Nonstandard limit theorems and large deviations for the Jacobi beta ensemble

Jan Nagel
May 11, 2014

Abstract

In this paper we show weak convergence of the empirical eigenvalue distribution and of the weighted spectral measure of the Jacobi ensemble, when one or both parameters grow faster than the dimension \( n \). In these cases the limit measure is given by the Marchenko-Pastur law and the semicircle law, respectively. For the weighted spectral measure we also prove large deviation principles under this scaling, where the rate functions are those of the other classical ensembles.

Keywords: Jacobi ensemble, spectral measure, random matrix theory, large deviations, semicircle law, Marchenko-Pastur law

1 Introduction

The Jacobi ensemble is one of the central distributions of Hermitian matrices and can be characterised by the eigenvalue density

\[
c_{a,b,\beta} \cdot \prod_{i<j} |\lambda_i - \lambda_j|^\beta \sum_{i=1}^n \lambda_i^{a-\beta} (1-\lambda_i)^{b-\beta} 1_{\{0<\lambda_i<1\}}
\]

with parameters \( a, b > 0 \), \( \beta > 0 \) and \( c_\beta = \frac{\beta}{2}(n-1) + 1 \) and \( c_{a,b,\beta} \) the normalisation constant. If \( X,Y \) are \( n \times \frac{2}{\beta} a \) and \( n \times \frac{2}{\beta} b \) \( (\frac{2}{\beta} a, \frac{2}{\beta} b \geq n) \) matrices with i.i.d. real (\( \beta = 1 \)), complex (\( \beta = 2 \)) or quaternion (\( \beta = 4 \)) standardnormal distributed entries, then the matrix \( A = (XX^* + YY^*)^{-1/2}XX^*(XX^* + YY^*)^{-1/2} \) has eigenvalues with density \( \frac{1}{c_{a,b,\beta}} \). Due to this construction, results about the spectral properties of the Jacobi ensemble can be applied in the
multivariate analysis of variance (MANOVA, see [Muirhead (1982)]). In statistical mechanics, the Jacobi ensemble arises as a model for a log gas with logarithmic interactions, confined to the interval $[0, 1]$ ([Dyson (1962), Forrester (2010)]). One central object in the asymptotic study of random matrix ensembles is the empirical eigenvalue distribution

\[ \hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}, \]  

when $\delta_z$ denotes the Dirac measure in $z$. Recently, a number of authors showed interest in the so-called spectral measure of random matrices, a weighted version of the empirical eigenvalue distribution. For the matrix $A$ the spectral measure $\mu_n$ is defined by the relation

\[ \int x^k \mu_n(dx) = \langle e_1, A^k e_1 \rangle \]  

for all $k \geq 1$, the functional calculus in the sense of the spectral theorem ([Dunford and Schwartz (1963)]). For a matrix of the classical ensembles, $\mu_n$ has the representation

\[ \mu_n = \sum_{i=1}^{n} w_i \delta_{\lambda_i}, \]  

with weights $w_i = |\langle e_1, u_i \rangle|^2$, when $u_1, \ldots, u_n$ is a corresponding system of eigenvectors. Due to the invariance of the ensembles under unitary conjugations, the weights are independent from the eigenvalues and follow a Dirichlet distribution with parameter $\frac{\beta}{2}n$ ([Dawid (1977)]). Since both the eigenvalue and the weight distribution can be defined not only for $\beta \in \{1, 2, 4\}$, we will consider general $\beta > 0$ and denote the joint distribution of eigenvalues (1.1) and independent Dirichlet weights by $J_\beta E(a, b)$.

The large $n$ behaviour of the eigenvalues with density (1.1) has been subject to intensive research. There are numerous results such as almost sure limits ([Collins (2005), Leff (1964)]), CLTs ([Jiang (2009), Nagao and Wadati (1993), Johnstone (2008)]) or large deviations ([Hiai and Petz (2006)]) about the eigenvalues in the bulk, i.e. for $\mu_n$ or of the edge behaviour of the largest eigenvalue. For a general overview we refer to the books of [Anderson et al. (2010), Forrester (2010) and Mehta (2004)]. For the weighted random measure, large deviation principles were proven by Gamboa and Rouault (2009, 2008) and central limit theorems can be found in the papers of Lytova and Pastur (2009a,b) and Dette and Nagel (2012).

The mentioned results assume that the parameters $a = a_n$ and $b = b_n$ satisfy a standard behaviour,

\[ \frac{a_n}{n} \xrightarrow{n \to \infty} a_0 \in [0, \infty), \quad \frac{b_n}{n} \xrightarrow{n \to \infty} b_0 \in [0, \infty). \]
Significantly less is known in the what we call nonstandard case, where \( a_n \) and/or \( b_n \) are of larger order than \( n \). Under restrictions on the minimal and maximal order of parameters, Dette and Nagel (2009) showed that the rescaled empirical eigenvalue distribution can converge to different classes of limit distributions. In particular, these classes include the semicircle law and the Marchenko-Pastur law, the large dimensional limit of the other two classical distributions of Hermitian matrices, the Gaussian and the Laguerre ensemble, respectively. Jiang (2012) showed weak convergence to the Marchenko-Pastur law as well as convergence at the edge under the assumption \( a_n = o(\sqrt{b_n}), n = o(\sqrt{b_n}) \). Both of these results rely on uniform estimates, for the difference between eigenvalues and zeros of orthogonal polynomials (Dette and Nagel (2009)), and for the difference between densities (Jiang (2012)).

In this paper, we use a different approach and show weak convergence of rescaled versions of \( \mu_n \) and \( \hat{\mu}_n \) to the semicircle law and the Marchenko-Pastur law without any additional assumptions on the rate of the parameters. The main idea is to study the spectral measure of a tridiagonal representation of the Jacobi ensemble. The weak convergence follows from the entrywise convergence of the tridiagonal matrix, thus eliminating any need for uniform estimates. Furthermore, we prove large deviation principles for the spectral measure \( \mu_n \) under nonstandard scaling. These LDPs illustrate further the connection between the classical ensembles, as the rate functions are given by the rate functions of the Gaussian and the Laguerre ensemble, respectively.

This paper is structured as follows: In Section 2, we state the results and introduce the other classical ensembles. Section 3 introduces the tridiagonal models and the relation to orthogonal polynomials. Finally, Section 4 contains the proofs and some auxiliary results.

## 2 Results

To formulate the results and put them into context, we define the two remaining classical ensembles. The Gaussian (beta-)ensemble is defined by the eigenvalue density

\[
f_G(\lambda) = c_G \prod_{i<j} |\lambda_i - \lambda_j|^{\beta} \prod_{i=1}^n e^{-\frac{1}{2} \lambda_i^2}.
\]

This is the eigenvalue distribution of a Hermitian matrix with independent Gaussian entries on and above the diagonal. For growing dimension, the celebrated result of Wigner (1955) states that the empirical eigenvalue measure \( \hat{\mu}_n \) of the rescaled eigenvalues \( \sqrt{\frac{2}{\beta n}} \lambda_i \) converges almost surely weakly to the semicircle law \( \text{SC} \) with density

\[
\frac{1}{2\pi} \sqrt{4 - x^2} 1_{(-2 < x < 2)}.
\]
The Laguerre (beta-)ensemble, or Wishart ensemble, has eigenvalue density

\[ f_L(\lambda) = c_L^\gamma \prod_{i<j} |\lambda_i - \lambda_j|^{\beta} \prod_{i=1}^{n} \lambda_i^{a-\gamma} e^{-\lambda_i} 1_{\{\lambda_i > 0\}}, \]  

with parameter \( a > 0 \), which corresponds to a square of a Gaussian matrix. As \( n \to \infty \) and \( \frac{\beta n}{2a} \to \tau \in (0, 1] \), the measure \( \hat{\mu}_n \) of the rescaled eigenvalues \( \tilde{\lambda}_i = \frac{2\tau}{n \beta} \lambda_i \) tends almost surely to the Marchenko-Pastur law \( \text{MP}(\tau) \) with parameter \( \tau \) and density

\[ \frac{1}{2\pi \tau x} \sqrt{(x - \tau^-)(\tau^+ - x)} 1_{\{\tau^- < x < \tau^+\}}, \]

where \( \tau^- = (\sqrt{\tau} - 1)^2 \) and \( \tau^+ = (\sqrt{\tau} + 1)^2 \).

### 2.1 Weak convergence

Our first two main theorems deal with the almost sure weak convergence of the rescaled eigenvalue distribution of the Jacobi ensemble. In this case, the limit is the same for the empirical measure \( \hat{\mu}_n \) and the spectral measure \( \mu_n \). In what follows, we regard the random measures as elements of the set \( \mathcal{M}_c \) of probability measures on \( \mathbb{R} \) with compact support, endowed with the weak topology and the corresponding Borel \( \sigma \)-algebra. To simplify notation, we write \( \beta' = \frac{1}{2} \beta \).

**Theorem 2.1** Suppose in the Jacobi ensemble \( \mathcal{J} \beta \mathcal{E}(a_n, b_n) \) the parameters satisfy

\[ \frac{a_n}{n} \to \infty, \quad \frac{a_n}{b_n} \to \sigma > 0 \]

and additionally

\[ \sqrt{\frac{b_n}{n}} \left( \sigma - \frac{a_n}{b_n} \right) \to 0. \]

Then the spectral measure \( \mu_n \) and the empirical eigenvalue distribution \( \hat{\mu}_n \) of the rescaled eigenvalues

\[ (\sigma + 1) \sqrt{b_n \frac{1+\sigma}{\sigma n \beta'}} \left( \lambda_i - \frac{\sigma}{\sigma + 1} \right) \]

converge weakly to the semicircle law \( \text{SC} \) almost surely.
Theorem 2.2 If the parameters of the Jacobi ensemble \( \mathcal{J} \beta \mathcal{E}(a_n, b_n) \) satisfy
\[
\frac{\beta' n}{a_n} \xrightarrow{n \to \infty} \tau \in (0, 1], \quad \frac{b_n}{n} \xrightarrow{n \to \infty} \infty,
\]
then the spectral measure \( \mu_n \) and the empirical eigenvalue distribution \( \hat{\mu}_n \) of the rescaled eigenvalues
\[
\frac{b_n}{a_n} \lambda_i
\]
converge weakly to the Marchenko-Pastur law \( \text{MP}(\tau) \) almost surely.

The fact that the rescaled Jacobi eigenvalue distribution can have the same large dimensional limit as the Gaussian and the Laguerre ensemble is closely related to the finite dimensional convergence to the other ensembles. To explain this connection, we state here the convergence for fixed dimension. These weak convergence results for fixed size are probably well-known (see e.g. Gamboa et al. (2012) for \( \beta = 2 \)), however, we could not find a general reference for them.

Proposition 2.3 (i) Let \( \lambda_N \) be an eigenvalue vector of the Jacobi ensemble \( \mathcal{J} \beta \mathcal{E}(a_N, b_N) \) with density (1.1). Suppose \( a_N \to \infty \) and \( a_N/b_N \to \sigma \in (0, \infty) \) while the dimension \( n \) is fixed and additionally
\[
\sqrt{a_N} \left( \sigma - \frac{a_N}{b_N} \right) \xrightarrow{N \to \infty} 0,
\]
then, with \( 1_n = (1, \ldots, 1) \),
\[
(\sigma + 1) \sqrt{b_N \frac{\sigma + 1}{\sigma}} \left( \lambda_N - \frac{\sigma}{\sigma + 1} 1_n \right)
\]
converges in distribution to a random vector with density (2.1).

(ii) Let \( \lambda_N \) be an eigenvalue vector of the \( \mathcal{J} \beta \mathcal{E}(a, b_N) \) with density (1.1). If \( b_N \to \infty \) with fixed dimension \( n \), then the rescaled eigenvalue vector
\[
b_N \lambda_N
\]
converges in distribution to a random vector with density (2.3).

Heuristically, Theorem 2.2 is now a combination of part (ii) of the Proposition with the large dimensional limit of the Laguerre ensemble. We let \( b_n \) tend to infinity faster than \( n \) to approach the Laguerre ensemble, while simultaneously increasing dimension and choosing \( a_n \) such that the Laguerre eigenvalues converge to the right limit. Theorem 2.1 has a similar explanation: parameters \( a_n \) and \( b_n \) of the same order growing faster than the dimension carry part (i) of the proposition over in the large dimensional limit.
2.2 Large deviations

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

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

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

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

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

The rate function $I$ is good if its level sets $\{ x \in U \mid I(x) \leq a \}$ are compact for all $a \geq 0$. In our case, the measures $P_n$ will be the distributions of the random spectral measures $\mu_n$ and we will say that the sequence of measures $\mu_n$ satisfies an LDP. Similar to the weak convergence results in Section 2.1, the rate function is determined by the large deviation behaviour of the two other classical ensembles. To formulate the rate functions we first need some definitions. The Kullback-Leibler distance between two probability measures $\mu$ and $\nu$ is given by

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

if $\mu$ is absolutely continuous with respect to $\nu$ and $K(\mu|\nu) = \infty$ otherwise. Let for $x \geq 2$

$$F_G(x) = \int_{x}^{\infty} \sqrt{t^2 - 4} dt = \frac{x}{2} \sqrt{x^2 - 4} - 2 \log \left( \frac{x + \sqrt{x^2 - 4}}{2} \right)$$

and define $F_G(x) = F_G(-x)$ for $x \leq -2$. Note that $F_G$ is the rate function for the largest eigenvalue of the Gaussian ensemble (see Anderson et al. (2010)). Following Simon (2005a), we say that a probability measure $\mu$ satisfies the Blumenthal-Weyl condition (B.W.c.) if

(i) $\text{supp}(\mu) = [-2, 2] \cup \{ E_j^- \}_{j=1}^{N^-} \cup \{ E_j^+ \}_{j=1}^{N^+}$, where $N^-, N^+$ may be 0, finite or infinite with

$$E_1^- < E_2^- < \cdots < -2 \quad \text{and} \quad E_1^+ > E_2^+ > \cdots > 2.$$

(ii) If $N^-$ or $N^+$ is infinite, then $E_j^-$ converges towards $-2$ and $E_j^+$ converges to 2, respectively.
Gamboa and Rouault (2008) showed that the sequence of weighted spectral measures \( \mu_n \) of the Gaussian ensemble satisfies the LDP with speed \( \beta'n \) and good rate function

\[
\mathcal{I}_G(\mu) = K(SC|\mu) + \sum_{j=1}^{N^-} \mathcal{F}_G(E_j^-) + \sum_{j=1}^{N^+} \mathcal{F}_G(E_j^+) 
\]

if \( \mu \) satisfies B.W.c. and \( \mathcal{I}_G(\mu) = \infty \) otherwise. Using the tridiagonal representation of spectral measures we introduce in Section 3, the rate function \( \mathcal{I}_G \) can be written in terms of recursion coefficients of the measure \( \mu \). This large deviation behaviour is fundamentally different from the one of the empirical measure \( \hat{\mu}_n \), which satisfies a LDP with speed \( n^2 \) and rate function related to Voiculescu’s entropy. Our first nonstandard LDP shows the approximation of the Gaussian ensemble.

**Theorem 2.4** Suppose that the parameters of the Jacobi ensemble \( J_{\beta \mathcal{E}}(a_n, b_n) \) satisfy

\[
\frac{a_n}{n} \xrightarrow{n \to \infty} \infty, \quad \frac{b_n}{n} \xrightarrow{n \to \infty} \infty, \quad \text{and} \quad \frac{a_n - b_n}{\sqrt{b_n n}} \xrightarrow{n \to \infty} 0.
\]

Then the spectral measure \( \mu_n \) of the rescaled eigenvalues

\[
4 \sqrt{\frac{b_n}{n \beta}} \left( \lambda_i - \frac{1}{2} \right)
\]

satisfies the LDP with speed \( \beta'n \) and good rate function \( \mathcal{I}_G \).

In the case of the Laguerre ensemble, the sequence of spectral measures \( \mu_n \) satisfies an LDP with speed \( \beta'n \) and a good rate function \( \mathcal{I}_L \), which is conjectured to be the Laguerre-analogue of \( \mathcal{I}_G \). However, there is an explicit formulation of \( \mathcal{I}_L \) in terms of the recursion variables introduced in the following section and we refer to equation (3.8) for the definition.

**Theorem 2.5** Suppose that the parameters of the Jacobi ensemble \( J_{\beta \mathcal{E}}(a_n, b_n) \) satisfy

\[
\frac{\beta'n}{a_n} \xrightarrow{n \to \infty} \tau \in (0,1], \quad \frac{b_n}{n} \xrightarrow{n \to \infty} \infty,
\]

then the spectral measure \( \mu_n \) of the rescaled eigenvalues

\[
\frac{b_n}{a_n} \lambda_i
\]

satisfies the LDP with speed \( a_n \) and good rate function \( \mathcal{I}_L \).
Note that the LDP in Theorem 2.5 for the spectral measure of the rescaled eigenvalues implies the almost sure convergence to the Marchenko-Pastur law. Thus, Theorem 2.2 is a direct consequence of Theorem 2.5 and Lemma 4.2. The LDP in Theorem 2.4 is however more restrictive concerning the behaviour of the parameters than our convergence result in Theorem 2.1. More precisely, the assumption \( \frac{a_n - b_n}{\sqrt{b_n n}} \to 0 \) implies \( \sigma = 1 \) in Theorem 2.1.

We remark that although the two theorems suggest an exponentially fast approximation of the Gaussian and Laguerre ensemble, this would require constraints on the order of \( a_n \) and \( b_n \). To completely avoid these constraints, we use a different method of proof.

### 3 Tridiagonal Representations

The proofs are based on tridiagonal matrix models for the classical ensembles, valid for all values of \( \beta > 0 \) and whose entries can be decomposed into independent random variables. In the following, we say that a random variable is Gamma(\( a \)) distributed, if it has the density

\[
x^{a-1} \frac{e^{-x}}{\Gamma(a)} 1_{\{x>0\}}
\]

with \( a > 0 \). The logarithmic moment generating function of the Gamma(\( a \)) distribution is \( \Lambda(t) = -a \log(1 - t) \) (for \( t < 1 \)) with Fenchel-Legendre transform \( \Lambda^*(x) = a \cdot g(a^{-1}x) \), where the function \( g \) is defined by

\[
g(x) = x - \log x - 1
\]

if \( x > 0 \) and \( g(x) = \infty \) otherwise.

For the Jacobi ensemble, the tridiagonal matrix model was found by Killip and Nenciu (2004) and is constructed as follows. Let \( p_1, \ldots, p_{2n-1} \) be independent random variables distributed as

\[
p_k \sim \begin{cases} 
\text{Beta} \left( \frac{2n-k}{2} \beta', a + b - \frac{2n-k-2}{2} \beta' \right) & k \text{ even}, \\
\text{Beta} \left( a - \frac{k-1}{2} \beta', b - \frac{k-1}{2} \beta' \right) & k \text{ odd}.
\end{cases}
\]

and define

\[
d_k = p_{2k-2}(1 - p_{2k-3}) + p_{2k-1}(1 - p_{2k-2})

c_k = \sqrt{p_{2k-1}(1 - p_{2k-2}) p_{2k}(1 - p_{2k-1})}
\]
with \( p_{-1} = p_0 = 0 \). Then the tridiagonal matrix

\[
J_n(\beta, a, b) = \begin{pmatrix}
d_1 & c_1 & & \\
c_1 & d_2 & \ddots & \\
& \ddots & \ddots & c_{n-1} \\
c_{n-1} & & & d_n
\end{pmatrix},
\]

is a matrix of the Jacobi ensemble, that is, the eigenvalues follow the density (1.1) and the square moduli of the top entries of the eigenvectors are Dirichlet distributed with parameter \( \beta' \) independent of the eigenvalues. Note that our parametrisation and scaling differs from the one used by Killip and Nenciu (2004), who consider eigenvalues in \([-2, 2]\) and beta distributions obtained from the one in (3.2) by the transformation \( x \mapsto 1 - 2x \).

Such a sparse matrix model is particularly convenient when we consider weak convergence of spectral measures. The \( k \)-th moment of the spectral measure \( \mu_n \) is given by the upper left entry of the \( k \)-th power of the matrix, compare (1.3). Thus, it depends only on a finite number of entries, and entrywise convergence of the tridiagonal matrix to a (possible infinite) matrix implies convergence of moments and then weak convergence of the spectral measure (given that the limit measure is uniquely determined by its moments). To prove convergence of moments of the empirical eigenvalue measure, we would need to control a growing number of entries and uniform statements.

The tridiagonal representation of the two other classical ensembles were proven by Dumitriu and Edelman (2002). The tridiagonal matrix \( G_n(\beta) \) of the Gaussian ensemble with notation as in (3.4) has standardnormal distributed diagonal entries and positive subdiagonal entries with \( c_k^2 \sim \text{Gamma}((n-k)\beta') \), such that \( d_1, \ldots, d_n, c_1, \ldots c_{n-1} \) are independent. As \( n \to \infty \), the standardised matrix \( \frac{1}{\sqrt{\beta} n} G_n(\beta) \) converges entrywise almost surely to the infinite tridiagonal matrix

\[
SC = \begin{pmatrix}
0 & 1 & & \\
1 & 0 & 1 & \\
& \ddots & \ddots & \ddots \\
& & \ddots & \\
& & & 1
\end{pmatrix},
\]

which has the semicircle distribution SC as spectral measure.

For the matrix \( L_n(\beta, a) \) of the Laguerre ensemble, the entries can be written as

\[
d_k = z_{2k-2} + z_{2k-1},
\]

\[
c_k = \sqrt{z_{2k-1}^2 z_{2k}},
\]

(3.5)
where \( z_0 = 0 \) and \( z_1, \ldots, z_{2n-1} \) are independent random variables with distributions

\[
\begin{align*}
z_{2k-1} &\sim \text{Gamma}(a - (k-1)\beta'), \\
z_{2k} &\sim \text{Gamma}((n-k)\beta').
\end{align*}
\]

Under the classical rescaling as in Section 2, the limit matrix is the tridiagonal matrix \( \mathcal{MP}(\tau) \) with diagonal entries \( d_1 = 1, d_k = 1 + \tau \) for \( k > 1 \) and subdiagonal entries \( c_k = \sqrt{\tau} \). The corresponding spectral measure is the Marchenko-Pastur law \( \text{MP}(\tau) \).

The weak convergence in Proposition 2.3 is a direct consequence of the tridiagonal representation of the ensembles and the weak convergence of the scalar beta distribution: if \( X_N \sim \text{Beta}(a, b_N) \), then \( b_N X_N \) converges in distribution to a Gamma\((a)\) distributed random variable and if \( X_N \sim \text{Beta}(a_N, b_N) \) with parameters satisfying the assumptions as in case (i), then

\[
(\sigma + 1) \sqrt{b_N \frac{\sigma + 1}{\sigma}} \left( X_N - \frac{\sigma}{\sigma + 1} \right) \xrightarrow{n \to \infty} N(0, 1).
\]

The componentwise convergence of the tridiagonal models gives then the weak convergence of the eigenvalues.

Besides parametrizing the classical ensembles by independent random variables, the tridiagonal models have another fundamental property: they represent the multiplication \( f(x) \mapsto xf(x) \) in \( L^2(\mu_n) \), when \( \mu_n \) is the spectral measure and the basis of \( L^2(\mu_n) \) is \( \{P_0, \ldots P_{n-1}\} \) with \( P_j \) the \( j \)-th orthonormal polynomial. As a consequence, the entries in (3.4) appear in the three term recursion of the orthogonal polynomials,

\[
x P_j(x) = c_{j+2} P_{j+1}(x) + d_{j+1} P_j(x) + c_j P_{j-1}(x)
\]

for \( 1 \leq j \leq n-3 \). For more on this relation, we refer to Simon (1997, 2005b). The measure \( \mu \) is supported on \([0, \infty)\) if and only if there are nonnegative \( z_1, z_2, \ldots \) such that the decomposition in (3.5) holds. Furthermore, if \( \mu \) is concentrated on \([0, 1]\), then \( z_k = p_k(1 - p_{k-1}) \) form a chain sequence with \( p_k \in [0, 1] \). In this case we get a decomposition of recursion coefficients as in (3.3).

It is possible to formulate the rate function \( \mathcal{I}_G \) of the Gaussian ensemble defined in (2.5) in terms of the recursion coefficients. The sum-rule by Killip and Simon (2003) states that

\[
\mathcal{I}_G(\mu) = \sum_{k=1}^{\infty} \frac{1}{2} d_k^2 + g(c_k^2),
\]

where \( d_k \) and \( c_k \) are the recursion coefficients of polynomials orthonormal with respect to \( \mu \), \( g \) is as in (3.1) and both sides may be equal to +\( \infty \) simultaneously. With the decomposition of recursion coefficients, we are able to formulate the rate function \( \mathcal{I}_L \) in the LDP of the spectral
measure of the Laguerre ensemble. In this case, all measures are supported by \([0, \infty)\) and we have (Gamboa and Rouault (2008))

\[
I_L(\mu) = \sum_{k=1}^{\infty} g(z_{2k-1}) + \tau g(z_{2k}/\tau),
\]

with \(z_k\) as in (3.5). If the support of \(\mu\) is not a subset of \([0, \infty)\), we set \(I_L(\mu) = \infty\). Note that the minimum of \(I_L\) is attained for \(z_{2k-1} = 1, z_{2k} = \tau\), which corresponds to the Marchenko-Pastur law \(\text{MP}(\tau)\).

4 Proofs

Before we start to prove the main theorems of Section 2, we first prove some auxiliary results for the concentration and large deviations of scalar beta and gamma distributed random variables.

Lemma 4.1 Let \(X \sim \text{Beta}(a, b)\) be a Beta-distributed random variable, then for any \(\frac{1}{2} > \epsilon > 0\),

\[
P\left(|X - E[X]| > \epsilon \right) \leq 4 \exp\left\{ -\frac{\epsilon^2}{28} \cdot \frac{a^3 + b^3}{ab} \right\}
\]

Proof: Let \(Y \sim \text{Gamma}(a), Z \sim \text{Gamma}(b)\) be independent gamma distributed random variables with mean \(a\) and \(b\), respectively. We will make use of the equality in distribution

\[
X \overset{d}{=} \frac{Y}{Y + Z}.
\]

The standard Chernoff bounds for gamma distributed random variables give for \(\epsilon > 0\) the inequalities

\[
P(Y > a(1 + \epsilon)) \leq \exp\left( a(\log(1 + \epsilon) - \epsilon) \right), \quad P(Y < a(1 - \epsilon)) \leq \exp\left( a(\log(1 - \epsilon) + \epsilon) \right).
\]

Using that \(\log(1 - \epsilon) + \epsilon \leq \log(1 + \epsilon) - \epsilon \leq -\frac{1}{4} \epsilon^2\) for \(0 < \epsilon < \frac{1}{2}\), this yields

\[
P\left(|\frac{1}{a} Y - 1| > \epsilon \right) \leq 2 \exp\left( -\frac{a^2 \epsilon^2}{4} \right)
\]

To get an inequality for the beta distribution note that \(|A - 1| < \epsilon\) and \(|B - 1| < \epsilon\) implies for any \(c \geq 0\)

\[
\left| \frac{A}{cA + B} - \frac{1}{1 + c} \right| \leq \frac{|A - 1| + |B - 1|}{(cA + B)(c + 1)} \leq \frac{2\epsilon}{(1 - \epsilon)(c + 1)^2} \leq 4\epsilon,
\]
where we used \( \varepsilon < \frac{1}{2} \) for the last inequality. Combining this implication with \( c = \frac{a}{b} \) with the Chernoff bounds for \( \frac{1}{a} Y \) and \( \frac{1}{b} Z \) gives

\[
P \left( \frac{b}{a} |X - E[X]| > \varepsilon \right) = P \left( \left| \frac{1}{a} Y - \frac{1}{b} Z - \frac{a}{b} + 1 \right| > \varepsilon \right) \leq 2 \exp \left( -a \frac{\varepsilon^2}{64} \right) + 2 \exp \left( -b \frac{\varepsilon^2}{64} \right).
\]

On the other hand, the symmetry of the beta distribution yields also the inequality

\[
P \left( \frac{a}{b} |X - E[X]| > \varepsilon \right) = P \left( \frac{a}{b} |(1 - X) - E[(1 - X)]| > \varepsilon \right)
= P \left( \left| \frac{1}{b} Z - \frac{b}{a} + 1 \right| > \varepsilon \right)
\leq 2 \exp \left( -a \frac{\varepsilon^2}{64} \right) + 2 \exp \left( -b \frac{\varepsilon^2}{64} \right).
\]

These two inequalities together give

\[
P \left( \max\{\frac{a}{b}, \frac{b}{a}\} |X - E[X]| > \varepsilon \right) \leq 4 \exp \left( -\min\{a, b\} \frac{\varepsilon^2}{64} \right),
\]

which implies

\[
P \left( |X - E[X]| > \varepsilon \right) \leq 4 \exp \left( -\frac{\varepsilon^2}{64} \min\{a, b\} \cdot (\max\{\frac{a}{b}, \frac{b}{a}\})^2 \right)
= 4 \exp \left( -\frac{\varepsilon^2}{64} \max\{\frac{a^2}{b^2}, \frac{b^2}{a^2}\} \right) \leq 4 \exp \left( -\frac{\varepsilon^2}{64} \cdot \frac{1}{2} \left( \frac{a^2}{b^2} + \frac{b^2}{a^2} \right) \right),
\]

and we obtain the stated inequality.

\[\Box\]

The following theorem shows that the distance between the empirical eigenvalue distribution and the corresponding spectral measure goes to zero almost surely. Here, we consider the Kolmogorov (or uniform) distance defined for measures \( \mu, \nu \) on the real line with distribution functions \( F_\mu \) and \( F_\nu \), respectively, by

\[
d_K(\mu, \nu) = \sup_{x \in \mathbb{R}} |F_\mu(x) - F_\nu(x)|.
\]

Note that convergence in the Kolmogorov metric implies weak convergence.

**Theorem 4.2** For the empirical eigenvalue distribution \( \hat{\mu}_n \) defined in (1.2) and the corresponding spectral measure \( \mu_n \) as in (1.4), we have

\[
d_K(\hat{\mu}_n, \mu_n) \xrightarrow{n \to \infty} 0
\]

almost surely.
**Proof:** The measures \( \hat{\mu}_n \) and \( \mu_n \) have the same support points \( \lambda_1, \ldots, \lambda_n \), which implies for the Kolmogorov distance

\[
d_K(\hat{\mu}_n, \mu_n) = \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} \left( w_i - \frac{1}{n} \right) \right|.
\]

Recall that the weight vector \((w_1, \ldots, w_n)\) is Dirichlet distributed with parameter \( \beta' \) and has the same distribution as

\[
\left( \frac{G_1}{G_1 + \cdots + G_n}, \frac{G_2}{G_1 + \cdots + G_n}, \ldots, \frac{G_1}{G_1 + \cdots + G_n} \right)
\]

where \( G_1, \ldots, G_n \) are independent Gamma(\( \beta' \)) distributed. Therefore, the sum \( w_1 + \cdots + w_k \) is Beta(\( k\beta', (n - k)\beta' \)) distributed with mean \( \frac{k}{n} \). Applying Lemma 4.1 gives for \( 0 < \varepsilon < \frac{1}{4} \)

\[
P \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} \left( w_i - \frac{1}{n} \right) \right| > \varepsilon \right) \leq \sum_{k=1}^{n} P \left( \left| w_1 + \cdots + w_k - \frac{k}{n} \right| > \varepsilon \right)
\]

\[
\leq \sum_{k=1}^{n} 4 \exp \left( -\frac{\varepsilon^2}{128} \cdot \beta' \cdot \frac{k^3 + (n - k)^3}{k(n - k)} \right)
\]

\[
= \sum_{k=1}^{n} 4 \exp \left( -\frac{\varepsilon^2}{128} \cdot \beta' \cdot \frac{k^3 + (1 - k)^3}{k(1 - k)} \right)
\]

\[
\leq 4n \exp \left( -\frac{\varepsilon^2}{128} \cdot \beta' n \right),
\]

where for the last inequality, note that \( x \mapsto x^3 + (1 - x)^3 \) is minimal on \([0, 1]\) in \( \frac{1}{2} \). The almost sure convergence of \( d_K(\hat{\mu}_n, \mu_n) \) follows then from an application of the Borel-Cantelli Lemma. \( \square \)

**Lemma 4.3** Let \( G_1 \sim \text{Gamma}(\tilde{\alpha}_n) \) and \( G_2 \sim \text{Gamma}(\tilde{\beta}_n) \) denote independent random variables and let \( \alpha_n \) and \( \beta_n \) be positive real numbers going to \( \infty \) such that

\[
\lim_{n \to \infty} \tilde{\alpha}_n = \alpha_0 \in (0, \infty), \quad \lim_{n \to \infty} \tilde{\beta}_n = \beta_0 \in (0, \infty).
\]

(i) If \( \beta_n / \alpha_n \to \infty \), then the vector

\[
G_n = \left( \frac{1}{\alpha_n} G_1, \frac{1}{\sqrt{\alpha_n \beta_n}} G_1, \frac{1}{\beta_n} G_1, \frac{1}{\beta_n} G_2 \right)
\]

satisfies an LDP with speed \( \alpha_n \) and good rate function

\[
I(x_1, x_2, x_3, x_4) = \alpha_0 \cdot g(\alpha_0^{-1} x_1)
\]

if \( x_1 > 0, x_2 = x_3 = 0, x_4 = \beta_0 \) and \( I(x_1, x_2, x_3, x_4) = \infty \) otherwise.

13
(ii) If for another set $\gamma_n$ of positive real numbers with $\gamma_n/\alpha_n \to 0$ the convergence

$$\frac{\tilde{\alpha}_n - \tilde{\beta}_n}{\sqrt{\gamma_n \alpha_n}} \xrightarrow{n \to \infty} 0$$

holds, then

$$\mathcal{G}_n = \left( \frac{1}{\alpha_n} G_1, \frac{1}{\alpha_n} G_2, \frac{1}{2 \sqrt{\gamma_n \alpha_n}} (G_1 - G_2) \right)$$

satisfies an LDP with speed $\gamma_n$ and good rate function

$$I(x_1, x_2, x_3) = \alpha_0^{-1} x_3^2$$

if $x_1 = x_2 = \alpha_0$ and $I(x_1, x_2, x_3) = \infty$ otherwise.

**Proof:** In the situation (i), the logarithmic moment generating of $\alpha_n \mathcal{G}_n$ is given by

$$\log E \left[ \exp \left\{ \alpha_n \langle t, \mathcal{G}_n \rangle \right\} \right] = \log E \left[ \exp \left\{ G_1 \left( t_1 + t_2 \sqrt{\frac{\alpha_n}{\beta_n}} + t_3 \frac{\alpha_n}{\beta_n} \right) + G_2 \cdot t_4 \frac{\alpha_n}{\beta_n} \right\} \right]$$

$$= \log \left( \frac{1}{1 - t_1 - t_2 \sqrt{\frac{\alpha_n}{\beta_n}} - t_3 \frac{\alpha_n}{\beta_n}} \right)^{-\alpha_n} + \frac{\tilde{\beta}_n}{\beta_n} \log \left( \frac{1}{1 - t_4 \frac{\alpha_n}{\beta_n}} \right)^{-\beta_n},$$

so we obtain the limit

$$\lim_{n \to \infty} \alpha_n^{-1} \log E \left[ \exp \left\{ \alpha_n \langle t, \mathcal{G}_n \rangle \right\} \right] = -\alpha_0 \log(1 - t_1) + \beta_0 t_4 =: \mathcal{C}(t).$$

The function $\mathcal{C}$ is finite and differentiable on $(-\infty, 1) \times \mathbb{R}^3$ and steep. Then the Gärtner-Ellis Theorem (see [Dembo and Zeitouni (1998)]) yields an LDP for $\mathcal{G}_n$ with speed $\alpha_n$ and good rate function

$$I(x) = \sup_{t \in \mathbb{R}^4} \{ \langle t, x \rangle - \mathcal{C}(t) \},$$

which is infinite unless $x_1 > 0, x_2 = x_3 = 0, x_4 = \beta_0$, in which case

$$I(x) = \sup_{t_1 \in \mathbb{R}} \{ t_1 x_1 + \alpha_0 \log(1 - t_1) \} = x_1 - \alpha_0 - \alpha_0 \log(\alpha_0^{-1} x_1).$$
This proves part (i). For part (ii), we get for the logarithmic moment generating function

\[
\gamma_n^{-1} \log E \left[ \exp \left\{ \gamma_n \langle t, G_n \rangle \right\} \right] \\
= \gamma_n^{-1} \log \left[ \exp \left\{ G_1 \left( t \frac{\gamma_n}{\alpha_n} + t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right) + G_2 \left( t_2 \frac{\gamma_n}{\alpha_n} - t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right) \right\} \right] \\
= \gamma_n^{-1} \log \left( 1 - t_1 \frac{\gamma_n}{\alpha_n} - t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right)^{-\tilde{\alpha}_n} + \gamma_n^{-1} \log \left( 1 - t_2 \frac{\gamma_n}{\alpha_n} + t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right)^{-\tilde{\beta}_n} \\
= \gamma_n^{-1} \log \left( \left( 1 - t_1 \frac{\gamma_n}{\alpha_n} - t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right) \left( 1 - t_2 \frac{\gamma_n}{\alpha_n} + t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right) \right)^{-\tilde{\alpha}_n} \\
+ \log \left( 1 - t_2 \frac{\gamma_n}{\alpha_n} + t_3 \frac{1}{2} \sqrt{\frac{\gamma_n}{\alpha_n}} \right)^{\gamma_n^{-1}(\tilde{\alpha}_n - \tilde{\beta}_n)}.
\]

Due to the assumption \( \frac{\tilde{\alpha}_n - \tilde{\beta}_n}{\sqrt{\gamma_n \alpha_n}} \to 0 \), the second term goes to zero and since \( \gamma_n / \alpha_n \to 0 \), the first term can be written as

\[
-\frac{\tilde{\alpha}_n}{\alpha_n} \log \left( 1 - t_1 \frac{\gamma_n}{\alpha_n} - t_2 \frac{\gamma_n}{\alpha_n} - t_3 \frac{1}{4} \frac{\gamma_n}{\alpha_n} + o \left( \frac{\gamma_n}{\alpha_n} \right) \right)^{\alpha_n / \gamma_n},
\]

which has the limit

\[
\mathcal{C}(t) := \alpha_0 (t_1 + t_2 + \frac{1}{4} t_3^2).
\]

As in (i), this implies an LDP for \( G_n \) with speed \( \gamma_n \) and the stated rate function. \( \square \)

### 4.1 Proof of Theorem 2.1

As described in Section 3, the weak convergence of \( \mu_n \) follows, if we show entrywise convergence of the rescaled tridiagonal matrix

\[
(1 + \sigma) \sqrt{b_n \frac{1 + \sigma}{\sigma n \beta'}} \left( \mathcal{J}_n(\beta, a_n, b_n) - \frac{\sigma}{1 + \sigma} I_n \right)
\]

to the tridiagonal representation \( \mathcal{SC} \) of the semicircle law. By Lemma 4.2, this is equivalent to the weak convergence of the empirical measure \( \hat{\mu}_n \). We start with the sub-diagonal entries, whose square is given by

\[
b_n \frac{(1 + \sigma)^3}{\sigma n \beta'} p_{2k-1} (1 - p_{2k-2}) p_{2k} (1 - p_{2k-1}).
\]
The random variable \( b_n \frac{1 + \sigma}{n^{\beta'}} p_{2k} \) has mean

\[
\frac{b_n}{n^{\beta'}} \cdot \frac{(n - k) \beta'}{a_n + b_n - (2k - 1) \beta'} \xrightarrow{n \to \infty} 1
\]

and by Lemma 4.1 for \( n \) sufficiently large,

\[
P \left( \left| \frac{b_n}{n} p_{2k} - E \left[ \frac{b_n}{n} p_{2k} \right] \right| > \varepsilon \right) \leq 4 \exp \left\{ -\frac{\varepsilon^2}{128} \cdot \frac{n^2}{b_n} \cdot \frac{((n - k) \beta')^3 + (a_n + b_n - (n + k - 1) \beta')^3}{((n - k) \beta') (a_n + b_n - (n + k - 1) \beta')} \right\}
\]

The exponent is of order \(-\varepsilon^2 n\), so the right hand side is summable. The Borel-Cantelli Lemma implies

\[
(4.2) \quad \frac{b_n}{n^{\beta'}} \frac{1 + \sigma}{p_{2k}} \xrightarrow{n \to \infty} 1
\]

almost surely. Similarly, the random variables \( p_{2k-1} \) and \( p_{2k-2} \) concentrate exponentially fast around their mean, which converges to \( \frac{\sigma}{1 + \sigma} \) and 0, respectively. The almost sure convergence of the subdiagonal entries to 1 follows. The diagonal entries of the rescaled matrix are

\[
\sqrt{\frac{b_n}{n}} (1 + \sigma) \frac{p_{2k-2}(1 - p_{2k-3})}{\sigma n^{\beta'}} + \sqrt{\frac{b_n}{n}} \frac{(1 + \sigma)^3}{\sigma n^{\beta'}} (p_{2k-1} - \frac{\sigma}{1 + \sigma}) + \sqrt{\frac{b_n}{n}} \frac{(1 + \sigma)^3}{\sigma n^{\beta'}} p_{2k-1} p_{2k-2}.
\]

The almost sure convergence in (4.2) implies that the first and last summand vanish. For the second summand, note that

\[
P \left( \left| \sqrt{\frac{b_n}{n}} |p_{2k-1} - E[p_{2k-1}]| > \varepsilon \right) \leq 4 \exp \left\{ -\frac{\varepsilon^2}{128} \cdot \frac{n}{b_n} \cdot \frac{(a_n - (k - 1) \beta')^3 + (b_n - (k - 1) \beta')^3}{(a_n - (k - 1) \beta')(b_n - (k - 1) \beta')} \right\}
\]

and the right hand side is again summable. The assumption on the convergence speed of \( a_n/b_n \) to \( \sigma \) in Theorem 2.1 guarantees that

\[
\sqrt{\frac{b_n}{n}} \left( E[p_{2k-1}] - \frac{\sigma}{1 + \sigma} \right) = \sqrt{\frac{b_n}{n}} \left( \frac{a_n - (k - 1) \beta'}{a_n + b_n - (2k - 2) \beta'} - \frac{\sigma}{\sigma + 1} \right) = \frac{\sqrt{b_n} (a_n/b_n - \sigma) + \sigma/b_n (k - 1) \beta'}{\sqrt{b_n} (a_n/b_n + 1 - 2k - 2 \beta') (1 + \sigma)}
\]

vanishes as \( n \to \infty \) and therefore the second summand goes to zero almost surely. Consequently all diagonal entries go to zero and the subdiagonal entries converge to 1. The entrywise limit of \((1 + \sigma) \sqrt{b_n^{1+\sigma}} (\mathcal{J}_n(\beta, a_n, b_n) - \frac{\sigma}{1 + \sigma} I_n)\) is therefore the infinite matrix with spectral measure \( SC \). This concludes the proof of Theorem 2.2. \(\square\)
4.2 Proof of Theorem 2.2

The almost sure weak convergence of \( \mu_n \) is a direct consequence of Theorem 2.5, since the rate function in Theorem 2.5 is good with the unique minimizer MP(\( \tau \)) and the topology on \( M_1 \) is metrizable. Therefore, \( \mu_n \xrightarrow{n \to \infty} \text{MP}(\tau) \) and the convergence of \( \hat{\mu}_n \) with the same limit follows from Lemma 4.2.

\[ \blacksquare \]

4.3 Proof of Theorem 2.4

We start by showing an LDP for the independent beta distributed entries of the rescaled matrix

\[ 4\sqrt{\frac{b_n}{n\beta}} \left( J_n(\beta, a_n, b_n) - \frac{1}{2} I_n \right). \]

with diagonal elements

\[ (4\sqrt{\frac{b_n}{n\beta}} p_{2k-2}) (1 - p_{2k-3}) + 4\sqrt{\frac{b_n}{n\beta}} (p_{2k-1} - \frac{1}{2}) + p_{2k-1} \left( 4\sqrt{\frac{b_n}{n\beta}} p_{2k-2} \right) \]

and off-diagonal elements the square root of

\[ p_{2k-1} (1 - p_{2k-2}) \left( 16\frac{b_n}{n\beta} p_{2k} \right) (1 - p_{2k-1}). \]

Recalling the definition of the random variables \( p_k \) in (3.2) and the representation of the beta distribution in (4.1), we have for the canonical moments of odd index

\[ \left( p_{2k-1}, 4\sqrt{\frac{b_n}{n\beta}} (p_{2k-1} - \frac{1}{2}) \right) \overset{d}{=} \psi \left( \frac{1}{b_n} G_1, \frac{1}{b_n} G_2, \frac{1}{2\sqrt{b_n n\beta'}} (G_1 - G_2) \right), \]

where \( G_1 \sim \text{Gamma}(a_n - (k - 1)\beta') \) independent of \( G_2 \sim \text{Gamma}(b_n - (k - 1)\beta') \) and

\[ \psi(x_1, x_2, x_3) = \left( \frac{x_1}{x_1 + x_2}, \frac{4}{\sqrt{2}} \frac{x_3}{x_1 + x_2} \right). \]

The assumptions of Lemma 4.3 (ii) are satisfies, if we set \( \gamma_n = n\beta' \) and \( \alpha_n = b_n \), such that \( \alpha_0 = 1 \) and

\[ \frac{\tilde{\alpha}_n - \tilde{\beta}_n}{\sqrt{\gamma_n \alpha_n}} = \frac{a_n - b_n}{\sqrt{n\beta' b_n}} \xrightarrow{n \to \infty} 0. \]
We get that \((\frac{1}{b_n} G_1, \frac{1}{b_n} G_2, \frac{1}{2\sqrt{b_n n^\beta}}(G_1 - G_2))\) satisfies an LDP with speed \(n\beta'\) and good rate function \(I_0(x_1, x_2, x_3) = x_3^2\) if \(x_1 = x_2 = 1\) and \(I_0(x_1, x_2, x_3) = \infty\) otherwise. By the contraction principle, \((p_{2k-1}, 4\sqrt{\frac{b_n}{n^\beta}}(p_{2k-1} - \frac{1}{2}))\) satisfies the LDP with the same speed and good rate function

\[
I_1(y_1, y_2) = \inf \{I_0(x_1, x_2, x_3) | \psi(x_1, x_2, x_3) = (y_1, y_2)\} = \frac{1}{2} y_2^2,
\]

if \(y_1 = \frac{1}{2}\) and \(I_1(y_1, y_2) = \infty\) otherwise. This is immediate from the fact that on the set where \(I_1\) (or \(I_0\)) is finite, \(y_2 = \frac{1}{2} x_3\). Among the canonical moments with even index, we need to control

\[
(y_1, y_2, y_3) := \left( p_{2k}, 4\sqrt{\frac{b_n}{n^\beta}} p_{2k}, 16\frac{b_n}{n^\beta} p_{2k} \right) \overset{d}{=} \psi \left( \frac{1}{n^\beta} G_1, \frac{1}{\sqrt{b_n n^\beta}} G_1, \frac{1}{b_n} G_1, \frac{1}{b_n} G_2 \right),
\]

again with independent \(G_1 \sim \text{Gamma}((n - k)\beta'), G_2 \sim \text{Gamma}(a_n + b_n - (n - k - 1)\beta')\) and

\[
\psi(x_1, x_2, x_3, x_4) = \frac{x_3}{x_3 + x_4}, \left( \frac{4}{\sqrt{2}} \frac{x_2}{x_3 + x_4}, 8\frac{x_1}{x_3 + x_4} \right).
\]

Lemma 4.3 part (i), this time with \(\alpha_n = n\beta', \beta_n = b_n, \alpha_0 = 1, \beta_0 = 2\), shows that the vector of rescaled gamma distributed random variables satisfies an LDP with speed \(n\beta'\) and rate function

\[
I_2(x_1, x_2, x_3, x_4) = g(x_1)
\]

if \(x_1 > 0, x_2 = x_3 = 0, x_4 = 2\) and \(I_2(x_1, x_2, x_3, x_4) = \infty\) otherwise. The contraction principle gives an LDP for \((y_1, y_2, y_3)\), where the rate function \(I_3\) is finite only if \(y_1 = y_2 = 0, y_3 > 0\) and in this case, \(y_3 = 4x_1\) and

\[
I_3(y_1, y_2, y_3) = g \left( \frac{y_3}{4} \right).
\]

Since \(p_1, p_2, \ldots\) are independent, the Dawson-Gärtner Theorem (Theorem 4.6.1 in Dembo and Zeitouni (1998)) yields an LDP for the infinite vector

\[
y = (y_1, y_1, y_2, y_2, y_2, y_2, y_2, y_3, y_3, \ldots) = \left( p_1, 4\sqrt{\frac{b_n}{n^\beta}} (p_1 - \frac{1}{2}), p_2, 4\sqrt{\frac{b_n}{n^\beta}} p_2, 16\frac{b_n}{n^\beta} p_2, p_3, \ldots \right)
\]

with speed \(n\beta'\) and good rate function

\[
I_y = \sum_{k=1}^{\infty} I_1(y_{2k-1}, y_{2k-1,2}) + I_3(y_{2k,1}, y_{2k,2}, y_{2k,3}),
\]

finite only if \(y_{2k-1,1} = \frac{1}{2}, y_{2k,1} = y_{2k,2} = 0\) and \(y_{2k,3} > 0\) and in this case,

\[
I_y(y) = \sum_{k=1}^{\infty} \frac{1}{2} y_{2k-1,2}^2 + g \left( \frac{y_{2k,3}}{4} \right).
\]
The entries of the rescaled tridiagonal matrix, which are the recursion coefficients of the spectral measure, are given by

\[ c_k^2 = y_{2k-1,1} (1 - y_{2k-2,1}) y_{2k,3} (1 - y_{2k-1,1}), \]
\[ d_k = y_{2k-2,2} (1 - y_{2k-3,1}) + y_{2k-2,2} y_{2k-1,1} + y_{2k-1,2}. \]

In particular, the recursion coefficients are a continuous function of \( y \). Applying the contraction principle once more, we can set \( c_k^2 = \frac{1}{4} y_{2k,3} \) and \( d_k = y_{2k-1,2} \) and obtain that the sequence \( r = (d_1, c_1, d_2, \ldots) \) of recursion coefficients satisfies the LDP with speed \( n \beta' \) and good rate

\[ I_r(d_1, c_1, d_2, \ldots) = \sum_{k=1}^{\infty} \frac{1}{2} d_k^2 + g(c_k^2). \]

To transfer this LDP to the spectral measure, let \( \mathcal{R}_c \) denote the set of sequences \( r = r(\mu) \) of measures \( \mu \in \mathcal{M}_c \) with compact support. The mapping \( r(\mu) \mapsto \mu \) is well-defined on \( \mathcal{R}_c \) and continuous and so the contraction principle yields the LDP for the spectral measure \( \mu_n \) with rate function \( \mathcal{I}(\mu) = I_r(r(\mu)) = \mathcal{I}_L(\mu) \).

4.4 Proof of Theorem 2.5:

We need to prove an LDP for the entries of the rescaled matrix

\[ \frac{b_n}{a_n} \mathcal{J}_n(\beta, a_n, b_n), \]

which has diagonal entries

\[ \frac{b_n}{a_n} p_{2k-2} (1 - p_{2k-3}) + \frac{b_n}{a_n} p_{2k-1} (1 - p_{2k-2}) \]

and the squared off-diagonal entries are

\[ \frac{b_n}{a_n} p_{2k-1} (1 - p_{2k-2}) \cdot \frac{b_n}{a_n} p_{2k} (1 - p_{2k-1}). \]

The canonical moments with odd index appear in the matrix in two ways, and can be written as

\[ \left( p_{2k-1}, \frac{b_n}{a_n} p_{2k-1} \right) \overset{d}{=} \phi \left( \frac{1}{a_n} G_1, \frac{1}{b_n} G_1, \frac{1}{b_n} G_2 \right), \]

where \( G_1 \sim \text{Gamma}(a_n - (k - 1)\beta') \) independent of \( G_2 \sim \text{Gamma}(b_n - (k - 1)\beta') \) and

\[ \phi(x_1, x_2, x_3) = \left( \frac{x_2}{x_2 + x_3}, \frac{x_1}{x_2 + x_3} \right). \]
From the first part of Lemma 4.3 with \( \alpha_0 = \beta_0 = 1 \), we see that the random vector 
\[
\left( \frac{1}{a_n} G_1, \frac{1}{b_n} G_2 \right)
\]
contains the recursion variables of the spectral measure \( \mu_n \) of the rescaled random matrix. It remains to apply the continuous mapping 
\[
(z_1, z_2, \ldots) \mapsto \left( \frac{1}{a_n} G_1, \frac{1}{b_n} G_2 \right),
\]
with \( G_1 \sim \text{Gamma}((n - k)\beta') \) and \( G_2 \sim \text{Gamma}(a_n + b_n - (n - k - 1)\beta') \) independent and \( \phi \) as above. Applying again the first part of Lemma 4.3 with \( \alpha_0 = \tau \) and \( \beta_0 \), we get an LDP for the vector of gamma distributed random variables with speed \( a_n \) and good rate function \( I_2(x_1, x_2, x_3) = \tau g(x_1/\tau) \) for \( x_1 > 0, x_2 = 0, x_3 = 1 \). The contraction principle gives the LDP for \( (p_{2k}, \frac{b_n}{a_n} p_{2k}) \) with rate function \( I_3(y_1, y_2) = \tau I_1(y_1, y_2/\tau) \). Collecting the independent canonical moments we get a projective LDP for
\[
I_y(y) = \sum_{k=1}^{\infty} I_1(y_{2k-1,1}, y_{2k-1,2}) + \tau I_1(y_{2k,1}, y_{2k,2}/\tau).
\]
We now perform a transformation to the random variables
\[
z = (z_1, z_2, \ldots) = \left( \frac{b_n}{a_n} p_1, p_2, \frac{b_n}{a_n} p_2, \ldots \right)
\]
with speed \( a_n \) and good rate function
\[
I_z(z) = \sum_{k=1}^{\infty} I_1(0, z_{2k-1}) + \tau I_1(0, z_{2k}/\tau).
\]
To complete the proof, note that \( z \) contains the recursion variables of the spectral measure \( \mu_n \) of the rescaled random matrix. It remains to apply the continuous mapping \( z(\mu) \mapsto \mu \) from the recursion variables to the spectral measure analogous to the step in Section 4.3 to complete the proof. \( \square \)
References

Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*. Cambridge University Press, Cambridge.

Collins, B. (2005). Product of random projections, Jacobi ensembles and universality problems arising from free probability. *Probab. Theory Relat. Fields*, 133:315–344.

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

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.

Dette, H. and Nagel, J. (2009). Some asymptotic properties of the spectrum of the jacobi ensemble. *SIAM J. Math. Anal.*, 41:1491–1507.

Dette, H. and Nagel, J. (2012). Distributions on unbounded moment spaces and random moment sequences. *Ann. Probab.*, 40:2690–2704.

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

Dunford, N. and Schwartz, J. T. (1963). *Linear Operators Part II: Spectral Theory*. Interscience Publishers, New York.

Dyson, F. J. (1962). Statistical theory of the energy levels of complex systems. i-iii. *J. Math. Phys.*, 3:140–175.

Forrester, P. J. (2010). *Log-gases and random matrices*. Princeton University Press, Princeton, NJ.

Gamboa, F., Nagel, J., Rouault, A., and Wagener, J. (2012). Large deviations for random matricial moment problems. *J. Multivar. Anal.*, 106:17–35.

Gamboa, F. and Rouault, A. (2008). Large deviations for random spectral measures and sum rules. *arXiv.org*, arXiv:0804.4322v1 [math.PR].

Gamboa, F. and Rouault, A. (2009). Canonical moments and random spectral measures. *J. Theoret. Probab.*, DOI 10.1007/s10959-009-0239-1.
Hiai, F. and Petz, D. (2006). Large deviations for functions of two random projection matrices. *Acta Sci. Math.*, 72(3):581–609.

Jiang, T. (2009). Approximation of haar distributed matrices and limiting distributions of eigenvalues of jacobi ensembles. *Probab. Theory Relat. Fields*, 144:221–246.

Jiang, T. (2012). Limit theorems for beta-jacobi ensembles. *arXiv.org*, arXiv:0911.2262 [math.PR].

Johnstone, I. M. (2008). Multivariate analysis and jacobi ensembles: Largest eigenvalue, tracy-widom limits and rates of convergence. *Ann. Stat.*, 36:2638–2716.

Killip, R. and Nenciu, I. (2004). Matrix models for circular ensembles. *Int. Math. Res. Not.*, 50:2665–2701.

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

Leff, H. S. (1964). Class of ensembles in the statistical theory of energy-level spectra. *J. Math. Phys.*, 5:763–768.

Lytova, A. and Pastur, L. (2009a). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37:1778–1840.

Lytova, A. and Pastur, L. (2009b). Fluctuations of matrix elements of regular functions of gaussian random matrices. *J. Stat. Phys.*, 134:147–159.

Mehta, M. L. (2004). *Random Matrices*. Elsevier/Academic Press, Amsterdam.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.

Nagao, T. and Wadati, M. (1993). Eigenvalue distribution of random matrices at the spectrum edge. *J. Phys. Soc. Jpn.*, 62:3845–3856.

Simon, B. (1997). The classical moment problem as a self-adjoint finite difference operator. *Adv. Math*, 137:82–203.

Simon, B. (2005a). *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. Colloquium Publications. American Mathematical Society 54, Part 1. Providence, RI.
Simon, B. (2005b). *Orthogonal polynomials on the unit circle. Part 2: spectral theory*. American Mathematical Society Colloquium Publications. Colloquium Publications. American Mathematical Society 54, Part 2. Providence, RI.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.*, 62:548–564.