([0, 1])$  such that*

$$(7.3) \quad \text{tr}\langle\langle P, P \rangle\rangle > 0$$

*for all matrix polynomials  $P$  of degree at most  $n - 1$ . Then  $(M_0, \dots, M_{2n-3})$  is in the interior of the moment space  $\mathfrak{M}_{p,2n-3}$ . If additionally  $\Sigma(\{0\}) = \Sigma(\{1\}) = \mathbf{0}$ , then  $(M_0, \dots, M_{2n-1})$  is in the interior of the moment space  $\mathfrak{M}_{p,2n-1}$ .*

By the above lemma, there are two sufficient conditions for the existence of the first  $2n - 1$  canonical moments: either (7.3) is satisfied for all polynomials up to degree  $n$ , or it holds for polynomials up to degree  $n - 1$  and the additional assumption  $\Sigma(\{0, 1\}) = \mathbf{0}$  is satisfied. If the condition (7.3) fails for some polynomial of degree  $n$ , then atoms at the boundary can indeed cause the moments to be more "extremal". This can be made more precise in the scalar case, for which we refer to [9], Theorem 1.2.5 and Definition 1.2.10. Suppose  $\mu$  is a scalar measure on  $[0, 1]$  with  $n$  support points, then any nonzero polynomial with degree less than  $n$  has positive  $L^2(\mu)$ -norm, but there is a polynomial of degree  $n$  with vanishing norm. Then the first  $2n - 3$  moments will be in the interior of the moment space. On the other hand, the fact that  $(M_0, \dots, M_{2n-1})$  lies in the boundary of the moment space is actually equivalent to the fact that  $\{0, 1\}$  has positive mass. If both 0 and 1 are in the support of  $\mu$ , then already  $(M_0, \dots, M_{2n-2})$  lies at the boundary of the moment space. If exactly one support point is equal to 0 or 1, then the first  $2n - 2$  moments are interior, but the first  $2n - 1$  ones are not. If the support contains 0, then  $M_{2n-1} = M_{2n-1}^-$ ,

whereas 1 in the support implies  $M_{2n-1} = M_{2n-1}^+$ . The two versions of  $\mu$  are then called the lower and upper principal representation of  $(M_0, \dots, M_{2n-2})$ , respectively. In the matrix case, the boundary of  $\mathfrak{M}_{p,2n-1}$  has a more complicated structure and there is no such equivalence.

**Proof of Lemma 7.1:** We again refer to [11]. Lemma 2.3 says that  $(M_0, \dots, M_m)$  is an element of  $\mathfrak{M}_{p,m}$  if and only if, for all matrices  $A_0, \dots, A_m$ , such that  $Q(x) = A_m x^m + \dots + A_0$  is nonnegative definite for all  $x \in [0, 1]$ , we have

$$(7.4) \quad \operatorname{tr} \sum_{k=0}^m A_k M_k \geq 0.$$

Note that the case  $A_m = \mathbf{0}$  is also included. Furthermore,  $(M_0, \dots, M_m)$  is an interior point of  $\mathfrak{M}_{p,m}$  if and only if, for all  $A_0, \dots, A_m$  for which such  $Q$  is nonnegative definite on  $[0, 1]$  and nonzero, we have

$$(7.5) \quad \operatorname{tr} \sum_{k=0}^m A_k M_k > 0.$$

Theorem 2.5 of [11] shows that if the degree of  $Q$  is even, say  $2\ell$ , then such a polynomial can be written as

$$(7.6) \quad Q(x) = B_1(x)B_1(x)^\dagger + x(1-x)B_2(x)B_2(x)^\dagger,$$

where  $B_1$  and  $B_2$  are matrix polynomials of degree  $\ell$  and  $\ell - 1$ , respectively. If the degree of  $Q$  is equal to  $2\ell - 1$ , then

$$(7.7) \quad Q(x) = xB_1(x)B_1(x)^\dagger + (1-x)B_2(x)B_2(x)^\dagger,$$

with  $B_1, B_2$  of degree  $\ell - 1$ . Let  $\Sigma \in \mathcal{M}_p([0, 1])$  with  $M_k$  the  $k$ -th moment of  $\Sigma$ . If  $m = 2\ell$  and  $A_m \neq \mathbf{0}$ , then, using the decomposition (7.6),

$$(7.8) \quad \begin{aligned} \operatorname{tr} \sum_{k=0}^m A_k M_k &= \operatorname{tr} \int Q(x) d\Sigma(x) \\ &= \operatorname{tr} \int B_1(x)B_1(x)^\dagger d\Sigma(x) + \operatorname{tr} \int x(1-x)B_2(x)B_2(x)^\dagger d\Sigma(x) \\ &= \operatorname{tr} \int B(x)^\dagger_1 d\Sigma(x) B_1(x) + \operatorname{tr} \int x(1-x)B_2(x)^\dagger d\Sigma(x) B_2(x) \\ &= \operatorname{tr} \langle \langle B_1, B_1 \rangle \rangle + \operatorname{tr} \langle \langle pB_2, B_2 \rangle \rangle, \end{aligned}$$

where  $p(x) = x(1-x)$ .

A similar calculation can be made if  $m = 2\ell - 1$  and  $A_m \neq \mathbf{0}$ . Together with the characterizations of the moment space by (7.4) and (7.5), this implies that for any  $\Sigma \in \mathcal{M}_p([0, 1])$  and matrix polynomial  $B$

$$(7.9) \quad \operatorname{tr} \langle \langle qB, B \rangle \rangle \geq 0,$$

when  $q(x)$  is the scalar polynomial  $1, x, 1 - x$  or  $x(1 - x)$ . Furthermore, the first  $2m - 1$  moments of  $\Sigma$  are in the interior of the moment space  $\mathfrak{M}_{p, 2m-1}$ , if

$$(7.10) \quad \operatorname{tr} \langle \langle qB, B \rangle \rangle > 0,$$

whenever  $B$  is nonzero and such that the degree of  $q(x)B(x)B(x)^\dagger$  is at most  $2m - 1$ . We remark that this is actually equivalent to the criterion given in [11] and stated in terms of Hankel matrices.

Now suppose that  $\Sigma$  is such that  $\operatorname{tr} \langle \langle P, P \rangle \rangle > 0$  for all nonzero polynomials  $P$  of degree at most  $n - 1$ . We show that then (7.10) is satisfied whenever the degree of  $qBB^\dagger$  is at most  $2n - 3$ . For  $q(x) = 1$  this is trivially true. In the other cases,

$$(7.11) \quad \operatorname{tr} \langle \langle qB, B \rangle \rangle = \operatorname{tr} \langle \langle qB, qB \rangle \rangle + \operatorname{tr} \langle \langle qB, (1 - q)B \rangle \rangle.$$

Since  $qB$  has degree at most  $n - 1$ , the first inner product on the right hand side of (7.11) is positive by assumption. The second one is nonnegative by (7.4), since  $q(1 - q)BB^\dagger$  is nonnegative definite on  $[0, 1]$ . This proves that  $(M_0, \dots, M_{2n-3})$  is in the interior of  $\mathfrak{M}_{p, 2m-1}$ .

Now assume  $\Sigma(\{0, 1\}) = \mathbf{0}$ , we show that then (7.10) is satisfied whenever  $qBB^\dagger$  has degree at most  $2n - 1$  and  $B$  is nonzero. In this case,  $B$  is of degree at most  $n - 1$ , and  $\operatorname{tr} \langle \langle B, B \rangle \rangle$  is positive. Using that  $\Sigma$  has no mass at  $0, 1$ ,

$$(7.12) \quad \operatorname{tr} \langle \langle B, B \rangle \rangle = \lim_{\varepsilon \rightarrow 0} \operatorname{tr} \int_{\varepsilon}^{1-\varepsilon} B(x)^\dagger d\Sigma(x)B(x),$$

and then there exists a  $\varepsilon > 0$ , such that the integral on the right hand side is positive. Since  $q(x) \geq \varepsilon(1 - \varepsilon)$  on  $[\varepsilon, 1 - \varepsilon]$ , and  $\int_A B(x)^\dagger d\Sigma B(x)$  is always nonnegative definite,

$$(7.13) \quad \operatorname{tr} \langle \langle qB, B \rangle \rangle \geq \operatorname{tr} \int_{\varepsilon}^{1-\varepsilon} q(x)B(x)^\dagger d\Sigma(x)B(x) \geq \varepsilon(1 - \varepsilon) \operatorname{tr} \int_{\varepsilon}^{1-\varepsilon} B(x)^\dagger d\Sigma(x)B(x),$$

which gives a positive lower bound. □

## 7.2 Proof of Lemma 4.1

Let us begin by noting that if  $z_1, \dots, z_N$  are random vectors in  $\mathbb{C}^p$ , independent and complex standard normal distributed, then almost surely, any  $p$  of these vectors span  $\mathbb{C}^p$ . This implies

that almost surely,  $H = z_1 z_1^\dagger + \cdots + z_N z_N^\dagger$  has full rank. Consider such a realization and let

$$P(x) = C_{n-1}x^{n-1} + \cdots + C_1x + C_0$$

be a matrix polynomial of degree at most  $n - 1$ . We have

$$\mathrm{tr}\langle\langle P, P \rangle\rangle = \mathrm{tr} \sum_{i=1}^N P(\lambda_i)^\dagger \mathbf{v}_i \mathbf{v}_i^\dagger P(\lambda_i) = \sum_{i=1}^N \mathbf{v}_i^\dagger P(\lambda_i)^\dagger P(\lambda_i) \mathbf{v}_i = \sum_{i=1}^N \|P(\lambda_i) \mathbf{v}_i\|^2.$$

Suppose that  $\mathrm{tr}\langle\langle P, P \rangle\rangle = 0$ , then the above calculation shows that for all  $i$ ,  $\mathbf{v}_i$  is in the kernel of  $P(\lambda_i)$ . We may rewrite this in matrix form by saying that

$$(7.14) \quad \mathbf{W}\mathbf{P} = \mathbf{0},$$

where  $\mathbf{P}$  is  $np \times p$  with  $\mathbf{P}^\dagger = (C_0, \dots, C_{n-1})$ , and  $\mathbf{W}$  is  $np \times np$  with

$$\mathbf{W} = \begin{pmatrix} \mathbf{v}_1^\dagger & \mathbf{v}_1^\dagger \lambda_1 & \cdots & \mathbf{v}_1^\dagger \lambda_1^{n-1} \\ \mathbf{v}_2^\dagger & \mathbf{v}_2^\dagger \lambda_2 & \cdots & \mathbf{v}_2^\dagger \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_{np}^\dagger & \mathbf{v}_{np}^\dagger \lambda_{np} & \cdots & \mathbf{v}_{np}^\dagger \lambda_{np}^{n-1} \end{pmatrix}.$$

Now, we show that  $\mathbf{W}$  is nonsingular, so that the only solution to (7.14) is  $\mathbf{P} = \mathbf{0}$ , that is,  $P$  is the zero polynomial. Let  $\mathbf{H}$  be the  $np \times np$  block-diagonal matrix with blocks  $H^{1/2}$  on the diagonal, then  $\mathbf{H}$  is nonsingular. The matrix  $\mathbf{Z} = \mathbf{W}\mathbf{H}$  has the same structure as  $\mathbf{W}$ , except that  $\mathbf{v}_i$  is replaced by  $z_i$ . We use an argument similar to what has been done in the proof of Lemma 2.2 in [24]. Conditionally on the eigenvalues, the determinant of  $\mathbf{Z}$  is a polynomial in the  $np^2$  entries of  $z_1, \dots, z_{np}$ . Since they are all independent standard Gaussians, they have a joint density and then either  $\det(\mathbf{Z})$  is 0 with probability 0 or it is the zero polynomial. Let us fix  $z_{kp+i} = e_i$  for  $k = 0, \dots, n-1$ ,  $i = 1, \dots, p$ , where  $e_1, \dots, e_p$  is the canonical basis of  $\mathbb{C}^p$ . In this case,

$$(7.15) \quad \mathbf{Z} = \begin{pmatrix} e_1^\dagger & e_1^\dagger \lambda_1 & \cdots & e_1^\dagger \lambda_1^{n-1} \\ e_2^\dagger & e_2^\dagger \lambda_2 & \cdots & e_2^\dagger \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ e_p^\dagger & e_p^\dagger \lambda_p & \cdots & e_p^\dagger \lambda_p^{n-1} \\ e_1^\dagger & e_1^\dagger \lambda_{p+1} & \cdots & e_1^\dagger \lambda_{p+1}^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ e_p^\dagger & e_p^\dagger \lambda_{np} & \cdots & e_p^\dagger \lambda_{np}^{n-1} \end{pmatrix}.$$

By reordering rows and columns, this matrix may be transformed into the block diagonal matrix  $\tilde{\mathbf{Z}}$  with  $n \times n$  Vandermonde-blocks,

$$(7.16) \quad \tilde{\mathbf{Z}} = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-1} & & & & & \\ 1 & \lambda_{p+1} & \cdots & \lambda_{p+1}^{n-1} & & & & & \\ \vdots & \vdots & & \vdots & & & & & \\ 1 & \lambda_{(n-1)p+1} & \cdots & \lambda_{(n-1)p+1}^{n-1} & & & & & \\ & & & & \cdots & & & & \\ & & & & & & & & \\ & & & & & & 1 & \lambda_p & \cdots & \lambda_p^{n-1} \\ & & & & & & 1 & \lambda_{2p} & \cdots & \lambda_{2p}^{n-1} \\ & & & & & & \vdots & \vdots & & \vdots \\ & & & & & & 1 & \lambda_{np} & \cdots & \lambda_{np}^{n-1} \end{pmatrix},$$

which has determinant

$$(7.17) \quad \det(\tilde{\mathbf{Z}}) = \prod_{k=1}^p \prod_{1 \leq j < j' < n} (\lambda_{jp+k} - \lambda_{j'p+k}).$$

Since the  $\lambda_i$  are almost surely disjoint, the matrix  $\tilde{\mathbf{Z}}$  is almost surely non-singular, which implies that  $\mathbf{W}$  is almost surely nonsingular.  $\square$

### 7.3 Proof of Lemma 6.1

We have to compute the Jacobian determinant of the mapping  $(\mathcal{A}, \mathcal{B}) \mapsto \mathcal{U}$ . We will do this by using the moments as intermediate variables. Let us begin by noting that  $u_{n-1}, U_{2n-1}$  depend on  $M_1, \dots, M_{2n-1}$ , but not on any higher moments and  $v_n, U_{2n}$  depend only on  $M_1, \dots, M_{2n}$ . Since for the similarity transforms in Section 2 we used only matrices depending on moments of strictly lower order, the same statements can be made for the Hermitian versions, where  $\mathcal{B}_n, \mathcal{U}_{2n-1}$  depend on  $M_1, \dots, M_{2n-1}$  and  $\mathcal{A}_n, \mathcal{U}_{2n}$  depend on  $M_1, \dots, M_{2n}$ . We have in particular

$$(7.18) \quad \frac{\partial(\mathcal{B}, \mathcal{A})}{\partial M} := \frac{\partial(\mathcal{B}_1, \mathcal{A}_1, \dots, \mathcal{B}_n)}{\partial(M_1, \dots, M_{2n-1})} = \frac{\partial \mathcal{B}_1}{\partial M_1} \times \frac{\partial \mathcal{A}_1}{\partial M_2} \times \cdots \times \frac{\partial \mathcal{B}_n}{\partial M_{2n-1}}.$$

Here, we denote by  $\frac{\partial F(M)}{\partial M}$  the Jacobian determinant of the mapping  $F : \mathcal{H}_p \rightarrow \mathcal{H}_p$ , seen as a mapping of all the  $p^2$  functionally independent real entries of a matrix in  $\mathcal{H}_p$ , and with the straightforward generalization to mappings with several such matricial coordinates, see [28]. In particular, Theorem 3.5 in [28] shows that for nonsingular  $A$ ,

$$(7.19) \quad \frac{\partial(AMA^\dagger)}{\partial M} = \det(A)^{2p}.$$

Recall that

$$(7.20) \quad \mathcal{A}_k = H_{2k-2}^{-1/2} H_{2k} H_{2k-2}^{-1/2}, \quad \mathcal{B}_k = H_{2k-2}^{-1/2} \langle \langle x P_{k-1}, P_{k-1} \rangle \rangle H_{2k-2}^{-1/2}$$

and  $H_{2k} = M_{2k} - M_{2k}^-$  depends only on  $M_1, \dots, M_{2k}$  (see (2.16)). Then by (7.19),

$$(7.21) \quad \frac{\partial \mathcal{A}_k}{\partial M_{2k}} = \det(H_{2k-2})^{-p}, \quad \frac{\partial \mathcal{B}_k}{\partial M_{2k-1}} = \det(H_{2k-2})^{-p}.$$

Putting these together, we get that (7.18) is given by

$$(7.22) \quad \frac{\partial(\mathcal{B}, \mathcal{A})}{\partial M} = \det(H_{2n-2})^{-p} \prod_{k=1}^{n-1} \det(H_{2k-2})^{-2p}.$$

To end this first step, we need to evaluate

$$(7.23) \quad \frac{\partial \mathcal{U}}{\partial M} := \frac{\partial(\mathcal{U}_1, \dots, \mathcal{U}_{2n-1})}{\partial(M_1, \dots, M_{2n-1})} = \frac{\partial \mathcal{U}_1}{\partial M_1} \times \dots \times \frac{\partial \mathcal{U}_{2n-1}}{\partial M_{2n-1}}.$$

where we have by (2.15)

$$(7.24) \quad \frac{\partial \mathcal{U}_k}{\partial M_k} = \frac{\partial \left( R_k^{-1/2} H_k R_k^{-1/2} \right)}{\partial M_k} = \det(R_k)^{-p}$$

and then

$$(7.25) \quad \frac{\partial \mathcal{U}}{\partial M} = \prod_{k=1}^{2n-1} \det(R_k)^{-p}.$$

Putting together (7.22) and (7.25), we have shown that

$$(7.26) \quad \frac{\partial(\mathcal{B}, \mathcal{A})}{\partial \mathcal{U}} = \det(H_{2n-2})^{-p} \prod_{k=1}^{n-1} \det(H_{2k-2})^{-2p} \prod_{k=1}^{2n-1} \det(R_k)^p.$$

To express this in terms of the canonical moments, we use

$$R_k = R_{k-1}(\mathbf{1} - U_{k-1})U_{k-1}, \quad H_k = R_k U_k,$$

(see [11] formulas (2.19) and (2.16)). Taking determinants, we obtain

$$(7.27) \quad \det R_k = \prod_{j=1}^{k-1} \det(\mathbf{1} - \mathcal{U}_j) \det \mathcal{U}_j, \quad \det H_{2k-2} = \det R_{2k-2} \det \mathcal{U}_{2k-2}.$$

We gather (7.26) and (7.27), to obtain that the pushforward of the measure (6.7) by the mapping  $(\mathcal{A}, \mathcal{B}) \mapsto \mathcal{U}$  has, up to a multiplicative constant, the density

$$\begin{aligned}
& \prod_{k=1}^{n-1} \det(\mathcal{A}_k)^{p(n-k-1)} \det(H_{2n-2})^{-p} \prod_{k=1}^{n-1} \det(H_{2k-2})^{-2p} \prod_{k=1}^{2n-1} \det(R_k)^p \\
&= \prod_{k=1}^{n-1} \det(H_{2k-2})^{-p(n-k-1)} \det(H_{2k})^{p(n-k-1)} \det(H_{2n-2})^{-p} \prod_{k=1}^{n-1} \det(H_{2k-2})^{-2p} \prod_{k=1}^{2n-1} \det(R_k)^p \\
&= \prod_{k=1}^{n-1} \det(H_{2k-2})^{-p(n-k)} \det(H_{2k})^{p(n-k-1)} \prod_{k=1}^n \det(H_{2k-2})^{-p} \prod_{k=1}^{2n-1} \det(R_k)^p \\
(7.28) \quad &= \prod_{k=1}^n \det(H_{2k-2})^{-p} \prod_{k=1}^{2n-1} \det(R_k)^p,
\end{aligned}$$

where for the second line we used (7.20), and then observe the telescopic product of the determinants of  $H_k$ .

It remains to express (7.28) in terms of the canonical moments. It's time to use (7.27) to get

$$\begin{aligned}
& \prod_{k=1}^n \det(H_{2k-2})^{-p} \prod_{k=1}^{2n-1} \det(R_k)^p = \prod_{k=1}^{n-1} \det(R_{2k})^{-p} \det(\mathcal{U}_{2k})^{-p} \prod_{k=1}^{2n-1} \det(R_k)^p \\
&= \prod_{k=1}^{n-1} \det(\mathcal{U}_{2k})^{-p} \prod_{k=1}^n \det(R_{2k-1})^p \\
&= \prod_{k=1}^{n-1} \det(\mathcal{U}_{2k})^{-p} \prod_{k=1}^n \prod_{i=1}^{2k-2} \det(\mathbf{1} - \mathcal{U}_i)^p \det(\mathcal{U}_i)^p \\
&= \prod_{k=1}^{n-1} \det(\mathcal{U}_{2k})^{-p} \prod_{k=1}^{n-1} \det((\mathbf{1} - \mathcal{U}_{2k-1})\mathcal{U}_{2k-1}(\mathbf{1} - \mathcal{U}_{2k})\mathcal{U}_{2k})^{p(n-k)} \\
(7.29) \quad &= \prod_{k=1}^{n-1} \det((\mathbf{1} - \mathcal{U}_{2k-1})\mathcal{U}_{2k-1})^{p(n-k)} \prod_{k=1}^{n-1} \det(-\mathcal{U}_{2k})^{p(n-k)} \det(\mathcal{U}_{2k})^{p(n-k)}.
\end{aligned}$$

This ends the proof of Lemma 6.1. □

## 7.4 Two proofs of Lemma 6.2

It follows from Lemma 2.1 of Duran, Lopez-Rodriguez [15], that the eigenvalues of  $J_n$  are precisely the zeros of the  $n$ -th polynomial orthogonal with respect to  $\Sigma$ . The quadrature formula of Sinap, van Assche [34] implies that the zeros of this polynomial are equal to the support of the spectral

measure. As a consequence,

$$(7.30) \quad \det(J_n) = \prod_{i=1}^{np} \lambda_i, \quad \det(I - J_n) = \prod_{i=1}^{np} (1 - \lambda_i).$$

In view of (7.30) we have to prove that

$$(7.31) \quad \det(I_n - J_n) = \prod_{k=1}^{2n-1} \det(\mathbf{1} - \mathcal{U}_k), \quad \det J_n = \left( \prod_{k=1}^n \det \mathcal{U}_{2k-1} \right) \left( \prod_{k=1}^{n-1} \det(\mathbf{1} - \mathcal{U}_{2k}) \right).$$

We give two proofs. The first one is matricial, using a recursion of Schur complements and the second one is based on the Szegő mapping and matrix polynomials on the unit circle.

#### 7.4.1 First proof

Using the Schur complement formula (see Theorem 1.1 in [25]),

$$(7.32) \quad \begin{aligned} \det(I_n - J_n) &= \det(I_{n-1} - J_{n-1}) \det(\mathbf{1} - \mathcal{B}_n - (0, \dots, 0, \tilde{\mathcal{A}}_{n-1}^\dagger)(I_{n-1} - J_{n-1})^{-1}(0, \dots, \tilde{\mathcal{A}}_{n-1}^\dagger)^\dagger) \\ &= \det(I_{n-1} - J_{n-1}) \det\left(\mathbf{1} - \mathcal{B}_n - \tilde{\mathcal{A}}_{n-1}^\dagger [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} \tilde{\mathcal{A}}_{n-1}\right) \\ &= \det(I_{n-1} - J_{n-1}) \det(\varphi_n), \end{aligned}$$

where we wrote  $[A]_{i,j}$  for the  $p \times p$  sub-block in position  $i, j$  and we define

$$(7.33) \quad \begin{aligned} \varphi_n &= \gamma_n^{-1/2} \left( \mathbf{1} - \mathcal{B}_n - \tilde{\mathcal{A}}_{n-1}^\dagger [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} \tilde{\mathcal{A}}_{n-1} \right) \gamma_n^{1/2} \\ &= \gamma_n^{-1/2} \left( \mathbf{1} - \gamma_n^{1/2} u_{n-1} \gamma_n^{-1/2} - \gamma_n^{1/2} \gamma_{n-1}^{-1/2} [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} \gamma_{n-1}^{-1/2} \gamma_n^{1/2} \right) \gamma_n^{1/2} \\ &= \left( \mathbf{1} - u_{n-1} - \gamma_{n-1}^{-1/2} [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} \gamma_{n-1}^{-1/2} \gamma_n \right) \\ &= \left( \mathbf{1} - u_{n-1} - \gamma_{n-1}^{-1/2} [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} \gamma_{n-1}^{1/2} v_{n-1} \right). \end{aligned}$$

Recall the non-Hermitian recursion coefficients  $u_n, v_n$  have been defined in (2.2) and (2.4). Using again the formula of Schur complements (see Theorem 1.2 in [25]),

$$(7.34) \quad \begin{aligned} [(I_{n-1} - J_{n-1})^{-1}]_{n-1, n-1} &= \left( \mathbf{1} - \mathcal{B}_{n-1} - (0, \dots, 0, \tilde{\mathcal{A}}_{n-2}^\dagger)(I_{n-2} - J_{n-2})^{-1}(0, \dots, \tilde{\mathcal{A}}_{n-2}^\dagger)^\dagger \right)^{-1} \\ &= \left( \mathbf{1} - \mathcal{B}_{n-1} - \tilde{\mathcal{A}}_{n-2}^\dagger [(I_{n-2} - J_{n-2})^{-1}]_{n-2, n-2} \tilde{\mathcal{A}}_{n-2} \right)^{-1} \\ &= \gamma_{n-1}^{1/2} \varphi_{n-1}^{-1} \gamma_{n-1}^{-1/2}. \end{aligned}$$

We see that  $\varphi_n$  satisfies the recursion

$$(7.35) \quad \varphi_1 = \mathbf{1} - u_0, \quad \varphi_n = \mathbf{1} - u_{n-1} - \varphi_{n-1}^{-1} v_{n-1}, \quad n \geq 2.$$



Let us write  $V_k = \mathbf{1} - U_k$ . Then we claim that the solution to this recursion is given by

$$(7.36) \quad \varphi_n = V_{2n-2}V_{2n-1}.$$

We prove (7.36) by induction. For  $n = 1$ , we have by (2.10)

$$(7.37) \quad \varphi_1 = \mathbf{1} - u_0 = \mathbf{1} - \zeta_1 = \mathbf{1} - U_1 = V_1,$$

which agrees with (7.36) since  $V_0 = \mathbf{1}$ . Then,

$$(7.38) \quad \begin{aligned} \varphi_{n+1} &= \mathbf{1} - u_n - \varphi_n^{-1}v_n \\ &= \varphi_n^{-1} [\varphi_n - \varphi_n(\zeta_{2n} + \zeta_{2n+1}) - \zeta_{2n-1}\zeta_{2n}] \\ &= \varphi_n^{-1} [V_{2n-2}V_{2n-1} - V_{2n-2}V_{2n-1}(V_{2n-1}U_{2n} + V_{2n}U_{2n+1}) - V_{2n-2}U_{2n-1}V_{2n-1}U_{2n}] \\ &= \varphi_n^{-1}V_{2n-2} [V_{2n-1} - V_{2n-1}^2U_{2n} - V_{2n-1}V_{2n}U_{2n+1} - U_{2n-1}V_{2n-1}U_{2n}]. \end{aligned}$$

In the last line, we write  $U_{2n-1}V_{2n-1}U_{2n} = V_{2n-1}U_{2n} - V_{2n-1}^2U_{2n}$  for the last term, which then cancels the second term in the brackets and leads to

$$(7.39) \quad \begin{aligned} \varphi_{n+1} &= \varphi_n^{-1}V_{2n-2} [V_{2n-1} - V_{2n-1}V_{2n}U_{2n+1} - V_{2n-1}U_{2n}] \\ &= \varphi_n^{-1}V_{2n-2}V_{2n-1} [\mathbf{1} - V_{2n}U_{2n+1} - U_{2n}] \\ &= \varphi_n^{-1}\varphi_n [V_{2n} - V_{2n}U_{2n+1}] \\ &= V_{2n}V_{2n+1}. \end{aligned}$$

This proves (7.36). We may then calculate recursively for (7.32)

$$\det(I_n - J_n) = \det(I_{n-1} - J_{n-1}) \det \varphi_n = \det(\varphi_1 \dots \varphi_n),$$

so that

$$(7.40) \quad \det(I_n - J_n) = \prod_{k=1}^{2n-1} \det V_k = \prod_{k=1}^{2n-1} \det(\mathbf{1} - U_k) = \prod_{k=1}^{2n-1} \det(\mathbf{1} - \mathcal{U}_k).$$

For the computation of  $\det J_n$ , we make use of a decomposition proven in Lemma 2.1 of [21]. There exists a block bi-diagonal matrix  $Z_n$ , such that  $J_n = Z_n Z_n^\dagger$  and (see the proof in [19]), the block  $D_k$  in position  $k, k$  of  $Z_n$  satisfies

$$(7.41) \quad \det(D_k) = \det(\zeta_{2n-1})^{1/2}.$$

Then, this implies

$$(7.42) \quad \begin{aligned} \det J_n &= (\det Z_n)^2 = \prod_{k=1}^n (\det D_k)^2 = \prod_{k=1}^n \det \zeta_{2k-1} \\ &= (\det U_1)(\det V_2) \cdots (\det U_{2n-2})(\det V_{2n-2})(\det U_{2n-1}), \end{aligned}$$

which gives the second identity in (7.31). □

### 7.4.2 Second proof of Lemma 6.2 via Szegő's mapping

It was tempting to extend to the matrix case the method used in the scalar one for the Jacobi ensemble. The main steps use successively:

- the inverse Szegő mapping to turn the problem on  $[0, 1]$  into a problem on the unit circle,
- the correspondence between orthogonal polynomials on the unit circle and on the real line,
- the Szegő recursion for polynomials on the unit circle.

To begin with, we transfer the measure on  $[0, 1]$  to a measure on  $[-2, 2]$  by the mapping  $x \mapsto 2-4x$ . The new Jacobi matrix  $\hat{J}_n$  is deduced from the original matrix  $J_n$  by

$$(7.43) \quad \hat{J}_n = 2I_n - 4\Omega J_n \Omega,$$

where  $\Omega$  is a diagonal matrix with alternating blocks  $\pm \mathbf{1}$ 's on the diagonal.

Let  $\hat{P}_0, \dots, \hat{P}_n$  be the monic orthogonal polynomials associated with  $\hat{J}_n$ . From [5] Section 2.9 (with reference in particular to [15] and [34])

$$(7.44) \quad \det \hat{P}_n(z) = \det(zI_n - \hat{J}_n),$$

so that

$$(7.45) \quad \det \hat{P}_n(z) = \det((z-2)I_n + 4\Omega J_n \Omega) = 4^{np} \det\left(\frac{z-2}{4}I_n + J_n\right)$$

and in particular,

$$(7.46) \quad \det J_n = 4^{-np} \det \hat{P}_n(2), \quad \det(I_n - J_n) = (-4)^{-np} \det \hat{P}_n(-2).$$

We refer to the definition of the Szegő mapping given in Section 2.2.2. In this Section, we write  $\Sigma_{\mathbb{R}}$  for a matrix measure on the real line and denote by  $\Sigma_{\mathbb{T}} = \tilde{S}z^{-1}(\Sigma_{\mathbb{R}})$  the preimage under the Szegő mapping. The correspondence between polynomials orthogonal with respect to  $\Sigma_{\mathbb{T}}$  and with respect to  $\Sigma_{\mathbb{R}}$  is ruled by the following theorem (see Proposition 1 in [37]). It is the matrix version of a famous theorem due to Szegő [29]. Since the notations are slightly different from the usual ones, we rewrite the proof in Section 7.4.3.

**Theorem 7.2 (Yakhlef-Marcellán)** *Let  $\Sigma_{\mathbb{R}} \in \mathcal{M}_{p,1}([-2, 2])$  be a nontrivial matrix measure and denote by  $\Sigma_{\mathbb{T}} = \tilde{S}z(\Sigma_{\mathbb{R}})$  the symmetric measure on  $\mathbb{T}$  obtained by the Szegő mapping.*

If  $\hat{P}_n$  is the  $n$ -th right monic orthogonal polynomial for  $\Sigma_{\mathbb{R}}$  and  $\Phi_{2n}$  the  $2n$ -th right monic orthogonal polynomial<sup>1</sup> for  $\Sigma_{\mathbb{T}}$ , then

$$(7.47) \quad \hat{P}_n(z + z^{-1}) = [z^{-n}\Phi_{2n}(z) + z^n\Phi_{2n}(z^{-1})] \tau_n^{-1},$$

where

$$(7.48) \quad \tau_n := \mathbf{1} + \Phi_{2n}(0) = \mathbf{1} - \kappa_{2n-1}\alpha_{2n-1}(\kappa_{2n-1})^{-1}$$

with

$$\kappa_k = (\rho_0 \cdots \rho_{k-1})^{-1}, \quad \rho_j = (\mathbf{1} - \alpha_j^2)^{1/2}.$$

From (7.46) and (7.48) we deduce taking  $z = \pm 1$ ,

$$(7.49) \quad \hat{P}_n(\pm 2) = 2(\pm 1)^n \Phi_{2n}(\pm 1) \tau_n^{-1},$$

hence

$$(7.50) \quad \det \hat{P}_n(\pm 2) = 2^p \det(1 - \alpha_{2n-1})^{-1} (\pm 1)^{np} \det \Phi_{2n}(\pm 1).$$

Recall that the recursion formula expressed for the monic polynomials on the unit circle, in this particular case, is

$$(7.51) \quad z\Phi_k(z) - \Phi_{k+1}(z) = z^k \Phi_k(z^{-1}) \kappa_k \alpha_k \kappa_k^{-1}$$

(see (3.11) in [5]), so that

$$\Phi_{2n}(1) = \prod_{j=0}^{2n-1} (1 - \kappa_j \alpha_j \kappa_j^{-1}), \quad \Phi_{2n}(-1) = \prod_{j=0}^{2n-1} (1 + (-1)^j \kappa_j \alpha_j \kappa_j^{-1})$$

and then

$$(7.52) \quad \det \Phi_{2n}(1) = \prod_{j=0}^{2n-1} \det(\mathbf{1} - \alpha_j), \quad \det \Phi_{2n}(-1) = \prod_{j=0}^{2n-1} \det(\mathbf{1} + (-1)^j \alpha_j).$$

These relations are the matrix extension of Lemma 5.2 of [26]. Coming back to (7.46) and (7.50), we get

$$(7.53) \quad \det(J_n) = 2^{-(2n-1)p} \prod_{j=0}^{2n-2} \det(\mathbf{1} - \alpha_j), \quad \det(I_n - J_n) = 2^{-(2n-1)p} \prod_{j=0}^{2n-2} \det(\mathbf{1} + (-1)^j \alpha_j).$$

The connection with the canonical moments follows then from (2.18). Note that this identity still holds if  $\Sigma$  is not nontrivial, as long as  $\mathbf{0} < \mathcal{U}_k < \mathbf{1}$ , or equivalently  $-\mathbf{1} < \alpha_{k-1} < \mathbf{1}$ .

---

<sup>1</sup>The right monic OP for  $\Sigma_{\mathbb{T}}$  are obtained by applying Gram-Schmidt to  $\{\mathbf{1}, z\mathbf{1}, \dots\}$ .

### 7.4.3 Proof of Theorem 7.2

In the scalar case, the proof is given in [32] Theorem 13.1.5 or in [33] Theorem 1.9.1, with references therein. In the matrix case, one can follow the same scheme.

Since  $\Sigma_{\mathbb{T}}$  is invariant, the Verblunsky coefficients are Hermitian (see Lemma 4.1 in [5]). The matrix Laurent polynomial  $z^{-n}\mathbf{\Phi}_{2n}(z) + z^n\mathbf{\Phi}_{2n}(z^{-1})$  is invariant by  $z \mapsto z^{-1}$ . Hence there exists a matrix polynomial  $\tilde{Q}_n$  of degree  $n$ , such that

$$(7.54) \quad z^{-n}\mathbf{\Phi}_{2n}(z) + z^n\mathbf{\Phi}_{2n}(z^{-1}) = \tilde{Q}_n(z + z^{-1}),$$

(see for instance Lemma 13.4.2 in [33]). Collecting terms with highest degrees, we have

$$\tilde{Q}_n(z + z^{-1}) = (z^n + z^{-n})\boldsymbol{\tau}_n + \dots$$

and then

$$(7.55) \quad \tilde{Q}_n(z + z^{-1})\boldsymbol{\tau}_n^{-1} = Q_n(z + z^{-1})$$

where now  $Q_n(x)$  is a monic polynomial of degree  $n$ . Now, let us check that the  $\tilde{Q}_k$  (hence  $Q_k$ ) are orthogonal polynomials for  $\Sigma_{\mathbb{R}}$ . First notice that

$$\tilde{Q}_k(z + z^{-1}) = z^{-k} (\mathbf{\Phi}_{2k}(z) + z^{2k}\mathbf{\Phi}_{2k}(z^{-1})).$$

From the Szegő mapping and (7.55), orthogonality of  $\tilde{Q}_n$  and  $\tilde{Q}_r$  (for  $n \neq r$ ) with respect to  $\Sigma_{\mathbb{R}}$  is equivalent to orthogonality (with respect to  $\Sigma_{\mathbb{T}}$ ) of  $\mathbf{\Phi}_{2n}(z) + z^{2n}\mathbf{\Phi}_{2n}(z^{-1})$  and  $H$  where

$$H(z) = z^{n-r} [\mathbf{\Phi}_{2r}(z) + z^{2r}\mathbf{\Phi}_{2r}(z^{-1})],$$

which is a polynomial of degree  $n+r$  without constant term. By definition,  $\mathbf{\Phi}_{2n}$  is orthogonal to  $z^j \mathbf{1}$  for all  $j = 0, \dots, 2n-1$ . Besides,  $z^{2n}\mathbf{\Phi}_{2n}(z^{-1})$  is (right) orthogonal to  $z^j \mathbf{1}$  for  $j = 1, \dots, 2n$ . Indeed,

$$\begin{aligned} \int [z^{2n}\mathbf{\Phi}_{2n}(\bar{z})]^\dagger d\Sigma_{\mathbb{T}}(z) z^j &= \int \mathbf{\Phi}_{2n}(\bar{z})^\dagger d\Sigma_{\mathbb{T}}(z) z^{j-2n} \\ &= \int \mathbf{\Phi}_{2n}(z)^\dagger d\Sigma_{\mathbb{T}}(z) z^{2n-j} \end{aligned}$$

(by invariance of  $\Sigma_{\mathbb{T}}$ ) and this last integral is  $\mathbf{0}$  for  $1 \leq j \leq 2n$  due to the orthogonality of  $\mathbf{\Phi}_{2n}$  with polynomials of degree at most  $2n-1$ .

One can then conclude that  $\mathbf{\Phi}_{2n}(z) + z^{2n}\mathbf{\Phi}_{2n}(z^{-1})$  is orthogonal to  $z^k$  for  $1 \leq k \leq 2n-1$ , and so to  $H$ . Summarizing, the  $Q_n$ 's are the monic polynomials orthogonal with respect to  $\Sigma_{\mathbb{R}}$ , and then  $\hat{P}_n = Q_n$  for every  $n$ , or in other words, by (7.54) and (7.55)

$$(7.56) \quad \hat{P}_n(z + z^{-1}) = [z^{-n}\mathbf{\Phi}_{2n}(z) + z^n\mathbf{\Phi}_{2n}(z^{-1})] \boldsymbol{\tau}_n^{-1}.$$

□

## References

- [1] G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, Cambridge, 2010.
- [2] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and sum rules for spectral theory - A pedagogical approach. *J. Spectr. Theory*, 8(4):1551–1581, 2018.
- [3] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and the Lukic conjecture. *Duke Math. J.*, 167(15):2857–2902, 2018.
- [4] D. Damanik, R. Killip, and B. Simon. Perturbations of orthogonal polynomials with periodic recursion coefficients. *Ann. of Math.*, 171(3):1931–2010, 2010.
- [5] D. Damanik, A. Pushnitski, and B. Simon. The analytic theory of matrix orthogonal polynomials. *Surv. Approx. Theory*, 4:1–85, 2008.
- [6] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.
- [7] H. Dette and J. Nagel. Some asymptotic properties of the spectrum of the Jacobi ensemble. *SIAM J. Math. Anal.*, 41(4):1491–1507, 2009.
- [8] H. Dette and J. Nagel. Matrix measures, random moments, and Gaussian ensembles. *J. Theoret. Probab.*, 25(1):25–49, 2012.
- [9] H. Dette and W. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics, 1997.
- [10] H. Dette and W.J. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*, volume 338. John Wiley & Sons, 1997.
- [11] H. Dette and W.J. Studden. Matrix measures, moment spaces and Favard’s theorem for the interval  $[0,1]$  and  $[0,\infty)$ . *Linear Algebra Appl.*, 345:169–193, 2002.
- [12] H. Dette, D. Tomecki, and M. Venker. Universality in random moment problems. *Electronic J. Probab.*, 23, 2018.
- [13] H. Dette and J. Wagoner. Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the Geronimus relations. *Linear Algebra Appl.*, 432:1609–1626, 2010.

- [14] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.
- [15] A.J. Duran and P. Lopez-Rodriguez. Orthogonal matrix polynomials: zeros and Blumenthal’s theorem. *J. Approx. Theory*, 84(1):96–118, 1996.
- [16] A.J. Durán and W. Van Assche. Orthogonal matrix polynomials and higher-order recurrence relations. *Linear Algebra Appl.*, 219:261–280, 1995.
- [17] A. Edelman and B. Sutton. The Beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems. *Found. Comput. Math.*, 8(2):259–285, 2008.
- [18] F. Gamboa, J. Nagel, and A. Rouault. Sum rules via large deviations. *J. Funct. Anal.*, 270(2):509 – 559, 2016.
- [19] F. Gamboa, J. Nagel, and A. Rouault. Supplement to ”sum rules and large deviations for spectral matrix measures”. arXiv 1610.02071, 2016.
- [20] F. Gamboa, J. Nagel, and A. Rouault. Sum rules and large deviations for spectral measures on the unit circle. *Random Matrices Theory Appl.*, 6(1), 2017.
- [21] F. Gamboa, J. Nagel, and A. Rouault. Sum rules and large deviations for spectral matrix measures. to appear in *Bernoulli*, 2018.
- [22] F. Gamboa and A. Rouault. Canonical moments and random spectral measures. *J. Theoret. Probab.*, 23:1015–1038, 2010. Erratum in the same journal (2015) doi 10.1007/s10959-015-0653-5.
- [23] F. Gamboa and A. Rouault. Large deviations for random spectral measures and sum rules. *Applied Mathematics Research eXpress*, 2011(2):281–307, 2011.
- [24] F. Gamboa and A. Rouault. Operator-valued spectral measures and large deviations. *J. Statist. Plann. Inference*, 154:72–86, 2014.
- [25] R. A. Horn and F. Zhang. Basic properties of the Schur complement. In *The Schur Complement and Its Applications*, pages 17–46. Springer US, Boston, MA, 2005.
- [26] R. Killip and I. Nenciu. Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701, 2004.

- [27] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math.*, 158(1):253–321, 2003.
- [28] A.M. Mathai. *Jacobians of matrix transformations and functions of matrix arguments*. World Scientific Publishing Company, 1997.
- [29] G. Szegő. *Orthogonal polynomials*, volume 23 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, New York, 1939.
- [30] B.D. McKay. Expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applic.*, 40:203–216, 1981.
- [31] M.L. Mehta. *Random matrices*, volume 142. Elsevier, 2004.
- [32] B. Simon. *Orthogonal polynomials on the unit circle. Part 2*, volume 54 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2005. Spectral theory.
- [33] B. Simon. *Szegő's theorem and its descendants*. M. B. Porter Lectures. Princeton University Press, Princeton, NJ, 2011.
- [34] A. Sinap and W. Van Assche. Orthogonal matrix polynomials and applications. *J. Comput. Appl. Math.*, 66(1-2):27–52, 1996.
- [35] B. Sutton. Computing the complete CS decomposition. *Numerical Algorithms*, 50(1):33–65, 2009.
- [36] J. Wagener. *Matrixwertige kanonische Momente auf dem Einheitskreis und ihre Anwendungen in der Stochastik*. PhD thesis, Ruhr-Universität Bochum, 2010.
- [37] H.O. Yakhlef and F. Marcellán. Orthogonal matrix polynomials, connection between recurrences on the unit circle and on a finite interval. In *Approximation, optimization and mathematical economics (Pointe-à-Pitre, 1999)*, pages 369–382. Physica, Heidelberg, 2001.