Journal of Multivariate Analysis 106 (2012) 17–35

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Large deviations for random matricial moment problems

Fabrice Gamboa [a], Jan Nagel [b,*], Alain Rouault [c], Jens Wagener [d]

[a] Université Paul Sabatier, Institut de Mathématiques de Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 9, France
[b] Technische Universität München, Zentrum Mathematik, 85747 Garching, Germany
[c] Université Versailles-Saint-Quentin, LMV UMR 8100, 45 Avenue des Etats-Unis, 78035-Versailles Cedex, France
[d] Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

ARTICLE INFO

Article history:
Received 1 November 2010
Available online 18 November 2011

AMS subject classifications:
15B52
60F10

Keywords:
Random matrices
Moment spaces
Canonical moments
Large deviations
Carathéodory functions
Schur functions

ABSTRACT

We consider the moment space $\mathcal{M}_n^K$ corresponding to $p \times p$ complex matrix measures defined on K ($K = [0, 1]$ or $K = \mathbb{T}$). We endow this set with the uniform distribution. We are mainly interested in large deviation principles (LDPs) when $n \to \infty$. First we fix an integer k and study the vector of the first k components of a random element of $\mathcal{M}_n^K$. We obtain an LDP in the set of k-arrays of $p \times p$ matrices. Then we lift a random element of $\mathcal{M}_n^K$ into a random measure and prove an LDP at the level of random measures. We end with an LDP on Carathéodory and Schur random functions. These last functions are well connected to the above random measure. In all these problems, we take advantage of the so-called canonical moments technique by introducing new (matricial) random variables that are independent and have explicit distributions.

1. Introduction

1.1. Preliminary: some notations

All along this article, p will be a positive integer, and $p = 1$ will be referred as the scalar case. We denote respectively by $\mathcal{S}_p(\mathbb{C})$ the set of all Hermitian $p \times p$ matrices and by $\mathcal{S}_p^+(\mathbb{C})$ the one of all Hermitian nonnegative $p \times p$ matrices. If $A, B \in \mathcal{S}_p(\mathbb{C})$ we write $A \le B$ (resp. $A < B$) if, and only if, $B - A$ is nonnegative (resp. positive) definite. This is the so-called Loewner partial order on $\mathcal{S}_p(\mathbb{C})$ (see for example [21]). We recall that every $A \in \mathcal{S}_p^+(\mathbb{C})$ has a unique nonnegative square root denoted by $A^{1/2} \in \mathcal{S}_p^+(\mathbb{C})$. The set of all $p \times p$ unitary matrices is denoted by $\mathbb{U}(p)$.

Let K be either $[0, 1]$ or $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$. A matrix-valued probability measure on K is a measure μ on K with values in $\mathcal{S}_p^+(\mathbb{C})$ such that

$$\int_K d\mu = I_p,$$

where I_p is the $p \times p$ identity matrix. We denote by $\mathcal{P}(K)$ the set of all matrix-valued probability measures on K. In general, if $(X, \mathcal{A})$ is a measurable space, we denote by $\mathbb{M}_1(X)$ the set of all probability measures on X. We equip it with

* Corresponding author.
E-mail addresses: gamboa@math.univ-toulouse.fr (F. Gamboa), jan.nagel@ma.tum.de, jan.nagel@rub.de (J. Nagel), alain.rouault@uvsq.fr (A. Rouault), jens.wagener@rub.de (J. Wagener).

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.11.006

the weak convergence topology. This is the coarsest topology such that the mappings $\mu \mapsto \int f(x)d\mu(x)$ are continuous. Here, $f \in \mathcal{C}_b(X)$ (the space of bounded continuous functions on X) is arbitrary (see [2] for completeness).

One of the main objects of interest in our work is, for $n \in \mathbb{N}$, the matricial moment space $\mathcal{M}_n^K$ defined by

$$\mathcal{M}_n^K := \left\{ \left(\int_K x^j d\mu(x) \right)_{j=1,\dots,n}, \ \mu \in \mathcal{P}(K) \right\}. \tag{1.1}$$

This is a compact set having a nonempty interior — denoted by Int$\mathcal{M}_n^K$ — (see [10] for $K = [0, 1]$ and [12] for $K = \mathbb{T}$).

1.2. What is done in this paper?

The aim of our work is to give a picture of the asymptotic behaviour of the set sequence $(\mathcal{M}_n^K)$. More precisely, we first equip the set $\mathcal{M}_n^K$ with the uniform distribution $\mathbb{P}_{K,n}$. Then, for $k \le n$, we consider $\mathbb{P}_{K,n,k}$ the pushforward probability of $\mathbb{P}_{K,n}$ under the projection on $\mathcal{M}_k^K$. We study, for fixed k, the exponential convergence of $(\mathbb{P}_{K,n,k})_n$ when n goes to infinity. The asymptotic behaviour of $(\mathbb{P}_{K,n,k})_n$ was widely studied in the scalar case beginning with the seminal paper of Chang et al. [4] where a central limit theorem (CLT) for $(\mathbb{P}_{[0,1],n,k})$ is proved. Roughly speaking, $(\mathbb{P}_{[0,1],n,k})_n$ converges to the degenerate distribution concentrated on the k first moments of the nonsymmetric arcsine law and there are Gaussian fluctuations around this limit. In the same frame, large deviations are studied in [17]. In these papers, the main ingredient for obtaining asymptotic results is a clever reparametrization of $\mathcal{M}_n^{[0,1]}$. The new parameters, defined recursively, are the so-called canonical moments (see [11] for a complete overview). Informally, given the $k-1$ first moments, the k-th canonical moment is the relative position of the k-th moment in the range (interval) of possible k-th moments. This allows for fixed n, to define a bijection between Int$\mathcal{M}_n^{[0,1]}$ and $(0, 1)^n$. The key property is that the pushforward of the rather involved probability measure $\mathbb{P}_{[0,1],n,k}$ under this mapping is a product measure, i.e. the canonical moments are independent. This is an old result first showed in [30] (a simple proof is given in the first chapter of Dette and Studden [11]). Moreover, extensions of the asymptotic results on $(\mathbb{P}_{K,n,k})_n$ at the level process are studied in [8]. Also in the scalar case, and using a suitable cousin reparametrization (also called canonical moments or Verblunsky coefficients) a CLT and large deviation are tackled for $(\mathbb{P}_{\mathbb{T},n,k})_n$ in [24]. In this last paper, a step towards a multidimensional setting, that is replacing $[0, 1]$ by $[0, 1]^d$ $(d \ge 1)$, is also done. In a more recent work Dette and Nagel [9] extend some of the asymptotic results previously described to the matricial moment problem on $[0, 1]$ $(p > 1)$. As a matter of fact, by using the right extension of canonical moments proposed and first studied in [10], it is shown there that a CLT holds. As before, the key property is the independence, under the uniform distribution on $\mathcal{M}_n^{[0,1]}$, of the matricial canonical moment vector. Here, we revisit these results and obtain new asymptotic result on $\mathcal{M}_n^K$. First, we obtain a CLT when $K = \mathbb{T}$. Further, we show large deviation principles (LDPs) in both cases, $K = [0, 1]$ and $K = \mathbb{T}$. These LDPs are at level 2, that means that they hold for sequences of distributions of random matricial measures having uniform matricial moments. The main tool is more or less similar as the one used in the scalar case, namely the stochastic independence of the matricial canonical moment. Nevertheless, the matricial case appears to be more technical and due to noncommutativity needs more care. Moreover, thanks to the general invariance Proposition 3.5 the complex case $(K = \mathbb{T})$ is tackled by using a polar decomposition argument.

Besides, it is well known that the truncated trigonometrical problem is connected to two problems of functional analysis on the disc: the so-called Carathéodory and Schur problems, respectively. Let us explain the setting in the scalar case, although our results will be in the general matrix case. An analytic function, F, on $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ is called a Carathéodory function iff $F(0) = 1$ and $\Re F(z) > 0$ for all $z \in \mathbb{D}$. Let $\mathcal{C}_1$ be the set composed by all these functions. An analytic function f on $\mathbb{D}$ is called a Schur function iff $\sup_{z\in\mathbb{D}} |f(z)| \le 1$. Let $\mathfrak{S}_1$ be the set of all Schur functions. The correspondence

$$F(z) = \frac{1 + zf(z)}{1 - zf(z)}, \qquad f(z) = \frac{1}{z}\frac{F(z) - 1}{F(z) + 1} \tag{1.2}$$

is one-to-one between $\mathcal{C}_1$ and $\mathfrak{S}_1$. Any $F \in \mathcal{C}_1$ has a representation

$$F(z) = \int_{\mathbb{T}} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \tag{1.3}$$

for a unique probability measure μ on $\mathbb{T}$ (Herglotz representation theorem). The Taylor expansion of F is

$$F(z) = 1 + 2\sum_1^{\infty} c_n(F) z^n, \tag{1.4}$$

where the c_n's are the conjugate moments of μ, i.e.

$$c_n(F) = \int_{\mathbb{T}} e^{-in\theta} d\mu(\theta) = \bar{\gamma}_n.$$

The classical Carathéodory problem is to find $F \in \mathcal{C}_1$ such that the first n Taylor coefficients coincide with given numbers $c_1, \ldots, c_n$. It is clearly equivalent to the truncated moment problem. The Taylor expansion of f is

$$f(z) = \sum_0^\infty s_n(f) z^n. \tag{1.5}$$

The Schur problem is to find a Schur function $f(z)$ such that the first n Taylor coefficients coincide with given numbers $s_0, \ldots, s_{n-1}$. The set

$$\mathscr{S}_n := \{(s_0(f), \ldots, s_{n-1}(f)); f \in \mathfrak{S}_1\}$$

is a compact subset of $\mathbb{C}^n$. In the general matrix case, we will study the impact of uniform sampling on the space of Taylor coefficients of these functions. These results are new, even in the scalar case.

One of the main objects of random matrix theory is to obtain asymptotic results in the limit of large size. Here, on the contrary, the size p of matrices is fixed but the dimension n of the array of matrices tends to infinity. At first insight, these two topics are very distinct. Nevertheless, even in the case $p = 1$, there is a connection between the random moment problem and the random matrix theory, as described in [18]. Let us formulate it shortly in the generic situation. The spectral measure of the pair consisting of a $n \times n$ matrix (unitary or Hermitian) and a fixed vector is a discrete measure. It can be described either by its locations (n points) and its weights, or by a convenient array of its moments. When the matrix is random, both representations have remarkable distributions, and the asymptotical behaviour can be considered from two points of view. If now we fix p orthonormal vectors instead of only one, we obtain a random matricial spectral measure and we may consider the array of its (matricial) moments. This asymptotics will be treated in a forthcoming paper.

The paper is organized as follows. Section 2 is devoted to the case $K = [0, 1]$. It begins with useful definitions and properties around LDPs and ends with the main result on level 2 LDP (Theorem 2.8). Section 3 is devoted to the case $K = \mathbb{T}$. We first show a CLT (Theorem 3.6 and Corollary 3.7) and then turn to large deviation results (Corollaries 3.8 and 3.9, Theorem 3.10). In Section 4, we establish an LDP for random Carathéodory functions and random Schur functions, respectively (Theorem 4.1). All technical proofs are postponed to Section 5.

2. Matrix measures on [0, 1]

Here, we will work on $K = [0, 1]$ and the set defined in (1.1) is

$$\mathcal{M}_n^{[0,1]} := \left\{ \mathbf{S}_n = (S_1, \ldots, S_n) \mid S_j := \int_0^1 x^j d\mu(x),\ j = 1, \ldots, n\ \mu \in \mathcal{P}([0, 1]) \right\} \subset (\mathcal{S}_p^+(\mathbb{C}))^n. \tag{2.1}$$

The moment space $\mathcal{M}_n^{[0,1]}$ is a compact subset of $(\mathcal{S}_p^+(\mathbb{C}))^n$ with nonempty interior [10]. Therefore the uniform distribution $\mathcal{U}(\mathcal{M}_n^{[0,1]})$ is well defined by the density

$$\left(\int_{\mathcal{M}_n^{[0,1]}} dS_1 \cdots dS_n \right)^{-1} I\{\mathbf{S}_n \in \mathcal{M}_n^{[0,1]}\} \tag{2.2}$$

with respect to $dS_1 \cdots dS_n$ where, if $S = (s_{ij})_{i,j=1}^n$

$$dS = \prod_{i \le j \le n} ds_{ij}^{\Re} \prod_{i < j \le n} ds_{ij}^{\Im}, \tag{2.3}$$

where for $s \in \mathbb{C}$, $s := s^{\Re} + i s^{\Im}$ is the standard decomposition of s in real and imaginary parts. The main tool to study random moments $\mathbf{S}_n \sim \mathcal{U}(\mathcal{M}_n^{[0,1]})$ are the canonical moments which are introduced in the next section.

2.1. Canonical moments for matrix measures on [0, 1]

For a moment vector $(S_1, \ldots, S_n) \in \mathcal{M}_n^{[0,1]}$ we build the block Hankel matrices

$$\underline{H}_{2m} := \begin{pmatrix} S_0 & \cdots & S_m \\ \vdots & & \vdots \\ S_m & \cdots & S_{2m} \end{pmatrix} \qquad \overline{H}_{2m} := \begin{pmatrix} S_1 - S_2 & \cdots & S_m - S_{m+1} \\ \vdots & & \vdots \\ S_m - S_{m+1} & \cdots & S_{2m-1} - S_{2m} \end{pmatrix} \tag{2.4}$$

and

$$\underline{H}_{2m+1} := \begin{pmatrix} S_1 & \cdots & S_{m+1} \\ \vdots & & \vdots \\ S_{m+1} & \cdots & S_{2m+1} \end{pmatrix} \qquad \overline{H}_{2m+1} := \begin{pmatrix} S_0 - S_1 & \cdots & S_m - S_{m+1} \\ \vdots & & \vdots \\ S_m - S_{m+1} & \cdots & S_{2m} - S_{2m+1} \end{pmatrix}. \tag{2.5}$$

Dette and Studden [10] showed that the point $(S_1, \ldots, S_n)$ is in $\mathrm{Int}\mathcal{M}_n^{[0,1]}$ if, and only if, the matrices $\underline{H}_n$ and $\overline{H}_n$ are both positive definite.

For $(S_1, \ldots, S_n) \in \mathrm{Int}(\mathcal{M}_n^{[0,1]})$ we define

$$\begin{aligned}
\underline{h}_{2m}^* &:= (S_{m+1}, \ldots, S_{2m})\\
\underline{h}_{2m-1}^* &:= (S_m, \ldots, S_{2m-1})\\
\bar{h}_{2m}^* &:= (S_m - S_{m+1}, \ldots, S_{2m-1} - S_{2m})\\
\bar{h}_{2m-1}^* &:= (S_m - S_{m+1}, \ldots, S_{2m-2} - S_{2m-1})
\end{aligned}$$

and consider the $p \times p$ matrices

$$S_{n+1}^- := \underline{h}_n^* \underline{H}_{n-1}^{-1} \underline{h}_n, \quad n \ge 1, \tag{2.6}$$

$$S_{n+1}^+ := S_n - \bar{h}_n^* \bar{H}_{n-1}^{-1} \bar{h}_n, \quad n \ge 2, \tag{2.7}$$

(for the sake of completeness we also define $S_1^- = 0$ and $S_1^+ = I_p, S_2^+ = S_1$). Note that S_{n+1}^- and S_{n+1}^+ are continuous functions of $(S_1, \ldots, S_n)$ and that $S_n^- < S_n < S_n^+$ if and only if $(S_1, \ldots, S_n) \in \mathrm{Int}\mathcal{M}_n^{[0,1]}$. These preliminary notations allow to introduce the canonical moments of a matrix measure on $[0, 1]$.

Definition 2.1. For $\mathbf{S}_n = (S_1, \ldots, S_n) \in \mathrm{Int}\, \mathcal{M}_n^{[0,1]}$ we define the canonical moments by

$$U_k = (S_k^+ - S_k^-)^{-1/2}(S_k - S_k^-)(S_k^+ - S_k^-)^{-1/2}, \quad k = 1, \ldots, n. \tag{2.8}$$

It is clear that each $U_k \in \mathcal{S}_p(\mathbb{C})$ and satisfies $0_p < U_k < I_p$. Therefore we can define a mapping

$$\begin{aligned}
\varphi^{(n)} &: \mathrm{Int}\mathcal{M}_n^{[0,1]} \longrightarrow (0_p, I_p)^n,\\
\varphi^{(n)}(\mathbf{S}_n) &= \mathbf{U}_n = (U_1, \ldots, U_n).
\end{aligned} \tag{2.9}$$

By Eq. (2.8), the ordinary moments can be recursively calculated from the canonical moments and the mapping $\varphi^{(n)}$ is one-to-one. Now consider a random vector of moments $\mathbf{S}_n \sim \mathcal{U}(\mathcal{M}_n^{[0,1]})$, then $\mathbf{S}_n \in \mathrm{Int}\mathcal{M}_n^{[0,1]}$ almost surely. Dette and Nagel [9] showed that the corresponding canonical moments $\mathbf{U}_n = \varphi^{(n)}(\mathbf{S}_n)$ are independent and that $U_k \in \mathcal{S}_p^+(\mathbb{C})$ follows a complex matricial distribution $\mathrm{Beta}_p(p(n-k+1), p(n-k+1))$ where for $a, b > p-1$ the distribution $\mathrm{Beta}_p(a, b)$ has the density (with respect to dX)

$$\mathcal{B}_p(a, b)^{-1}(\det X)^{a-p}(\det(I_p - X))^{b-p} \tag{2.10}$$

(see [23] or [27]). The normalizing constant $\mathcal{B}_p(a, b)$ is defined by

$$\mathcal{B}_p(a, b) := \frac{\Gamma_p(a)\Gamma_p(b)}{\Gamma_p(a+b)}, \quad a, b > p-1. \tag{2.11}$$

Here $\Gamma_p(a)$ denotes the complex multivariate Gamma function

$$\Gamma_p(a) := \pi^{p(p-1)/2} \prod_{i=1}^p \Gamma(a - i + 1), \quad a > p - 1.$$

The matricial Beta distribution is one of the three main distributions of complex Hermitian matrices, together with the Gaussian unitary ensemble GUE_p having the density

$$(2\pi^p)^{-p/2} e^{-\mathrm{tr}\frac{1}{2}X^2} \tag{2.12}$$

and the complex Wishart distribution $W_p(a)$ with density

$$\Gamma_p(a)^{-1}(\det X)^{a-p} e^{-\mathrm{tr}X}, \quad a > p-1. \tag{2.13}$$

We refer to [26,15] for more on these distributions. The following result shows that the Wishart distribution and the Gaussian distribution appear as weak limits of the matricial Beta distribution when the parameters tend to infinity.

Theorem 2.2. *Let $(a_n)_n$ be a sequence of positive parameters such that* $\lim_{n\to\infty} a_n = \infty$.

(i) *If $X_n \sim \mathrm{Beta}_p(a_n, a_n)$, then*

$$\sqrt{8a_n}\left(X_n - \frac{1}{2}I_p\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathrm{GUE}_p.$$

(ii) *Let $c > p-1$. If $X_n \sim \mathrm{Beta}_p(c, a_n)$ then*

$$a_n X_n \xrightarrow[n\to\infty]{\mathcal{D}} W_p(c).$$

The first statement shows that the centred rescaled canonical moments converge in distribution to the GUE_p. This is the keystone to obtain a CLT in [9]. Notice also, that this implies that the sequence (X_n) converges in probability towards $\frac{1}{2}I_p$. The second statement will play an important role in the study of matrix measures on $\mathbb{T}$.

2.2. Large deviations

To make this paper self contained let us first recall what is an LDP. For more on LDP we refer to [7]. Let $(u_n)_n$ be an increasing positive sequence of real numbers going to infinity with n.

Definition 2.3. Let U be a Hausdorff topological space and $\mathcal{B}(U)$ its Borel σ-field. We say that a sequence $(Q_n)_n$ of probability measures on $(U, \mathcal{B}(U))$ satisfies an LDP with speed (u_n) and rate function $I : U \to [0, \infty]$ if:

(i) I is lower semicontinuous.
(ii) For any measurable set A of U:

$$-I(\mathrm{Int}\,A) \le \liminf_{n\to\infty} u_n^{-1} \log Q_n(A) \le \limsup_{n\to\infty} u_n^{-1} \log Q_n(A) \le -I(\mathrm{Clo}\,A),$$

where $I(A) = \inf_{\xi\in A} I(\xi)$ and $\mathrm{Clo}\,A$ is the closure of A.

If we omit to give the speed it means that $u_n = n$. We say that the rate function I is good if its level sets $\{x \in U : I(x) \le a\}$ are compact for any $a \ge 0$. More generally, a sequence of U-valued random variables is said to satisfy an LDP if their distributions satisfy an LDP.

We will need the following well known large deviation result (see e.g. [7, Chapter 4 p. 126 and 130]).

Contraction principle. Assume that $(Q_n)_n$ satisfies an LDP on $(U, \mathcal{B}(U))$ with good rate function I and speed (u_n). Let T be a continuous mapping from U to another Hausdorff topological space V. Then $Q_n \circ T^{-1}$ satisfies an LDP on $(V, \mathcal{B}(V))$ with speed (u_n) and good rate function

$$I'(y) = \inf_{x:T(x)=y} I(x), \quad (y \in V).$$

The so-called cross entropy (or Kullback information) plays an important role in the interpretation of some of our results, for the sake of completeness we recall its definition.

Kullback Information. Let P and Q be probability distributions on $(U, \mathcal{B}(U))$. The Kullback information of P with respect to Q is

$$\mathcal{K}(P; Q) := \begin{cases} \displaystyle\int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q \text{ and } \quad \log \dfrac{dP}{dQ} \in L^1(P) \\ \infty & \text{otherwise.} \end{cases}$$

Our first result is an LDP for matricial beta distributions. For the case where the matrix dimension tends to infinity, various LDPs can be found in the literature, see for example [20]. Here we are interested in the case of fixed dimension and growing parameters.

Theorem 2.4. *Let $a_0, a > 0$ and $c > p - 1$. Further set, for $n \ge 1$, $a_n := a_0 + an$.*

(i) *Let $B_n \sim \mathrm{Beta}_p(a_n, a_n)$. Then B_n satisfies an LDP with good rate function*

$$\mathcal{I}_B^{(1)}(B) = \begin{cases} -a \log\det(B - B^2) - 2ap \log 2, & \text{if } 0_p < B < I_p, \\ \infty & \text{otherwise}. \end{cases} \tag{2.14}$$

(ii) *Let $B_n \sim \mathrm{Beta}_p(c, a_n)$. Then B_n satisfies an LDP with good rate function*

$$\mathcal{I}_B^{(2)}(B) = \begin{cases} -a \log\det(I_p - B), & \text{if } 0_p < B < I_p, \\ \infty & \text{otherwise}. \end{cases} \tag{2.15}$$

Remark 2.5. For the sake of simplicity we show an LDP only for very special sequences of parameters. This is enough to obtain our further results. However, the result holds for arbitrary sequences $a_n \nearrow \infty$.

As a consequence of the last theorem, an LDP for the random matricial vector $\mathbf{U}_k^{(n)} = (U_1, \ldots, U_k)$ of the first k canonical moments associated to a random matricial vector $\mathbf{S}_n$ uniformly drawn holds. Indeed, as mentioned before, the components of $\mathbf{U}_k^{(n)} = (U_1, \ldots, U_k)$ are independent, so that we obtain:

Corollary 2.6. *Let $\mathbf{S}_n \sim \mathcal{U}(\mathcal{M}_n^{[0,1]})$ and for k fixed, let $\mathbf{U}_k^{(n)}$ denote the projection of $\mathbf{U}_n = \varphi^{(n)}(\mathbf{S}_n)$ onto the first k coordinates. Then the sequence $\left(\mathbf{U}_k^{(n)}\right)_n$ satisfies an LDP in $(\mathcal{S}_p^+(\mathbb{C}))^k$ with good rate function*

$$\mathcal{I}_{\mathbf{U}}(\mathbf{U}_k) = \begin{cases} -\displaystyle\sum_{i=1}^{k} p \log\det(U_i - U_i^2) - 2kp^2 \log 2, & \text{if } \mathbf{U}_k \in (0_p, I_p)^k, \\ \infty & \text{otherwise}. \end{cases} \tag{2.16}$$

Obviously the rate function $\mathcal{I}_{\mathbf{U}}$ achieves its minimum value 0 at $\mathbf{U}_k = \left(\frac{1}{2}I_p, \ldots, \frac{1}{2}I_p\right)$ that appears as discussed before for general sequences of matricial beta distributed random matrices, see Theorem 2.2) as the limit of $\mathbf{U}_k^{(n)}$. Notice also that the constant infinite sequence $U_k = \frac{1}{2}I_p$, $k \geq 1$ is the moment sequence of the matrix arcsine law ν_p defined by

$$d\nu_1(x) = \frac{dx}{\pi\sqrt{x(1-x)}}, \qquad d\nu_p(x) = d\nu_1(x)I_p, \quad (p > 1), \tag{2.17}$$

see [9].

Now, the vector of ordinary moments $(S_1, \ldots, S_k)$ is a continuous function of the canonical moment vector $\mathbf{U}_k^{(n)}$. So we obtain the following Corollary from Corollary 2.6 by a simple application of the contraction principle and the identity

$$\det(S_{k+1}^+ - S_{k+1}^-) = \det \prod_{i=1}^{k} U_i(I_p - U_i) \tag{2.18}$$

(see [10]).

Corollary 2.7. *Let $\mathbf{S}_n \sim \mathcal{U}(\mathcal{M}_n^{[0,1]})$ and for $k < n$ let $\mathbf{S}_k^{(n)}$ denote the projection of $\mathbf{S}_n$ onto the first k coordinates. Then $\mathbf{S}_k^{(n)}$ satisfies an LDP with good rate function*

$$\mathcal{I}_{\mathbf{S}}(\mathbf{S}_k) = \begin{cases} -p \log\det(S_{k+1}^+ - S_{k+1}^-) - 2kp^2 \log 2, & \text{if } \mathbf{S}_k \in \operatorname{Int} \mathcal{M}_n^{[0,1]}, \\ \infty & \text{otherwise}. \end{cases} \tag{2.19}$$

We end this section with an LDP for random matrix measures on [0, 1]. For this purpose, for every n let $\mathbb{P}_n$ denote any probability measure on $\mathcal{P}([0, 1])$ such that the pushforward by the mapping

$$\mu \in \mathcal{P}([0,1]) \mapsto \mathbf{S}_n(\mu) = (S_1(\mu), \ldots, S_n(\mu)) \in \mathcal{M}_n^{[0,1]}$$

is $\mathcal{U}(\mathcal{M}_n^{[0,1]})$.

Theorem 2.8. *The sequence $(\mathbb{P}_n)_n$ satisfies an LDP in $\mathbb{M}_1(\mathcal{P}([0, 1]))$ with good rate function*

$$\mathcal{I}_{[0,1]}(\mu) = \begin{cases} -p \displaystyle\int_0^1 \log\det W(x)\, d\nu_1(x), & \text{if } \nu_1\{\det W = 0\} = 0, \\ \infty & \text{otherwise}, \end{cases} \tag{2.20}$$

where $d\mu(x) = W(x)d\nu_p(x) + d\mu^s(x)$ is the Lebesgue decomposition[1] *of μ with respect to ν_p as matricial measures on* [0, 1] *(ν_1 and ν_p are the arcsine measures defined by (2.17)).*

Remark 2.9. 1. When $p = 1$ (scalar case) the rate function is also

$$\mathcal{I}_{[0,1]}(\mu) = \mathcal{K}(\nu_1; \mu). \tag{2.21}$$

The matricial case has also an interpretation in terms of cross-entropy which we hope to address in a future work.

2. A cousin result of Theorem 2.8 holds in the frame of real matrix measures. In this case the constant p in the rate function is replaced by $\frac{p+1}{2}$. All arguments remain essentially unchanged and we refer to [9] for the underlying results on real matrix valued random moments and the corresponding canonical moments.
3. From Theorem 2.8 and Corollary 2.7 together with the contraction principle one easily obtains the following identity of rate functions. For $\mathbf{S}_k = (S_1, \ldots, S_k) \in \operatorname{Int} \mathcal{M}_n^{[0,1]}$ we have

$$\mathcal{I}_{\mathbf{S}}(\mathbf{S}_k) = -p \log\det(S_{k+1}^+ - S_{k+1}^-) - 2kp^2 \log 2 = \inf_{\mathcal{D}(\mathbf{S}_k)} -p \int_0^1 \log\det W(x) d\nu_1(x), \tag{2.22}$$

where

$$\mathcal{D}(\mathbf{S}_k) = \left\{ \mu \in \mathcal{P}([0,1]) \,\middle|\, \int_0^1 x^j d\mu(x) = S_j,\ j = 1, \ldots, k \right\} \tag{2.23}$$

and W is defined as in Theorem 2.8.

[1] See [28] on Lebesgue decomposition for matricial measures.

3. Matrix measures on $\mathbb{T}$: the trigonometric case

In this section, we consider the space $\mathcal{P}(\mathbb{T})$ of matrix-valued probability measures on the unit circle $\mathbb{T}$. In what follows Γ_j denotes the j-th trigonometric moment of a matrix measure $\mu \in \mathcal{P}(\mathbb{T})$, that is

$$\Gamma_j = \Gamma_j(\mu) = \int_{-\pi}^{\pi} e^{ij\theta} d\mu(\theta) \tag{3.1}$$

and for $n \in \mathbb{N}$ and $p \geq 1$ the set defined in (1.1) is

$$\mathcal{M}_n^{\mathbb{T}} := \left\{ (\Gamma_1, \ldots, \Gamma_n) \mid \Gamma_j = \Gamma_j(\mu), \mu \in \mathcal{P}(\mathbb{T}) \right\} \subset \left(\mathbb{C}^{p\times p}\right)^n . \tag{3.2}$$

Unlike to moments of matrix measures on [0, 1], the moment Γ_j is no more Hermitian. Therefore we use the following Lebesgue measure on $\mathbb{C}^{p\times p}$. For $X \in \mathbb{C}^{p\times p}$ define

$$dX = \prod_{1\leq i,j\leq p} dx_{ij}^{\Re} dx_{ij}^{\Im}. \tag{3.3}$$

3.1. Canonical moments on $\mathbb{T}$

As in the above section we use a notion of canonical moments to study $\mathcal{M}_n^{\mathbb{T}}$. First, for $(\Gamma_1, \ldots, \Gamma_n) \in \mathcal{M}_n^{\mathbb{T}}$, we build the block Toeplitz matrix

$$T_n := \left(\Gamma_{i-j}\right)_{i,j=0,\ldots,n} . \tag{3.4}$$

Dette and Wagener [12] showed that $(\Gamma_1, \ldots, \Gamma_n) \in \operatorname{Int} \mathcal{M}_n^{\mathbb{T}}$ if and only if $T_n > 0$. Therefore this interior is nonempty. Furthermore they proved that for $(\Gamma_1, \ldots, \Gamma_n) \in \operatorname{Int} \mathcal{M}_n^{\mathbb{T}}$ the range of the moment Γ_{n+1} is the set

$$K_n = \left\{ W \in \mathbb{C}^{p\times p} \mid L_n^{-1/2}(W - M_n)R_n^{-1/2} = U, UU^* \leq I_p \right\}, \tag{3.5}$$

where the matrices L_n, R_n and M_n are defined by

$$L_n := \left[I_p - (\Gamma_1, \ldots, \Gamma_n) T_{n-1}^{-1} (\Gamma_1, \ldots, \Gamma_n)^* \right], \tag{3.6}$$

$$R_n := \left[I_p - (\Gamma_{-n}, \ldots, \Gamma_{-1}) T_{n-1}^{-1} (\Gamma_{-n}, \ldots, \Gamma_{-1})^* \right], \tag{3.7}$$

$$M_n := (\Gamma_1, \ldots, \Gamma_n) T_{n-1}^{-1} (\Gamma_{-n}, \ldots, \Gamma_{-1})^* , \tag{3.8}$$

respectively. In this frame, canonical moments are defined by normalizing the moments in the following way.

Definition 3.1. For $(\Gamma_1, \ldots, \Gamma_n) \in \operatorname{Int} \mathcal{M}_n^{\mathbb{T}}$ we define the canonical moments $A_j, j = 1, \ldots, n$ setting

$$A_1 := \Gamma_1, \qquad A_j := L_{j-1}^{-1/2}(\Gamma_j - M_{j-1})R_{j-1}^{-1/2} \quad (j = 2, \ldots, n). \tag{3.9}$$

The canonical moments of a matrix measure always lie in the set

$$\mathbb{D}_p = \{U \in \mathbb{C}^{p\times p} \mid UU^* \leq I_p\} \tag{3.10}$$

and coincide with the well known Verblunsky coefficients appearing in the Szegö recursion of orthonormal matrix polynomials (see e.g. [29, Section 2.13]). They are connected to the trigonometric moments by a one-to-one mapping $\psi^{(n)} : \operatorname{Int} \mathcal{M}_n^{\mathbb{T}} \to \operatorname{Int} \mathbb{D}_p^n$ recursively defined by Definition 3.1.

We now state a Taylor expansion of the inverse of the mapping $\psi^{(n)}$. Here and in the following $\|M\|$ always denotes the Frobenius norm of the complex entries matrix M, that is

$$\|M\| := \operatorname{tr}(M^*M)^{1/2}.$$

Lemma 3.2. *Let $n \in \mathbb{N}^+$ and $\mathbf{A}_n = (A_1, \ldots, A_n) \in \operatorname{Int} \mathbb{D}_p^n$. The mapping $(\psi^{(n)})^{-1} : \mathbf{A}_n \mapsto \mathbf{X}_n = (\Gamma_1, \ldots, \Gamma_n)$ induced by the definition of canonical moments has an order one Taylor expansion at 0. Namely,*

$$\mathbf{X}_n = \mathbf{A}_n + o(\|\mathbf{A}_n\|). \tag{3.11}$$

In the following this Taylor expansion will be used to derive results concerning trigonometric moments from results obtained for canonical moments.

3.2. Weak convergence in the trigonometrical case

As in the real case we define a uniform distribution $\mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$ on $\mathcal{M}_n^{\mathbb{T}}$ by the density

$$\left(\int_{\mathcal{M}_n^{\mathbb{T}}} d\Gamma_1 \cdots d\Gamma_n\right)^{-1} I\left\{\mathbf{X}_n \in \mathcal{M}_n^{\mathbb{T}}\right\}, \tag{3.12}$$

now with respect to the measure (3.3). We first state a result on the distribution of the canonical moments when the corresponding trigonometric moments are uniformly distributed.

Lemma 3.3. *Let $\mathbf{X}_n \sim \mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$ and $\mathbf{A}_n = (A_1, \ldots, A_n) = \psi^{(n)}(\mathbf{X}_n) \in (\mathbb{D}_p)^n$ denote the corresponding vector of canonical moments. Then $A_1, \ldots, A_n$ are independent and for $k = 1, \ldots, n$, A_k has density*

$$\frac{1}{c_k^{(n)}} \det\left(I_p - A_k^* A_k\right)^{2p(n-k)} \tag{3.13}$$

with respect to (3.3), where $c_k^{(n)}$ is a normalizing constant.

We now establish a relation between the Hermitian random matrices from Section 2 and matricial random variables without symmetry condition:

Theorem 3.4. *If A_k is a random matrix with density (3.13), then*

$$A_k \stackrel{(d)}{=} V B_k^{1/2}, \tag{3.14}$$

where V and B_k are independent, V is Haar distributed in $\mathbb{U}(p)$ and B_k follows a multivariate complex Beta distribution $\mathrm{Beta}_p(p, 2p(n-k)+p)$ (see (2.10)).

The previous theorem is a particular case of the following general variable change result. It is quite natural and useful in other asymptotical problems involving random complex matrices. Similar arguments have been used recently by Fischmann et al. [14] to generate matrices of the Ginibre ensemble.

Proposition 3.5. *Let M be a $p \times p$ random matrix with complex entries whose density with respect to (3.3) is $f(x_1^2(M), \ldots, x_p^2(M))$ where $x_1(M), \ldots, x_p(M)$ are the (positive) singular values, and f is a symmetric function. Then, the random matrices $H = M^*M$ and $U = (M^*M)^{-1/2}M$ are independent, U is Haar distributed in $\mathbb{U}(p)$ and the density of $H \in \mathcal{S}_p^+(\mathbb{C})$ with respect to (2.3) is proportional to $f(\lambda_1(H), \ldots, \lambda_p(H))$ where $\lambda_1(H), \ldots, \lambda_p(H)$ are the eigenvalues of H.*

We are now in the position to give our first limit theorem in the trigonometrical case.

Theorem 3.6. *Let $\mathbf{X}_n \sim \mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$, $\mathbf{A}_n = \psi^{(n)}(\mathbf{X}_n)$ and $\mathbf{A}_n^k$ denote the projection onto the first k coordinates (k is fixed). Then for $n \to \infty$ the weak convergence*

$$\sqrt{2pn}\mathbf{A}_n^k \xrightarrow{\mathcal{D}} \mathcal{G}_k \tag{3.15}$$

holds, where $\mathcal{G}_k = (G_1, \ldots, G_k)$ and $G_1, \ldots, G_k$ are complex iid random matrices of the Ginibre complex ensemble (see [19]), or, in other words, having density

$$g(G) = \pi^{-p^2} \exp(-\|G\|^2) \tag{3.16}$$

with respect to (3.3).

As a consequence, using the Taylor expansion of Lemma 3.2 and the δ-method (see for example [31]), we obtain a weak convergence theorem for the rescaled random trigonometric moments. This is the subject of the next corollary.

Corollary 3.7. *Let $\mathbf{X}_n \sim \mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$ and $\mathbf{X}_n^k$ denote the projection onto the first k coordinates (k is fixed). Then when $n \to \infty$*

$$\sqrt{2pn}\mathbf{X}_n^k \xrightarrow{\mathcal{D}} \mathcal{G}_k, \tag{3.17}$$

(here $\mathcal{G}_k$ is as in Theorem 3.6).

3.3. Large deviations in the trigonometrical case

Our final results concern LDPs for random moments and matrix measures on the unit circle. The large deviations in the scalar trigonometrical case are due to [24, Theorems 4.2 and 4.4]. Nevertheless, in that paper, there was a mistake in the computation of the Jacobian. A power 2 is missing.

The proof of the next Corollary follows directly from part (ii) of Theorem 2.4 (applying the contraction principle). We again use the equality $A_k \stackrel{(d)}{=} VB_k^{1/2}$, where $B_k \sim \mathrm{Beta}_p(p, 2p(n-k)+p)$ and V is Haar distributed on the unitary group. By Lemma 3.3 the canonical moments are independent, giving the final form of the rate function.

Corollary 3.8. *Let $\mathbf{X}_n \sim \mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$, $\mathbf{A}_n = \psi^{(n)}(\mathbf{X}_n)$ and $\mathbf{A}_n^k$ denote the projection onto the first k coordinates (k is fixed). Then $\mathbf{A}_n^k$ satisfies an LDP with good rate function*

$$\mathcal{I}_{\mathbf{A}}(\mathbf{Z}) = \mathcal{I}_{\mathbf{A}}(Z_1, \dots, Z_k) = \begin{cases} -2p \displaystyle\sum_{i=1}^{k} \log\det\left(I_p - Z_i^* Z_i\right), & \textit{if } \mathbf{Z} \in \mathrm{Int}\, \mathbb{D}_p^k, \\ \infty & \textit{otherwise}. \end{cases} \tag{3.18}$$

Another application of the contraction principle for the mapping $\psi^{(n)}$ yields the following LDP for the trigonometric moments.

Corollary 3.9. *Let $\mathbf{X}_n \sim \mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$ and $\mathbf{X}_n^k$ denote the projection onto the first k coordinates (k is fixed). Then $\mathbf{X}_n^k$ satisfies an LDP with good rate function*

$$\mathcal{I}_{\Gamma}(\mathbf{X}) = \mathcal{I}_{\Gamma}(\Gamma_1, \dots, \Gamma_k) = \begin{cases} -2p \log \dfrac{\det(T_k)}{\det(T_{k-1})}, & \textit{if } \mathbf{X} \in \mathrm{Int}\, \mathcal{M}_k^{\mathbb{T}}, \\ \infty & \textit{otherwise}. \end{cases} \tag{3.19}$$

Here, T_k denotes the block Toeplitz matrix (3.4) defined by $(\Gamma_1, \dots, \Gamma_k)$.

Finally we state an LDP for a sequence of random matrix measures on $\mathbb{T}$. For every n, let $\mathbb{Q}_n$ denote a probability measure on the set $\mathcal{P}(\mathbb{T})$ such that the pushforward by the mapping

$$\mu \in \mathcal{P}(\mathbb{T}) \mapsto \mathbf{X}_n(\mu) = (\Gamma_1(\mu), \dots, \Gamma_n(\mu)) \in \mathcal{M}_n^{\mathbb{T}}$$

is $\mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$.

Theorem 3.10. *The sequence $(\mathbb{Q}_n)_n$ satisfies an LDP in $\mathbb{M}_1(\mathcal{P}(\mathbb{T}))$ with good rate function*

$$\mathcal{I}_{\mathbb{T}}(\mu) = \begin{cases} -\dfrac{p}{\pi} \displaystyle\int_{\mathbb{T}} \log\det(W(\theta))d\theta, & \textit{if } \det W(\theta) \neq 0 \textit{ a.e.}, \\ \infty & \textit{otherwise}, \end{cases} \tag{3.20}$$

where $d\mu(\theta) = W(\theta)\frac{d\theta}{2\pi} + d\mu^s(\theta)$ is the Lebesgue decomposition of μ with respect to $\frac{d\theta}{2\pi} I_p$ as matricial measures on $\mathbb{T}$.

The proof is very similar to that one of Theorem 2.8 and therefore omitted.

Remark 3.11. 1. For $p = 1$ the rate function is also

$$\mathcal{I}_{\mathbb{T}}(\mu) = 2\mathcal{K}\left(\frac{d\theta}{2\pi}; \mu\right). \tag{3.21}$$

It is the content of Theorem 4.4 in [24] but a factor 2 was missing in that paper, owing to a mistake in the Jacobian (7.2).

2. As in Remark 2.9 we see, from Theorem 3.10 and Corollary 3.9 together with the contraction principle, the following identity of rate functions. For $\mathbf{X}_k = (\Gamma_1, \dots, \Gamma_k) \in \mathrm{Int}\, \mathcal{M}_k^{\mathbb{T}}$ we have

$$\mathcal{I}_{\Gamma}(\mathbf{X}_k) = -2p \log \frac{\det(T_k)}{\det(T_{k-1})} = \inf_{\mathcal{C}(\mathbf{X}_k)} -\frac{p}{\pi} \int_{\mathbb{T}} \log\det(W(\theta))d\theta, \tag{3.22}$$

where

$$\mathcal{C}(\mathbf{X}_k) = \left\{ \mu \in \mathcal{P}(\mathbb{T}) \middle| \int_{-\pi}^{\pi} e^{ij\theta} d\mu(\theta) = \Gamma_j, j = 1, \dots, k \right\} \tag{3.23}$$

and W is defined as in Theorem 3.10.

4. Application: random Carathéodory and Schur matrix functions

In the above Theorem 3.10, we studied a family of random measures. Since the truncated trigonometrical moment problem is closely connected to the Carathéodory problem, which is itself connected to the Schur problem, it may be natural to look at the corresponding random functions. In this section we study the impact of uniform sampling on the space of Taylor coefficients of these functions. We first give the framework, which can be seen in [6] or [13] and then we give our results. It seems to be new, even in the scalar case.

4.1. Carathéodory and Schur matrix-valued functions

As before, let p be a given positive integer. By a $\mathbb{C}^{p\times p}$-valued Carathéodory matrix function $F(z)$, one means a $p\times p$ matrix-valued function which is holomorphic in $\mathbb{D}$, has a nonnegative real part there

$$F^{\Re}(z) \equiv \frac{1}{2}(F(z)+F(z)^*) \geq 0, \quad z\in\mathbb{D},$$

and such that $F(0)=I_p$. We use the notation $\mathcal{C}_p$ to designate the class of such $\mathbb{C}^{p\times p}$-valued Carathéodory matrix functions. We also define the class $\mathfrak{S}_p$ of $\mathbb{C}^{p\times p}$-matrix valued functions f analytic in $\mathbb{D}$ and contractive there, i.e. such that $f(z)\in\bar{\mathbb{D}}_p$ for $z\in\mathbb{D}$, which are called matrix valued Schur functions.

The correspondence

$$F(z)=(I_p+zf(z))(I_p-zf(z))^{-1} \quad\text{and}\quad f(z)=z^{-1}(F(z)-I_p)(F(z)+I_p)^{-1} \tag{4.1}$$

is one-to-one between $\mathcal{C}_p$ and $\mathfrak{S}_p$. Any $F\in\mathcal{C}_p$ has a representation

$$F(z)=\int_{\mathbb{T}}\frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\mu(e^{i\theta}), \quad z\in\mathbb{D},$$

for a unique $\mu\in\mathcal{P}(\mathbb{T})$. Any $F\in\mathcal{C}_p$ has a finite radial limit $\lim_{r\uparrow 1}F(re^{i\theta}) =: F(e^{i\theta})$ for almost every θ. The corresponding value of f in such a point $e^{i\theta}$ will be denoted by $f(e^{i\theta})$. If

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

is the Lebesgue decomposition of μ one has the identity

$$W(\theta)=F^{\Re}(e^{i\theta})=(I_p-e^{-i\theta}f(e^{i\theta})^*)^{-1}(I_p-f(e^{i\theta})^*f(e^{i\theta}))(I_p-e^{i\theta}f(e^{i\theta}))^{-1} \tag{4.2}$$

a.e. and for a.e. θ, $\det W(\theta)\neq 0$ iff $f(e^{i\theta})^*f(e^{i\theta})<1$ (Proposition 3.16 in [6]).

The Taylor expansion of F is given by

$$F(z)=I_p+2\sum_{k=1}^{\infty}C_k(F)z^k,$$

where the coefficients are the conjugate trigonometric moments of the matrix measure μ associated to F, i.e.

$$C_k(F)=\int_{\mathbb{T}}e^{-ik\theta}d\mu(\theta)=\Gamma_k^*.$$

The classical Carathéodory problem is to find $F\in\mathcal{C}_p$ such that the first n Taylor coefficients coincide with given $p\times p$ matrices $C_1,\dots,C_n$. It is clearly equivalent to the truncated moment problem.

Each Schur function in $\mathfrak{S}_p$ is associated to a matrix measure $\mu\in\mathcal{P}(\mathbb{T})$, hence to the sequence of its canonical moments $(A_k)_{k\geq 1}$. For every $j\geq 1$, let f_j be the Schur function corresponding to the shifted sequence $(A_k)_{k\geq j+1}$, and set $f_0=f$. From Theorem 3.19 of [6], we have the recursive relations:

$$f_k(z)=z^{-1}(B_k^R)^{-1}[f_{k-1}(z)-A_k^*][I_p-A_kf_{k-1}(z)]^{-1}B_k^L, \tag{4.3}$$

$$f_k(z)=(B_{k+1}^R)^{-1}[zf_{k+1}(z)+A_{k+1}^*][I_p+zA_{k+1}f_{k+1}]^{-1}B_{k+1}^L, \tag{4.4}$$

where

$$B_k^R:=[I_p-A_k^*A_k]^{1/2}, \qquad B_k^L:=[I_p-A_kA_k^*]^{1/2}. \tag{4.5}$$

The Taylor expansion of f is

$$f(z)=\sum_0^{\infty}G_k(f)z^k. \tag{4.6}$$

The Schur problem is to find a Schur function $f \in \mathfrak{S}_p$ such that the first n Taylor coefficients coincide with given numbers $G_0, \dots, G_{n-1}$. A solution exists if and only if the block matrix

$$\begin{pmatrix} G_0 & 0 & 0 & \dots & 0 \\ G_1 & G_0 & 0 & \dots & 0 \\ G_2 & G_1 & G_0 & \dots & 0 \\ & & \dots & & \\ G_{n-1} & G_{n-2} & G_{n-3} & \dots & G_0 \end{pmatrix}$$

is contractive, i.e. if it satisfies $GG^* \le I_{np}$ (see [13, Theorem 3.1.1]). The set

$$\mathscr{S}_n := \{(G_0(f), \dots, G_{n-1}(f)); f \in \mathfrak{S}_p\}$$

is a relatively compact subset of $(\mathbb{C}^{p\times p})^n$.

In both problems, the system of canonical moments (alias Verblunsky coefficients, alias Schur coefficients) plays a prominent role. In Section 3.3 we saw that the dependence between the moments (hence the C_k's) and the canonical moments is triangular. The relation between the Taylor coefficients of a Schur function and its Schur coefficients (i.e. the canonical moments of the associated measure) is also triangular. We postpone the presentation of this point in the proof of Theorem 4.1.

4.2. Randomization: large deviations

For every n let $\mathbb{P}_n^c$ denote a probability measure on the set $\mathcal{C}_p$ such that the pushforward by the mapping

$$F \in \mathcal{C}_p \mapsto \mathbf{C}_n(F) = (C_1(F), \dots, C_n(F)) \in \mathcal{M}_n^{\mathbb{T}}$$

is $\mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$. Let also $\mathbb{P}_n^s$ denote a probability measure on the set $\mathfrak{S}_p$ such that the pushforward by the mapping

$$f \in \mathfrak{S}_p \mapsto \mathbf{G}_n(f) := (G_0(f), \dots, G_{n-1}(f)) \in \mathscr{S}_n$$

is $\mathcal{U}(\mathscr{S}_n)$.

One gets the following LDP for matrix valued Carathéodory and Schur functions.

Theorem 4.1. *The sequence $(\mathbb{P}_n^c)_n$ satisfies an LDP in $\mathbb{M}_1(\mathcal{C}_p)$ with good rate function*

$$\mathcal{I}_p^C(F) = \begin{cases} -\dfrac{p}{\pi}\displaystyle\int_{\mathbb{T}} \log\det F^{\Re}(e^{i\theta})d\theta, & \text{if } \det F^{\Re}(e^{i\theta}) \ne 0 \text{ a.e.,} \\ \infty & \text{otherwise.} \end{cases} \tag{4.7}$$

The sequence $(\mathbb{P}_n^s)_n$ satisfies an LDP in $\mathbb{M}_1(\mathfrak{S}_p)$ with good rate function

$$\mathcal{I}_p^S(f) = \begin{cases} -\dfrac{p}{\pi}\displaystyle\int_{\mathbb{T}} \log\det(I_p - f(e^{i\theta})^* f(e^{i\theta}))d\theta, & \text{if } \det(I_p - f(e^{i\theta})^* f(e^{i\theta})) \ne 0 \text{ a.e.,} \\ \infty & \text{otherwise.} \end{cases} \tag{4.8}$$

Remark 4.2. Behind Theorems 3.10 and 4.1 (and as will be seen in the proofs), there is a triple identity, which holds true in the generic case:

$$\begin{aligned} \sum_n \log\det(I_p - A_n A_n^*) &= \int_{\mathbb{T}} \log\det W(\theta)\frac{d\theta}{2\pi} = \int_{\mathbb{T}} \log\det F^{\Re}(e^{i\theta})\frac{d\theta}{2\pi} \\ &= \int_{\mathbb{T}} \log\det(I_p - f(e^{i\theta})^* f(e^{i\theta}))\frac{d\theta}{2\pi}, \end{aligned} \tag{4.9}$$

say

$$(1) = (2) = (3) = (4).$$

Equality $(1) = (2)$ is the Szegö Theorem for matrix-valued measures (see Theorem 2.13.5 in [29]), and $(1) = (4)$ is the matricial version of Boyd's theorem (see 2.7.7 of [29] in the scalar case).

5. Proofs

5.1. *Proof of Theorem 2.2*

If X is $\mathrm{Beta}_p(\alpha, \beta)$ distributed, then

$$X \overset{(d)}{=} (W_1 + W_2)^{-1/2} W_1 (W_1 + W_2)^{-1/2},$$

where $W_1 \sim W_p(\alpha)$ and $W_2 \sim W_p(\beta)$ are independent and Wishart distributed.

For (i), we choose $\alpha = \beta = a_n$ and observe that

$$X_n - \frac{1}{2} I_p \overset{(d)}{=} \frac{1}{2} (W_1 + W_2)^{-1/2} \left[\left(W_1 - a_n I_p\right) + \left(a_n I_p - W_2\right) \right] (W_1 + W_2)^{-1/2}$$

then we apply Proposition A.1(i) and (ii).

For (ii), it is enough to take $\alpha = c$ and $\beta = a_n$ and apply Proposition A.1(i). □

5.2. *Proof of Theorem 2.4*

We give a proof only for $a_n = an$.

To prove (i) let $B_n \sim \mathrm{Beta}_p(an, an)$, then again the following equality in distribution holds

$$B_n \overset{(d)}{=} \left(\sum_{i=1}^{2n} W_i \right)^{-1/2} \left(\sum_{i=1}^{n} W_i \right) \left(\sum_{i=1}^{2n} W_i \right)^{-1/2}, \tag{5.1}$$

where the random variables are independent and $W_p(a)$ distributed. (see e.g. [27]). By Proposition A.2 each component $V_n^{(1)}, V_n^{(2)}$ of the vector

$$\begin{pmatrix} V_n^{(1)} \\ V_n^{(2)} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} W_i \\ \dfrac{1}{n} \displaystyle\sum_{i=n+1}^{2n} W_i \end{pmatrix}$$

satisfies an LDP with good rate function $\Lambda^\star$ given by (A.2).

The independence of the random variables W_i now yields an LDP for $(V_n^{(1)}, V_n^{(2)})$ with good rate function $\Lambda^\star(X) + \Lambda^\star(Y)$. By the contraction principle and equality (5.1) the random variable B_n satisfies an LDP on $(0_p, I_p)$ with good rate function

$$\begin{aligned} \mathcal{I}(Z) &= \inf_{\mathcal{Z}} \left(\Lambda^\star(X) + \Lambda^\star(Y) \right) \\ &= \inf_{\mathcal{Z}} \left(\mathrm{tr}(X+Y) - a \log \det(XY) - 2pa + 2pa \log a \right), \end{aligned}$$

where the infimum is taken over the set

$$\mathcal{Z} = \left\{ (X, Y) \in \mathcal{S}_p^+(\mathbb{C})^2 \mid Z = (X+Y)^{-1/2} X (X+Y)^{-1/2} \right\}.$$

On $\mathcal{Z}$ we have $\det(XY) = \det(Z(I_p - Z)\det(X+Y)^2)$ and we can write the rate function as

$$\mathcal{I}(Z) = -a \log \det \left(Z(I_p - Z) \right) - 2pa + 2pa \log a + \inf_{\mathcal{Z}} \left(\mathrm{tr}(X+Y) - 2a \log \det(X+Y) \right).$$

Appealing to (A.4) with $L = (2a)^{-1}(X+Y)$, we see that

$$\mathcal{I}(Z) = -a \log \det(Z(I_p - Z)) - 2pa \log 2.$$

To prove (ii) let $B_n \sim \mathrm{Beta}_p(c, an)$. Then we have

$$B_n \overset{(d)}{=} \left(\frac{X}{n} + \frac{1}{n} \sum_{i=1}^{n} W_i \right)^{-1/2} \frac{X}{n} \left(\frac{X}{n} + \frac{1}{n} \sum_{i=1}^{n} W_i \right)^{-1/2},$$

where $X \sim W_p(c)$, $(W_i)_{i=1,\ldots,n}$ are iid $W_p(a)$ distributed and X and $(W_i)_{i=1,\ldots,n}$ are independent. By Propositions A.2 and A.3, we get for $\left(\frac{X}{n}, \frac{1}{n}\sum_{i=1}^n W_i\right)$ an LDP with rate function the sum of rate functions and by the contraction principle, we get an LDP with rate function

$$\mathcal{I}(Z) = \inf_{\mathcal{Z}} (\mathrm{tr}\, X + \mathrm{tr}\, Y - a \log \det Y - ap + ap \log a),$$

where $\mathcal{Z}$ is as in the proof of Theorem 2.4(i). On $\mathcal{Z}$ we have $\det(Y) = \det(X+Y)\det(I_p - Z)$, hence

$$\mathrm{tr}\, X + \mathrm{tr}\, Y - a \log \det Y = \mathrm{tr}(X+Y) - a \log \det(X+Y) - a \log \det(I_p - Z)$$

and the infimum is achieved for $(X+Y) = aI_p$ by (A.4). This completes the proof. □

5.3. Proof of Theorem 2.8

We follow here the proof given in [17] concerning the scalar case. Let $\widetilde{\mathbb{P}}_n$ be the probability measure on the infinite dimensional moment space

$$\mathcal{M}_\infty^{[0,1]} = \left\{ \mathbf{S} = (S_1, S_2, \dots) \mid S_j = \int_0^1 x^j d\mu(x),\ \mu \in \mathcal{P}([0,1]) \right\}$$

induced by the bijection $\mathbf{S} \mapsto \mu_{\mathbf{S}}$. Now if $\prod_k^\infty$ denotes the canonical projection $\mathcal{M}_\infty^{[0,1]} \to \mathcal{M}_k^{[0,1]}$, then the measure $\tilde{\mathbb{P}}_n \circ \left(\prod_k^\infty\right)^{-1}$ is the law of $\mathbf{S}_k^{(n)}$. Therefore, Corollary 2.7 yields an LDP for the sequence $\left(\tilde{\mathbb{P}}_n \circ \left(\prod_k^\infty\right)^{-1}\right)_n$ with speed n and good rate function

$$\widetilde{\mathcal{I}}_k(\mathbf{S}_k) = -p \log \det(S_{k+1}^+ - S_{k+1}^-) - 2kp^2 \log 2.$$

By Dawson–Gärtner's Theorem (see [7]) the sequence $\widetilde{\mathbb{P}}_n$ satisfies an LDP with good rate function

$$\widetilde{\mathcal{I}}(\mathbf{S}) = \sup_{k\in\mathbb{N}} \widetilde{\mathcal{I}}_k(\mathbf{S}_k).$$

It remains to calculate the right hand side of the last equality, which is given by

$$\sup_{k\in\mathbb{N}} -p \log\left(4^{pk} \det(S_{k+1}^+ - S_{k+1}^-)\right).$$

Let μ denote a matrix measure corresponding to the sequence $\mathbf{S}_k$ and let $\tilde{\mu}$ denote the image measure on $[-1, 1]$ obtained from μ by the affine transformation $x \mapsto 2\left(x - \frac{1}{2}\right)$. Since canonical moments are invariant under affine transformations, i.e., $U_i(\mu) = U_i(\tilde{\mu})$ (see for example [9, Lemma 3.1]), we have

$$\det(S_{k+1}^+(\mu) - S_{k+1}^-(\mu)) = \prod_{i=1}^k \det(U_i(\mu) - U_i^2(\mu)) = \prod_{i=1}^k \det(U_i(\tilde{\mu}) - U_i^2(\tilde{\mu})),$$

where the first identity is again (2.18). Now denote by μ_C the symmetric matrix measure on $\mathbb{T}$ associated with $\tilde{\mu}$, that is

$$\int_{-1}^1 f(x) d\tilde{\mu}(x) = \int_{-\pi}^{\pi} f(\cos(\theta)) d\mu_C(\theta). \tag{5.2}$$

The canonical moments $U_i(\tilde{\mu})$ are related to the canonical moments $A_i(\mu_C)$ by the relation (see [12])

$$U_i(\tilde{\mu}) = \frac{1}{2}(A_i(\mu_C) + I_p).$$

This gives for the range

$$\det(S_{k+1}^+(\mu) - S_{k+1}^-(\mu)) = \prod_{i=1}^k 4^{-p} \det(I_p - A_i(\mu_C)^2).$$

Since $0 \le \det(I_p - A_i(\mu_C)^2) \le 1$, the sequence $\widetilde{\mathcal{I}}_k(\mathbf{S}_k)$ is increasing in k which yields

$$\sup_{k\in\mathbb{N}} -p \log\left(4^{pk} \det(S_{k+1}^+ - S_{k+1}^-)\right) = \lim_{k\to\infty} -p \log\left(\prod_{i=1}^k \det(I_p - A_i(\mu_C)^2)\right).$$

Then the Szegö Theorem for matrix-valued measures (Theorem 2.13.5 in [29]) yields

$$\begin{aligned}\widetilde{\mathcal{I}}(\mathbf{S}(\mu)) &= \lim_{k\to\infty} -p \log\left(\prod_{i=1}^k \det(I_p - A_n(\mu_C)^2)\right)\\ &= -\frac{p}{2\pi} \int_{-\pi}^{\pi} \log\det W(\theta) d\theta,\end{aligned}$$

where $d\mu_C(\theta) = W(\theta)\frac{d\theta}{2\pi} + d\mu_S$ is the Lebesgue decomposition of μ_C. Since μ_C is symmetric, W is an even function

$$\widetilde{\mathcal{I}}(\mathbf{S}(\boldsymbol{\mu})) = -\frac{p}{\pi} \int_0^{\pi} \log\det W(\theta) d\theta$$

which, after projection on $[0, 1]$ yields

$$\widetilde{\mathcal{I}}(\mathbf{S}(\boldsymbol{\mu})) = -\frac{p}{\pi} \int_0^1 \log\det V(x) \frac{dx}{\sqrt{x(1-x)}},$$

where $V(x) = W(\arccos(2x - 1))$ is the Radon–Nikodym derivative of μ with respect to the arcsine matricial measure. The result follows from the contraction principle and the continuity of the mapping $\mathbf{S} \mapsto \mu_{\mathbf{S}}$.

5.4. Proof of Lemma 3.2

First we recall the notion of Fréchet differentiability (see for example [3]).

Let $\mathcal{U}$ be an open subset of a complex Banach space X and Φ a continuous map from $\mathcal{U}$ to a complex Banach space Y. The map Φ is called differentiable at $U \in \mathcal{U}$, if there exists a bounded linear operator L from X to Y such that

$$\lim_{V\to 0} \frac{\|\Phi(U+V)-\Phi(U)-LV\|}{\|V\|} = 0.$$

We denote L by $D\Phi(U)$ and call it differential of Φ at U.

For this notion of differentiability we have the following rules:

- (*chain-rule*) Let Z be a Banach space, $\mathcal{V}$ be an open subset of Y and $\psi : \mathcal{V} \to Z$ be a continuous mapping from $\mathcal{V}$ to Z. If $\Phi(U) \in \mathcal{V}$, if Φ is differentiable at U and if ψ is differentiable at $\Phi(U)$ then $\psi \circ \Phi$ is differentiable at U and
$$D(\psi \circ \Phi)(U) = D\psi(\Phi(U)) \circ D\Phi(U). \tag{5.3}$$
- (*product-rule*) If we have a multiplicative structure on Y and if Φ and ψ are continuous maps from $\mathcal{U}$ to Y, both differentiable at U_0 then the map $\Phi\psi : U \mapsto \Phi(U)\cdot\psi(U)$ is differentiable at U_0 and for every V
$$D(\Phi\psi)(U_0)V = [D\Phi(U_0)V]\cdot\psi(U_0) + \Phi(U_0)\cdot[D\psi(U_0)V]. \tag{5.4}$$

We note that the mapping $M \mapsto M^{1/2}$ is differentiable at I_p. Further, the action of the differential at that point is the multiplication by $\frac{1}{2}$. Lemma 3.2 now follows using the above mentioned rules and the following lemma.

Lemma 5.1. *Let* $(\Gamma_1,\ldots,\Gamma_n) \in \operatorname{Int}\mathcal{M}_n^{\mathbb{T}}$. *For the matrices* L_n *and* R_n *defined in* (3.6) *and* (3.7), *respectively, the following recursions hold*

$$L_n = L_{n-1}^{1/2}\left(I_p - A_nA_n^*\right)L_{n-1}^{1/2} \quad \text{and} \quad R_n = R_{n-1}^{1/2}\left(I_p - A_n^*A_n\right)R_{n-1}^{1/2}. \tag{5.5}$$

Proof. We only show the result for L_n. For R_n, the proof is left for the reader.

Here we use the notation of Dette and Wagener [12]. Let ϕ_n^L and ϕ_n^R be the orthonormal matrix polynomials. Using the Szegö recursion (compare e.g. [29, Section 2.13]) and the fact that $L_n^{-1/2}$ is Hermitian we obtain

$$\begin{aligned} I_p &= \langle z\phi_n^L, z\phi_n^L\rangle_L \\ &= \langle L_n^{-1/2}L_{n+1}^{1/2}\phi_{n+1}^L + A_{n+1}\tilde{\phi}_n^R,\ L_n^{-1/2}L_{n+1}^{1/2}\phi_{n+1}^L + A_{n+1}\tilde{\phi}_n^R\rangle_L \\ &= L_n^{-1/2}L_{n+1}L_n^{-1/2} + A_{n+1}A_{n+1}^*. \end{aligned}$$

Indeed the definition of the inner products directly yields

$$\langle\tilde{\phi}_n^R, \tilde{\phi}_n^R\rangle_L = \langle\phi_n^R, \phi_n^R\rangle_R = I_p.$$

The assertion of the lemma follows. □

In the following we will differentiate mappings from $\mathbb{C}^{np\times p}$ to $\mathbb{C}^{p\times p}$. We have from the definition of canonical moments

$$\Gamma_k = L_{k-1}^{1/2}A_kR_{k-1}^{1/2} + M_{k-1} \quad (1 \le k \le n), \tag{5.6}$$

where the matrices L_{k-1}, R_{k-1} and M_{k-1} are defined in (3.6)–(3.8). The differentiability of $\mathbf{A}_n \mapsto L_{k-1}^{1/2}A_kR_{k-1}^{1/2}$ at $0_p^{(n)} = (0_p,\ldots,0_p) \in \mathbb{C}^{np\times p}$ follows obviously using the product rule. Indeed, first the linear map $\mathbf{A}_n \mapsto A_k$ is obviously differentiable in $0_p^{(n)}$. The action of the differential is the multiplication by the map itself. The differentiability of $\mathbf{A}_n \mapsto L_k$ and $\mathbf{A}_n \mapsto R_k$ can be established using induction on k and Lemma 5.1 together with chain and product rules. Again by induction one obtains $L_k(0_p^{(n)}) = R_k(0_p^{(n)}) = I_p$. Now the product rule yields, for every $V \in \mathbb{C}^p$

$$\begin{aligned} D(L_{k-1}^{1/2}A_kR_{k-1}^{1/2})(0_p^{(n)})V &= \left[DL_{k-1}^{1/2}(0_p^{(n)})V\right]\cdot A_k(0_p^{(n)})\cdot R_{k-1}^{1/2}(0_p^{(n)}) + L_{k-1}^{1/2}(0_p^{(n)})\cdot A_kVR_{k-1}^{1/2}(0_p^{(n)}) \\ &\quad + L_{k-1}^{1/2}(0_p^{(n)})\cdot A_k(0_p^{(n)})\cdot\left[DR_{k-1}^{1/2}(0_p^{(n)})V\right] \\ &= A_kV. \end{aligned}$$

It remains to show that $M_{k-1} = o(\|\mathbf{A}_n\|)$ for $k = 1,\ldots,n$. It is done by induction with respect to k together with an appeal to the continuity of the inversion at $I_{(k-1)p}$. This yields the conclusion of Lemma 3.2.

5.5. Proof of Lemma 3.3

We have by definition of the canonical moments that A_k depends only on $\Gamma_1, \ldots, \Gamma_k$ so that the Jacobian of $\psi^{(n)}$ is the product of the Jacobians of $(\Gamma_1, \ldots, \Gamma_k) \mapsto A_k$ $(k = 1, \ldots, n)$. As

$$A_k = L_{k-1}^{-1/2}(\Gamma_k - M_{k-1})R_{k-1}^{-1/2}$$

and because L_{k-1}, R_{k-1} and M_{k-1} are independent of Γ_k, Theorem 3.2 from [25] gives the following Jacobian J_k for the mapping $\Gamma_k \mapsto A_k$:

$$\begin{aligned} J_k &= \det\left(L_{k-1}^{-1/2}\left(L_{k-1}^{-1/2}\right)^*\right)^p \det\left(R_{k-1}^{-1/2}\left(R_{k-1}^{-1/2}\right)^*\right)^p \\ &= \det(L_{k-1})^{-p}\det(R_{k-1})^{-p}, \end{aligned}$$

where the last equality follows because L_{k-1} and R_{k-1} are Hermitian. From Lemma 5.1 we obtain

$$\det(L_{k-1})^{-p}\det(R_{k-1})^{-p} = \prod_{j=1}^{k-1}\det(I_p - A_j^*A_j)^{-p}\det(I_p - A_j^*A_j)^{-p} = \prod_{j=1}^{k-1}\det(I_p - A_jA_j^*)^{-2p}.$$

Consequently, the Jacobian of $\psi^{(n)}$ is the product

$$\prod_{k=1}^{n}\prod_{j=1}^{k-1}\det\left(I_p - A_j^*A_j\right)^{2p} = \prod_{k=1}^{n-1}\det\left(I_p - A_k^*A_k\right)^{2p(n-k)}.$$

This yields exactly the assertion of the lemma.

5.6. Proof of Proposition 3.5

The proof of this proposition uses the following lemma.

Lemma 5.2. *Let A be a $p \times p$ matrix of full rank and $A = UH^{1/2}$ its polar decomposition with $H = A^*A \in \mathcal{S}_p(\mathbb{C})$ and $U = A(A^*A)^{-1/2} \in \mathbb{U}(p)$. If A is random and if*

$$\forall V \in \mathbb{U}(p) \quad A \overset{(d)}{=} VA \tag{5.7}$$

then U and H are independent, and U is Haar distributed.

Proof of Lemma 5.2. We have for all bounded measurable functions f_1, f_2

$$\begin{aligned} \mathbb{E}\left(f_1(U)f_2(H)\right) &= \mathbb{E}f_1\left(A(A^*A)^{-1/2}\right)f_2\left((A^*A)\right) \\ &= \mathbb{E}f_1\left(VA(A^*A)^{-1/2}\right)f_2\left((A^*A)\right) && (5.8) \\ &= \int_{\mathbb{U}(p)}\left[\mathbb{E}f_1\left(VA(A^*A)^{-1/2}\right)f_2\left((A^*A)\right)\right]d_{Haar}(V) && (5.9) \\ &= \mathbb{E}\left(\left[\int_{\mathbb{U}(p)} f_1\left(VA(A^*A)^{-1/2}\right)d_{Haar}(V)\right]f_2\left((A^*A)\right)\right) && (5.10) \\ &= \mathbb{E}\left(\left[\int_{\mathbb{U}(p)} f_1(V)d_{Haar}(V)\right]f_2\left((A^*A)\right)\right) && (5.11) \\ &= \left[\int_{\mathbb{U}(p)} f_1(V)d_{Haar}(V)\right]\mathbb{E}\left(f_2\left((A^*A)\right)\right), && (5.12) \end{aligned}$$

where in (5.8) we take into account the invariance by left multiplication, in (5.9) the fact that V is arbitrary in $\mathbb{U}(p)$, in (5.10) Fubini's theorem, and in (5.11) the invariance of Haar by right multiplication. $\square$

Proof of Proposition 3.5. The assumption (5.7) is trivially verified since VA and A have the same singular values. It remains to determine the distribution of $H = M^*M$. By a simple application of Proposition 4.1.3 of [1], we see that the singular values of M have on $(0, \infty)^p$ a joint density proportional to

$$|\Delta(x_1^2, \ldots, x_p^2)|^2 f(x_1^2, \ldots, x_p^2)(x_1 \cdots x_p),$$

where Δ is the Vandermonde function. This implies directly that the eigenvalues of H have on $(0, \infty)^p$ a joint density proportional to

$$|\Delta(\lambda_1, \ldots, \lambda_p)|^2 f(\lambda_1, \ldots, \lambda_p).$$

Now it is easy to lift to the matrix H by Proposition 4.1.1 of [1]. $\square$

Proof of Theorem 3.4. If A_k has density $f(A_k)$ it fulfils the assumptions of Proposition 3.5, with

$$f(\lambda_1, \dots, \lambda_p) = \frac{1}{c_k^{(n)}} \prod_{j=1}^{p} (1-\lambda_j)^{2p(n-k)}$$

and the density of B_k is proportional to

$$\det(I_p - B_k)^{2p(n-k)}.$$

This expression fits with (2.10) with $a = p$ and $b = 2p(n-k) + p$. □

5.7. Proof of Theorem 3.6

One proof of Theorem 3.6 directly follows from two applications of Theorem 3.4 together with Lemma 3.3, Theorem 2.2 and the continuous mapping theorem. We give a second proof here.

5.7.1. Alternative proof: Gaussian approximation

We use two clever results. The first one will give a representation of the law of A_k.

Theorem 5.3 (*Collins [5, Theorem 5.1] or Forrester and Krishnapur [16]*). *The top $p \times p$ sub-block of a Haar distributed matrix from $\mathbb{U}(p+q)$, where $q \geq p$, has a density in $\mathbb{D}_p$ proportional to*

$$A \mapsto \det\left(I_p - AA^*\right)^{q-p}.$$

The second one is the following "Borel theorem".

Theorem 5.4 (*Jiang [22, Corollary 1]*). *There exists two $N \times N$ random matrices $\prod_N = (\pi_{i,j})_{1\leq i,j\leq N}$ and $Y_N = (y_{i,j})_{1\leq i,j\leq N}$ defined on the same probability space such that*

(i) *$\prod_N$ is Haar distributed in $\mathbb{U}(N)$*
(ii) *all the $y_{i,j}$, $1 \leq i, j \leq N$ are independent and standard complex Gaussian distributed.*
(iii) *For $m_N = [N/(\log N)^2]$*

$$\max_{i\leq N, j\leq m_N} \left| \sqrt{N}\pi_{i,j} - y_{i,j} \right| \to 0$$

in probability as $N \to \infty$.

From the above notation and Lemma 3.3, A_k is distributed as the top $p \times p$ sub-block of $\prod_N$ with $N = 2p(n-k+1)$. Up to a change of probability space we have then for $i, j \leq p$

$$\sqrt{2p(n-k+1)}(A_k)_{i,j} - y_{i,j} \to 0$$

in probability as $n \to \infty$, which leads easily to the conclusion since k is fixed.

5.8. Proof of Corollary 3.9

By the contraction principle and Corollary 3.8, $(\mathbf{X}_n^k)_n$ satisfies an LDP with good rate function

$$\widetilde{\mathcal{I}}_\Gamma(\Gamma_1, \dots, \Gamma_k) = \begin{cases} -2p \displaystyle\sum_{i=1}^{k} \log\det(I_p - A_i^* A_i), & \text{if } (\Gamma_1, \dots, \Gamma_k) \in \operatorname{Int} \mathcal{M}_k^{\mathbb{T}}, \\ \infty & \text{otherwise,} \end{cases}$$

where $(A_1, \dots, A_k) = \psi^{(k)}(\Gamma_1, \dots, \Gamma_k)$. An application of the formula for determinants of block matrices (see for example [21]) yields

$$\det(T_k) = \det(T_{k-1}) \det(R_k) = \det(T_{k-1}) \det(L_k),$$

because L_k and R_k are Schur complements in T_k. From Lemma 5.1 we obtain

$$\det(R_k) = \prod_{i=1}^{k} \det(I_p - A_i^* A_i)$$

and so

$$\sum_{i=1}^{k} \log\det(I_p - A_i^* A_i) = \log \frac{\det(T_k)}{\det(T_{k-1})},$$

which is the assertion of Corollary 3.9.

5.9. Proof of Theorem 4.1

For ($\mathbb{P}_n^c$) (Carathéodory problem), the assertion is a consequence of Theorem 3.10, the contraction principle and (4.2). Recall the main point: under $\mathcal{U}(\mathcal{M}_n^{\mathbb{T}})$, the variables $A_1, \dots, A_n$ are independent, and A_k has a density proportional to $\det\left(I_p - A_j^* A_j\right)^{2(n-j)p}$.

For ($\mathbb{P}_n^s$) (Schur problem), we first remark from (4.3) that the mapping $(G_0(f), \dots, G_{n-1}(f)) \mapsto (A_1, \dots, A_n)$ is triangular, i.e. that $G_k(f)$ depends only on $A_1, \dots, A_{k+1}$. Let us give details. In the scalar case, it is 1.3.48 in [29] and we follow the same scheme, up to change due to noncommutativity. Relation (4.4) for $k = 0$ implies

$$f(z)(B_1^L)^{-1}[I_p + zA_1 f_1(z)] = (B_1^R)^{-1}[zf_1(z) + A_1^*].$$

Identifying the powers of z^n on both sides yields

$$G_0(f) = (B_1^R)^{-1} A_1^* B_1^L$$

$$G_n(f) = (B_1^R)^{-1} G_{n-1}(f_1) B_1^L - G_0(f)(B_1^L)^{-1} A_1 G_{n-1}(f_1) - \sum_{j=1}^{n-1} G_j(f)(B_1^L)^{-1} A_1 G_{n-1-j}(f_1).$$

Lemma 1.3 in [6] (see also formula (2.13.52) in [29]) says that

$$A_j^* B_j^L = B_j^R A_j^*$$

for every $j \geq 1$ so that we get $G_0(f) = A_1$ and identifying the powers of z^n on both sides yields:

$$G_0(f) = A_1^*$$

$$G_n(f) = (B_1^R)^{-1} G_{n-1}(f_1) B_1^L - \sum_{j=0}^{n-1} G_j(f)(B_1^L)^{-1} A_1 G_{n-1-j}(f_1) \quad (n \geq 1). \tag{5.13}$$

Induction on n leads to

$$G_n(f) = V_n A_{n+1}^* W_n + \text{polynomial in } (A_1, A_1^*, \dots, A_n, A_n^*), \tag{5.14}$$

where

$$V_n = B_1^R B_2^R \cdots B_n^R, \qquad W_n = B_n^L B_{n-1}^L \cdots B_1^L.$$

From this relation, we see that, if we froze $A_1, \dots, A_n$ the Jacobian of the mapping $G_n(f) \mapsto A_{n+1}$ is (Theorem 3.2 of [25])

$$|\det(V_n V_n^*)|^p |\det(W_n W_n^*)|^p = \prod_{k=1}^{n} [\det(I_p - A_k^* A_k)]^{2p}.$$

Like in the proof of Lemma 3.3, it turns out that the Jacobian of the mapping

$$(G_0(f), \dots, G_{n-1}(f)) \mapsto (A_1, \dots, A_n)$$

is then

$$\prod_{k=1}^{n-1} \det(I_p - A_k^* A_k)^{2(n-k)}.$$

We conclude that the distribution of $(A_1, \dots, A_n)$ under $\mathbb{P}_n^s$ is the same as the distribution of $(A_1, \dots, A_n)$ under $\mathbb{P}_n^c$. Applying again the contraction principle, we see that ($\mathbb{P}_n^s$) satisfies an LDP with good rate function

$$I_p^s(f) = -\frac{p}{\pi} \int_{\mathbb{T}} \log\det W(\theta) d\theta,$$

where W is related to μ the underlying matrix measure. To have a rate function depending explicitly on f, we go back to the correspondence (4.2) between W and f so that

$$\log\det W(\theta) = \log\det(I_p - f(e^{i\theta})^* f(e^{i\theta})) - 2\log|\det(I_p - e^{i\theta} f(e^{i\theta}))|$$

and apply Jensen's formula to the function $\det(I_p - zf(z))$. This yields (4.8). □

Acknowledgements

The authors would like to thank two anonymous referees for their constructive comments on an earlier version of this paper.

The work of the authors was supported by the Deutsche Forschungsgemeinschaft: (Sonderforschungsbereich Tr/12; project C2, Fluctuations and universality of invariant random matrix ensembles). A.R.'s work was partly supported by the ANR project Grandes Matrices Aléatoires ANR-08-BLAN-0311-01.

Appendix. Some properties of the Wishart distribution

For $a > 0$, the Laplace transform of the complex Wishart distribution $W_p(a)$ is given for $K \in \mathcal{S}_p$ by

$$\Lambda(K) = \log \mathrm{E}\left[e^{\operatorname{tr}(KW)}\right] = -a \log \det(I_p - K) \tag{A.1}$$

if $K < I_p$ and infinite otherwise. From the divisibility of the family of Wishart distributions (indexed by a), we deduce the following easy results (law of large numbers and CLT).

Proposition A.1. *As $a_n \to \infty$ we have for $W_n \sim W_p(a_n)$*

(i) $\lim_{n\to\infty} \frac{1}{a_n} W_n = I_p$ *(in probability)*,

(ii) $(a_n)^{-1/2}\left(W_n - a_n I_p\right) \xrightarrow{\mathcal{D}} \mathrm{GUE}_p$.

Since the following large deviation result is not so obvious, we give a proof.

Proposition A.2. *For fixed p and $a > 0$, if the variables X_k, $k \geq 1$ are independent and $W_p(a)$ distributed, then $\frac{1}{n}(X_1 + \cdots + X_n)$ satisfies an LDP in $\mathcal{S}_p^+(\mathbb{C})$ with good rate function*

$$\Lambda^\star(X) = \begin{cases} \operatorname{tr} X - a \log \det X - ap(1 - \log a) & \text{if } \det X > 0, \\ \infty & \text{otherwise}. \end{cases} \tag{A.2}$$

Proof. The multidimensional Cramér theorem gives an LDP with good rate function

$$\Lambda^\star(X) = \sup_{K \in \mathcal{S}_p(\mathbb{C})} \operatorname{tr}(KX) - \Lambda(K). \tag{A.3}$$

We first give a nonvariational expression of $\Lambda^\star(X)$.

If $\det X = 0$, for every n we choose $K_n \in \mathcal{S}_p(\mathbb{C})$ such that $K_n x = 0$ for x in the range of X and such that the restriction of K_n to the kernel of X is $-nI_d$, where $d \geq 1$ is the dimension of this kernel. We have $\operatorname{tr}(K_n X) - \Lambda(K_n) = ad \log(n + 1)$ and the supremum in (A.3) is infinite.

If $\det X \neq 0$, make the variable change $K = I_p - aX^{-1}L$ and observe that

$$\log \det L \leq \operatorname{tr}(L - I_p) \tag{A.4}$$

with equality only at $L = I_p$. □

At last, we have another LDP for rescaled Wishart distributions. Its proof is left to the reader and uses directly the density (2.13).

Proposition A.3. *Let p and a be fixed. If X is $W_p(a)$ distributed then X/n satisfies an LDP in $\mathcal{S}_p^+(\mathbb{C})$ with good rate function*

$$\mathcal{I}_s(X) = \operatorname{tr} X. \tag{A.5}$$

References

[1] G. Anderson, A. Guionnet, O. Zeitouni, An Introduction to Random Matrices, Cambridge University Press, Cambridge, 2010.
[2] C. Berg, The matrix moment problem, in: A. Moren, A. Branquinho (Eds.), Coimbra Lecture Notes on Orthogonal Polynomials, Nova Science Pub Inc., 2008, pp. 1–56.
[3] H. Cartan, Calcul Différentiel, Hermann, Paris, 1967.
[4] F. Chang, J. Kempermann, W. Studden, A normal limit theorem for moment sequences, Ann. Probab. 21 (3) (1993) 1295–1309.
[5] B. Collins, Product of random projections, Jacobi ensembles and universality problems arising from free probability, Probab. Theory Related Fields 133 (3) (2005) 315–344.
[6] D. Damanik, A. Pushnitski, B. Simon, The analytic theory of matrix orthogonal polynomials, Surv. Approx. Theory 4 (2008) 1–85.
[7] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer, 1998.
[8] H. Dette, F. Gamboa, Asymptotic properties of the algebraic moment range process, Acta Math. Hungar. 116 (3) (2007) 247–264.
[9] H. Dette, J. Nagel, Matrix measures, random moments and Gaussian ensembles, J. Theoret. Probab. (2010) doi:10.1007/s10959-011-0370-7.
[10] H. Dette, W.J. Studden, Matrix measures, moment spaces and Favard's theorem for the interval [0, 1] and [0, ∞), Linear Algebra Appl. 345 (2002) 169–193.
[11] H. Dette, W. Studden, The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis, in: Wiley Series in Probability and Statistics, 1997.
[12] H. Dette, J. Wagener, Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the Geronimus relations, Linear Algebra Appl. 432 (2010) 1609–1626.
[13] V. Dubovoj, B. Fritzsche, B. Kirstein, Matricial Version of the Classical Schur Problem, BG Teubner Gmbh, 1992.
[14] J. Fischmann, W. Bruzda, B.A. Khoruzhenko, H.-J. Sommers, K. Zyczkowski, Induced ginibre ensemble of random matrices and quantum operations, 2011. arXiv.org, arXiv:1107.5019v1 [math-ph].
[15] P. Forrester, Log-Gases and Random Matrices, Princeton University Press, 2010.
[16] P. Forrester, M. Krishnapur, Derivation of an eigenvalue probability density function relating to the Poincaré disk, J. Phys. A 42 (2009) 385204.
[17] F. Gamboa, L.-V. Lozada-Chang, Large deviations for random power moment problem, Ann. Probab. 32 (3B) (2004) 2819–2837.

[18] F. Gamboa, A. Rouault, Canonical moments and random spectral measures, J. Theoret. Probab. (2010) doi:10.1007/s10959-009-0239-1.
[19] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, J. Math. Phys. 6 (1965) 440–449.
[20] F. Hiai, D. Petz, Large deviations for functions of two random projections, Acta Sci. Math. (Szeged) 72 (2006) 581–609.
[21] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 1985.
[22] T. Jiang, Maxima of entries of Haar distributed matrices, Probab. Theory Related Fields 131 (1) (2005) 121–144.
[23] C.G. Khatri, Classical statistical analysis based on a certain multivariate complex Gaussian distribution, Ann. Math. Statist. 36 (1965) 98–114.
[24] L. Lozada-Chang, Large deviations on moment spaces, Electron. J. Probab. 10 (2005) 662–690.
[25] A. Mathai, Jacobians of Matrix Transformations and Functions of Matrix Argument, World Scientific Publ., 1997.
[26] M. Mehta, Random Matrices, Pure and Applied Mathematics, Elsevier/Acacemic Press, Amsterdam, 2004.
[27] K.C.S. Pillai, G.M. Jouris, Some distribution problems in the multivariate complex Gaussian case, Ann. Math. Statist. 42 (1971) 517–525.
[28] J. Robertson, M. Rosenberg, The decomposition of matrix-valued measures, Michigan Math. J. 15 (1968) 353–368.
[29] B. Simon, Orthogonal Polynomials on the Unit Circle. Part 1: Classical Theory, Colloquium Publications, 2005, American Mathematical Society 54, Part 1. Providence, RI: American Mathematical Society (AMS).
[30] M. Skibinsky, Some striking properties of binomial and beta moments, Ann. Math. Statist. 40 (1969) 1753–1764.
[31] A.W. van der Vaart, Asymptotic Statistics, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, 1998.