

Sum rules and large deviations for spectral matrix measures

FABRICE GAMBOA¹, JAN NAGEL² and ALAIN ROUAULT³

¹*Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse Cedex 9, France.
E-mail: gamboa@math.univ-toulouse.fr*

²*Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands. E-mail: j.h.nagel@tue.nl*

³*Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035-Versailles Cedex, France. E-mail: alain.rouault@uvsq.fr*

In the paradigm of random matrices, one of the most classical object under study is the empirical spectral distribution. This random measure is the uniform distribution supported by the eigenvalues of the random matrix. In this paper, we give large deviation theorems for another popular object built on Hermitian random matrices: the spectral measure. This last probability measure is a random weighted version of the empirical spectral distribution. The weights involve the eigenvectors of the random matrix. We have previously studied the large deviations of the spectral measure in the case of scalar weights. Here, we will focus on matrix valued weights. Our probabilistic results lead to deterministic ones called “sum rules” in spectral theory. 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. By using only probabilistic tools of large deviations, we extend the sum rules to the case of Hermitian matrix-valued measures.

Keywords: large deviations; matrix-valued measures; orthogonal matrix polynomials; random matrices; sum rules

1. Introduction

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 $A \subset \mathbb{R}$. Here, \mathbb{M}_p is the set of all $p \times p$ matrices with complex entries and $\mathcal{H}_p \subset \mathbb{M}_p$ is the subset of Hermitian matrices. Further, if for any A , $\Sigma(A)$ is a nonnegative definite matrix we say that Σ is nonnegative. We denote by $\mathcal{M}_p(T)$ the set of $p \times p$ nonnegative matrix measures with support in $T \subset \mathbb{R}$, that is, all the scalar measures $\Sigma_{i,j}$ have support in T . Further, $\mathcal{M}_{p,1}(T)$ is the subset of $\mathcal{M}_p(T)$ of normalized measures Σ satisfying $\Sigma(T) = \mathbf{1}$, where $\mathbf{1}$ is the $p \times p$ identity matrix.

Notice that a matrix measure Σ may be identified with the collection of scalar measures $x^* \Sigma x$, with $\|x\| = 1$. Hence, Σ_n converges weakly to Σ if and only if $x^* \Sigma_n x$ converges weakly to Σ for all unit vectors x .

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 $D = \text{diag}(\lambda_i)$ contains the eigenvalues of X and U is the matrix formed by an orthonormal basis of eigenvectors. Fix $1 \leq p \leq N$ and assume that the system (e_1, \dots, e_p) of the p first vectors of

the canonical basis of \mathbb{C}^N is cyclic, that is, $\text{Span}\{A^k e_i, k \geq 0, i \leq p\} = \mathbb{C}^N$. Then, there exists a unique spectral matrix measure $\Sigma^X \in \mathcal{M}_p(\mathbb{R})$ associated with $(X; e_1, \dots, e_p)$ supported by the spectrum of X , such that for all $k > 0$ and $1 \leq i, j \leq p$

$$\langle e_i, X^k e_j \rangle = \int_{\mathbb{R}} x^k d\Sigma_{i,j}^X(x). \tag{1.1}$$

For $j = 1, \dots, N$, let $u_j := (U_{i,j})_{i=1, \dots, p}$ be the j th truncated column of U . Then

$$\Sigma_p^X(dx) = \sum_{j=1}^N u_j u_j^* \delta_{\lambda_j}(dx), \tag{1.2}$$

where δ_a is the Dirac measure in a . That is, $\{\lambda_1, \dots, \lambda_N\}$ is the support of Σ^X and $u_1 u_1^*, \dots, u_N u_N^*$ are the weights. Furthermore, as U is unitary we have $\sum_j u_j u_j^* = \mathbf{1}$ so that $\Sigma^X \in \mathcal{M}_{p,1}(T)$.

In this paper, we deal with random spectral matrix measures. We draw a random matrix X in \mathcal{H}_N (with $N \geq p$) and consider the spectral measure Σ^X associated with $(X; e_1, \dots, e_p)$. The random matrix X is sampled in \mathcal{H}_N from the distribution

$$\frac{1}{Z_V^N} e^{-\text{tr} NV(X)} dX. \tag{1.3}$$

V is a confining potential and dX is the Lebesgue measure on \mathcal{H}_N . The eigenvalues of X behave as a Coulomb gas with distribution

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

Here, $d\lambda$ denotes the Lebesgue measure on \mathbb{R}^N . Moreover, the distribution of X is invariant under unitary conjugations, such that the matrix of eigenvectors of X follows the Haar distribution in the set of $N \times N$ unitary matrices, independent of its eigenvalues. In particular, the system (e_1, \dots, e_p) is almost surely cyclic and the distribution of the spectral measure does not depend on this choice of unit vectors. Therefore, we may omit the reference to these vectors in the definition of Σ^X . The most popular models are the Gaussian Ensemble, corresponding to the potential $V(x) = x^2/2$, the Laguerre Ensemble, with $V(x) = \tau^{-1}x - (\tau^{-1} - 1) \log x$ ($x > 0, 0 < \tau \leq 1$) and the Jacobi Ensemble with $V(x) = -\kappa_1 \log x - \kappa_2 \log(1 - x)$ ($x \in (0, 1), \kappa_1, \kappa_2 > 0$).

Under convenient assumptions on V , as N goes to infinity, the random spectral matrix measure Σ^X converges to some equilibrium matrix measure depending on V . In this paper, we provide precise results on the rate of convergence. Indeed, we show that this random object satisfies a large deviation principle (LDP) and compute precisely the rate function. To be self contained, we recall in Section 3.2 the definition and useful facts on LDP. The rate function of this LDP is not obvious. It involves a matrix extension of the reversed Kullback–Leibler information with respect to the equilibrium matrix measure (see equation (2.5) and Theorem 5.2) and a contribution of the

outlying eigenvalues. Furthermore, the probabilistic frame of large deviations allows to obtain important results in mathematical analysis. Namely, it allows to show sum rules. Let us give in a nutshell what is a sum rule.

One of the most popular sum rule is the Verblunsky’s form of Szegő’s theorem for a probability measure μ on the unit circle \mathbb{T} . The Verblunsky coefficients $(\alpha_k)_{k \geq 0}$ of μ are complex number lying in the unit disk related to the orthogonal polynomials in $L^2(\mu)$. The Szegő–Verblunsky sum rule¹ may be written as

$$\mathcal{K}\left(\frac{d\theta}{2\pi} \middle| \mu\right) = - \sum_{k \geq 0} \log(1 - |\alpha_k|^2), \tag{1.4}$$

where \mathcal{K} is the relative entropy or Kullback–Leibler information, defined by

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

Identity (1.4) connects an entropy with a functional of the recursion coefficients, and remains one of the most important results of the theory of orthogonal polynomials on the unit circle (OPUC) (see [24] 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 [20]. The reference measure is no more the uniform probability but the semicircle distribution

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

The right-hand side of the sum rule involves now the Jacobi coefficients used in the construction of the orthogonal polynomial. Further, in the left-hand side appears an extra term corresponding to a contribution of isolated masses of μ outside $[-2, 2]$ (bound states).

In the previous paper [12], 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

$$\text{MP}(\tau)(dx) = \frac{\sqrt{(\tau^+ - x)(x - \tau^-)}}{2\pi \tau x} \mathbb{1}_{(\tau^-, \tau^+)}(x) dx, \tag{1.7}$$

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

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

¹ See the preface of [24].

where C_{u^-,u^+} is the normalizing constant. Recently, Breuer et al. [2] published a paper exposing the method for a non probabilistic audience and applied the probabilistic approach to tackle half of the Lukic conjecture [3].

Besides, known extensions of Szegő’s theorem [7] and of the Killip–Simon sum rule [4] are available in the context of matrix-valued measures and the Matrix Orthogonal Polynomials 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 matrix framework. We give here a positive answer. For the MOPUC context, see [17] and [14]. 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 sum rule and also to state a matrix version of the sum rule relative to the Marchenko–Pastur distribution.

Let us explain the main features of our method. As mentioned by several authors ([19], [26]), the main characteristic of the above sum rules is that both sides are nonnegative (possibly infinite) functionals. We will identify them as rate function of large deviations for some random measures. Roughly speaking, that means that this sequence of random measures converges to a deterministic limit exponentially fast and the probability 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 a rate functions yields the equality $I_A = I_B$, i.e., the sum rule.

For a Hermitian matrix X of size $N \times N$, let us come back to the spectral measure defined in (1.1) and (1.2). If we assume further that $N = pn$ for some positive integer n then it is possible to build a block tridiagonal matrix

$$J_n = \begin{pmatrix} B_1 & A_1 & & & \\ A_1^* & B_2 & \ddots & & \\ & \ddots & \ddots & & \\ & & & A_{n-1} & \\ & & & A_{n-1}^* & B_n \end{pmatrix} \tag{1.8}$$

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, [24]). The general case is more complicated and requires technical tools from the theory of MOPRL (see [25]). In Section 2, we will recall the construction of such tridiagonal representations.

As a result, we have two encodings of Σ^X : (1.2) and (1.8). 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 classical ensembles have very nice properties [10]. 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 the so-called χ distributions [10]. We will give in Lemmas 6.1 and 6.2 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 in terms of the blocks involved in the Jacobi representation. Further, the uniqueness of a rate function leads to our two main Theorems 2.1 and 2.2 that are sum rules. Theorem 2.1 is the Damanik–Killip–Simon sum rule which has been proved in [4] by strong analysis tools. For $p = 1$, it has been proved earlier in [20]. We recover the result of [4] by using only probabilistic arguments. Up to our knowledge,

our Theorem 2.2 is new and is the matrix extension of the one we have obtained in [12] for $p = 1$. As it is the case for the Damanik–Killip–Simon sum rule, we obtain in Corollary 2.2 consequences for perturbations of positive semidefinite Jacobi operators.

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. The proofs of the LDP’s are in Sections 5 and 6. Finally, some preliminaries on matrix measures and technical proofs are deferred to a supplementary file [13].

2. Block Jacobi matrices

As for scalar measures, we may also consider a representation of a matrix measure Σ by a sequence of Jacobi- or recursion coefficients. In the matrix case, they are now the coefficients appearing in the recursion of orthogonal matrix polynomials. For details of this construction, we refer to the supplementary file [13].

First, suppose that Σ has infinite support. We still have the property that 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 of orthonormal matrix polynomials by the matrix

$$J = \begin{pmatrix} B_1 & A_1 & & \\ A_1^* & B_2 & \ddots & \\ & \ddots & \ddots & \\ & & & \ddots \end{pmatrix} \tag{2.1}$$

with B_k Hermitian and A_k non-singular the coefficients of the polynomial recursion. Moreover the measure Σ is again the spectral measure of the matrix J defined as in (1.1) (Theorem 2.11 of [5]). Let us remark that to each Σ corresponds a whole equivalence class of Jacobi coefficients, but there is exactly one representative such that all A_k are Hermitian positive definite (Theorem 2.8 in [5]) and any other choice A'_k, B'_k can be written as

$$A'_k = \sigma_{k-1}^* A_k \sigma_k, \quad B'_k = \sigma_{k-1}^* B_k \sigma_{k-1} \tag{2.2}$$

with unitary $\sigma_k, k \geq 1$ and $\sigma_0 = \mathbf{1}$.

For a finite dimensional matrix $X \in \mathcal{H}_N$ of dimension $N = pn$, the spectral matrix measure of Σ^X is supported by at most N points and is in particular not nontrivial. However, if there

are actually N distinct eigenvalues of X , the recursion coefficients A_1, \dots, A_{n-1} and B_1, \dots, B_n may still be defined for Σ^X and there is a Jacobi matrix J_n as in (1.8), such that $\Sigma^X = \Sigma^{J_n}$.

When the measures are supported by $[0, \infty)$, there is a specific form of the Jacobi coefficients following from a decomposition developed in [9]. This leads to a particularly interesting parametrization, which will be crucial for the Laguerre ensemble. Since it is also of independent interest, we formulate it as a lemma.

Lemma 2.1. *Suppose Σ is a matrix measure with infinite support in $[0, \infty)$. Then, there exist non-singular $D_k, C_k \in \mathbb{M}_p$, such that the Jacobi operator J of Σ can be decomposed as $J = ZZ^*$, where Z is the bidiagonal matrix*

$$Z = \begin{pmatrix} D_1 & \mathbf{0} & & \\ C_1 & D_2 & \ddots & \\ & \ddots & \ddots & \\ & & & \ddots \end{pmatrix}. \tag{2.3}$$

Moreover, if D'_k, C'_k is another sequence defining the same spectral measure, then there are unitary matrices $\sigma_k, \tau_k \in \mathbb{M}_p$, such that

$$D'_k = \sigma_k D_k \tau_k, \quad C'_k = \sigma_{k+1} C_k \tau_k \tag{2.4}$$

and there is a unique decomposition such that D_k, C_k are Hermitian and positive definite.

Suppose Σ is a quasi scalar measure, that is if $\Sigma = \sigma \cdot \mathbf{1}$ with σ a scalar measure and Π is a positive matrix measure with Lebesgue decomposition

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

where h has values in the set of positive semi-definite $p \times p$ Hermitian matrices and the matrix measure Π^s is singular with respect to Σ . We extend the definition (1.5) by

$$\mathcal{K}(\Sigma | \Pi) := - \int \log \det h(x)\sigma(dx). \tag{2.5}$$

We remark that it is possible to rewrite the above quantity in the flavour of Kullback–Leibler information (or relative entropy) with the notation of [21] or [22], see also Corollary 8 in [17].

2.1. Sum rules

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

$$\Sigma = \Sigma_I + \sum_{i=1}^{N^+} \Gamma_i^+ \delta_{\lambda_i^+} + \sum_{i=1}^{N^-} \Gamma_i^- \delta_{\lambda_i^-}, \tag{2.6}$$

where $\text{supp}(\Sigma_I) \subset I = [\alpha^-, \alpha^+]$, $N^-, N^+ \in \mathbb{N} \cup \{\infty\}$, $\Gamma_i^\pm \in \mathcal{H}_p$ are of rank 1 and

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

If N^- (resp. N^+) is infinite, then λ_j^- converges towards α^- (resp. λ_j^+ converges to α^+). Note that we count atoms outside of $[\alpha^-, \alpha^+]$ multiple times if their respective weight has rank larger than 1 and we assume some order on the rank 1 decompositions. Further, we define $\mathcal{S}_{p,1} = \mathcal{S}_{p,1}(\alpha^-, \alpha^+) := \{\Sigma \in \mathcal{S}_p(\alpha^-, \alpha^+) \mid \Sigma(\mathbb{R}) = \mathbf{1}\}$.

2.1.1. *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 [20] and [26], Chapter 3). In the more general case $p \geq 2$, it gives two forms for the divergence with respect to

$$\Sigma_{\text{SC}} = \text{SC} \cdot \mathbf{1}, \tag{2.7}$$

where SC is defined in (1.6) and supported by $[\alpha^-, \alpha^+] = [-2, 2]$. We refer to [4], formula (10.4) and [26], formula (4.6.13) for this matrix sum rule. The block Jacobi matrix associated with Σ_{SC} has entries

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

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

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

and $\mathcal{F}_H^-(x) = \mathcal{F}_H^+(-x)$. 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

$$\text{tr } G(X) = \text{tr } X - \log \det X - p. \tag{2.8}$$

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

Theorem 2.1. *Let $\Sigma \in \mathcal{M}_{p,1}(\mathbb{R})$ be a spectral measure with Jacobi matrix (2.1). If $\Sigma \in \mathcal{S}_{p,1}(-2, 2)$, then*

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

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 $\text{tr } G(\sigma AA^* \sigma^*) = \text{tr } G(AA^*)$ for any unitary σ , the value of the right hand side in Theorem 2.1 is independent of the choice of σ_n 's in (2.2). As described in the Introduction, an important consequence of a sum rule are equivalent conditions for finiteness of both sides. It is a *gem*, as defined on p.19 of [26]. The following statement is the gem implied by Theorem 2.1, it can also be found in Simon's book [26], combining Theorem 4.6.1 and Theorem 4.6.3 therein.

Corollary 2.1. *Assume the Jacobi matrix J is nonnegative definite and let Σ be the spectral measure of J . Then*

$$\sum_{n=1}^{\infty} \text{tr}((A_n A_n^*)^{1/2} - \mathbf{1})^2 + \text{tr } B_n^2 < \infty \tag{2.9}$$

if and only if

1. $\Sigma \in \mathcal{S}_{p,1}(-2, 2)$.
2. $\sum_{i=1}^{N^+} (\lambda_k - 2)^{3/2} + \sum_{i=1}^{N^-} (|\lambda_i| - 2)^{3/2} < \infty$.
3. If Σ admits the decomposition

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

where $d\Sigma_s$ is singular with respect to dx and f is an integrable function of $[-2, 2]$ to the set of nonnegative Hermitian $p \times p$ matrices, then

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

2.1.2. *Our new sum rule: The Laguerre case*

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

$$\Sigma_{\text{MP}(\tau)} = \text{MP}(\tau) \cdot \mathbf{1}, \tag{2.10}$$

where MP is defined in (1.7) and supported by $[\alpha^-, \alpha^+]$ with $\alpha^\pm = \tau^\pm = (1 \pm \sqrt{\tau})^2$. The block Jacobi matrix associated with $\Sigma_{\text{MP}(\tau)}$ has entries:

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

which corresponds to $D_k = \mathbf{1}, C_k = \tau \cdot \mathbf{1}$.

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

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

and

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

One of our main results is Theorem 2.2. The proof is again in Section 4.

Theorem 2.2. *Assume that the Jacobi matrix J is nonnegative definite and let Σ be the spectral measure associated with J . Then for any $\tau \in (0, 1]$, if $\Sigma \in \mathcal{S}_{p,1}(\tau^-, \tau^+)$,*

$$\begin{aligned} \mathcal{K}(\Sigma_{\text{MP}(\tau)} \mid \Sigma) &+ \sum_{n=1}^{N^+} \mathcal{F}_L^+(\lambda_n^+) + \sum_{n=1}^{N^-} \mathcal{F}_L^-(\lambda_n^-) \\ &= \sum_{k=1}^{\infty} \tau^{-1} \text{tr} G(D_k D_k^*) + \text{tr} G(\tau^{-1} C_k C_k^*), \end{aligned}$$

where both sides may be infinite simultaneously and D_k, C_k are as (2.3). If $\Sigma \notin \mathcal{S}_{p,1}(\tau^-, \tau^+)$, the right hand side equals $+\infty$.

It follows from (2.4), that although the values for D_k, C_k in the decomposition (2.3) are not unique, the value of the rate function is independent of the choice of representatives for the measure Σ . So in particular, they may be chosen Hermitian positive definite. Moreover, for any choice for D_k, C_k , we remark that the matrices $D_k D_k^*$ and $C_k C_k^*$ are in fact similar to a Hermitian matrix, hence the sum on the right-hand side in Theorem 2.2 is always real valued.

Similar to the matrix *gem*, Theorem 2.1, we can formulate equivalent conditions on the matrices D_k, C_k and the spectral measure, which characterize finiteness of the sum. The following corollary is the matrix counterpart of Corollary 2.4 in [12]. It follows immediately from Theorem 2.2, since $\mathcal{F}_L^-(0) = \infty$ and

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

and, for H similar to a Hermitian matrix,

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

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

Corollary 2.2. *Assume the Jacobi matrix J is nonnegative definite and let Σ be the spectral measure of J . Then*

$$\sum_{k=1}^{\infty} [\text{tr}(D_k D_k^* - \mathbf{1})^2 + \text{tr}(C_k C_k^* - \tau \mathbf{1})^2] < \infty \tag{2.11}$$

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 Σ 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

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 $\sigma^2 > 0$. A random variable X taking values in \mathcal{H}_N is distributed according to the Gaussian unitary ensemble GUE_N , if all real diagonal entries are distributed as $\mathcal{N}(0, 1)$ and the real and imaginary parts of off-diagonal variables are independent and $\mathcal{N}(0, 1/2)$ distributed (also called complex standard normal distribution). All entries are assumed to be independent up to symmetry and conjugation. The random matrix X/\sqrt{N} has then the distribution given by (1.3) and the joint density of the (real) eigenvalues $\lambda = (\lambda_1, \dots, \lambda_N)$ of X is (see, for example, [1])

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

In analogy to the scalar χ^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 $\text{LUE}_N(\gamma)$ with parameter γ . If $\gamma \geq N$, the eigenvalues have the density (see, for example, [1])

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

It is a well-known consequence of invariance under unitary conjugation, that in the classical ensembles (1.3), the array of random eigenvalues and the random eigenvector (unitary) matrix are independent. Further, this latter matrix is Haar distributed ([6]). This implies the following equality in distribution for the weights given in Lemma 3.1 (see Proposition 3.1 in [17]), 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 is 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 3.1. *If (V_1, \dots, V_N) follows the distribution $\mathbb{G}_{p,N}$ and if $H = \sum_{k=1}^N V_k$, then $(H^{-1/2} V_1 H^{-1/2}, \dots, H^{-1/2} V_N H^{-1/2})$ follows the distribution $\mathbb{D}_{p,N}$.*

Our main 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

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

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 $\lim_{x \rightarrow b^\pm} V(x) = V(b^\pm)$ with possible limit $V(b^\pm) = +\infty$. Under the assumption

(A1) Confinement:

$$\liminf_{x \rightarrow b^\pm} \frac{V(x)}{\log|x|} > 2,$$

the empirical distribution $\mu_u^{(N)}$ of eigenvalues $\lambda_1, \dots, \lambda_N$ has a limit μ_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 [18] or [1]). This convergence can be viewed as a consequence of the LDP for the sequence $(\mu_u^{(N)})_N$. We need two additional assumptions on μ_V :

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

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

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

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 μ_V is the semicircle law SC as in (2.7). 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, μ_V is the Marchenko-Pastur law $\text{MP}(\tau)$ as in (2.10). In both the Hermite and the Laguerre cases, the assumptions (A1), (A2) and (A3) are satisfied. We need one more definition related to outlying eigenvalues:

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

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

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 (see, for example, Proposition 3.2 in [12]).

3.2. Large deviations: A general theorem

In order to be self-contained, let us recall the definition of large deviations and we refer to [8] for more details.

Definition 3.2. Let E be a topological Hausdorff space and let $\mathcal{I} : E \rightarrow [0, \infty]$ 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 ($a_n \nearrow \infty$), if:

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

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

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

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

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

To prove our main large deviation principle, we will use a special extension of Baldi’s theorem, which extends also Bryc’s lemma. A proof of this new theorem is in [11]. It is one of the key tool for the statements of Section 4 in [12] and it is also useful in [14].

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 continuous dual space of \mathcal{X} . We denote by $C_b(\mathcal{Y})$ the set of all bounded continuous functions $\varphi : \mathcal{Y} \rightarrow \mathbb{R}$. Recall that 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

$$F(x) - \langle x^*, x \rangle < F(z) - \langle x^*, z \rangle \tag{3.8}$$

for all $z \neq x$.

Theorem 3.1. Assume that $X_n \in \mathcal{X}$ and $Y_n \in \mathcal{Y}$ are defined on the same probability space and that the two sequences (X_n) and (Y_n) are exponentially tight. Assume further that

1. There is a set $D \subset \mathcal{X}^*$ containing 0 and functions $\Lambda : D \rightarrow \mathbb{R}$, $J : \mathcal{C}_b(\mathcal{Y}) \rightarrow \mathbb{R}$ such that for all $x^* \in D$ and $\varphi \in \mathcal{C}_b(\mathcal{Y})$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \exp(n \langle x^*, X_n \rangle + n \varphi(Y_n)) = \Lambda(x^*) + J(\varphi). \tag{3.9}$$

2. If \mathcal{F} denotes the set of exposed points x of

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

with an exposing hyperplane x^* satisfying $x^* \in D$ and $\gamma x^* \in D$ for some $\gamma > 1$, then for every $x \in \{ \Lambda^* < \infty \}$ there exists a sequence $(x_k)_k$ with $x_k \in \mathcal{F}$ such that $\lim_{k \rightarrow \infty} x_k = x$ and

$$\lim_{k \rightarrow \infty} \Lambda^*(x_k) = \Lambda^*(x).$$

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

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

where

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

To apply this theorem in our setting, we remark that \mathcal{X} may be taken as the vector space $\mathcal{M}_p^s(T)$ of bounded signed matrix measures which can be written as $\Sigma = \Sigma_1 - \Sigma_2$ with $\Sigma_1, \Sigma_2 \in \mathcal{M}_p(T)$ and equipped with the weak topology. The continuous dual \mathcal{X}^* contains the space $\mathcal{C}_p(T)$ of bounded continuous functions $f : T \rightarrow \mathcal{H}_p$ with the pairing

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

Then Theorem 3.1 will give the LDP in $\mathcal{M}_p^s(T)$, and since $\mathcal{M}_p(T)$ is a closed subset, the LDP in $\mathcal{M}_p(T)$ follows from Lemma 4.1.5 of [8].

3.3. Large deviations: Results

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 $(\lambda_1, \dots, \lambda_N)$ is \mathbb{P}_N^V distributed and where the distribution of weights (W_1, \dots, W_N) is $\mathbb{D}_{p,N}$ as in the case of classical ensembles. As explained in Section 3.1, this is precisely the distribution of the spectral measure of $(X_N; e_1, \dots, e_p)$, with X_N an $N \times N$ matrix drawn from the distribution (1.3). Recall that under assumption (A1), the empirical measure of the eigenvalues converges to an equilibrium measure μ_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 [12]. 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 3.2. *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 in $\mathcal{M}_{p,1}(\mathbb{R})$ in the weak topology 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 3.1. A natural extension of Theorem 3.2 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 $V_N : \mathbb{R} \rightarrow (-\infty, +\infty]$ 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 (B_1, A_1, \dots) and (D_1, C_1, \dots) respectively, allows us to prove further LDP's for the spectral measure, independently of Theorem 3.2. Since we need a specific block structure, we assume $N = np$. Let $\mathcal{M}_{p,1,c}$ be the set of all compactly supported normalized matrix measures.

Theorem 3.3. *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 GUE_N with $N = np$. Then the sequence $(\Sigma^{(n)})_n$ satisfies the LDP in $\mathcal{M}_{p,1,c}(\mathbb{R})$ in the weak topology with speed pn and good rate function*

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

where B_k, A_k are the Jacobi coefficients of Σ as in (2.1) if Σ is non-trivial and $\mathcal{I}_H(\Sigma) = \infty$ if Σ is trivial.

Theorem 3.4. *Let Y_n be distributed according to the Laguerre ensemble $\text{LUE}_N(p\gamma_n)$, with $N = np$ and $\gamma_n \geq n$ an integer sequence such that $\frac{n}{\gamma_n} \rightarrow \tau \in (0, 1]$ and let $\Sigma^{(n)}$ be the spectral measure of $\frac{1}{p\gamma_n}Y_n$. Then the sequence $(\Sigma^{(n)})_n$ satisfies the LDP in $\mathcal{M}_{p,1,c}([0, \infty))$ in the weak topology with speed pn and good rate function*

$$\mathcal{I}_L(\Sigma) = \sum_{k=1}^{\infty} [\tau^{-1}G(D_k D_k^*) + G(\tau^{-1}C_k C_k^*)],$$

with D_k, C_k as in Lemma 2.1. If Σ 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 [15], with small changes to allow a general sequence of parameters.

Lemma 3.2. (i) *If $X \sim \text{GUE}_p$ with p fixed, then the sequence $(\frac{1}{\sqrt{n}}X)_n$ satisfies the LDP in \mathcal{H}_p with speed n and good rate function*

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

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

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

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

(iii) *Let $Z \sim \text{LUE}_p(1)$ 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 in \mathcal{H}_p with speed n and good rate function*

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

if Z is Hermitian of rank 1 and nonnegative and $\mathcal{I}_3(Z) = \infty$ otherwise.

4. Proofs of Theorems 2.1 and 2.2: From large deviations to sum rules

Proof of Theorem 2.1. Consider the matrix measure $\Sigma^{(n)}$ with $N = np$ support points with density proportional to

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

and weight distribution $\mathbb{D}_{p,N}$ independent of the support points. By Theorem 3.2, the sequence $(\Sigma^{(n)})_n$ satisfies the LDP with speed np and rate function \mathcal{I}_V , where μ_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 2.1. On the other hand, $\Sigma^{(n)}$ is also the spectral measure of the random matrix $\frac{1}{\sqrt{pn}} X_n$, where $X_n \sim \text{GUE}_{pn}$. By Theorem 3.3, the sequence $(\Sigma^{(n)})_n$ satisfies also the LDP in the space of compactly supported matrix measures with speed np and rate function \mathcal{I}_H , the right-hand side of the equation in Theorem 2.1. 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 Σ 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 $\text{tr} B^2 + \text{tr} G(A)$ diverges as $\|A\| \rightarrow \infty$ or $\|B\| \rightarrow \infty$ and so the right-hand side in Theorem 2.1 equals $+\infty$. \square

Proof of Theorem 2.2. Fix $\tau \in (0, 1]$ and let $V(x) = \tau^{-1}x - (\tau^{-1} - 1) \log x$ for $x \geq 0$ and $V(x) = +\infty$ if $x < 0$. From Theorem 3.2, 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 $\text{MP}(\tau)$ 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 2.2. We would like to combine this result with the LDP in Theorem 3.4, 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 proportional to

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

This is the eigenvalue distribution of the matrix $\frac{1}{p\gamma_n} Y_n$, when $Y_n \sim \text{LUE}_{pn}(p\gamma_n)$. By Theorem 3.4, 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 2.2. We may as well write the previous density as

$$\frac{1}{Z_{V_n}^N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^2 \prod_{i=1}^N e^{-NV_n(\lambda_i)}, \tag{4.2}$$

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 $V_n(x) \geq V(x)$ for all x and on the sets $\{x \mid V(x) \leq M\}$, 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 3.1, 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 Σ . The extension to measures with non-compact support follows as in the proof of Theorem 2.1. \square

5. Proof of Theorem 3.2: Spectral LDP for a general potential

This section is devoted to the proof of Theorem 3.2. We will follow the track of the proof developed for the scalar case in [12] 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

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

as in (2.6), we define the j th projection π_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 (allowing ties). Note that π_j is not continuous in the weak topology. For this reason, we need to change our topology on \mathcal{S}_p by identifying Σ as in (5.1) with the vector

$$(\Sigma_I, (\lambda_i^+)_{i \geq 1}, (\lambda_i^-)_{i \geq 1}, (\Gamma_i^+)_{i \geq 1}, (\Gamma_i^-)_{i \geq 1}) \tag{5.2}$$

with $\lambda_i^+ = \alpha^+$ and $\Gamma_i^+ = 0$ if $i > N^+$ and $\lambda_i^- = \alpha^-$ and $\Gamma_i^- = 0$ if $i > N^-$. Then (5.2) is an element of the space $\mathcal{M}_p(I) \times \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \times \mathcal{H}_p^{\mathbb{N}} \times \mathcal{H}_p^{\mathbb{N}}$, equipped with the product topology. Analogously to Lemma 4.5 in [12], 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 $N > 2j$

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

Then the following joint LDP (Theorem 4.1 in [12]) 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 α^\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 3.2 and the assumptions (A1), (A2) and (A3) are satisfied. By Lemma 3.1, we may decouple the weights and consider the (non-normalized) measure

$$\tilde{\Sigma}^{(N)} = \frac{1}{N} \sum_{k=1}^N v_k v_k^* \delta_{\lambda_k}, \tag{5.3}$$

where the entries of $v_1, \dots, v_N \in \mathbb{C}^P$ 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 $I(j) := I \setminus \{\lambda_1^+, \lambda_1^-, \dots, \lambda_j^+, \lambda_j^-\}$ denote the interval I without the j largest and smallest eigenvalues. Analogously, let $I^+(j) := I \setminus \{\lambda_1^+, \dots, \lambda_j^+\}$ and $I^-(j) := I \setminus \{\lambda_1^-, \dots, \lambda_j^-\}$. Then we write $\tilde{\Sigma}_{I(j)}^{(N)}$ 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 induced by (5.2).*

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

$$\mathcal{I}(\Sigma, x^+, x^-) = \mathcal{K}(\Sigma_V \mid \Sigma) + \text{tr } \Sigma(I) - p + \mathcal{I}_{\lambda^\pm}(x^+, x^-).$$

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

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

Proof. We show here only the first part of the theorem, for the other cases just omit the largest or smallest eigenvalues. To further simplify the notations, set $\lambda^\pm(j) := (\lambda_1^+, \dots, \lambda_j^+, \lambda_1^-, \dots, \lambda_j^-)$. First, to show that the sequence $(\tilde{\Sigma}_{I(j)}^{(N)}, \lambda^\pm(j))$ is exponentially tight, define the set

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

which is compact by the remark in Section 1 (here $A \leq B$ is in the Loewner order). For every $\gamma > 2$

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\tilde{\Sigma}_{I(j)}^{(N)}(I) > \gamma) &\leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log P\left(\frac{1}{N} \sum_{k=1}^N v_k v_k^* \geq \gamma/2 \cdot \mathbf{1}\right) \\ &\leq -\text{tr } G(\gamma/2 \cdot \mathbf{1}) = -pG(\gamma/2), \end{aligned}$$

by Lemma 3.2(ii), where we used the fact that $\sum_{k=1}^N v_k v_k^*$ follows the LUE $_p(N)$ distribution. Furthermore, by Theorem 5.1 and Assumption (A1), the sequence of $\lambda^\pm(j)$ is exponentially tight, such that

$$\lim_{\gamma, M \rightarrow \infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \log P((\tilde{\Sigma}_{I(j)}^{(N)}, \lambda^\pm(j)) \notin K_{\gamma, M}) = -\infty.$$

We prove the joint LDP by applying Theorem 3.1. For this, let D be the set of continuous $f : [\alpha^-, \alpha^+] \rightarrow \mathcal{H}_p$ 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 φ a bounded continuous function from \mathbb{R}^{2j} to \mathbb{R} , we consider the joint moment generating function

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

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

$$\mathcal{G}_N(f, \varphi) = \mathbb{E} \left[\exp(N\varphi(\lambda^\pm(j))) \prod_{i:\lambda_i \in I(j)} \mathbb{E}[\exp\{\text{tr}(f(\lambda_i)v_k v_k^*)\} \mid \lambda_1, \dots, \lambda_n] \right] \tag{5.4}$$

$$= \mathbb{E} \left[\exp(N\varphi(\lambda^\pm(j))) \prod_{i:\lambda_i \in I(j)} \mathbb{E}[\exp\{v_k^* f(\lambda_i)v_k\} \mid \lambda_1, \dots, \lambda_n] \right]. \tag{5.5}$$

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

$$\log \mathbb{E}[\exp(v^* A v)] = -\log \det(\mathbf{1} - A) =: L(A) \tag{5.6}$$

so that (5.4) becomes,

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

where $\mu_{u, I(j)}^{(N)}$ 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 $\mathcal{I}_{\lambda^\pm}$. Since φ is bounded, we can conclude from Varadhan's Integral Lemma ([8], p. 137)

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[\exp\{N\varphi(\lambda^\pm(j))\}] = J(\varphi) := \sup_{y \in \mathbb{R}^{2j}} \{\varphi(y) - \mathcal{I}_{\lambda^\pm}(y)\}. \tag{5.7}$$

Since $\mu_u^{(N)}$ 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 μ_V at a negligible cost. For the exact estimates, we may follow along the lines of [12] to conclude

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

where

$$\Lambda(f) = \int L \circ f d\mu_V$$

and L is given in (5.6). Let us notice that in [12], $A(\eta)$ has to be replaced by

$$A(\eta) = \{|\mu_{u,I(j)}^{(n)}(L(f)) - \mu_V(L(f))| < \eta\}.$$

Theorem 3.1 yields the LDP for $(\tilde{\Sigma}_I^{(n)}, \lambda^\pm)$ with good rate

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

once we show the second assumption therein is satisfied. Theorem 5 of [23] identifies Λ^* as

$$\Lambda^*(\Sigma) = \int L^*(h) d\mu_V + \int r\left(\frac{d\Sigma^S}{d\theta}\right) d\theta, \tag{5.8}$$

where:

- L^* is the convex dual of L ,
- r its recession function,
- the Lebesgue-decomposition of Σ with respect to μ_V is

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

- θ is any scalar measure such that Σ^S is absolutely continuous with respect to θ .

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

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

The function L is convex (as in the scalar case, apply Hölder’s inequality in the definition (5.6)) and analytic. The supremum is then reached at a critical value. We denote by $\mathcal{D}[F(Y)]$ the Fréchet derivative of a function $F : \mathcal{H}_p \rightarrow \mathbb{R}$ at Y and look for Y such that

$$\mathcal{D}[\text{tr}(XY) - L(Y)](Z) = 0 \tag{5.9}$$

for every Z . It is well known that, as functions of Y for X fixed, $\mathcal{D}[\text{tr}(XY)](Z) = \text{tr}(XZ)$ and $\mathcal{D}[\log \det Y](Z) = \text{tr}(Y^{-1}Z)$ so that (5.9) becomes, by the chain rule,

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

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

$$L^*(X) = \text{tr}(X - \mathbf{1}) + \log \det(X^{-1}) = \text{tr} X - p + \log \det(X^{-1}) = \text{tr} G(X). \tag{5.10}$$

The recession function is

$$r(X) = \sup\{\text{tr}(XW) : W < \mathbf{1}\}.$$

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

$$r(X) = \text{tr } X. \tag{5.11}$$

Gathering (5.10) and (5.11) and plugging into (5.8) we get

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

To show that Λ^* satisfies the second assumption of Theorem 3.1, we refer to Section A.2 in [13]. Then $(\tilde{\Sigma}_{I(j)}^{(N)}, \lambda^\pm(j))$ satisfies the LDP with speed n and rate function

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

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 induced by (5.2) satisfies the LDP with speed N and rate function*

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

Proof. The proof is similar to the proof of Theorem 4.3 in [12] 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 $\Gamma_i \sim \text{LUE}_p(1)$, so by Lemma 3.2(iii), each individual weight $\frac{1}{N} \Gamma_k$ satisfies the LDP with speed N and rate function $\mathcal{I}_3(X) = \text{tr } X$ for nonnegative definite X of rank 1 and $\mathcal{I}_3(X) = \infty$ otherwise. 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

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

with the rate function

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

if $\tilde{\Sigma}(\mathbb{R}) = Z$ and $\bar{\mathcal{I}}_j(\tilde{\Sigma}, Z) = \infty$ otherwise. Keeping the weights along the way, we may apply the projective method (the Dawson–Gärtner Theorem, p. 162 in the book of [8]) 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

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

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

$$\bar{\mathcal{I}}(\tilde{\Sigma}, Z) = \mathcal{K}(\Sigma_V \mid \tilde{\Sigma}) + \text{tr } Z - p + \sum_{i=1}^{N^+} \mathcal{F}(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}(\lambda_i^-). \tag{5.12}$$

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 (5.2). 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 (5.12), we need to minimize over positive definite $Z \in \mathcal{H}_p$ the function

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

The term $\mathcal{I}_2(Z)$ comes from Lemma 3.2(ii) with $\gamma = 1$ and attains its minimal value 0 for $Z = \mathbf{1}$.

We have obtained the LDP claimed in Theorem 3.2 on the subset $\mathcal{S}_{p,1}$ in the topology induced by (5.2). As in Section 4.5 of [12], 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 3.2.

5.3. Proof of Remark 3.1

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

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

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

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) \leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log(\mathbb{P}_N^V \otimes \mathbb{D}_{p,N})(A_N) \tag{5.13}$$

and

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \log(\mathbb{P}_N^V \otimes \mathbb{D}_{p,N})(A_N). \tag{5.14}$$

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

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

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

$$\Gamma_N^{V_N}(A_N) \leq \Gamma_N^V(A_N). \tag{5.15}$$

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 $\{x \mid V(x) \leq M\}$, this implies

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{\Gamma_N^{V_N}(A_N)}{\Gamma_N^V(A_N)} \geq \lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{\Gamma_N^V(A_N \cap K_{N,M})}{\Gamma_N^V(A_N)}. \tag{5.16}$$

If we take now $A = \mathcal{M}_{p,1}$, then $\Gamma_N^V(A_N) = Z_N^V$ and the right-hand side of (5.16) becomes

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

By the LDP for the extreme eigenvalues, Theorem 5.1, and taking into account Assumption (A1), this limit tends to 0 as $M \rightarrow \infty$. Together with (5.15), we have shown that for $A = \mathcal{M}_{p,1}$

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

Since $(\mathbb{P}_N^{V_N} \otimes \mathbb{D}_{p,N})(A_N) = (Z_N^V)^{-1} \Gamma_N^V(A_N)$, the inequality (5.15) leads to the inequality (5.13) and the arguments for (5.16) yield

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

for any $M \geq 0$. Letting $M \rightarrow \infty$, this implies inequality (5.14), as by the LDP for the extreme eigenvalues we have

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

6. Proof of Theorems 3.3 and 3.4

6.1. Hermite block case

The starting point for the proof of Theorem 3.3 is the following block-tridiagonal representation of the Gaussian ensemble. It is a matrix extension of a famous result of Dumitriu and Edelman [10]. The proof is deferred to the supplementary file [13].

Lemma 6.1. *Let $D_k \sim \text{GUE}_p$ and C_k be Hermitian non-negative definite such that $C_k^2 \sim \text{LUE}_p(p(n - k))$ for $k = 1, \dots, n$ 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 GUE_{pn} .

Proof of Theorem 3.3. 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 3.2. 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 $(D_1^{(n)}, C_1^{(n)}, \dots, D_k^{(n)})$ with rate given by the corresponding sum of rate functions \mathcal{I}_1 and \mathcal{I}'_2 .

Now, we follow the strategy developed in [16] for the scalar case. The random matrix measure $\Sigma^{(n)}$ belongs to $\mathcal{M}_{p,c}^1(\mathbb{R})$ and may be identified with its sequence of moments. Since the product topology \mathcal{T}_m of the convergence of moments on $\mathcal{M}_{p,1,c}(\mathbb{R})$ 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

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

being a continuous injection, the LDP of $\Sigma^{(N)}$ in $\mathcal{M}_{p,1,c}(\mathbb{R})$ is then a consequence of the following LDP on the sequence of moments and of the inverse contraction principle (see [8], Theorem 4.2.4 and the subsequent Remark (a)). \square

Proposition 6.1. *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}_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*

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

and such that

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

where J is the infinite block Jacobi matrix with blocks (B_1, A_1, \dots) as in (2.1). In that case

$$\mathcal{I}_m(m_1, \dots) = \sum_{k=1}^{\infty} \frac{1}{2} \operatorname{tr} B_k^2 + \operatorname{tr} G(A_k^2). \tag{6.3}$$

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

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

If J is the $kp \times kp$ Jacobi matrix built 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 $\mathcal{I}_m^{(k)}$ 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

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

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

The Dawson–Gärtner theorem (see [8]) yields the LDP for the whole moment sequence $m(\Sigma^{(n)})$ in $\mathcal{H}_p^{\mathbb{N}_0}$ with good rate

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

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 Σ . In this case,

$$\mathcal{I}_m(m_1, \dots) = \sup_{k \geq 1} \mathcal{I}^{(k)}(B_1, A_1, \dots, B_k) = \sum_{k=1}^{\infty} \left(\frac{1}{2} \text{tr } B_k^2 + \text{tr } G(A_k^2) \right). \quad \square$$

6.2. Laguerre block case

To prove Theorem 3.3, we need the following block-bidiagonal representation, again a matrix extension of a result of Dumitriu and Edelman [10].

Lemma 6.2. *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*

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

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 $\text{LUE}_{pn}(pm)$ ($m \geq n$).

Proof of Theorem 3.4. As in the proof of Theorem 3.3, we start by looking at the individual entries of the tridiagonal representation of Lemma 6.2, 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 3.2 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 $\text{tr } G(\tau^{-1} \cdot)$ 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

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

for C positive definite and $\mathcal{I}_C(C) = \infty$ otherwise. Similarly, 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 3.2 with $\gamma = 1$. If we consider the speed pn and the square root $D_k^{(n)}$, this changes the rate to

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

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

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

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

Since D_k, C_k determine the moments of the spectral measure of L_n as in Lemma 6.2, we may conclude as in the Hermite case a LDP for a finite collection of moments of $\Sigma^{(n)}$ 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}_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 Lemma 2.1, 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 (2.3). In this case

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

Given another sequence D'_k, C'_k generating the same moment sequence (m_1, \dots) as the positive definite D_k, C_k , we may use Lemma 2.1 to see

$$\begin{aligned} \operatorname{tr} G(D_k^2) &= \operatorname{tr} G(D'_k (D'_k)^*), \\ \operatorname{tr} G(\tau^{-1} C_k^2) &= \operatorname{tr}(\tau^{-1} C'_k (C'_k)^*). \end{aligned}$$

That is, the value of the rate function is independent of the choice of Jacobi coefficients. The inverse contraction principle implies then the LDP for the spectral measure $\Sigma^{(n)}$. □

Acknowledgments

We warmly thank the referee for a very careful read of this paper and helpful suggestions.

Supplementary Material

Preliminaries on matrix measures (DOI: [10.3150/17-BEJ1003SUPP](https://doi.org/10.3150/17-BEJ1003SUPP); .pdf). We refer to the supplementary file [13] for more details on matrix measures. It also explains how some of the results from the scalar case extend to the matrix case.

References

[1] Anderson, G.W., Guionnet, A. and Zeitouni, O. (2010). *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics **118**. Cambridge: Cambridge Univ. Press. MR2760897

- [2] Breuer, J., Simon, B. and Zeitouni, O. (2017). Large deviations and sum rules for spectral theory – A pedagogical approach. *J. Spectr. Theory*. To appear.
- [3] Breuer, J., Simon, B. and Zeitouni, O. (2017). Large deviations and the Lukic conjecture. Preprint. Available at [arXiv:1703.00653](https://arxiv.org/abs/1703.00653).
- [4] Damanik, D., Killip, R. and Simon, B. (2010). Perturbations of orthogonal polynomials with periodic recursion coefficients. *Ann. of Math. (2)* **171** 1931–2010. [MR2680401](https://doi.org/10.2307/2680401)
- [5] Damanik, D., Pushnitski, A. and Simon, B. (2008). The analytic theory of matrix orthogonal polynomials. *Surv. Approx. Theory* **4** 1–85. [MR2379691](https://doi.org/10.1007/s11464-008-0001-1)
- [6] Dawid, A.P. (1977). Spherical matrix distributions and a multivariate model. *J. Roy. Statist. Soc. Ser. B* **39** 254–261. [MR0518916](https://doi.org/10.2307/2346381)
- [7] Delsarte, P., Genin, Y.V. and Kamp, Y.G. (1978). Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits Syst.* **25** 149–160. [MR0481886](https://doi.org/10.1109/TCS.1978.1075188)
- [8] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. New York: Springer. [MR1619036](https://doi.org/10.1007/978-1-4612-2769-0)
- [9] Dette, H. and Studden, W.J. (2002). Matrix measures, moment spaces and Favard’s theorem for the interval $[0, 1]$ and $[0, \infty)$. *Linear Algebra Appl.* **345** 169–193. [MR1883272](https://doi.org/10.1016/S0024-3795(02)00001-1)
- [10] Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. [MR1936554](https://doi.org/10.1063/1.1431302)
- [11] Gamboa, F., Nagel, J. and Rouault, A. (2016). Addendum to “Sum rules via large deviations.” Preprint. Available at [arXiv:1610.02071](https://arxiv.org/abs/1610.02071).
- [12] Gamboa, F., Nagel, J. and Rouault, A. (2016). Sum rules via large deviations. *J. Funct. Anal.* **270** 509–559. [MR3425894](https://doi.org/10.1016/j.jfa.2016.05.014)
- [13] Gamboa, F., Nagel, J. and Rouault, A. (2017). Supplement to “Sum rules and large deviations for spectral matrix measures.” DOI:10.3150/17-BEJ1003SUPP.
- [14] Gamboa, F., Nagel, J. and Rouault, A. (2017). Sum rules and large deviations for spectral measures on the unit circle. *Random Matrices Theory Appl.* **6** 1750005. [MR3612269](https://doi.org/10.1142/S175137581750005)
- [15] Gamboa, F., Nagel, J., Rouault, A. and Wagener, J. (2012). Large deviations for random matricial moment problems. *J. Multivariate Anal.* **106** 17–35. [MR2887677](https://doi.org/10.1016/j.jmva.2012.02.001)
- [16] Gamboa, F. and Rouault, A. (2011). Large deviations for random spectral measures and sum rules. *Appl. Math. Res. Express. AMRX* **2011** 281–307. [MR2835992](https://doi.org/10.1016/j.amrx.2011.08.001)
- [17] Gamboa, F. and Rouault, A. (2014). Operator-valued spectral measures and large deviations. *J. Statist. Plann. Inference* **154** 72–86. [MR3258407](https://doi.org/10.1016/j.jspi.2014.01.001)
- [18] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. [MR1487983](https://doi.org/10.2307/121017)
- [19] Killip, R. (2007). Spectral theory via sum rules. In *Spectral Theory and Mathematical Physics: A Festschrift in Honor of Barry Simon’s 60th Birthday. Proc. Sympos. Pure Math.* **76** 907–930. Amer. Math. Soc., Providence, RI. [MR2310217](https://doi.org/10.1090/S0033-540X-07-00000-0)
- [20] Killip, R. and Simon, B. (2003). Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math. (2)* **158** 253–321. [MR1999923](https://doi.org/10.2307/2689923)
- [21] Mandrekar, V. and Salehi, H. (1971). On singularity and Lebesgue type decomposition for operator-valued measures. *J. Multivariate Anal.* **1** 167–185. [MR0301165](https://doi.org/10.1016/0047-1494(71)90001-1)
- [22] Robertson, J.B. and Rosenberg, M. (1968). The decomposition of matrix-valued measures. *Michigan Math. J.* **15** 353–368. [MR0239044](https://doi.org/10.2307/239044)
- [23] Rockafellar, R.T. (1971). Integrals which are convex functionals. II. *Pacific J. Math.* **39** 439–469. [MR0310612](https://doi.org/10.2307/2340612)
- [24] Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle. Part 1. Classical Theory. American Mathematical Society Colloquium Publications* **54**. Providence, RI: Amer. Math. Soc. [MR2105088](https://doi.org/10.1090/S0025-5718-2005-015088-0)

- [25] Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle. Part 2. Spectral Theory*. American Mathematical Society Colloquium Publications **54**. Providence, RI: Amer. Math. Soc. [MR2105089](#)
- [26] Simon, B. (2011). *Szegő's Theorem and Its Descendants. Spectral Theory for L^2 Perturbations of Orthogonal Polynomials*. *M. B. Porter Lectures*. Princeton, NJ: Princeton Univ. Press. [MR2743058](#)

Received February 2017 and revised November 2017