# Sum rules and large deviations for spectral matrix measures

Fabrice Gamboa, Jan Nagel, Alain Rouault

# Sum rules and large deviations for spectral matrix measures

Fabrice Gamboa* Jan Nagel† Alain Rouault‡

September 20, 2016

**Abstract**

A sum rule relative to a reference measure on $\mathbb{R}$ is a relationship between the reversed Kullback-Leibler divergence of a positive measure on $\mathbb{R}$ and some non-linear functional built on spectral elements related to this measure (see for example Killip and Simon 2003). In this paper, using only probabilistic tools of large deviations, we extend the sum rules obtained in Gamboa, Nagel and Rouault (2015) to the case of Hermitian matrix-valued measures. We recover the earlier result of Damanik, Killip and Simon (2010) when the reference measure is the (matrix-valued) semicircle law and obtain a new sum rule when the reference measure is the (matrix-valued) Marchenko-Pastur law.

**Keywords :** Sum rules, block Jacobi matrix, matrix orthogonal polynomials, matrix-valued spectral measures, large deviations, random matrices.

**MSC 2010:** 60F10, 42C05, 47B36, 15B52

## 1 Introduction

For a probability measure $\mu$ on the unit circle $\mathbb{T}$ with Verblunsky coefficients $(\alpha_k)_{k\geq 0}$, Verblunsky's form of Szegő's theorem may be written as

$$\mathcal{K}\left(\tfrac{d\theta}{2\pi}|\mu\right) = -\sum_{k\geq 0}\log(1-|\alpha_k|^2)\,, \tag{1.1}$$

*Université Paul Sabatier, Institut de Mathématiques de Toulouse, 31062 Toulouse Cedex 9, France, gamboa@math.univ-toulouse.fr

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

‡Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035-Versailles Cedex France, e-mail: alain.rouault@uvsq.fr

where $\mathcal{K}$ is the relative entropy, defined by

$$\mathcal{K}(\nu \mid \mu) = \begin{cases} \displaystyle\int_{\mathbb{T}} \log \frac{d\nu}{d\mu}\, d\nu & \text{if } \nu \text{ is absolutely continuous with respect to } \mu, \\ \infty & \text{otherwise.} \end{cases} \tag{1.2}$$

Identity (1.1) is a sum rule[1] connecting an entropy and a functional of the recursion coefficients, and remains one of the most important result of the theory of orthogonal polynomials on the unit circle (OPUC) (see [Sim05a] for extensive history and bibliography).
The corresponding result in the theory of orthogonal polynomials on the real line (OPRL) is the Killip-Simon sum-rule ([KS03]). The reference measure is the semicircle distribution

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

The right hand side involves the Jacobi coefficients and in the left hand side appears an extra term corresponding to a contribution of isolated masses of $\mu$ outside $[-2,2]$ (bounds states).
In a previous paper ([GNR16b]), we gave an interpretation and a new proof of this result from a probabilistic point of view. This approach allowed us to prove new sum rules, when the reference measure is the Marchenko-Pastur distribution

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

where $\tau \in (0,1]$, $\tau^\pm = (1 \pm \sqrt{\tau})^2$, and also the Kesten-McKay distribution

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

where $C_{u^-,u^+}$ is the normalizing constant.

Besides, known extensions of Szegő's theorem ([DGK78]) and of the Killip-Simon sum rule ([DKS10]) are available in the context of matrix-valued measures and the Matrix Orthogonal Pololynomials on the Unit Circle (MOPUC) or Matrix Orthogonal polynomials on the Real Line (MOPRL).
It seems natural to see if the probabilistic methods are robust enough to encompass this matricial framework. Actually, the answer is positive. For the MOPUC context see [GR14] and [GNR16a]. The aim of the present article is precisely to treat the MOPRL case. This allows to give interpretation and new proof of the Damanik-Killip-Simon's sum rule and also to state a matrix version of the sum rule relative to the Marchenko-Pastur distribution.

[1] see the preface of [Sim05a]

Let us explain the main features of our method[2]. As mentioned by several authors ([Kil07], [Sim11]), the main characteristic of the above sum rules both sides are nonegative (possibly infinite) functionals. We will identify them as rate function of large deviations for random measures. Roughly speaking, that means that a sequence of random measures converges to a deterministic limit exponentially fast and the probabilities of deviating from the limit is measured by the rate function. We give two different encodings of the randomization with two rate functions $I_A$ and $I_B$, for which the uniqueness of rate functions yields the equality $I_A = I_B$.
To be more specific, let us give some notation. For $p \geq 1$ (fixed in all the sequel), let us denote by $\mathbb{M}_p$ the set of all $p \times p$ matrices with complex entries and by $\mathcal{H}_p \subset \mathbb{M}_p$ the subset of Hermitian matrices.

A matrix measure $\Sigma = (\Sigma_{i,j})_{i,j}$ of size $p$ on $\mathbb{R}$ is a matrix of signed complex measures, such that $\Sigma(A) = (\Sigma_{i,j}(A))_{i,j} \in \mathcal{H}_p$ for any Borel set $A \subset \mathbb{R}$. Further, if for any $A$, $\Sigma(A)$ is a nonnegative matrix we say that $\Sigma$ is nonnegative. We denote by $\mathcal{M}_p(T)$ the set of $p \times p$ nonnegative matrix measures with support in $T \subset \mathbb{R}$. Further, $\mathcal{M}_{p,1}(T)$ is the subset of $\mathcal{M}_p(T)$ of normalized measures $\Sigma$ satisfying $\Sigma(T) = \mathbf{1}$, where $\mathbf{1}$ is the $p \times p$ identity matrix..
A natural example of matrix measures comes from an application of the spectral theorem. More precisely, a Hermitian matrix $X$ of size $N \times N$ may be written as $UDU^*$ where $D = \mathrm{diag}(\lambda_i)$ contains the eigenvalues of $X$ and $U$ is the matrix formed by an orthonormal basis of eigenvectors. Assume that the system $(e_1, \ldots, e_p)$ of the $p$ first vectors of the canonical basis of $\mathbb{C}^N$ is cyclic, i.e. that $\mathrm{Span}\{A^k e_i, k \geq 0, i \leq N\} = \mathbb{C}^N$. Then, for $p \leq N$, there exists a unique spectral matrix measure $\Sigma^X \in \mathcal{M}_p(\mathbb{R})$ supported by the spectrum of $X$, such that for all $k > 0$ and $1 \leq i, j \leq p$

$$(X^k)_{i,j} = \int_{\mathbb{R}} x^k d\Sigma^X_{i,j}(x), \tag{1.3}$$

Let us assume that all the eigenvalues of $X$ have a single multiplicity. For $j = 1, \cdots, N$, let $u^j := (U_{i,j})_{i=1,\cdots,p}$ be the $j^{th}$ truncated column of $U$. Then, obviously

$$\Sigma^X_p(dx) = \sum_{j=1}^{N} u_j u_j^* \delta_{\lambda_j}(dx), \tag{1.4}$$

where $\delta_a$ is the Dirac measure in $a$. That is, $\{\lambda_1, \ldots, \lambda_N\}$ is the support of $\Sigma^X$ and $u_1 u_1^*, \ldots u_N u_N^*$ are the weights. Furthermore, as $U$ is unitary we have $\sum_j u_j u_j^* = \mathbf{1}$ so that $\Sigma^X \in \mathcal{M}_{p,1}(T)$.
If we assume further that $N = pn$ for some positive integer $n$ then it is possible to build a block

[2]Note that recently Breuer et al. [BSZ16] posted on arxiv a paper exposing the method to a non probabilistic audience and giving some developments.

tridiagonal matrix

$$J_n = \begin{pmatrix} B_1 & A_1 & & \\ A_1^* & B_2 & \ddots & \\ & \ddots & \ddots & A_{n-1} \\ & & A_{n-1}^* & B_n \end{pmatrix} \tag{1.5}$$

such that $\Sigma^X = \Sigma^{J_n}$. Here, all the blocks of $J_n$ are elements of $\mathbb{M}_p$. The case $p = 1$ is the most classical and relies on the construction of the OPRL in $L^2(\Sigma^X)$ (see for example [Sim05a]). The general case is more complicated and requires technical tools from the theory of MOPRL (see [Sim05b]). In Section 2.1, we will recall the construction of such tridiagonal representations.
As a result, we have two encodings of $\Sigma^X$: (1.4) and (1.5).

Our measures are random in the sense that we first draw a random matrix $X$ in $\mathcal{H}_N$ (with $N \geq p$) and then considering its spectral measure $\Sigma^X$. In a large class of random matrix models, invariance by unitary transformations is postulated, which means that $X$ is sampled in $\mathcal{H}_N$ from the distribution

$$Z_N^{-1} e^{-\mathrm{tr} N V(X)} dX, \tag{1.6}$$

for $V$ a confining potential and $dX$ is the Lebesgue measure on $\mathcal{H}_N$. Its eigenvalues behave as a Coulomb gas (see formula (3.3)) and the matrix of eigenvectors follows independently the Haar distribution in the set of $N \times N$ unitary matrices. The most popular models are the Gaussian Ensemble, corresponding to the potential $V(x) = x^2/2$, the Laguerre Ensemble, with $V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log x$ $(x > 0)$ and the Jacobi Ensemble with $V(x) = -\kappa_1 \log x - \kappa_2 \log(1 - x)$ $(x \in (0, 1))$.

We will prove in this paper that, under convenient assumptions on $V$, as $N$ goes to infinity, the random spectral matrix measure $\Sigma^X$ converges to some equilibrium matrix measure (depending on $V$) at exponential rate. Indeed, we show that this random object satisfies a large deviation principle (LDP). To be self contained, we recall in Section 3.2 the definition and useful facts on LDP. Furthermore, the rate function of this LDP involves a matrix extension of the reversed Kullback-Leibler information with respect to the equilibrium matrix measure (see equation (2.12) and Theorem 5.2) and a contribution of the outlying eigenvalues.

Looking for the right hand side of a possible sum rule is equivalent to look for a LDP for the encoding by means of the sequence of blocks $A_k$ and $B_k$. In the scalar case ($p = 1$), it is well known that the above classical ensembles have very nice properties [DE02]. For the Gaussian Ensemble

the coefficients appearing in the tridiagonal matrix are independent with simple distributions. The diagonal terms have Gaussian distribution while the subdiagonal ones have so-called $\chi$ distributions [DE02]. We will give in Lemmas 6.1 and 6.3 results in the same spirit in the general block case both for the Gaussian and Laguerre ensembles. These properties allow to compute the rate function of the LDP by the way of the blocks involved in the Jacobi representation. Further, the uniqueness of a rate function leads to our two main Theorems 2.3 and 2.5 that are sum rules. Theorem 2.3 has been proved in [DKS10] by strong analysis tools. We recover this result by using only probabilistic arguments. For $p = 1$, it has been proved earlier in [KS03]. Up to our knowledge, our Theorem 2.5 is new and is the matrix extension of the one we have obtained in [GNR16b] for $p = 1$.

We stress that one of the main differences with purely functional analysis methods is that, for us, the SC distribution and the free semi-infinite matrix (corresponding to $A_k \equiv \mathbf{1}$ and $B_k \equiv \mathbf{0}$, with $\mathbf{0}$ the $p \times p$ matrix of zeros) do not have a central role. Additionally, the non-negativity of both sides of the sum rule is automatic.

For the Jacobi ensemble, the method used in the scalar case, based on the Szegő mapping, is not directly extendible. So, getting a sum rule needs a more careful study. To avoid long developments here, we keep this point for a forthcoming paper.

Our paper is organized as follows. In Section 2, we first give definitions and tools to handle MOPRL and spectral measures. Then, we state our main results concerning sum rules for spectral matrix measures. In Section 3, we introduce random models and state Large Deviations Principles, first for random spectral measures drawn by using a general potential and then for block Jacobi coefficients in the Hermite and Laguerre cases. Section 4 is devoted to proofs of both sum rules, up to the LDP's. The most technical proofs of LDP's are postponed to Sections 5 and 6.

# 2 MOPRL and block Jacobi matrices

## 2.1 Construction

We will need to work with polynomials with coefficients in $\mathbb{M}_p$. They will be *orthogonal* with respect to some matrix measure on $\mathbb{R}$. Let us give some notation and recall some useful facts (see [DPS08] for details). Let $\Sigma \in \mathcal{M}_p(\mathbb{R})$ be a compactly supported matrix measure. Further, let $F$ and $G$ be continuous matrix valued functions $F, G : \mathbb{R} \to \mathbb{M}_p$. We define the two pseudo-scalar

products $p \times p$ (which are elements of $\mathbb{M}_p$) by setting

$$\begin{aligned} \langle\langle F, G\rangle\rangle_R &= \int_{\mathbb{R}} F(z)^* d\Sigma(z) G(z)\,, \\ \langle\langle F, G\rangle\rangle_L &= \int_{\mathbb{R}} G(z) d\Sigma(z) F(z)^*\,. \end{aligned}$$

A sequence of functions on $\mathbb{R}$, $(\varphi_j)$ with values in $\mathbb{M}_p$ is called right-orthonormal if

$$\langle\langle \varphi_i, \varphi_j\rangle\rangle_R = \delta_{ij}\mathbf{1}\ .$$

The orthogonal polynomial recursion is built as follows. First, assume that $\Sigma$ is nontrivial, that is,

$$\mathrm{tr}\langle\langle P, P\rangle\rangle_R > 0 \tag{2.1}$$

for any non zero matrix polynomial $P$ (see Lemma 2.1 of [DPS08] for equivalent characterizations of nontriviality). Applying the Gram-Schmidt procedure to $\{\mathbf{1}, x\mathbf{1}, x^2\mathbf{1}, \dots\}$, we obtain a sequence $(P_n^R)_n$ of right monic matrix orthogonal polynomials. In other words, $P_n^R$ is the unique matrix polynomial $P_n^R(x) = x^n\mathbf{1}+$ lower order terms such that $\langle\langle x^k\mathbf{1}, P_n^R\rangle\rangle_R = \mathbf{0}$ for $k = 0, \dots, n-1$. For nontrivial matrix measures, this is possible for any $n \geq 0$ and this sequence satisfies the recurrence relation

$$xP_n^R(x) = P_{n+1}^R(x) + P_n^R(x)u_n^R + P_{n-1}^R v_n^R\,. \tag{2.2}$$

If we set

$$\gamma_n := \langle\langle P_n^R, P_n^R\rangle\rangle_R,$$

then $\gamma_n$ is positive definite and we have

$$v_n^R = \gamma_{n-1}^{-1}\gamma_n\,.$$

To get normalized orthogonal polynomials $p_n^R$ we set

$$p_0^R = \mathbf{1}\quad,\quad p_n^R = P_n^R\kappa_n^R \tag{2.3}$$

where for every $n$, $\kappa_n^R \in \mathbb{M}_p$ has to satisfy

$$\gamma_n = \left(\kappa_n^R\left(\kappa_n^R\right)^*\right)^{-1}\ . \tag{2.4}$$

This constraint opens several choices for $\kappa_n^R$ (see Section 2.1.5 of [DPS08]). Let us leave the choice open, setting

$$\kappa_n^R = \gamma_n^{-1/2}\sigma_n \tag{2.5}$$

with $\sigma_n$ unitary and $\sigma_0 = \mathbf{1}$.

**Remark 2.1** *We can define similarly the sequence of monic polynomials $P_n^L$ and the sequence of left-orthonormal polynomials $p_n^L$ in the same way. We have*

$$P_n^R = (P_n^L)^*$$

*and*

$$\langle\langle P_n^L, P_n^L\rangle\rangle_L = \gamma_n$$

*and the recurrence relation:*

$$xP_n^L(x) = P_{n+1}^L(x) + u_n^L P_n^L(x) + v_n^L P_{n-1}^L \tag{2.6}$$

*with*

$$v_n^L = \gamma_n\gamma_{n-1}^{-1}\,.$$

*The above condition (2.5) is replaced by $p_n^L = \tau_n\gamma_n^{-1/2}P_n^L$.*

To formulate the recursion in terms of orthonormal polynomials, we use (2.3) and get

$$xp_n^R = p_{n+1}^R(\kappa_{n+1}^R)^{-1}\kappa_n^R + p_n^R(\kappa_n^R)^{-1}u_n^R\kappa_n^R + p_{n-1}^R(\kappa_{n-1}^R)^{-1}v_n^R\kappa_n^R \tag{2.7}$$

i.e.

$$xp_n^R = p_{n+1}^R A_{n+1}^* + p_n^R B_{n+1} + p_{n-1}^R A_n \tag{2.8}$$

with

$$\begin{aligned} A_n &= (\kappa_{n-1}^R)^{-1}v_n^R\kappa_n^R = \sigma_{n-1}^*\gamma_{n-1}^{1/2}v_n^R\gamma_n^{-1/2}\sigma_n = \sigma_{n-1}^*\gamma_{n-1}^{-1/2}\gamma_n^{1/2}\sigma_n \\ B_{n+1} &= (\kappa_n^R)^{-1}u_n^R\kappa_n^R = \sigma_n^*\gamma_n^{1/2}u_n^R\gamma_n^{-1/2}\sigma_n\,. \end{aligned} \tag{2.9}$$

Note that (2.9) yields

$$A_nA_n^* = \sigma_{n-1}^*\gamma_{n-1}^{-1/2}\gamma_n\gamma_{n-1}^{-1/2}\sigma_{n-1}\,. \tag{2.10}$$

In other terms the map $f \mapsto (x \mapsto xf(x))$ defined on the space of matrix polynomials is a right homomorphism and is represented in the (right-module) basis $\{p_0^R, p_1^R, \dots\}$ by the matrix

$$J = \begin{pmatrix} B_1 & A_1 & \\ A_1^* & B_2 & \ddots \\ & \ddots & \ddots \end{pmatrix} \tag{2.11}$$

with $B_k$ Hermitian and $A_k$ non-singular. Moreover the measure $\Sigma$ is again the spectral measure of the matrix $J$ defined as in (1.3) (Theorem 2.11 of [DPS08]). Let us remark that although to each $\Sigma$ corresponds a whole equivalence class of Jacobi coefficients given by the different $\sigma_n$, there is exactly one representative such that all $A_k$ are Hermitian positive definite (Theorem 2.8 in [DPS08]).

Starting with a finite dimensional Jacobi matrix $J_n$ as in (1.5), the spectral matrix measure of $J_n$ is supported by at most $n$ points and is in particular not nontrivial. However, we may still define $p_1^R, \dots, p_{n-1}^R$ by the recursion (2.8). As long as the $A_k$'s are invertible, these polynomials are orthonormal with respect to the spectral measure of $J_n$ and (2.1) holds for all polynomials up to degree $n-1$.

If $\Sigma$ is a *quasi scalar* measure, that is if $\Sigma = \sigma \cdot \mathbf{1}$ with $\sigma$ a scalar measure and if $\Pi$ is a positive matrix measure with Lebesgue decomposition

$$\Pi(dx) = h(x)\sigma(dx) + \Pi^s(dx)\,,$$

we extend the definition (1.2) by

$$\mathcal{K}(\Sigma|\Pi) := -\int \log\det h(x)\ \sigma(dx)\,. \tag{2.12}$$

**Remark 2.2** *It is possible to rewrite the above quantity in the flavour of Kullback-Leibler information (or relative entropy) with the notation of [MS71] or [RR68], i.e.*

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

*if $\Sigma$ is strongly absolutely continuous with respect to $\Pi$, and infinity otherwise. See Corollary 8 in [GR14].*

## 2.2 Measures on $[0,\infty)$

When the measures are supported by $[0,\infty)$, there is a specific form of the Jacobi coefficients, leading to a particularly interesting parametrization, which will be crucial in the Laguerre ensemble.

In [DS02], it is proved that if a nontrivial matrix measure $\Sigma$ has a support included in $[0,\infty)$ then there exists a sequence $(\zeta_n)_n$ of non-singular elements of $\mathbb{M}_p$ such that

$$u_n^R = \zeta_{2n+1} + \zeta_{2n} \quad , \quad v_n^R = \zeta_{2n-1}\zeta_{2n} \tag{2.13}$$

with $\zeta_0 = \mathbf{0}$ and moreover,

$$\zeta_n = h_{n-1}^{-1} h_n, (n \geq 1) \tag{2.14}$$

with $h_0 = \mathbf{1}$ and for $n \geq 1$, $h_n \in \mathcal{H}_p$ is positive definite. Note that this implies

$$\gamma_n = \zeta_1 \dots \zeta_{2n} = h_{2n} \,.$$

From (2.9) we then have the representations of the Jacobi coefficients

$$\begin{aligned} B_{n+1} &= \sigma_n^* \gamma_n^{1/2} (\zeta_{2n+1} + \zeta_{2n}) \gamma_n^{-1/2} \sigma_n \\ A_n &= \sigma_{n-1}^* \gamma_{n-1}^{1/2} \zeta_{2n-1} \zeta_{2n} \gamma_n^{-1/2} \sigma_n \,. \end{aligned} \tag{2.15}$$

In the scalar case, this yields $B_1 = \zeta_1$ and for $n \geq 1$

$$B_{n+1} = \zeta_{2n+1} + \zeta_{2n}, \qquad (A_n)^2 = \zeta_{2n-1}\zeta_{2n} \,. \tag{2.16}$$

In the matrix case, we may set

$$Z_{2n+1} = \sigma_n^* \gamma_n^{1/2} \zeta_{2n+1} \gamma_n^{-1/2} \sigma_n \quad , \quad Z_{2n} = \sigma_n^* \gamma_n^{1/2} \zeta_{2n} \gamma_n^{-1/2} \sigma_n \,. \tag{2.17}$$

To highlight a further decomposition, we set

$$C_n = \sigma_n^* h_{2n}^{1/2} h_{2n-1}^{-1/2}, \qquad D_{n+1} = \sigma_n^* h_{2n}^{-1/2} h_{2n+1}^{1/2} \,.$$

With these definitions and (2.14) we see that the matrices

$$Z_{2n+1} = D_{n+1} D_{n+1}^*, \qquad Z_{2n} = C_n C_n^* \tag{2.18}$$

are in fact Hermitian positive definite. For the recursion coefficients we get the following matrix analogues of (2.16), $B_1 = D_1 D_1^*$, and for $n \geq 1$

$$B_{n+1} = D_{n+1} D_{n+1}^* + C_n C_n^* = Z_{2n+1} + Z_{2n}, \qquad A_n = D_n C_n^* \,. \tag{2.19}$$

In other words, the Jacobi operator $J$ can in fact be decomposed as $J = XX^*$, where $X$ is the bidiagonal matrix

$$X = \begin{pmatrix} D_1 & \mathbf{0} & \\ C_1 & D_2 & \ddots \\ & \ddots & \ddots \end{pmatrix} . \tag{2.20}$$

Moreover, the entries of $X$ can be chosen to be Hermitian positive definite. We are still free to choose the unitary matrices $\sigma_n$ (although we have to fix $\sigma_1 = \mathbf{1}$) in the definition of orthonormal

polynomials and we let $U$ be the block-diagonal matrix with $\sigma_1, \sigma_2, \dots$ on the diagonal. Moreover, let $P$ denote a block-diagonal matrix with unitary $p \times p$ matrices $\tau_1, \tau_2, \dots$ on the diagonal. Then our measure $\Sigma$ is also the spectral matrix measure of $UXPP^*X^*U^* = (UXP)(UXP)^*$. The matrix $UXP$ has the form

$$UXP = \begin{pmatrix} \sigma_1 D_1 \tau_1 & \mathbf{0} & & \\ \sigma_2 C_1 \tau_1 & \sigma_2 D_2 \tau_2 & \mathbf{0} & \\ & \sigma_3 C_2 \tau_2 & \sigma_3 D_3 \tau_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}.$$

For the first entry, $\sigma_1 = \mathbf{1}$ and $D_1$ is always Hermitian positive definite, so we may set $\tau_1 = \mathbf{1}$. Recall that for $A$ a non-singular matrix, there exists a unique unitary $\sigma$ such that $A\sigma$ is Hermitian positive definite, and if $\Sigma$ is nontrivial, all $D_k, C_k$ are non-singular. Therefore, we can recursively choose $\sigma_{k+1}$ such that $\sigma_{k+1} C_k \tau_k$ is Hermitian positive definite and then $\tau_{k+1}$ such that $\sigma_{k+1} D_{k+1} \tau_{k+1}$ is positive definite. This yields a unique decomposition with positive definite blocks.

## 2.3 Sum rules

For $\alpha^- < \alpha^+$, let $\mathcal{S}_p = \mathcal{S}_p(\alpha^-, \alpha^+)$ be the set of all bounded nonnegative measures $\Sigma \in \mathcal{M}_p(\mathbb{R})$ with

(i) $\operatorname{supp}(\Sigma) = K \cup \{\lambda_i^-\}_{i=1}^{N^-} \cup \{\lambda_i^+\}_{i=1}^{N^+}$, where $K \subset I = [\alpha^-, \alpha^+]$, $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$ and

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

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

Such a measure $\Sigma$ can be written as

$$\Sigma = \Sigma_{|I} + \sum_{i=1}^{N^+} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \Gamma_i^- \delta_{\lambda_i^-}, \tag{2.21}$$

for some nonnegative Hermitian matrices $\Gamma_1^+, \cdots, \Gamma_{N^+}^+, \Gamma_1^-, \cdots, \Gamma_{N^-}^-$. Further, we define $\mathcal{S}_{p,1} = \mathcal{S}_{p,1}(\alpha^-, \alpha^+) := \{\Sigma \in \mathcal{S}_p(\alpha^-, \alpha^+) | \, \Sigma(\mathbb{R}) = \mathbf{1}\}$.

### 2.3.1 The Hermite case revisited

In the scalar frame ($p = 1$), the Killip-Simon sum rule gives two different expressions for the divergence between a probability measure and the semicircle distribution (see [KS03] and [Sim11],

Chapter 3). In the more general case $p \geq 2$, it gives two forms for the divergence with respect to

$$\Sigma_{SC} = \mathrm{SC} \cdot \mathbf{1} \quad \text{with} \quad \mathrm{SC}(dx) = \frac{\sqrt{4-x^2}}{2\pi} 1_{[-2,2]}(x)\, dx$$

supported by $[\alpha^-, \alpha^+] = [-2, 2]$. We refer to [DKS10] Formula (10.4) and [Sim11], Formula (4.6.13) for this matrix sum rule. The block Jacobi matrix associated with $\Sigma_{SC}$ has entries

$$B_k^{\mathrm{SC}} = \mathbf{0}, \quad A_k^{\mathrm{SC}} = \mathbf{1},$$

for all $k \geq 1$. The spectral side of the sum rule involves a contribution of outlying eigenvalues, for which we define

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

and $\mathcal{F}_H^-(x) = \mathcal{F}_H^+(-x)$. Let $G$ be the very popular function (Cramér transform of the exponential distribution)

$$G(x) = x - 1 - \log x \quad (x > 0)\,.$$

We adopt the convention of the functional calculus, so that for $X \in \mathcal{H}_p$ positive, we have

$$\mathrm{tr}\, G(X) = \mathrm{tr} X - \log\det X - p\,. \tag{2.22}$$

Here is the first sum rule. This remarkable equality has been first proven by [DKS10]. In Section 4, we give a probabilistic proof. Indeed, we show that this sum rule is a consequence of two large deviation results.

**Theorem 2.3** *Let $\Sigma \in \mathcal{M}_{p,1}(\mathbb{R})$ be a spectral measure with Jacobi matrix (2.11). If $\Sigma \in \mathcal{S}_{p,1}(-2,2)$, then*

$$\mathcal{K}(\Sigma_{SC} \,|\, \Sigma) + \sum_{k=1}^{N^+} \mathcal{F}_H^+(\lambda_k^+) + \sum_{k=1}^{N^-} \mathcal{F}_H^-(\lambda_k^-) = \sum_{k=1}^{\infty} \tfrac{1}{2}\mathrm{tr} B_k^2 + \mathrm{tr}\, G(A_k A_k^*),$$

*where both sides may be infinite simultaneously. If $\Sigma \notin \mathcal{S}_{p,1}(-2,2)$, the right hand side equals $+\infty$.*

We remark that since $\mathrm{tr}\, G(\sigma A A^* \sigma^*) = \mathrm{tr}\, G(AA^*)$ for any unitary $\sigma$, the value of the right hand side in Theorem 2.3 is independent of the choice of $\sigma_n$'s in (2.9).

Let us restate the sum rule as in the notation of [Sim11]. To a spectral measure $\Sigma$ supported by $[-2,2]$, we associate the $m$-function (Stieltjès transform)

$$m(z) = \int \frac{1}{x-z} d\Sigma(x), \ (z \in \mathbb{C} \setminus [-2,2]).$$

For $z \in \mathbb{D}$ (interior of the unit disk), the function $M(z) = -m(z+z^{-1})$ admits radial limits : for almost all $\theta \in [0, 2\pi]$, the limit $M(e^{i\theta}) = \lim_{r\uparrow 1} M(re^{i\theta})$ exists and is neither vanishing nor infinite. Finally, let

$$\mathcal{Q}(\Sigma) = \frac{1}{\pi} \int_0^{2\pi} \log \left( \frac{\sin^p \theta}{\det(\operatorname{Im} M(e^{i\theta}))} \right) \sin^2 \theta \, d\theta \,. \tag{2.23}$$

Then the following statement is the combination of Theorem 4.6.1 and Theorem 4.6.3 in Simon's book [Sim11]. It is a *gem*, as defined on p.19 of [Sim11].

**Theorem 2.4** *Let $A_n, B_n$ be the entries of the block Jacobi matrix $J$ and let $\Sigma$ denote the spectral measure of $J$. Then*

$$\sum_{n=1}^{\infty} \operatorname{tr}((A_n A_n^*)^{1/2} - I)^2 + \operatorname{tr} B_n^2 < \infty$$

*if and only if*

*(a) The essential support of $J$ satisfies*

$$\sigma_{ess}(J) \subset [-2,2]$$

*(b) The eigenvalues $\{\lambda_j\}_{j=1}^{\infty} \notin \sigma_{ess}(J)$ satisfy*

$$\sum_{k=1}^{\infty} (|\lambda_k| - 2)^{3/2} < \infty$$

*(c) If $\Sigma$ admits the decomposition*

$$d\Sigma(x) = f(x)dx + d\Sigma_s(x)$$

*with $f \in$ and $d\Sigma_s$ singular with respect to $dx$, then*

$$\int_{-2}^{2} \sqrt{4-x^2} \log \det f(x) dx > -\infty \,.$$

*In this case, we have*

$$\sum_{n=1}^{\infty} \left(\tfrac{1}{2}\mathrm{tr}(B_n^2) + \mathrm{tr}(G(A_n A_n^*))\right) = \mathcal{Q}(\Sigma) + \sum_{\lambda \notin \sigma_{ess}(J)} F(\lambda). \tag{2.24}$$

*Here, $F$ is defined to be equal to $\mathcal{F}_H^+$ on $[2,\infty)$ and equal to $\mathcal{F}_H^-$ on $(-\infty, 2]$.*

Note that the integral $\mathcal{Q}(\Sigma)$ appearing in (2.24) and defined in (2.23) can be interpreted as a relative entropy, like in the scalar case treated in [Sim11], Lemma 3.5.1. For a measure $\Sigma$ supported on $[-2,2]$, the (inverse) Szegő mapping pushes forward $\Sigma$ to a measure $\Sigma_{Sz}$ on the unit circle symmetric with respect to complex conjugation, such that for all measureable and bounded $\varphi$,

$$\int_0^\pi \varphi(2\cos(\theta)) d\Sigma_{Sz}(e^{i\theta}) = \int_{-2}^2 \varphi(x)\, d\Sigma(x).$$

A straightforward generalization of the arguments in the above reference show

$$\mathrm{Im}\, M(e^{i\theta}) = \mathrm{Im}\, m(2\cos\theta) = \pi f(2\cos\theta)$$

for $\theta \in [0,\pi]$. Then, using the symmetry and setting $x = 2\cos\theta$, we obtain

$$\begin{aligned}
\mathcal{Q}(\Sigma) =& \frac{2}{\pi}\int_{-2}^2 \log\left(\frac{2^{-p}(4-x^2)^{-p/2}}{\det \pi f(x)}\right)\frac{1}{4}\sqrt{4-x^2}\, dx \\
=& \int_{-2}^2 \log\det\left(\frac{1}{2\pi}\sqrt{4-x^2} f(x)^{-1}\right)\frac{1}{2\pi}\sqrt{4-x^2}\, dx \\
=& \int \log\det\left(\frac{d\Sigma_{SC}}{d\Sigma}\right) d\,\mathrm{SC} = \mathcal{K}(\Sigma_{SC}\,|\,\Sigma).
\end{aligned}$$

### 2.3.2 Our new sum rule: the Laguerre case

In the Laguerre case, the central measure is the matrix Marchenko-Pastur law with scalar version

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

where $\tau \in (0,1]$, $\alpha^\pm = \tau^\pm = (1 \pm \sqrt{\tau})^2$ and we set $\Sigma_{\mathrm{MP}(\tau)} = \mathrm{MP}(\tau)\cdot \mathbf{1}$. The block Jacobi matrix associated with $\Sigma_{\mathrm{MP}}(\tau)$ has entries:

$$A_k^{MP} = \sqrt{\tau}\cdot \mathbf{1}\ ,\ (k \geq 1)\ ,\ B_1^{MP} = \mathbf{1}\ ,\ B_k^{MP} = (1+\tau)\cdot \mathbf{1}\ ,\ (k \geq 2)\ ,$$

which corresponds to $\zeta_{2k-1} = \mathbf{1}$ , $\zeta_{2k} = \tau \cdot \mathbf{1}$.

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

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

and

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

One of our main results is Theorem 2.5. Up to our knowledge, this result is new. The proof is again in Section 4.

**Theorem 2.5** *Assume the Jacobi matrix $J$ is nonnegative definite and let $\Sigma$ be the spectral measure associated with $J$. Then for any $\tau \in (0, 1]$, if $\Sigma \in \mathcal{S}_{p,1}(\tau^-, \tau^+)$,*

$$\mathcal{K}(\Sigma_{MP(\tau)} \,|\, \Sigma) + \sum_{n=1}^{N^+} \mathcal{F}_L^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_L^-(\lambda_n^-) = \sum_{k=1}^{\infty} \tau^{-1} \mathrm{tr}\, G(\zeta_{2k-1}) + \mathrm{tr}\, G(\tau^{-1}\zeta_{2k})$$

*where both sides may be infinite simultaneously and $\zeta_k$ is defined as in* (2.13)*. If $\Sigma \notin \mathcal{S}_{p,1}(\tau^-, \tau^+)$, the right hand side equals $+\infty$.*

Since the matrices $\zeta_k$ can be decomposed as in (2.14), they are in fact similar to a Hermitian matrix

$$\zeta_k = h_{k-1}^{-1/2} \left[ h_{k-1}^{-1/2} h_k h_{k-1}^{-1/2} \right] h_{k-1}^{1/2},$$

hence the sum on the right hand side in Theorem 2.5 is real valued.

Similar to the matrix *gem* , Theorem 2.4, we can formulate equivalent conditions on the matrices $\zeta_k$ and the spectral measure, which characterize finiteness of the sum. The following corollary is the matrix counterpart of Corollary 2.4 in [GNR16b]. It follows immediately from Theorem 2.5, since $\mathcal{F}_L^-(0) = \infty$ and

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

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

$$\mathrm{tr}\, G(\mathbf{1} + H) = \frac{1}{2}\mathrm{tr} H^2 + o(||H||) \quad (||H|| \to 0).$$

Here, $||\cdot||$ is any matrix norm.

**Corollary 2.6** *Assume the Jacobi matrix $J$ is nonnegative definite and let $\Sigma$ be the spectral measure of $J$. Then*

$$\sum_{k=1}^{\infty}[\mathrm{tr}(\zeta_{2k-1}-\mathbf{1})^2+\mathrm{tr}(\zeta_{2k}-\tau\mathbf{1})^2]<\infty \tag{2.25}$$

*if and only if*

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

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

# 3 Randomization and large deviations

## 3.1 Matrix random models

The results of the previous section rely on two classical distributions of random Hermitian matrices: the Gaussian (or Hermite) and the Laguerre (or Wishart) ensemble. We denote by $\mathcal{N}(0,\sigma^2)$ the centered Gaussian distribution with variance $\sigma^2$. A random variable $X$ taking values in $\mathcal{H}_N$ is distributed according to the Gaussian unitary ensemble $\mathrm{GUE}_N$, if all real diagonal entries are distributed as $(0,1)$ and the if the real and imaginary parts of off-diagonal variables are independent and $\mathcal{N}(0,1/2)$ distributed (this is called complex standard normal distribution). All entries are assumed to be independent up to symmetry and conjugation. The random matrix $X/\sqrt{N}$ has then the distribution given by (1.6) and the joint density of the (real) eigenvalues $\lambda=(\lambda_1,\ldots,\lambda_N)$ of $X$ is (see for example [AGZ10])

$$g_G(\lambda)=c_G\prod_{1\le i<j\le N}|\lambda_i-\lambda_j|^2\prod_{i=1}^{N}e^{-\lambda_i^2/2}. \tag{3.1}$$

In analogy to the scalar $\chi^2$ distribution, the Laguerre ensemble is the distribution of the square of Gaussian matrices as follows. If $G$ denotes a $N\times\gamma$ matrix with independent complex standard normal entries, then $GG^*$ is said to be distributed according to the Laguerre ensemble $\mathrm{LUE}_N(\gamma)$ with parameter $\gamma$. If $\gamma\ge N$, the eigenvalues have the density (see for example [AGZ10])

$$g_L(\lambda)=c_L^{\gamma}\prod_{1\le i<j\le N}|\lambda_i-\lambda_j|^2\prod_{i=1}^{N}\lambda_i^{\gamma-N}e^{-\lambda_i}1_{\{\lambda_i>0\}}. \tag{3.2}$$

It is a well-known consequence of the invariance under unitary conjugation, that in the classical ensembles (1.6), the array of random eigenvalues and the random eigenvector (unitary) matrix are independent. Further, this latter matrix is Haar distributed ([Daw77]). This implies the following equality in distribution for the weights given in Lemma 3.2 (see Proposition 3.1 in [GR14]), which is a matrix version of the beta-gamma relation for scalar random variables. First we need a definition.

**Definition 3.1** *1. If $v_1, \dots, v_N$ are independent complex standard normal distributed vectors in $\mathbb{C}^p$, set $V_j = v_j v_j^*$ for $j \leq N$. We say that $(V_1, \dots, V_N)$ follows the distribution $\mathbb{G}_{p,N}$ on $(\mathcal{H}_p)^N$.*

*2. If $U$ be Haar distributed in the set of $N \times N$ unitary matrices, set $u_j = (U_{i,j})_{1 \leq i \leq p} \in \mathbb{C}^p$ and $W_j = u_j u_j^*$ for $j \leq N$. We say that $(W_1, \dots, W_N)$ follows the $\mathbb{D}_{p,N}$ distribution on $(\mathcal{H}_p)^N$.*

In the scalar case, the array of weights $W_j = u_j^2$ is uniformly distributed on the simplex $\{w_1, \dots w_N \in [0,1] : \sum_i w_i = 1\}$.

**Lemma 3.2** *If $(V_1, \dots, V_N)$ follows the distribution $\mathbb{G}_{p,N}$ and if $H = \sum_{k=1}^N V_k$, then $\left(H^{-1/2} V_1 H^{-1/2}, \dots, H^{-1/2} V_N H^{-1/2}\right)$ follows the distribution $\mathbb{D}_{p,N}$.*

Our first large deviation principle will hold for a general class of $p \times p$ matrix measures. We draw the random eigenvalues $\lambda_1, \dots, \lambda_N$ from the absolute continuous distribution $\mathbb{P}_N^V$

$$d\mathbb{P}_N^V(\lambda) = \frac{1}{Z_V^N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^N e^{-NV(\lambda_i)}. \tag{3.3}$$

We suppose that the potential $V$ is continuous and real valued on the interval $(b^-, b^+)$ ($-\infty \leq b^- < b^+ \leq +\infty$), infinite outside of $[b^-, b^+]$ and $\lim_{x \to b^\pm} V(x) = V(b^\pm)$ with possible limit $V(b^\pm) = +\infty$. Under the assumption

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

the empirical distribution $\mu_{\mathtt{u}}^{(N)}$ of eigenvalues $\lambda_1, \dots, \lambda_N$ has a limit $\mu_V$ in probability, which is the unique minimizer of

$$\mu \mapsto \mathcal{E}(\mu) := \int V(x) d\mu(x) - \iint \log|x-y| d\mu(x) d\mu(y). \tag{3.4}$$

and which has a compact support (see [Joh98] or [AGZ10]). This convergence can be viewed as a consequence of the LDP for the sequence $(\mu_{\mathtt{u}}^{(N)})_N$. We need two additional assumptions on $\mu_V$:

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

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

$$\mathcal{J}_V(x) := V(x) - 2\int \log|x-\xi|\, d\mu_V(\xi) \tag{3.5}$$

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

In the Hermite case, we have $V(x) = \frac{1}{2}x^2$ and the equilibrium measure $\mu_V$ is the semicircle law. In the Laguerre case, we may set $V(x) = \tau^{-1}x - (\tau^{-1}-1)\log(x)$ for $\tau \in (0,1]$ and $V(x) = +\infty$ for negative $x$. In this case, $\mu_V$ is the Marchenko-Pastur law $\mathrm{MP}(\tau)$. In both the Hermite and the Laguerre case, the assumption (A1), (A2) and (A3) are satisfied. We need one more definition related to outlying eigenvalues:

$$\mathcal{F}_V^+(x) = \begin{cases} \mathcal{J}_V(x) - \inf_{\xi\in\mathbb{R}} \mathcal{J}_V(\xi) & \text{if } \alpha^+ \le x \le b^+, \\ \infty & \text{otherwise,} \end{cases} \tag{3.6}$$

$$\mathcal{F}_V^-(x) = \begin{cases} \mathcal{J}_{\mathcal{V}}(x) - \inf_{\xi\in\mathbb{R}} \mathcal{J}_V(\xi) & \text{if } b^- \le x \le \alpha^-, \\ \infty & \text{otherwise.} \end{cases} \tag{3.7}$$

One may check that in the Hermite case, $\mathcal{F}_V^\pm = \mathcal{F}_H^\pm$ and in the Laguerre case, $\mathcal{F}_V^\pm = \mathcal{F}_L^\pm$, where $\mathcal{F}_H^\pm$ and $\mathcal{F}_L^\pm$ have been defined in the previous section.

## 3.2 Basics on Large Deviations

In order to be self-contained, let us recall the definition of large deviations an some important tools ([DZ98]).

**Definition 3.3** *Let $E$ be a topological Hausdorff space and let $\mathcal{I} : E \to [0,\infty]$ be a lower semi-continuous function. We say that a sequence $(P_n)_n$ of probability measures on $(E, \mathcal{B}(E))$ satisfies a large deviation principle (LDP) with rate function $\mathcal{I}$ and speed $a_n$ i f:*

*(i) For all closed sets $F \subset E$:*

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

*(ii) For all open sets $O \subset E$:*

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

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

We will frequently use the following principle ([DZ98], p. 126).

**Contraction principle**. *Suppose that $(P_n)_n$ satisfies an LDP on $(E, \mathcal{B}(E))$ with good rate function $\mathcal{I}$ and speed $a_n$. Let $f$ be a continuous mapping from $E$ to another topological Hausdorff space $F$. Then $P_n \circ f^{-1}$ satisfies a LDP on $(F, \mathcal{B}(F))$ with speed $a_n$ and good rate function*

$$\mathcal{I}'(y) = \inf_{\{x \in E | f(x) = y\}} \mathcal{I}(x), \quad (y \in F).$$

To prove our main large deviation principle, we will use a special extension of Baldi's theorem, which extends also Bryc's lemma. This new theorem is given in the Appendix.

To apply this theorem in our setting, we remark that the topological dual of $\mathcal{M}_p(T)$ is the space $\mathcal{C}_p(T)$ of bounded continuous functions $f : T \to \mathcal{H}_p$ with the pairing

$$\langle \Sigma, f \rangle = \mathrm{tr} \int f d\Sigma.$$

## 3.3 Large Deviations

### 3.3.1 Random measures

Our first LDP holds for $p \times p$ matrix measures

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

whose support $(\lambda_1, \cdots, \lambda_N)$ is $\mathbb{P}_N^V$ distributed and where the distribution of weights $(W_1, \cdots, W_N)$ is $\mathbb{D}_{p,N}$ as in the case of classical ensembles. As explained in the introduction, this is precisely the distribution of the spectral measure of an $N \times N$ matrix $X_N$, drawn from the distribution (1.6). Recall that under assumption (A1), the empirical measure of the eigenvalues converges to an equilibrium measure $\mu_V$, supported by $[\alpha^-, \alpha^+]$. The rate function of our large deviation principle involves the reference matrix measure

$$\Sigma_V = \mu_V \cdot \mathbf{1}\,.$$

We recall that $\mathcal{F}_V^{\pm}$ has been defined in (3.6) and (3.7). The following theorem is the matrix counterpart of Theorem 3.1 in [GNR16b]. Note that in the scalar case, we had an additional parameter $\beta > 0$, corresponding to the inverse temperature of the log-gas. In the matrix case, we choose to fix $\beta = 2$ (for complex matrices) due to the nature of the matrix spaces.

**Theorem 3.4** *Assume that the potential $V$ satisfies the assumptions (A1), (A2) and (A3). Then the sequence of spectral measures $\Sigma^{(N)}$ under $\mathbb{P}_N^V \otimes \mathbb{D}_{p,N}$ satisfies the LDP with speed $N$ and good rate function*

$$\mathcal{I}_V(\Sigma) = \mathcal{K}(\Sigma_V \mid \Sigma) + \sum_{k=1}^{N^+} \mathcal{F}_V^+(\lambda_k^+) + \sum_{k=1}^{N^-} \mathcal{F}_V^-(\lambda_k^-)$$

*if $\Sigma \in \mathcal{S}_{p,1}(\alpha^-, \alpha^+)$ and $\mathcal{I}_V(\Sigma) = \infty$ otherwise.*

**Remark 3.5** *A natural extension of Theorem 3.4 holds for potentials $V = V_N$ depending on $N$, provided that $V_N$ converges to a deterministic potential $V$ in an appropriate sense. For example, it holds if we suppose that $V_N : \mathbb{R} \to (-\infty, +\infty]$ is a sequence converging to $V$ uniformly on the level sets $\{V \leq M\}$, where $V$ satisfies assumptions (A1), (A2) and (A3) and such that $V_N(x) \geq V(x)$.*

### 3.3.2 Jacobi coefficients

In the cases of Hermite and Laguerre ensembles, the particular form of the distribution of the parameters $(A_1, B_1, \dots)$ and $(\zeta_1, \zeta_2, \cdots)$ respectively, allows us to prove further LDP's for the spectral measure, independently of Theorem 3.4. They are in the subset $\mathcal{M}_{p,1,c}$ of compactly supported normalized matrix measures. Since we need a specific block structure, we assume $N = np$.

**Theorem 3.6** *Let $\Sigma^{(n)}$ be the spectral measure of $\frac{1}{\sqrt{np}} X_n$. Assume that $X_n$ is distributed according to the Hermite ensemble $\mathrm{GUE}_N$ ($N = np$). Then the sequence $(\Sigma^{(n)})_n$ satisfies the LDP in $\mathcal{M}_{p,1,c}(\mathbb{R})$ with speed $pn$ and good rate function*

$$\mathcal{I}_H(\Sigma) = \sum_{k=1}^{\infty} \left[ \tfrac{1}{2} \mathrm{tr}\, B_k^2 + \mathrm{tr}\, G(A_k A_k^*) \right]$$

*where $B_k, A_k$ are the recursion coefficients of $\Sigma$ as in (2.8) if $\Sigma$ is non-trivial and $\mathcal{I}_H(\Sigma) = \infty$ if $\Sigma$ is trivial.*

**Theorem 3.7** *Let $Y_n$ be distributed according to the Laguerre ensemble $\mathrm{LUE}_N(p\gamma_n)$, with $(N = np)$, $\gamma_n \geq n$ an integer sequence such that $\frac{n}{\gamma_n} \to \tau \in (0,1]$ and let $\Sigma^{(n)}$ be the spectral measure of $\frac{1}{p\gamma_n}Y_n$ with a decomposition of recursion coefficients as in Section 2. Then the sequence $(\Sigma^{(n)})_n$ satisfies the LDP in $\mathcal{M}_{p,1,c}([0,\infty))$ with speed $pn$ and good rate function*

$$\mathcal{I}_L(\Sigma) = \sum_{k=1}^{\infty} \left[\tau^{-1}G(\zeta_{2k-1}) + G(\tau^{-1}\zeta_{2k})\right],$$

*with $\zeta_k$ as in (2.14). If $\Sigma$ is a trivial measure, we have $\mathcal{I}_L(\Sigma) = \infty$.*

In order to prove LDP's for the spectral measures in terms of the recursion coefficients we need the following results for matrices of fixed size. The first and third are straightforward extensions of the scalar case, the second one can be found in [GNRW12], with small changes to allow a general sequence of parameters.

**Lemma 3.8** *(i) If $X \sim \mathrm{GUE}_p$ with $p$ fixed, then the sequence $(\frac{1}{\sqrt{n}}X)_n$ satisfies the LDP with speed $n$ and good rate function*

$$\mathcal{I}_1(X) = \tfrac{1}{2}\operatorname{tr} X^2.$$

*(ii) Let $Y_n \sim \mathrm{LUE}_p(\gamma_n)$ with a positive sequence $(\gamma_n)_n$ such that $\frac{\gamma_n}{n} \to \gamma > 0$, then the sequence $(\frac{1}{n}Y_n)_n$ satisfies the LDP with speed $n$ and good rate function*

$$\mathcal{I}_2(Y) = \gamma \operatorname{tr} G(\gamma^{-1}Y)$$

*if $Y$ is Hermitian and nonnegative and $\mathcal{I}_2(Y) = \infty$ otherwise.*

*(iii) Let $Z \sim \mathrm{LUE}_p(1)$ with $p$ fixed, that is, $Z = vv^*$ when $v$ is a vector of independent complex standard normal random variables. Then the sequence $(\frac{1}{n}Z)_n$ satisfies the LDP with speed $n$ and good rate function*

$$\mathcal{I}_3(Z) = \operatorname{tr} Z$$

*if $Z$ is Hermitian and nonnegative and $\mathcal{I}_3(Z) = \infty$ otherwise.*

# 4 Proof of Theorems 2.3 and 2.5: From large deviations to sum rules

**Proof of Theorem 2.3:**
Consider the matrix measure $\Sigma^{(n)}$ with $N = np$ support points with density

$$\frac{1}{Z_V^N} \prod_{1 \le i < j \le N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{N} e^{-N\lambda_i^2/2}$$

and weight distribution $\mathbb{D}_{p,N}$ independent of the support points. By Theorem 3.4, the sequence $(\Sigma^{(n)})_n$ satisfies the LDP with speed $np$ and rate function $\mathcal{I}_V$, where $\mu_V$ is the semicircle law and furthermore $\mathcal{F}_V^{\pm} = \mathcal{F}_H^{\pm}$. That is, the rate function is precisely the left hand side of the equation in Theorem 2.3. On the other hand, $\Sigma^{(n)}$ is also the spectral measure of the random matrix $\frac{1}{\sqrt{pn}} X_n$, where $X_n \sim \mathrm{GUE}_{pn}$. By Theorem 3.6, the sequence $(\Sigma^{(n)})_n$ satisfies also the LDP in the space of compactly supported measures with speed $np$ and rate function $\mathcal{I}_H$, the right hand side of the equation in Theorem 2.3. Since a large deviation rate function is unique, we must have $\mathcal{I}_V(\Sigma) = \mathcal{I}_H(\Sigma)$ for any compactly supported $\Sigma \in \mathcal{M}_{p,1,c}$. If $\Sigma$ is not compactly supported, it suffices to remark that the recursion coefficients cannot satisfy $\sup_n (||A_n|| + ||B_n||) < \infty$, as otherwise $J$ would be a bounded operator. But $\operatorname{tr} B^2 + \operatorname{tr} G(A)$ diverges as $||A|| \to \infty$ or $||B|| \to \infty$ and so the right hand side in Theorem 2.3 equals $+\infty$. □

**Proof of Theorem 2.5:**
Fix $\tau \in (0, 1]$ and let $V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log x$ for $x \ge 0$ and $V(x) = +\infty$ if $x < 0$. From Theorem 3.4 we get that under the distribution $\mathbb{P}_N^V \otimes \mathbb{D}_{p,N}$ the sequence $(\Sigma^{(n)})_n$ satisfies the LDP with speed $N = np$ and rate function $\mathcal{I}_V$. In this case, the equilibrium measure is the Marchenko-Pastur law $\mathrm{MP}(\tau)$ multiplied by $\mathbf{1}$. Further, we have $\mathcal{F}_V^{\pm} = \mathcal{F}_L^{\pm}$. So that, $\mathcal{I}_V$ is nothing more than the left hand side of the sum rule in Theorem 2.5. We would like to combine this result with the LDP in Theorem 3.7, but since this requires integer parameters, we need to modify the potential slightly. Define $\gamma_n = \lceil n\tau^{-1} \rceil$ and consider the eigenvalue distribution with density

$$\frac{1}{Z_V^N} \prod_{1 \le i < j \le N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{N} \lambda_i^{p\gamma_n - pn} e^{-p\gamma_n \lambda_i} 1_{[0,\infty)}(\lambda_i) \tag{4.1}$$

This is the eigenvalue distribution of the matrix $\frac{1}{p\gamma_n} Y_n$, when $Y_n \sim \mathrm{LUE}_{pn}(p\gamma_n)$. By Theorem 3.7, the spectral measure of this matrix satisfies the LDP with speed $pn$ and rate function $\mathcal{I}_L$ which is the right hand side of the sum rule in Theorem 2.5. We may as well write the eigenvalue

density (4.1) as

$$\frac{1}{Z_V^N} \prod_{1\le i<j\le N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^N e^{-pnV_n(\lambda_i)}, \tag{4.2}$$

where

$$V_n(x) = \frac{\lceil n\tau^{-1}\rceil}{n} x - \left(\frac{\lceil n\tau^{-1}\rceil}{n} - 1\right) \log x \tag{4.3}$$

for nonnegative $x$. Then $V_n(x) \ge V(x)$ for all $x$ and on the sets $\{x|\, V(x) \le M\}$, the potentials $V_n$ converge uniformly to $V$. Note that the point 0 is included in the level sets of $V$ only if $\tau = 1$. Therefore, by Remark 3.5, the spectral measure with support point density (4.2) satisfies the same LDP as under the density $\mathbb{P}_{pn}^V$ and then with rate function $\mathcal{I}_V$. This yields $\mathcal{I}_V(\Sigma) = \mathcal{I}_H(\Sigma)$ for any compactly supported measure $\Sigma$. The extension to measures with non-compact support follows as in the proof of Theorem 2.3. □

# 5 Proof of Theorem 3.4: Spectral LDP for a general potential

This section is devoted to the proof of Theorem 3.4. We will follow the track of the proof developed for the scalar case in [GNR16b] and will often refer to this paper for more details. The main idea is to apply the projective method and study a family of matrix measures restricted to the support $I = [\alpha^-, \alpha^+]$ of the equilibrium measure and a fixed number of extremal eigenvalues. For $\Sigma \in \mathcal{S}_p$ with

$$\Sigma = \Sigma_{|I} + \sum_{i=1}^{N^+} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \Gamma_i^- \delta_{\lambda_i^-} \tag{5.1}$$

we define the $j$-th projector $\pi_j$ by

$$\pi_j(\Sigma) = \Sigma_{|I} + \sum_{i=1}^{N^+\wedge j} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-\wedge j} \Gamma_i^- \delta_{\lambda_i^-},$$

that is, all but the $j$-th largest and smallest eigenvalues outside of $I = [\alpha^-, \alpha^+]$ are omitted. Note that $\pi_j$ is not continuous in the weak topology. For this reason we need to change our topology on $\mathcal{S}_p$ by identifying $\Sigma$ as in (5.1) with the vector

$$\big(\Sigma_{|I}, (\lambda_i^+)_{i\ge1}, (\lambda_i^-)_{i\ge1}, (\Gamma_i^+)_{i\ge1}, (\Gamma_i^-)_{i\ge1}\big) \tag{5.2}$$

with $\lambda_i^+ = \alpha^+$ and $\Gamma_i^+ = 0$ if $i > N^+$ and $\lambda_i^- = \alpha^-$ and $\Gamma_i^- = 0$ if $i > N^-$. Then we say that $\Sigma^{(n)}$ converges to $\Sigma$ if:

$$\begin{aligned} &\Sigma^{(n)}_{|I} \xrightarrow[n\to\infty]{} \Sigma_{|I} \text{ weakly and for every } i \geq 1 \\ &\big(\lambda_i^+(\Sigma^{(n)}), \lambda_i^-(\Sigma^{(n)}), \Gamma_i^+(\Sigma^{(n)}), \Gamma_i^-(\Sigma^{(n)})\big) \xrightarrow[n\to\infty]{} \big(\lambda_i^+(\Sigma), \lambda_i^-(\Sigma), \Gamma_i^+(\Sigma), \Gamma_i^-(\Sigma)\big). \end{aligned} \tag{5.3}$$

Analogously to Lemma 4.5 in [GNR16b], one can show that on the smaller set $\mathcal{S}_{p,1}$ of normalized measures, this topology is (strictly) stronger than the weak topology.

Let for $j$ fixed and $N > 2j$

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

Then the following joint LDP holds for the largest and/or smallest eigenvalues, where we write $\mathbb{R}^{\uparrow j}$ (resp. $\mathbb{R}^{\downarrow j}$) for the subset of $\mathbb{R}^j$ of all vectors with non-decreasing (resp. non-increasing) entries and, with a slight abuse of notation, we write $\alpha^\pm$ for the vector $(\alpha^\pm, \dots, \alpha^\pm) \in \mathbb{R}^j$.

**Theorem 5.1** *Let $j$ be a fixed integer and the potential $V$ such that (A1), (A2) and the control condition (A3) are satisfied.*

1. *If $b^- < \alpha^-$ and $\alpha^+ < b^+$, then the law of $(\lambda^+(j), \lambda^-(j))$ under $\mathbb{P}_N^V$ satisfies the LDP in $\mathbb{R}^{2j}$ with speed $N$ and good rate function*

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

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

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

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

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

## 5.1 LDP for the restricted measure and extremal eigenvalues

Suppose now that the distribution of $\Sigma^{(N)}$ is as in Theorem 3.4 and the assumptions (A1), (A2) and (A3) are satisfied. By Lemma 3.2, we may decouple the weights and consider the (non-normalized) measure

$$\tilde{\Sigma}^{(N)} = \frac{1}{N}\sum_{k=1}^{N} v_k v_k^* \delta_{\lambda_k}, \tag{5.4}$$

where the entries of $v_1, \ldots, v_N \in \mathbb{C}^p$ are independent complex standard normal distributed random vectors. The original distribution can then be recovered as the pushforward under

$$\tilde{\Sigma} \mapsto \tilde{\Sigma}(\mathbb{R})^{-1/2} \cdot \tilde{\Sigma} \cdot \tilde{\Sigma}(\mathbb{R})^{-1/2}.$$

Let $I(j) := I \setminus \{\lambda_1^+, \lambda_1^-, \ldots, \lambda_j^+, \lambda_j^-\}$ denote the interval $I$ without the $j$ largest and smallest eigenvalues. Analogously, let $I^+(j) := I \setminus \{\lambda_1^+, \ldots, \lambda_j^+\}$ and $I^-(j) := I \setminus \{\lambda_1^-, \ldots, \lambda_j^-\}$. Then we write $\tilde{\Sigma}^{(N)}_{I(j)}$ for the restriction of $\tilde{\Sigma}^{(N)}$ to $I(j)$. We use the analogous notation for the restrictions to $I^+(j), I^-(j)$ and $I$. The main result in this section is a joint LDP for the restricted measure and the collection of extremal eigenvalues.

**Theorem 5.2** *Suppose that the law of eigenvalues and weights is given by $\mathbb{P}_N^V \otimes \mathbb{G}_{p,N}$ and consider $\tilde{\Sigma}^{(N)}$ as a random element in $\mathcal{S}_p$ with topology* (5.3).

1. *If $b^- < \alpha^- < \alpha^+ < b^+$, then for any fixed $j \in \mathbb{N}$, the sequence $\big(\tilde{\Sigma}^{(N)}_{I(j)}, \lambda^+(j), \lambda^-(j)\big)$ satisfies the joint LDP with speed $N$ and good rate function*

$$\mathcal{I}(\Sigma, x^+, x^-) = \mathcal{K}(\Sigma_V \,|\, \Sigma) + \operatorname{tr}\Sigma(\mathbb{R}) - p + \mathcal{I}_{\lambda^\pm}(x^+, x^-)$$

2. *If $b^- = \alpha^-$, but $\alpha^+ < b^+$ (or $b^+ = \alpha^+$, but $\alpha^- > b^-$), then, with the same notation as in Theorem 5.1, $\big(\tilde{\Sigma}^{(N)}_{I^+(j)}, \lambda^+(j)\big)$(or $\big(\tilde{\Sigma}^{(N)}_{I^-(j)}, \lambda^-(j)\big)$ respectively,) satisfies the LDP with speed $N$ and good rate function*

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

**Proof:** We show here only the first part of the theorem, for the other cases just omit the largest or smallest eigenvalues. To begin with, for $M > \max\{|\alpha^+|, |\alpha^-|\}$, let $\lambda_M^+(j)$ (resp.$\lambda_M^-(j)$) be the collection of truncated eigenvalues

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

for $i = 1, \ldots, j$. To further simplify the notations, set $\lambda_M^{\pm}(j) := (\lambda_{M,1}^+, \ldots, \lambda_{M,j}^+, \lambda_{M,1}^-, \ldots, \lambda_{M,j}^-)$. The sequence $(\tilde{\Sigma}_{I(j)}^{(N)}, \lambda_M^{\pm}(j))$ is exponentially tight, since, with the compact set

$$K_{\gamma,M} = \left\{(\Sigma, \lambda) \in \mathcal{M}_p(I) \times \mathbb{R}^{2j} |\ \Sigma(I) \leq \gamma \cdot \mathbf{1}, \lambda \in [-M, M]^{2j}\right\},$$

we have

$$\limsup_{N\to\infty} \frac{1}{N} \log P\left((\tilde{\Sigma}_{I(j)}^{(N)}, \lambda_M^{\pm}(j)) \notin K_{\gamma,M}\right) \leq \limsup_{N\to\infty} \frac{1}{N} \log P\left(\frac{1}{N}\sum_{k=1}^{N} v_k v_k^* > \gamma \mathbf{1}\right) \leq -c\gamma,$$

where we used the fact that $\sum_{k=1}^{N} v_k v_k^*$ follows the $\mathrm{LUE}_p(N)$ distribution. We prove the joint LDP by applying Theorem 7.3. For this, let $D$ be the set of continuous $f : [\alpha^-, \alpha^+] \to \mathcal{H}_p$ such that for all $x \in [\alpha^-, \alpha^+]$, $f(x) < \mathbf{1}$, i.e., the eigenvalues of $f(x)$ are smaller than 1. For $f \in D$ and $\varphi$ a bounded continuous function from $\mathbb{R}^{2j}$ to $\mathbb{R}$ , we consider the joint moment generating function

$$\mathcal{G}_N(f, \varphi) = \mathbb{E}\left[\exp\left\{N\left(\mathrm{tr}\int f d\tilde{\Sigma}_{I(j)}^{(N)} + \varphi(\lambda_M^{\pm}(j))\right)\right\}\right].$$

Since the weights $v_k v_k^*$ of $\tilde{\Sigma}_{I(j)}^{(N)}$ are independent, we may first integrate with respect to the $v_k$'s, so that

$$\mathcal{G}_N(f, \varphi) = \mathbb{E}\left[\exp\left(N\varphi(\lambda_M^{\pm}(j))\right) \prod_{i:\lambda_i \in I(j)} \mathbb{E}\left[\exp\left\{\mathrm{tr}(f(\lambda_i) v_k v_k^*)\right\} | \lambda_1, \ldots, \lambda_n\right]\right] \tag{5.5}$$

$$= \mathbb{E}\left[\exp\left(N\varphi(\lambda_M^{\pm}(j))\right) \prod_{i:\lambda_i \in I(j)} \mathbb{E}\left[\exp\left\{v_k^* f(\lambda_i) v_k\right\} | \lambda_1, \ldots, \lambda_n\right]\right]. \tag{5.6}$$

Now, it is clear that for $v$ a standard normal complex vector in $\mathbb{C}^p$ and $A \in \mathcal{H}_p$ such that $A < 1$, we have

$$\log \mathbb{E}\left[\exp\left(v^* A v\right)\right] = -\log\det(\mathbf{1} - A) =: L(A) \tag{5.7}$$

so that (5.5) becomes,

$$\mathcal{G}_N(f, \varphi) = \mathbb{E}\left[\exp\left\{N\left(\mu_{\mathrm{u},I(j)}^{(N)}(L \circ f) + \varphi(\lambda_M^{\pm}(j))\right)\right\}\right],$$

where $\mu_{\mathrm{u},I(j)}^{(N)}$ is the restriction of the (scalar) empirical eigenvalue distribution to $I(j)$. It remains to calculate the expectation with respect to $\mathbb{P}_N^V$.

By Theorem 5.1, the extremal eigenvalues $\lambda^{\pm}(j)$ of the spectral measure satisfy the LDP with speed $N$ rate function $\mathcal{I}_{\lambda^{\pm}}$. By the contraction principle (see [DZ98] p.126), the truncated eigenvalues satisfy the LDP with rate function

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

Since the truncated eigenvalues are bounded, we can conclude from Varadhan's Integral Lemma ([DZ98] p. 137)

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left\{N\varphi(\lambda_M^{\pm}(j))\right\}\right] = J(\varphi) := \sup_{y\in\mathbb{R}^{2j}} \left\{\varphi(y) - \mathcal{I}_{M,\lambda^{\pm}}(y)\right\}. \tag{5.8}$$

Since $\mu_{\mathfrak{u}}^{(N)}$ satisfies a LDP with speed $N^2$, but we consider only the slower scale at speed $N$, we may replace it by the limit measure $\mu_V$ at a negligible cost. For the exact estimates, we may follow along the lines of [GNR16b] to conclude

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \log \mathcal{G}_N(f, \varphi) &= \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left\{N\left(\mu_V(L \circ f) + \varphi(\lambda_M^{\pm}(j))\right)\right\}\right] \\ &= G(f) + J(\varphi) \end{aligned}$$

where

$$G(f) = \int L \circ f \, d\mu_V$$

and $L$ is given in (5.7). Theorem 7.3 yields the LDP for $(\tilde{\Sigma}_I^{(n)}, \lambda_M^{\pm})$ with good rate

$$\mathcal{I}(\Sigma, \lambda) = G^*(\Sigma) + \mathcal{I}_{M,\lambda^{\pm}}(\lambda),$$

once we show the second assumption therein is satisfied.

Theorem 5 of [Roc71] identifies $G^*$ as

$$G^*(\Sigma) = \int L^*(h) d\mu_V + \int r\left(\frac{d\Sigma^S}{d\theta}\right) d\theta, \tag{5.9}$$

where:

- $L^*$ is the convex dual of $L$
- $r$ its recession function
- the Lebesgue-decomposition of $\Sigma$ with respect to $\mu_V$ is

$$d\Sigma(x) = h(x) d\mu_V(x) + d\Sigma^S(x)$$

- $\theta$ is any scalar measure such that $\Sigma^S$ is absolutely continuous with respect to $\theta$.

We begin by calculating $L^*$ and $r$. By definition,

$$L^*(X) = \sup_{Y \in \mathcal{H}_p} \{\mathrm{tr}(XY) - L(Y)\}\,.$$

The recession function is

$$r(X) = \sup\{\mathrm{tr}(XW) \mid ||W||_\infty < 1\}\,.$$

The function $L$ is convex (as in the scalar case, apply Hölder's inequality in the definition (5.7)) and analytic. The supremum is then reached at a critical value. We denote by $\mathcal{D}[F(Y)]$ the Fréchet derivative of a function $F : \mathcal{H}_p \to \mathbb{R}$ at $Y$ and look for $Y$ such that

$$\mathcal{D}[\mathrm{tr}(XY) - L(Y)](Z) = 0 \tag{5.10}$$

for every $Z$. It is well known that, as functions of $Y$ for $X$ fixed, $\mathcal{D}[\mathrm{tr}(XY)](Z) = \mathrm{tr}(XZ)$ and $\mathcal{D}[\log\det Y](Z) = \mathrm{tr}(Y^{-1}Z)$ so that (5.10) becomes, by the chain rule,

$$\mathrm{tr}(XZ) - \mathrm{tr}((\mathbf{1} - Y)^{-1}Z) = 0$$

for every $Z$ i.e. $X - (I - Y)^{-1} = 0$ hence $Y = \mathbf{1} - X^{-1}$ and

$$L^*(X) = \mathrm{tr}\,(X - \mathbf{1}) + \log\det(X^{-1}) = \mathrm{tr}X - p + \log\det(X^{-1}) = \mathrm{tr}\,G(X)\,. \tag{5.11}$$

If $X$ has a negative eigenvalue, then $r(X) = \infty$. For $X$ nonnegative definite, the supremum is attained for $W = \mathbf{1}$, so that

$$r(X) = \mathrm{tr}X. \tag{5.12}$$

Gathering (5.11) and (5.12) and plugging into (5.9) we get

$$\begin{aligned} G^*(\Sigma) &= \mathrm{tr}\int h d\mu_V - \int \log\det h\, d\mu_V - p + \mathrm{tr}\int d\Sigma^S \\ &= \mathcal{K}(\Sigma_V|\Sigma) + \mathrm{tr}\,\Sigma(I) - p. \end{aligned}$$

It remains to show that $G^*$ is sufficiently convex. A measure $\Sigma \in \mathcal{M}_p([\alpha^-, \alpha^+])$ is a point of strict convexity (an exposed point) of $G^*$ if there exists an exposing hyperplane $f \in C_p([\alpha^-, \alpha^+])$, such that

$$\mathrm{tr}\int f d\Sigma - G^*(\Sigma) > \mathrm{tr}\int f d\zeta - G^*(\zeta) \tag{5.13}$$

for all $\zeta \neq \Sigma$ (see (7.1)). Let $\Sigma = h_\Sigma \mu_V$ be absolutely continuous with respect to $\mu_V$ with positive definite continuous density $h_\Sigma$ and choose

$$f = \mathbf{1} - h_\Sigma^{-1} .$$

Then $f \in D$ and, by continuity and compactness of $[\alpha^-, \alpha^+]$, there exists a $\gamma > 1$ such that $\gamma f \in D$. Let $\zeta = h_\zeta \mu_V + \zeta^S$ the Lebesgue-decomposition of $\zeta$. Recalling the representation (5.9) and (5.11), inequality (5.13) is satisfied as soon as

$$\int \log\det h_\Sigma \, d\mu_V - p > \int \log\det h_\zeta \, d\mu_V - \mathrm{tr} \int h_\Sigma^{-1} d\zeta \, . \tag{5.14}$$

Since $\mathrm{tr} \int h_\Sigma^{-1} d\zeta \geq \mathrm{tr} \int h_\Sigma^{-1} h_\zeta d\mu_V$, it is enough to prove

$$\int \log\det h_\Sigma \, d\mu_V - p > \int \log\det h_\zeta \, d\mu_V - \mathrm{tr} \int h_\Sigma^{-1} h_\zeta d\mu_V \, . \tag{5.15}$$

This inequality follows from

$$\log\det A - \log\det B > p - \mathrm{tr}(A^{-1}B) \tag{5.16}$$

for Hermitian positive $A \neq B$. In order to prove (5.16), write

$$\log\det A - \log\det B = \sum_{i=1}^{p} \log\left(\lambda_i(A)\lambda_i(B)^{-1}\right) \geq p - \sum_{i=1}^{p} \lambda_i(A^{-1})\lambda_i(B) \tag{5.17}$$

with $\lambda_i(A), \lambda_i(B)$ the eigenvalues of $A, B$ written in any order. If we choose to order the eigenvalues of $A^{-1}$ in decreasing order (i.e. those of $A$ increasing) and those of $B$ in increasing order, it follows from the Hardy-Littlewood rearrangement inequality (see [Mir59]) that

$$\sum_{i=1}^{p} \lambda_i(A^{-1})\lambda_i(B) \leq \mathrm{tr}(A^{-1}B).$$

With this ordering of eigenvalues, (5.17) is strict unless $A, B$ have the same eigenvalues. If all eigenvalues of $A$ and $B$ coincide, then the left hand side of (5.16) is 0, while the right hand side is $p - \mathrm{tr}H$ with $\det H = 1$. The minimum value of $\mathrm{tr}H$ is $p$ which is achieved only for $H = \mathbf{1}$, in which case $A = B$. We get that $\Lambda^*$ is strictly convex at all points $\Sigma = h\mu_V$ with $h$ positive definite and continuous.

It remains to show that the set of exposed points of $G^*$ is dense in $\mathcal{M}_p([\alpha^-, \alpha^+])$. For a given $\Sigma \in \mathcal{M}_p([\alpha^-, \alpha^+])$, we divide $[\alpha^-, \alpha^+]$ by dyadic points into intervals

$$I_{k,n} = [\alpha^- + (k-1)(\alpha^+ - \alpha^-)/2^n, \alpha^- + k(\alpha^+ - \alpha^-)/2^n]$$

and put $\overline{I}_{k,n} = [\alpha^- + (k-1)(\alpha^+ - \alpha^-)/2^n + 2^{-2n}, \alpha^- + k(\alpha^+ - \alpha^-)/2^n - 2^{-2n}]$, i.e., $\overline{I}_{k,n}$ is constructed from $I_{k,n}$ by cutting off subintervals of length $2^{-2n}$. Define $\overline{h}_n$ on $\overline{I}_{k,n}$ as

$$\overline{h}_n\big|_{\overline{I}_{k,n}} \equiv \big((1-2^{-2n})\Sigma(I_{k,n}) + 2^{-2n}\cdot I\big)\,\Sigma_V(I_{k,n})^{-1}$$

and let $h_n$ be the continuous function on $[\alpha^-, \alpha^+]$ obtained by linear interpolation of the step function $\overline{h}_n$. Then $h_n$ is positive definite and continuous on $[\alpha^-, \alpha^+]$ and as in the scalar case, $h_n \cdot \Sigma_V$ converges weakly to $\Sigma$. This concludes the proof of the LDP for $\big(\tilde{\Sigma}^{(N)}_{I(j)}, \lambda^{\pm}_M(j)\big)$.

In order to extend the LDP to the untruncated eigenvalues, note that the LDP for $(\lambda^+(j), \lambda^-(j))$ implies the exponential tightness of the (unrestricted) extremal eigenvalues that is, for every $K > 0$ there exists a $M$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log P\big(\lambda_1^+ > M \text{ or } \lambda_1^- < -M\big) \le -K.$$

In particular,

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

so that as $M \to \infty$, the truncated eigenvalues are exponentially good approximation of the unrestricted ones. Moreover, $(\tilde{\Sigma}^{(N)}_{I(j)}, \lambda^{\pm}_M(j))$ are exponentially good approximations of $(\tilde{\Sigma}^{(N)}_{I(j)}, \lambda^{\pm}(j))$. By Theorem 4.2.16 in [DZ98] $(\tilde{\Sigma}^{(N)}_{I(j)}, \lambda^{\pm}(j))$ satisfies the LDP with speed $n$ and rate function

$$\begin{aligned}\mathcal{I}(\Sigma, x^{\pm}) &= \mathcal{K}(\Sigma_V \,|\, \Sigma) + \operatorname{tr}\Sigma(I) - p + \mathcal{I}_{\lambda^{\pm}}(x^{\pm}) \\ &= \mathcal{K}(\Sigma_V \,|\, \Sigma) + \operatorname{tr}\Sigma(I) - p + \sum_{i=1}^{j} \mathcal{F}^+(x_i^+) + \mathcal{F}^-(x_i^-)\,,\end{aligned}$$

which ends the proof of Theorem 5.2. □

## 5.2 LDP for the projective family

**Theorem 5.3** *For any fixed $j$, the sequence of projected spectral measures $\pi_j(\tilde{\Sigma}^{(N)})$ as elements of $\mathcal{S}_p$ with topology* (5.3) *satisfies the LDP with speed $N$ and rate function*

$$\tilde{\mathcal{I}}_j(\tilde{\Sigma}) = \mathcal{K}(\Sigma_V \,|\, \tilde{\Sigma}) + \operatorname{tr}\tilde{\Sigma}(I) - p + \sum_{i=1}^{N^+\wedge j} \big(\mathcal{F}_V^+(\lambda_i^+) + \operatorname{tr}\Gamma_i^+\big) + \sum_{i=1}^{N^-\wedge j} \big(\mathcal{F}_V^-(\lambda_i^-) + \operatorname{tr}\Gamma_i^-\big)\,.$$

**Proof:** The proof is similar to the proof of Theorem 4.3 in [GNR16b] and we omit the details for the sake of brevity. It is a combination of the LDP in Theorem 5.2 and the LDP of the weights $\frac{1}{N}\Gamma_k = \frac{1}{N} v_k v_k^*$ corresponding to the extreme eigenvalues. Indeed $\Gamma_i \sim \mathrm{LUE}_p(1)$, so by Proposition 3.8 (iii), each individual weight $\frac{1}{N}\Gamma_k$ satisfies the LDP with speed $N$ and rate function

$$\mathcal{I}_3(X) = \begin{cases} \operatorname{tr} X & \text{if } X \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

The independence of the weights and an application of the contraction principle complete then the proof. □

In order to come back to a normalized matrix measure in $\mathcal{S}_{p,1}$, we note that the LDP for $\pi_j(\tilde{\Sigma}^{(N)})$ also implies the joint LDP for

$$\big(\pi_j(\tilde{\Sigma}^{(N)}), \pi_j(\tilde{\Sigma}^{(N)})(\mathbb{R})\big),$$

with the rate function

$$\overline{\mathcal{I}}_j(\tilde{\Sigma}, Z) = \tilde{\mathcal{I}}_j(\tilde{\Sigma})$$

if $\tilde{\Sigma}(\mathbb{R}) = Z$ and $\overline{\mathcal{I}}_j(\tilde{\Sigma}, Z) = \infty$ otherwise. Keeping the weights along the way, we may apply the projective method (the Dawson-Gärtner Theorem, p. 162 in the book of [DZ98]) to the family of projections $(\pi_j(\tilde{\Sigma}^{(N)}), \pi_j(\tilde{\Sigma}^{(N)})(\mathbb{R}))_j$ and get a LDP for the pair $(\tilde{\Sigma}^{(N)}, \tilde{\Sigma}^{(N)}(\mathbb{R}))$ with rate function

$$\overline{\mathcal{I}}(\tilde{\Sigma}, Z) = \sup_j \overline{\mathcal{I}}_j(\tilde{\Sigma}, Z)\,.$$

This rate function equals $+\infty$ unless $\tilde{\Sigma}(\mathbb{R}) = Z$ and in this case is given by

$$\overline{\mathcal{I}}(\tilde{\Sigma}, Z) = \mathcal{K}(\Sigma_V \mid \tilde{\Sigma}) + \operatorname{tr} Z - p + \sum_{i=1}^{N^+} \mathcal{F}(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}(\lambda_i^-). \tag{5.18}$$

We remark that normalizing the matrix measure is not possible unless we keep track of the total mass for any $j$, as the mapping $\tilde{\Sigma} \mapsto \tilde{\Sigma}(\mathbb{R})^{-1/2}\tilde{\Sigma}\tilde{\Sigma}(\mathbb{R})^{-1/2}$ is not continuous in the topology (5.3). However, we may now apply the continuous mapping $(\tilde{\Sigma}, Z) \mapsto Z^{-1/2}\tilde{\Sigma}Z^{-1/2}$ and obtain a LDP for the sequence of measures $\Sigma^{(N)}$ in $\mathcal{S}_{p,1}$. The rate function is

$$\mathcal{I}(\Sigma) = \inf_{\tilde{\Sigma} = Z^{1/2}\Sigma Z^{1/2},\, Z>0} \tilde{\mathcal{I}}(\tilde{\Sigma}) = \inf_{Z>0} \tilde{\mathcal{I}}(Z^{1/2}\Sigma Z^{1/2}).$$

By (5.18), we need to minimize over positive definite $Z \in \mathcal{H}_p$ the function

$$\begin{aligned}
&- \int \log\det\left(\frac{d(Z^{1/2}\Sigma Z^{1/2})}{d\mu_V}\right) d\mu_V + \mathrm{tr} Z - p \\
&= - \int \log\det\left(Z^{1/2}\frac{d\Sigma}{d\mu_V}Z^{1/2}\right) d\mu_V + \mathrm{tr} Z - p \\
&= - \int \log\det\left(\frac{d\Sigma}{d\mu_V}\right) d\mu_V - \log\det Z + \mathrm{tr} Z - p \\
&= - \int \log\det\left(\frac{d\Sigma}{d\mu_V}\right) d\mu_V + \mathcal{I}_2(Z)\,.
\end{aligned}$$

The term $\mathcal{I}_2(Z)$ comes from Lemma 3.4 (ii) with $\gamma = 1$ and attains its minimal value 0 for $Z = \mathbf{1}$. □

We have obtained the LDP claimed in Theorem 3.4 on the subset $\mathcal{S}_{p,1}$ in the topology induced by (5.3). On $\mathcal{S}_{p,1}$ this is stronger than the weak topology and the rate function can be extended to $\mathcal{M}_{p,1}$ by setting $\mathcal{I}(\Sigma) = \infty$ for $\Sigma \notin \mathcal{S}_{p,1}$. This yields Theorem 3.4.

## 5.3 Proof of Remark 3.5

Let $A$ be a measurable subset of $\mathcal{M}_{p,1}$ and set

$$A_N = \left\{ (\lambda, W) \in \mathbb{R}^N \times \mathcal{H}_p^N \,\middle|\, \sum_{k=1}^N W_k \delta_{\lambda_k} \in A \right\}.$$

The LDP for $\Sigma^{(N)}$ with eigenvalue distribution $\mathbb{P}_N^{V_N}$ will follow from the LDP for eigenvalue distribution $\mathbb{P}_N^V$ once we show

$$\limsup_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) \le \limsup_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^{V} \otimes \mathbb{D}_{p,N})(A_N) \tag{5.19}$$

and

$$\liminf_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) \ge \liminf_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^{V} \otimes \mathbb{D}_{p,N})(A_N). \tag{5.20}$$

In fact, this does not require $A$ to be closed or open, respectively. For this, let

$$\Gamma_N^V(A_N) = \iint_{A_N} \prod_{1\le i<j\le N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^N e^{-NV(\lambda_i)} d\lambda \, d\mathbb{D}_{p,N}(W)\,,$$

and define $\Gamma_N^{V_N}(A_N)$ analogously, with $V$ replaced by $V_N$. Since $V_N \geq V$, we have

$$\Gamma_N^{V_N}(A_N) \leq \Gamma_N^V(A_N). \tag{5.21}$$

To get a reverse inequality, let $K_{N,M}$ be the set of $(\lambda, W) \in \mathbb{R}^N \times \mathcal{H}_p^N$, where $V(\lambda_i) \leq M$ for all $i$. Then

$$\Gamma_N^{V_N}(A_N) \geq \Gamma_N^{V_N}(A_N \cap K_{N,M}) \geq \left( \inf_{x:V(x)\leq M} e^{V(x)-V_N(x)} \right)^N \Gamma_N^V(A_N \cap K_{N,M}).$$

Since by assumption $e^{V(x)-V_N(x)}$ converges to 1 uniformly on $\{x|\, V(x) \leq M\}$, this implies

$$\lim_{N\to\infty} \frac{1}{N} \log \frac{\Gamma_N^{V_N}(A_N)}{\Gamma_N^V(A_N)} \geq \lim_{N\to\infty} \frac{1}{N} \log \frac{\Gamma_N^V(A_N \cap K_{N,M})}{\Gamma_N^V(A_N)}. \tag{5.22}$$

If we take now $A = \mathcal{M}_{p,1}$, then $\Gamma_N^V(A_N) = Z_N^V$ and the right hand side of (5.22) becomes

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N^V(\forall\, i :\, V(\lambda_i) \leq M).$$

By the LDP for the extreme eigenvalues, Theorem 5.1, this limit tends to 0 as $M \to \infty$. Together with (5.21), we have shown that for $A = \mathcal{M}_{p,1}$

$$\lim_{N\to\infty} \frac{1}{N} \log \frac{\Gamma_N^{V_N}(A_N)}{\Gamma_N^V(A_N)} = \lim_{N\to\infty} \frac{1}{N} \log \frac{Z_N^{V_N}}{Z_N^V} = 0.$$

Since $(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) = (Z_N^V)^{-1}\Gamma_N^V(A_N)$, the inequality (5.21) leads to the inequality (5.19) and the arguments for (5.22) yield

$$\liminf_{n\to\infty} \frac{1}{N} \log(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) \geq \liminf_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^V \otimes \mathbb{D}_{p,N})(A_N \cap K_{N,M})$$

for any $M \geq 0$. Letting $M \to \infty$, this implies inequality (5.20), as by the LDP for the extreme eigenvalues we have

$$\lim_{M\to\infty} \limsup_{N\to\infty} \frac{1}{N} \log(\mathbb{P}_N^V \otimes \mathbb{D}_{p,N})(K_{N,M}^c) = -\infty.$$

□

# 6 Proof of Theorems 3.6 and 3.7

## 6.1 Hermite block case

The starting point for the proof of Theorem 3.6 is the following block-tridiagonal representation of the Gaussian ensemble. It is a matrix extension of a famous result of Dumitriu and Edelman [DE02].

**Lemma 6.1** *Let $D_k \sim \mathrm{GUE}_p$ and $C_k$ be Hermitian non-negative definite such that $C_k^2 \sim \mathrm{LUE}_p(p(n-k))$ for $k = 1, \ldots, n$ and let all these matrices be independent. Then the $p \times p$ spectral measure of the block-tridiagonal matrix*

$$\mathcal{G}_n = \begin{pmatrix} D_1 & C_1 & & \\ C_1 & D_2 & \ddots & \\ & \ddots & \ddots & C_{n-1} \\ & & C_{n-1} & D_n \end{pmatrix} \tag{6.1}$$

*has the same distribution as the spectral measure of the Hermite ensemble $\mathrm{GUE}_{pn}$.*

**Proof:** Starting from a matrix $X_n$ distributed according to the Hermite ensemble $\mathrm{GUE}_{pn}$, we can construct the tridiagonal matrix $\mathcal{G}_n$ as

$$\mathcal{G}_n = T X_n T^*,$$

where $T$ is unitary and leaves invariant the subspaces spanned by the first $p$ unit vectors. Consequently, the spectral measure of $X_n$ is also the spectral measure of $\mathcal{G}_n$. The transformation $T$ is a composition of unitaries $T_1, \ldots, T_{n-1}$ analogous to the ones used by [DE02]. To illustrate the procedure, we construct the first transformation $T_1$. By $x_{i,j}$ we denotes the $p \times p$ subblock of $X_n$ in position $i, j$, let $\bar{x}_1 = (x_{1,2}, \ldots, x_{1,n})^*$ and $\bar{X} = (x_{i,j})_{2 \le i,j \le n}$. With this notation, $X_n$ can be structured as

$$X_n = \begin{pmatrix} x_{1,1} & \bar{x}_1^* \\ \bar{x}_1 & \bar{X} \end{pmatrix}.$$

Note that the Gaussian distribution implies that all (square) blocks are almost surely invertible. Then, set

$$\begin{aligned} \xi &= [(x_{2,1}^*)^{-1}(\bar{x}_1^* \bar{x}_1)(x_{2,1})^{-1}]^{1/2} x_{2,1} \in \mathbb{M}_{p,p} \\ \Gamma &= (\xi^*, \mathbf{0}, \ldots, \mathbf{0})^* \in \mathbb{M}_{(n-1)p,p} \end{aligned}$$

and define for $Z \in \mathbb{M}_{(n-1)p,p}$ the block-Householder reflection

$$H(Z) = I_{(n-1)p} - 2Z(Z^*Z)^{-1}Z^*.$$

If we set $Z = \Gamma - \bar{x}_1$ one may check that

$$\Gamma^*\Gamma = \xi^*\xi = \bar{x}_1^*\bar{x}_1, \quad \Gamma^*\bar{x}_1 = \xi^* x_{2,1} = x_{2,1}^*\xi = \bar{x}_1^*\Gamma, \quad Z^*Z = -2Z^*\bar{x}_1$$

and $H(Z)\bar{x}_1 = \Gamma$. We extend $H(Z)$ to an operator $\hat{H}$ on $\mathbb{C}^{np}$ leaving the first $p$ unit vectors invariant, which yields

$$\hat{H}X_n\hat{H}^* = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & H(Z) \end{pmatrix} \begin{pmatrix} x_{1,1} & \bar{x}_1^* \\ \bar{x}_1 & \bar{X} \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & H(Z)^* \end{pmatrix} = \begin{pmatrix} x_{1,1} & \Gamma^* \\ \Gamma & H(Z)\bar{X}H(Z)^* \end{pmatrix}.$$

Finally, let $W \in \mathbb{M}_{(n-1)p,(n-1)p}$ be the unitary block-diagonal matrix with the blocks $((\xi^*\xi)^{1/2}\xi^{-1}, \mathbf{1}, \ldots, \mathbf{1})$ on the diagonal and extend $W$ to an operator $\hat{W}$ on $\mathbb{C}^{np}$ as we did with $H(Z)$. Then $T_1 = \hat{W}\hat{H}$ is unitary, leaves the subspace spanned by the first $p$ unit vectors invariant and

$$T_1 X_n T_1^* = \begin{pmatrix} x_{1,1} & \tilde{\Gamma} \\ \tilde{\Gamma}^* & WH(Z)\bar{X}H(Z)^*W^* \end{pmatrix}.$$

with

$$\tilde{\Gamma} = ((\bar{x}_1^*\bar{x}_1)^{1/2}, \mathbf{0}, \ldots, \mathbf{0}).$$

The block $\bar{x}_1^*\bar{x}_1$ is distributed according to $\mathrm{LUE}_p(p(n-1))$ and since the definition of $W$ and $H(Z)$ is independent of $\bar{X}$, the block $WH(Z)\bar{X}H(Z)^*W^*$ is again a matrix of the Gaussian ensemble $\mathrm{GUE}_{p(n-1)}$. The assertion follows then from an iteration of these reflections. □

**Proof of Theorem 3.6:**

By Lemma 6.1, the spectral measure $\Sigma^{(n)}$ is also the spectral measure of the rescaled matrix $\frac{1}{\sqrt{np}}\mathcal{G}_n$. If we consider each block entry of this matrix separately, we are up to a linear change of the speed in the setting of Lemma 3.8. Thus, for any fixed $k$, the block $D_k^{(n)} := D_k/\sqrt{np}$ of the matrix in (6.1) satisfies the LDP in $\mathcal{H}_p$ with speed $pn$ and rate function $\mathcal{I}_1$. Similarly, if we let $C_k^{(n)} = C_k/\sqrt{np}$, then the block $(C_k^{(n)})^2$ satisfies the LDP with speed $pn$ and rate function $\mathcal{I}_2$ or equivalently, $C_k^{(n)}$ satisfies the LDP with speed $np$ and good rate

$$\mathcal{I}_2'(Y) = \mathcal{I}_2(Y^2)$$

if $Y$ is nonnegative definite and $\mathcal{I}_2'(Y) = \infty$ otherwise. Since the block entries are independent, we get a joint LDP for any fixed collection $(D_1^{(n)}, C_1^{(n)}, \ldots, D_k^{(n)})$ with rate given by the corresponding sum of rate functions $\mathcal{I}_1$ and $\mathcal{I}_2'$.

Now, we follow the strategy developed in [GR11] for the scalar case. The random matrix measure $\Sigma^{(n)}$ belongs to $\mathcal{M}_{p,c}^1(\mathbb{R})$. Since the topology $\mathcal{T}_m$ of the convergence of moments on $\mathcal{M}_{p,c}^1(\mathbb{R})$ is stronger than the trace $\mathcal{T}_w$ of the weak topology, it is enough to prove the LDP with respect to $\mathcal{T}_w$.

For each $k > 0$, the subset $X_k$ of matrix probability measures with support in $[-k, k]$ is compact for $\mathcal{T}_m$. Since the extremal eigenvalues satisfy the LDP with speed $N$ and a rate function tending to infinity, we deduce that $\Sigma^{(N)}$ is exponentially tight in $\mathcal{T}_m$.

The mapping

$$m : \mathcal{M}_{p,c}^1(\mathbb{R}) \to \mathcal{H}_p^{\mathbb{N}_0}, \qquad m(\Sigma) := \left( m_k(\Sigma) := \int x^k d\Sigma(x) \right)_{k \geq 1}$$

being a continuous injection, the LDP of $\Sigma^{(N)}$ in $\mathcal{M}_{p,c}^1(\mathbb{R})$ is then a consequence of the following LDP on the sequence of moments and of the inverse contraction principle (see [DZ98] Theorem 4.2.4 and the subsequent Remark (a)).

**Proposition 6.2** *The sequence $(m(\Sigma^{(n)})_n)$ satisfies the LDP in $\mathcal{H}_p^{\mathbb{N}_0}$ with speed $np$ and good rate function $\mathcal{I}_{\mathrm{m}}$ defined as follows. This function is finite in $(m_1, m_2, \dots)$ if and only if there exists a sequence $(B_1, A_1, \dots) \in \mathcal{H}_p^{\mathbb{N}_0}$ with $A_k > 0$, such that*

$$\sum_{k=1}^{\infty} \frac{1}{2} \mathrm{tr}\, B_k^2 + \mathrm{tr} G(A_k^2) < \infty$$

*and such that*

$$(m_r)_{i,j} = \langle e_i J^r e_j \rangle \ , \ i, j = 1, \dots, p, \ r \geq 1 \tag{6.2}$$

*where $J$ is the infinite block Jacobi matrix with blocks $(B_1, A_1, \dots)$ as in* (2.11).

*In that case*

$$\mathcal{I}_{\mathrm{m}} = \sum_{k=1}^{\infty} \frac{1}{2} \mathrm{tr} B_k^2 + \mathrm{tr} G(A_k^2)\,. \tag{6.3}$$

**Proof:** First, as we said in the beginning of this proof, for fixed $k$, $(D_1^{(n)}, C_1^{(n)}, \dots, D_k^{(n)})$ satisfies the LDP in $\mathcal{H}_p^{2k-1}$ with speed $np$ and good rate function

$$\mathcal{I}^{(k)}(D_1, C_1, \dots, D_k) = \sum_{j=1}^{k} \frac{1}{2} \mathrm{tr} D_j^2 + \sum_{j=1}^{k-1} \mathrm{tr} G(C_k^2)\,.$$

If $J$ is the $kp \times kp$ Jacobi matrix build from the blocks $D_1, C_1, \dots, D_k$, then the moments $(m_1(\Sigma^{(n)}), \dots, m_{2k-1}(\Sigma^{(n)}))$ of the spectral measure of $J$ are given by (6.2) and depend continuously on $D_j, C_j$. By the contraction principle, the sequence $(m_1(\Sigma^{(n)}), \dots, m_{2k-1}(\Sigma^{(n)}))$ satisfies the LDP with speed $np$ and good rate function $\mathcal{I}_{\mathrm{m}}^{(k)}$ defined as follows. It is infinite in

$(m_1, \dots, m_{2k-1})$ unless there exist block coefficients $(B_1, A_1, \dots, B_k)$ of the $kp \times kp$ matrix $J$ with $A_k > 0$ such that (6.2) holds. In this case the coefficients are necessarily unique and

$$\mathcal{I}_{\mathrm{m}}^{(k)}(m_1, \dots m_{2k-1}) = \mathcal{I}^{(k)}(B_1, A_1, \dots, B_k).$$

As in the scalar case, we do not consider the even case, since there is no injectivity there.

The Dawson-Gärtner theorem (see [DZ98]) yields the LDP for the whole moment sequence $m(\Sigma^{(n)})$ in $\mathcal{H}_p^{\mathbb{N}_0}$ with good rate

$$\mathcal{I}_{\mathrm{m}}(m_1, \dots) = \sup_{k \geq 1} \mathcal{I}_{\mathrm{m}}^{(k)}(m_1, \dots m_{2k-1}).$$

This supremum is finite if and only if there exists a (unique) sequence $(B_1, A_1, \dots)$ of coefficients satisfying $A_k > 0$ and (6.2). Note that this implies in particular that $(m_1, \dots)$ is the moment sequence of a nontrivial measure $\Sigma$. In this case

$$\begin{aligned}
\mathcal{I}_{\mathrm{m}}(m_1, \dots) &= \sup_{k \geq 1} \mathcal{I}^{(k)}(B_1, A_1, \dots, B_k) \\
&= \sup_k \sum_{j=1}^{k} \frac{1}{2} \mathrm{tr} B_j^2 + \sum_{j=1}^{k-1} \mathrm{tr} G(A_j^2) \\
&= \sum_{k=1}^{\infty} \left( \frac{1}{2} \mathrm{tr} B_k^2 + \mathrm{tr}\, G(A_k^2) \right).
\end{aligned}$$

□

## 6.2 Laguerre block case

The starting point for the proof of Theorem 3.6 is the following block-bidiagonal representation. It is a matrix extension of a famous result of Dumitriu and Edelman [DE02].

**Lemma 6.3** *Let $m \geq n$ and for $k = 1, \dots, n$ let $D_k$ and $C_k$ for $k = 1, \dots, n$ be independent random non-negative definite matrices in $\mathcal{H}_p$ such that*

$$C_k^2 \sim \mathrm{LUE}_p(p(n-k)) \quad , \quad D_k^2 \sim \mathrm{LUE}_p(p(m-k+1))$$

*and define the block matrix*

$$Z_n = \begin{pmatrix} D_1 & 0 & & \\ C_1 & D_2 & \ddots & \\ & \ddots & \ddots & 0 \\ & & C_{n-1} & D_n \end{pmatrix}.$$

*Then the $p \times p$ spectral matrix measure of $L_n = Z_n Z_n^*$ has the same distribution as the spectral matrix measure of a $pn \times pn$ matrix, distributed according to the* $\mathrm{LUE}_{pn}(pm)$ *($m \geq n$).*

**Proof:** We use the construction of the Laguerre ensemble $L_n = W_n W_n^*$, with $W_n$ a $pn \times pm$ matrix with independent complex Gaussian entries. Writing $w_{i,j}$ for the $p \times p$ block of $W_n$ in position $i, j$, let $R$ be a $pm \times pm$ unitary matrix constructed analogously to the matrix $\hat{W}\hat{H}$ in the proof of Lemma 6.1, such that

$$W_n R = \begin{pmatrix} \tilde{w} \\ \widetilde{W} \end{pmatrix}$$

with

$$\tilde{w} = (w_{1,1}, \ldots, w_{1,m})R = \left( \left( \sum_{i=1}^{m} w_{1,i}^* w_{1,i} \right)^{1/2}, 0_{p,p}, \ldots, 0_{p,p} \right).$$

The matrix $R$ can be chosen independently of $w_{i,j}, i \geq 2$ such that the entries of $\widetilde{W}$ are again independent complex Gaussian, independent of $\tilde{w}$. Similarly, write $z_{i,j}$ for the $p \times p$ block of $W_n R$ in position $i, j$ and let $L$ be a $p(n-1) \times p(n-1)$ unitary matrix such that

$$L(z_{2,1}^*, \ldots, z_{n,1}^*)^* = \left( \left( \sum_{i=2}^{n} z_{1,i}^* z_{1,i} \right)^{1/2}, 0_{p,p}, \ldots, 0_{p,p} \right)^*$$

If $\widetilde{L} = I \oplus L$ is the extension of $L$ to an operator on $\mathbb{C}^{pn}$, leaving the subspace of the first $p$ unit vectors invariant, then

$$\widetilde{L} W_n R = \begin{pmatrix} D_1 & 0 & \ldots & 0 \\ C_1 & & & \\ 0 & & L\widetilde{W}R & \\ \vdots & & & \end{pmatrix}.$$

The first blocks satisfy $D_1^2 \sim \mathrm{LUE}_p(pm), C_1^2 \sim \mathrm{LUE}_p(p(n-1))$ and by the invariance of the Gaussian distribution, the entries of $L\widetilde{W}R$ are again Gaussian distributed. Since we started with independent entries, all blocks in $\widetilde{L}W_n R$ are independent. The conjugation by $\widetilde{L}$ leaves the first $p$ eigenvector rows invariant, so $\widetilde{L}L_n\widetilde{L}^* = \widetilde{L}W_n R R^* W_n^* \widetilde{L}^*$ has the same spectral measure as $L_n$. This yields the first step in the reduction, the iterations are straightforward. □

**Proof of Theorem 3.7:**
As in the proof of Theorem 3.6, we start by looking at the individual entries of the tridiagonal

representation of Lemma 6.3, now multiplied by $\frac{1}{p\gamma_n}$. For any $k$, the rescaled block $\frac{1}{p\gamma_n}C_k^2$ satisfies by Lemma 3.8 the LDP with speed $p\gamma_n$ and rate $\mathcal{I}_2$ with $\gamma = \tau$. With the speed $pn$ we would like to consider, $\frac{1}{p\gamma_n}C_k^2$ satisfies then the LDP with rate $\operatorname{tr} G(\tau^{-1}\,\cdot)$ and, taking the square root, $C_k^{(n)} := \frac{1}{\sqrt{p\gamma_n}}C_k$ satisfies the LDP with speed $pn$ and rate function

$$\mathcal{I}_C(C) = \operatorname{tr} G(\tau^{-1}C^2)$$

for $C$ positive definite and $\mathcal{I}_C(C) = \infty$ otherwise. Similarily, if we let $D_k^{(n)} := \frac{1}{\sqrt{p\gamma_n}}D_k$, then $(D_k^{(n)})^2$ satisfies the LDP with speed $p\gamma_n$ and rate function $\mathcal{I}_2$ of Lemma 3.8 with $\gamma = 1$. If we consider the speed $pn$ and the square root $D_k^{(n)}$, this changes the rate to

$$\mathcal{I}_D(D) = \tau^{-1}\operatorname{tr} G(D^2)$$

for $D$ positive definite and $\mathcal{I}_D(D) = \infty$ otherwise.

Then we follow the same way as for the Hermite model. By the independence of the matrices $C_k, D_k$, this yields the LDP for any finite sequence $(D_1^{(n)}, C_1^{(n)}, D_2^{(n)}, \dots, D_k^{(n)}, C_k^{(n)})$ in the sequence space of Hermitian non-negative definite matrices with speed $pn$ and good rate

$$\mathcal{I}_{D,C}(D_1, C_1, \dots, D_k, C_k) = \sum_{j=1}^{k} \tau^{-1}\operatorname{tr} G(D_j^2) + \operatorname{tr} G(\tau^{-1}C_j^2). \tag{6.4}$$

From (2.19), this yields the LDP for $(B_1^{(n)}, A_1^{(n)}, \dots, B_k^{(n)})$ with $k$ fixed. As in the Hermite case, we may conclude a LDP for a finite collection of moments and then for the complete sequence of moments $m(\Sigma^{(n)})$ by application of the Dawson-Gärtner theorem. The resulting good rate function $\mathcal{I}_{\mathrm{m}}$ is finite in $(m_1, \dots) \in \mathcal{H}_p^{\mathbb{N}_0}$ only if $(m_1, \dots)$ is the moment sequence of a nontrivial measure with support in $[0, \infty)$. By the discussion in Section 2.2, this is equivalent to the existence of a sequence of positive definite matrices $D_1, C_1, D_2, \dots$ such that $(m_1, \dots)$ is the moment sequence of the spectral measure of $J = XX^*$ with $X$ as in (2.20). In this case

$$\mathcal{I}_{\mathrm{m}}(m_1, \dots) = \sum_{j=1}^{\infty} \tau^{-1}\operatorname{tr} G(D_j^2) + \operatorname{tr} G(\tau^{-1}C_j^2).$$

We use the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ and $\det(AB) = \det(BA)$ to get

$$\operatorname{tr} G(D_k^2) = \operatorname{tr} G(D_k D_k^*) = \operatorname{tr} G(Z_{2k-1}) = \operatorname{tr} G(\zeta_{2k-1})$$

with $Z_{2k-1}$ as in (2.17) and (2.18), and

$$\operatorname{tr} G(\tau^{-1}C_k^2) = \operatorname{tr}(\tau^{-1}C_k C_k^*) = \operatorname{tr}(\tau^{-1}Z_{2k}) = \operatorname{tr}(\tau^{-1}\zeta_{2k}).$$

So the value of the rate function does not depend on the unitary matrices $\sigma_n$ and $\tau_n$ in the construction of $D_k, C_k$, but only on the matrices $\zeta_k$, which in particular are uniquely determined by $(m_1, \dots)$. The inverse contraction principle implies then the LDP for the spectral measure $\Sigma^{(n)}$. □

# 7 Appendix: Extension of Baldi's theorem and Bryc's lemma

In this part we prove a theorem which combines a LDP with a convex rate function and a LDP with a non-convex one. It is one of the key tool for the statements in Section 4 in [GNR16b] and it will be used in [GNR16a]. The first LDP deals with a random spectral measure restricted to the support of the equilibrium measure and the second LDP deals with a subset of outliers.

To give the theorem in a general setting, assume that $\mathcal{X}$ and $\mathcal{Y}$ are Hausdorff topological vector spaces. Let $\mathcal{X}^*$ be the topological dual of $\mathcal{X}$ and equip $\mathcal{X}$ with the weak topology. We denote by $C_b(\mathcal{Y})$ the set of all bounded continuous functions $\varphi : \mathcal{Y} \to \mathbb{R}$. A point $x \in \mathcal{X}$ is called an exposed point of a function $F$ on $\mathcal{X}$, if there exists $x^* \in \mathcal{X}^*$ (called an exposing hyperplane for $x$) such that

$$F(x) - \langle x^*, x\rangle < F(z) - \langle x^*, z\rangle \tag{7.1}$$

for all $z \neq x$.

## 7.1 Some classical results in large deviations

Let us recall two well known results in the theory of large deviations, which have to be combined carefully in order to get our general theorem. The first one is the inverse of Varadhan's lemma (Theorem 4.4.2 in [DZ98]), the second one is a version of the so-called Baldi's theorem (Theorem 4.5.20 in [DZ98]). The latter differs from the version in [DZ98] in a straightforward condition to identify the rate function, which was applied for instance in [GRZ99] (see also [DG07]). The proof of our Theorem 7.3 will be quite similar to the proof of these two classical theorems.

**Theorem 7.1 (Bryc's Inverse Varadhan Lemma)** *Suppose that the sequence $(Y_n)$ of random variables in $\mathcal{Y}$ is exponentially tight and that the limit*

$$\Lambda(\varphi) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} e^{n\varphi(Y_n)}$$

*exists for every $\varphi \in \mathcal{C}_b(\mathcal{Y})$. Then $(Y_n)$ satisfies the LDP with the good rate function*

$$\mathcal{I}(y) = \sup_{\varphi \in \mathcal{C}_b(\mathcal{Y})} \{\varphi(y) - \Lambda(\varphi)\}\,.$$

*Furthermore, for every $\varphi \in \mathcal{C}_b(\mathcal{Y})$,*

$$\Lambda(\varphi) = \sup_{y \in \mathcal{Y}} \{\varphi(y) - \mathcal{I}(y)\}\,.$$

**Theorem 7.2 (A version of Baldi's Theorem)** *Suppose that the sequence $(X_n)$ of random variables in $\mathcal{X}$ is exponentially tight and that*

1. *There is a set $D \subset \mathcal{X}^*$ and a function $G_X : D \to \mathbb{R}$ such that for all $x^* \in D$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp\left(n\langle x^*, X_n\rangle\right) = G_X(x^*)\,; \tag{7.2}$$

2. *The set $\mathcal{F}$ of exposed points $x$ of*

$$G_X^*(x) = \sup_{x^* \in D} \{\langle x^*, x\rangle - G_X(x^*)\}$$

*with an exposing hyperplane $x^*$ satisfying $x^* \in D$ and $\gamma x^* \in D$ for some $\gamma > 1$, is dense in $\{G_X^* < \infty\}$.*

*Then $(X_n)$ satisfies the LDP with good rate function $G_X^*$.*

## 7.2 A general theorem

Our extension is the following combination of the two above theorems. The main point is that the rate function does not need to be convex, but we still only need to control linear functionals of $X_n$.

**Theorem 7.3** *Assume that $X_n \in \mathcal{X}$ and $Y_n \in \mathcal{Y}$ are defined on the same probabilistic space and that the two sequences $(X_n)$ and $(Y_n)$ are exponentially tight. Assume further that*

1. *There is a set $D \subset \mathcal{X}^*$ and functions $G_X : D \to \mathbb{R}$, $J : C_b(\mathcal{Y}) \to \mathbb{R}$ such that for all $x^* \in D$ and $\varphi \in C_b(\mathcal{Y})$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp\left(n\langle x^*, X_n\rangle + n\varphi(Y_n)\right) = G_X(x^*) + J(\varphi)\,; \tag{7.3}$$

2. *The set $\mathcal{F}$ of exposed points $x$ of*

$$G_X^*(x) = \sup_{x^* \in D} \{\langle x^*, x\rangle - G_X(x^*)\},$$

*with an exposing hyperplane $x^*$ satisfying $x^* \in D$ and $\gamma x^* \in D$ for some $\gamma > 1$, is dense in $\{G_X^* < \infty\}$.*

*Then, the sequence $(X_n, Y_n)$ satisfies the LDP with speed $n$ and good rate function*

$$\mathcal{I}(x, y) = G_X^*(x) + \mathcal{I}_Y(y)\,,$$

*where*

$$\mathcal{I}_Y(y) = \sup_{\varphi \in C_b(\mathcal{Y})} \{\varphi(y) - J(\varphi)\}.$$

Let us note that in view of Varadhan's Lemma we have

$$J(\varphi) = \sup_{y \in \mathcal{Y}} \{\varphi(y) - \mathcal{I}_Y(y)\}.$$

**Proof:**

**Upperbound**: The proof follows the lines of the proof of part (b) of Theorem 4.5.3 in [DZ98]. Note that since the sequence $(X_n, Y_n)$ is exponentially tight it suffices to show the upper bound for compact sets.

**Lowerbound**: As usual, it is enough to consider a neighbourhood $\Delta_1 \times \Delta_2$ of $(x, y)$ where $\mathcal{I}(x, y) < \infty$. Take $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}((X_n, Y_n) \in \Delta_1 \times \Delta_2)$ and get a lower bound tending to $\mathcal{I}(x, y)$ when the size of the neighbourhood tends to zero. Actually, due to the density assumption 2. it is enough to study the lower bound of $\mathbb{P}(X_n \in \Delta_1, Y_n \in \Delta_2)$ when $x \in \mathcal{F}$ and $\mathcal{I}_Y(y) < \infty$.

As in [DZ98] (Proof of Lemma 4.4.6), let $\varphi : \mathcal{Y} \to [0, 1]$ be a continuous function, such that $\varphi(y) = 1$ and $\varphi$ vanishes on the complement $\Delta_2^c$ of $\Delta_2$. For $m > 0$, define $\varphi_m := m(\varphi - 1)$. Note that

$$J(\varphi_m) \geq -\mathcal{I}_Y(y)\,.$$

We have

$$\mathbb{P}(X_n \in \Delta_1, Y_n \in \Delta_2) = \mathbb{E}\left[1_{\{X_n \in \Delta_1\}} 1_{\{Y_n \in \Delta_2\}} e^{n\langle x^*, X_n\rangle + n\varphi_m(Y_n)} e^{-n\langle x^*, X_n\rangle - n\varphi_m(Y_n)}\right]\,.$$

Now $-\varphi_m \geq 0$ and on $\Delta_1$, $-\langle x^*, X_n\rangle \geq -\langle x^*, x\rangle - \delta$ for a $\delta > 0$, so that

$$\mathbb{P}(X_n \in \Delta_1, Y_n \in \Delta_2) \geq \mathbb{E}\left[1_{\{X_n \in \Delta_1\}} 1_{\{Y_n \in \Delta_2\}} e^{n\langle x^*, X_n\rangle + n\varphi_m(Y_n)}\right]\, e^{-n\langle x^*, x\rangle - n\delta}\,. \tag{7.4}$$

Denoting

$$\ell_n = \frac{1}{n}\log \mathbb{E}e^{n\langle x^*, X_n\rangle} \ , \ \mathcal{L}_n := \frac{1}{n}\log \mathbb{E}e^{n\langle x^*, X_n\rangle + n\varphi_m(Y_n)}$$

and $\widetilde{\mathbb{P}}$ the new probability on $\mathcal{X}\times\mathcal{Y}$ such that

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = e^{n\langle x^*, X_n\rangle + n\varphi_m(Y_n) - n\mathcal{L}_n},$$

we get

$$\mathbb{P}(X_n \in \Delta_1, Y_n \in \Delta_2) \geq \widetilde{\mathbb{P}}(X_n \in \Delta_1, Y_n \in \Delta_2) e^{-n\langle x^*, x\rangle - n\delta + n\mathcal{L}_n}. \tag{7.5}$$

For the exponential term we have

$$\liminf_{n\to\infty} \frac{1}{n}\log e^{-n\langle x^*, x\rangle - n\delta + n\mathcal{L}_n} \geq \langle x^*, x\rangle - \delta + G_X(x^*) + J(\varphi_m) \geq -G_X^*(x) - \mathcal{I}_Y(y) - \delta. \tag{7.6}$$

We may choose $\delta$ arbitrarily small by choosing $\Delta_1$ sufficiently small, so that it will be enough to prove that

$$\widetilde{\mathbb{P}}(X_n \in \Delta_1, Y_n \in \Delta_2) \xrightarrow[n\to\infty]{} 1 \tag{7.7}$$

or equivalently, that

$$\widetilde{\mathbb{P}}(X_n \in \Delta_1^c) + \widetilde{\mathbb{P}}(Y_n \in \Delta_2^c) \xrightarrow[n\to\infty]{} 0\,. \tag{7.8}$$

For the first term, note that under $\widetilde{\mathbb{P}}$ the moment generating function of $X_n$ satisfies

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log \widetilde{\mathbb{E}}[e^{n\langle z^*, X_n\rangle}] &= \lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}[e^{n\langle z^*+x^*, X_n\rangle + \varphi_m(Y_n) - n\mathcal{L}_n}] \\
&= G_X(z^*+x^*) + J(\varphi_m) - G_X(x^*) - J(\varphi_m) \\
&= G_X(z^*+x^*) - G_X(x^*) \\
&=: \widetilde{G}_X(z^*),
\end{aligned}$$

for $z^* \in \widetilde{D} := \{z^* : x^* + z^* \in D\}$. We may then follow the argument on p.159-160 in [DZ98] (as an auxiliary result in their proof of the lower bound). Using that $x^* \in D$ is an exposing hyperplane, we get

$$\limsup_{n\to\infty} \frac{1}{n}\log \widetilde{\mathbb{P}}(X_n \in \Delta_1^c) < 0.$$

Considering the second term in (7.8), we have, on $\Delta_2^c$

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = e^{-nm + n\langle x^*, X_n\rangle - n\mathcal{L}_n}$$

so that

$$\widetilde{\mathbb{P}}(Y_n \in \Delta_2^c) \leq e^{-nm+n\ell_n - n\mathcal{L}_n}\,.$$

Taking the logarithm, this implies

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \widetilde{\mathbb{P}}(Y_n \in \Delta_2^c) &\leq -m + G_X(x^*) - G_X(x^*) - J(\varphi_m) \\ &= -m - \sup_{z\in\mathcal{Y}}\{\varphi_m(z) - \mathcal{I}_Y(z)\} \leq -m + \mathcal{I}_Y(y)\end{aligned}$$

which tends to $-\infty$ when $m \to \infty$.

To summarize, we have proved (7.8), i.e. (7.7), which with (7.5) and (7.6) gives

$$\lim_{\Delta_1 \downarrow x, \Delta_2 \downarrow y} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n \in \Delta_1, Y_n \in \Delta_2) \geq -G_X^*(x) - \mathcal{I}_Y(y)\,,$$

which leads to the lower bound of the LDP. □

**Remark 7.4** *In Section 4 of [GNR16b], and similar to Section 5.1 of the present paper, we studied the joint moment generating function of* $(\tilde{\mu}_I^{(n)}, \lambda_M^{\pm})$. *For* $s \in \mathbb{R}^{2j}$ *we introduced*

$$G_n(f,s) = \mathbb{E}\left[\exp\left\{n \int f\, d\tilde{\mu}_I^{(n)} + n\langle s, \lambda_M^{\pm}\rangle\right\}\right],$$

*and proved that for all* $f$ *such that* $\log(1-f)$ *is continuous and bounded (and all* $s \in \mathbb{R}^{2j}$*),*

$$\lim_{n\to\infty} \frac{1}{n} \log G_n(f,s) = G(f) + H(s)\,. \tag{7.9}$$

*Actually* $H$ *is the Legendre dual of* $\mathcal{I}_{M,\lambda^{\pm}}$*, i.e.*

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

*However, the rate* $\mathcal{I}_{M,\lambda^{\pm}}$ *might be non-convex (when* $V$ *is not convex) and hence the dual* $H^*$ *is not strictly convex on a dense set, and then the Assumption 2 of Theorem 7.2 may not be verified. The convergence in* (7.9) *is therefore not enough to conclude the joint LDP for* $(\tilde{\mu}_I^{(n)}, \lambda_M^{\pm})$ *directly from the Theorem 7.2 above. A complete proof of Theorem 3.1 in [GNR16b] needs actually an application of Theorem 7.3.*

# References

[AGZ10] G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices.* Cambridge University Press, Cambridge, 2010.

[BSZ16] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and sum rules for spectral theory - A pedagogical approach. arXiv preprint arXiv:1608.01467, 2016.

[Daw77] A. P. Dawid. Spherical matrix distributions and a multivariate model. *J. R. Stat. Soc., Ser. B*, 39:254–261, 1977.

[DE02] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.

[DG07] H. Dette and F. Gamboa. Asymptotic properties of the algebraic moment range process. *Acta Mathematica Hungarica*, 116:247–264, 2007.

[DGK78] P. Delsarte, Y.V. Genin, and Y.G. Kamp. Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits and Systems*, pages 149–160, 1978.

[DKS10] D. Damanik, R. Killip, and B. Simon. Perturbations of orthogonal polynomials with periodic recursion coefficients. *Ann. of Math.*, 171(3):1931–2010, 2010.

[DPS08] D. Damanik, A. Pushnitski, and B. Simon. The analytic theory of matrix orthogonal polynomials. *Surv. Approx.Theory*, 4:1–85, 2008.

[DS02] 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.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.

[GNR16a] F. Gamboa, J. Nagel, and A. Rouault. Sum rules and large deviations for spectral measures on the unit circle. arXiv preprint arXiv: 1604.06934, 2016.

[GNR16b] F. Gamboa, J. Nagel, and A. Rouault. Sum rules via large deviations. *J. Funct. Anal.*, 270(2):509 – 559, 2016.

[GNRW12] F. Gamboa, J. Nagel, A. Rouault, and J. Wagener. Large deviations for random matricial moment problems. *J. Multivariate Anal.*, 106:17–35, 2012.

[GR11] F. Gamboa and A. Rouault. Large deviations for random spectral measures and sum rules. *Applied Mathematics Research eXpress*, 2011(2):281–307, 2011.

[GR14] F. Gamboa and A. Rouault. Operator-valued spectral measures and large deviations. *J. Statist. Plann. Inference*, 154:72–86, 2014.

[GRZ99] F. Gamboa, A. Rouault, and M. Zani. A functional large deviation principle for quadratic forms of Gaussian stationary processes. *Stat. and Probab. Letters*, 43:299–308, 1999.

[Joh98] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[Kil07] R. Killip. Spectral theory via sum rules. In *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, volume 76 of *Proc. Sympos. Pure Math.*, pages 907–930. Amer. Math. Soc., Providence, RI, 2007.

[KS03] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math.*, 158(1):253–321, 2003.

[Mir59] L. Mirsky. On the trace of matrix products. *Math. Nachr.*, 20:171–174, 1959.

[MS71] V. Mandrekar and H. Salehi. On singularity and Lebesgue type decomposition for operator-valued measures. *J. Multivariate Anal.*, 1(2):167–185, 1971.

[Roc71] R.T. Rockafellar. Integrals which are convex functionals, II. *Pacific J. Math.*, 39(2):439–469, 1971.

[RR68] J.B. Robertson and M. Rosenberg. The decomposition of matrix-valued measures. *Michigan Math. J*, 15:353–368, 1968.

[Sim05a] B. Simon. *Orthogonal polynomials on the unit circle. Part 1*, volume 54 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2005. Classical theory.

[Sim05b] 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.

[Sim11] B. Simon. *Szegő's theorem and its descendants*. M. B. Porter Lectures. Princeton University Press, Princeton, NJ, 2011.