A matrix version of a higher-order Szegö theorem
Alain Rouault

To cite this version:
Alain Rouault. A matrix version of a higher-order Szegö theorem. 2020. hal-02776087v2

HAL Id: hal-02776087
https://hal.archives-ouvertes.fr/hal-02776087v2
Preprint submitted on 26 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A matrix version of a higher-order Szegő theorem

Alain Rouault

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

Abstract

We extend a higher-order sum rule proved by B. Simon to matrix valued measures on the unit circle and their matrix Verblunsky coefficients.

Keywords: Sum rules, Szegő’s theorem, Verblunsky coefficients, matrix measures on the unit circle, relative entropy

1. Introduction

A probability measure $\mu$ on the unit circle $T$ with infinite support is characterized by its Verblunsky coefficients $(\alpha_j(\mu))_{j \geq 0}$, elements in the interior of the unit disc. They are associated with the Szegő recursion of orthogonal polynomials in $L^2(T, d\mu)$. A sum rule is an identity between an entropy-like functional of this measure and a functional of the sequence of its Verblunsky coefficients (for short, we say ”V-coefficients” in the sequel). The most famous is Szegő’s theorem.

Theorem 1.1. Let $d\mu = w(\theta) \frac{d\theta}{2\pi} + d\mu_s$ be the Lebesgue decomposition of a probability measure on $T$ and let $(\alpha_n)_{n \geq 0}$ its V-coefficients. Then

$$\int_0^{2\pi} \log w(\theta) \frac{d\theta}{2\pi} = \sum_{k=0}^{\infty} \frac{1}{2} \log (1 - |\alpha_k|^2),$$

where both members can be simultaneously finite or $-\infty$.

In his book [7], B. Simon proved the following statement (higher-order Szegő theorem).

Theorem 1.2 ([7] Th. 2.8.1). Let $d\mu = w(\theta) \frac{d\theta}{2\pi} + d\mu_s$ be a probability measure on $T$ and let $(\alpha_n)_{n \geq 0}$ its V-coefficients. Then

$$\int_0^{2\pi} (1 - \cos \theta) \log w(\theta) \frac{d\theta}{2\pi} = \frac{1}{2} (1 - |1 + \alpha_0|^2) - \frac{1}{2} \sum_{k=0}^{\infty} |\alpha_{k+1} - \alpha_k|^2$$

$$+ \sum_{k=0}^{\infty} \left( \log (1 - |\alpha_k|^2) + |\alpha_k|^2 \right),$$

where both members can be simultaneously finite or $-\infty$. 
Actually this formula may be written in terms of entropies. For probability
measures \( \nu \) and \( \mu \) on \( T \), let \( \mathcal{K}(\nu|\mu) \) denote the Kullback-Leibler divergence or relative entropy of \( \nu \) with respect to \( \mu \):

\[
\mathcal{K}(\nu|\mu) = \begin{cases} 
\int_T \log \frac{d\nu}{d\mu} \, d\nu & \text{if } \nu \text{ is absolutely continuous with respect to } \mu, \\
\infty & \text{otherwise.}
\end{cases}
\]  

(1.3)

Usually, \( \mu \) is the reference measure. Here the spectral side will involve the reversed Kullback-Leibler divergence, where \( \nu \) is the reference measure and \( \mu \) is the argument. In this case, we have that \( \mathcal{K}(\nu|\mu) \) is finite if and only if

\[
\int_0^{2\pi} \log w(\theta) \, d\nu(\theta) > -\infty,
\]  

(1.4)

where \( d\mu = w(\theta)d\nu(\theta) + d\mu_s \) is the Lebesgue decomposition of \( \mu \) with respect to \( \nu \). If we denote

\[
d\lambda_0(\theta) = \frac{d\theta}{2\pi}, \quad d\lambda_1(\theta) = (1 - \cos \theta) \frac{d\theta}{2\pi}
\]  

(1.5)

the sum rule (1.1) may be written

\[
\mathcal{K}(\lambda_0|\mu) = -\sum_0^{\infty} \log (1 - |\alpha_k|^2),
\]  

(1.6)

and the sum rule (1.2) may be written

\[
\mathcal{K}(\lambda_1|\mu) = \mathcal{K}(\lambda_1|\lambda_0) + \text{Re} \alpha_0 + \frac{|\alpha_0|^2}{2} + \frac{1}{2} \sum_0^{\infty} |\alpha_{k+1} - \alpha_k|^2 \\
- \sum_0^{\infty} \left( \log (1 - |\alpha_k|^2) + |\alpha_k|^2 \right)
\]  

(1.7)

with

\[
\mathcal{K}(\lambda_1|\lambda_0) = \int_0^{2\pi} (1 - \cos \theta) \log(1 - \cos \theta) \frac{d\theta}{2\pi} = 1 - \log 2.
\]

In (1.7) both sides may be infinite simultaneously, and they are finite if and only if

\[
\sum_k \alpha_k^4 + |\alpha_{k+1} - \alpha_k|^2 < \infty.
\]  

(1.8)

Actually, it is easy to include (1.7) and (1.6) into a family of sum rules depending on a parameter \( g \) such that \( |g| \leq 1 \). Let

\[
d\lambda_g(\theta) = (1 - g \cos \theta) \, d\lambda_0(\theta)
\]  

(1.9)
(called one single nontrivial moment in [7] p. 86). Combining (1.7) and Szgő’s formula, we get, as mentioned in [5] Cor. 5.4:

$$K(\lambda g | \mu) = K(\lambda g | \lambda_0) + g \left( \text{Re } \alpha_0 + \frac{|\alpha_0|^2}{2} + \frac{1}{2} \sum_{k=1}^{\infty} |\alpha_k - \alpha_{k-1}|^2 \right)$$

$$+ \sum_{0}^{\infty} - \log(1 - |\alpha_k|^2) - g|\alpha_k|^2,$$

(1.10)

where

$$K(\lambda g | \lambda_0) = \int (1 - g \cos \theta) \log(1 - g \cos \theta) \frac{d\theta}{2\pi} = 1 - \sqrt{1 - g^2} + \log \frac{1 + \sqrt{1 - g^2}}{2}.$$  

(1.11)

It may be called GW sum rule, since $\lambda_g$ is the equilibrium measure in a random matrix model due to Gross and Witten ([6]).

For $g = 0$, we recover (1.1) formula and when $g = 1$, we recover (1.7).

Simon’s proof of Theorem 1.2 (see Sect. 2.8 in [7]) was based on the use of the Szegő function

$$D(z) = \exp \int_{0}^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log w(\theta) \frac{d\theta}{4\pi},$$

the asymptotics of the orthogonal polynomial and Szegő’s theorem. Later on, Simon gave another proof of this theorem in Sect. 2.8 of [9]. The new proof uses a relative Szegő function and a step-by-step sum rule provided by the coefficient stripping.

In a series of papers, Gamboa et al. tackled sum rules on the real line and on the unit circle on a probabilistic way, using large deviations techniques. The main argument is the uniqueness of the rate function when the large deviations of a random measure are considered under two different encodings. In particular, in [9], they (re)proved Szegő’s theorem as a sum rule, stated a new sum rule for the Hua-Pickrell measure, and asked for a possible probabilistic proof of the higher-order sum rule quoted above. Shortly after, Simon et al. [1] gave that proof.

It turns out that probabilistic tools are robust enough to be extended to matrix measures, which allowed Gamboa et al. to give a probabilistic proof of the famous matrix Szegő’s theorem of Delsarte et al. [3] involving matrix V-coefficients. With the notations of the following section, this theorem says that if $d\mu = w(\theta)d\lambda_0 + d\mu_\ast$ is a non-trivial matrix-measure, then

$$\int_{0}^{2\pi} \log \det w(\theta)d\lambda_0(\theta) = \sum_{0}^{\infty} \log \det(1 - \alpha_k \alpha_k^\dagger).$$  

(1.12)

---

1See references in [3].

2We use $\dagger$ for matrix adjoint, keeping the notation $^\ast$ for reversed polynomials.
In [5] the authors proved also a matrix version of the Hua-Pickrell sum rule and conjectured a matrix version of the GW sum rule (1.10).

These considerations open the way to two challenges: analytical proof and probabilistic proof. The second way seems accessible by combining the machinery of [5] and of [1], i.e. a large deviation for a random measure encoded by its V-coefficients, but it seems more natural to begin with the first way, which will be done in this note. Of course, a possible issue comes from the non-commutativity of the product of matrices, but as usual, the story ends well.

We present the notations and main results in Sect. 2.1. Theorem 2.2 is a matrix-version of (1.10) and Prop. 2.3 is a gem i.e. a condition of finiteness of the entropy. In Sect. 3, we give the proof of the first result, involving the coefficient stripping method and a limiting argument. In Sect. 4 we give the proof of the gem. Finally Sect. 5 is devoted to the proofs of intermediate results.

2. Notations and main result

2.1. Notations

Let us begin with some introductory elements on matrix measures. For a more detailed exposition, see [2] Sect. 1, [4] Sect. 4, [5] Sect. 6.

Let $p > 1$ be an integer and let $\mathcal{M}_p$ be the set of complex $p \times p$ matrix measures $\mu$ on $T$ which are Hermitian, nonnegative and normalized by $\mu(T) = 1$ (the $p \times p$ identity matrix). A matrix measure is called quasi-scalar if it may be written $1 \cdot \sigma$ with $\sigma$ a probability measure on $T$. A $p \times p$ matrix polynomial is a polynomial with coefficients in $\mathbb{C}^{p \times p}$. Given a measure $\mu \in \mathcal{M}_p$, we define two inner products on the space of $p \times p$ matrix polynomials by setting

$$\langle\langle f, g \rangle\rangle_R = \int f(e^{i\theta})^\dagger d\mu(\theta) g(e^{i\theta})$$
$$\langle\langle f, g \rangle\rangle_L = \int g(e^{i\theta}) d\mu(\theta) f(e^{i\theta})^\dagger .$$

A sequence of matrix polynomials $(\varphi_j)$ is called right-orthonormal if, and only if,

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

A matrix measure is called non-trivial if

$$\text{tr} \langle\langle f, f \rangle\rangle_R > 0$$

for every non-zero polynomial $f$. We define the right monic matrix orthogonal polynomials $\Phi^R_k$ by applying the block Gram-Schmidt algorithm to the sequence $\{1, z^1, z^2 1, \ldots\}$. In other words, $\Phi^R_k$ is the unique matrix polynomial $\Phi^R_k(z) = z^k 1 + $ lower order terms, such that $\langle\langle z^j 1, \Phi^R_k \rangle\rangle_R = 0$ for $j = 0, \ldots, k - 1$. The normalized orthogonal polynomials are defined by

$$\varphi^R_0 = 1 , \quad \varphi^R_k = \Phi^R_k \kappa^R_k .$$
Here the sequence of $p \times p$ matrices $(\kappa_k^R)$ satisfies, for all $k$, the condition 
$(\kappa_k^R)^{-1} \kappa_{k+1}^R > 0_p$ and is such that the sequence $(\varphi_k^R)$ is orthonormal. We 
define the sequence of left-orthonormal polynomials $(\varphi_k^L)$ in the same way except that the above condition is replaced by $\kappa_k^L (\kappa_k^L)^{-1} > 0$. The matrix Szegő recursion is then

\begin{align*}
 z\varphi_k^L - \rho_k^L \varphi_{k+1}^L &= \alpha_k^\dagger (\varphi_k^R)^* \quad (2.1) \\
 z\varphi_k^R - \rho_k^R \varphi_{k+1}^R &= (\varphi_k^L)^* \alpha_k^\dagger, \quad (2.2)
\end{align*}

where for all $k \in \mathbb{N}_0$,

- $\alpha_k$ belongs to $\mathcal{B}_p$, the closed unit ball of $\mathbb{C}^{p \times p}$ defined by
  \[ \mathcal{B}_p := \{ M \in \mathbb{C}^{p \times p} : MM^\dagger \leq 1 \}, \quad (2.3) \]

- $\rho_k^R$ and $\rho_k^L$ are the so-called defect matrices defined by
  \[ \rho_k^R := \left( 1 - \alpha_k \alpha_k^\dagger \right)^{1/2}, \quad \rho_k^L = \left( 1 - \alpha_k^\dagger \alpha_k \right)^{1/2} \quad (2.4) \]

- for a matrix polynomial $P$ with degree $k$, the reversed polynomial $P^*$ is defined by
  \[ P^*(z) := z^k P(1/z)^\dagger. \]

Verblunsky’s theorem establishes a one-to-one correspondence between non-trivial (normalized) matrix measures on $\mathbb{T}$ and sequences of elements in the interior of $\mathcal{B}_p$ (Theorem 3.12 in [2]).

In an alternative way, these V-coefficients may be introduced as matrix Schur coefficients as follows. Let $F$ be the Caratheodory (or Herglotz) transform of $\mu$ defined by:

\[ F(z) = \int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta), \quad z \in \mathbb{D} = \{ z : |z| < 1 \}, \]

and $f$ the Schur transform defined by:

\[ f(z) = z^{-1}(F(z) - 1)(F(z) + 1)^{-1}, \]

which is equivalent to

\[ F(z) = (1 + zf(z))(1 - zf(z))^{-1}. \quad (2.5) \]

The Schur recursion is defined as follows. At step 0 we set

\[ \alpha_0 = f(0), \]

which gives the first V-coefficient. We define the defect matrices (right and left) by

\[ \rho_0^R = (1 - \alpha_0 \alpha_0^\dagger)^{1/2}, \quad \rho_0^L = (1 - \alpha_0^\dagger \alpha_0)^{1/2}, \quad (2.6) \]
and then, at step 1 we set

$$Sf := f_1 = z^{-1}(\rho_0^R)^{-1}(f(z) - \alpha_0) \left(1 - \alpha_0^\dag f(z)\right)^{-1} \rho_0^L \quad (2.7)$$

and the second V-coefficient is

$$\alpha_1 = f_1(0).$$

The other coefficients are defined with the same algorithm

$$f_{k+1} = Sf_k, \quad \alpha_{k+1} = f_{k+1}(0), \ldots.$$  

The following theorem gives the connection between $F$ and the absolutely continuous part of $\mu$.

**Theorem 2.1** ([2] Prop. 3.16). For $z \in \mathbb{D}$, we have

$$\text{Re } F(z) = (1 - \bar{z}f(z)^\dag)^{-1}(1 - |z|^2 f(z)f(z)^\dag)(1 - z f(z))^{-1}. \quad (2.8)$$

and the non-tangential boundary values $\text{Re } F(e^{i\theta})$ and $f(e^{i\theta})$ exist for a.e. $\theta$.

If $\mu$ is a normalized matrix measure with Lebesgue decomposition

$$d\mu(\theta) = w(\theta)d\lambda_0(\theta) + d\mu_s(\theta)$$

(where $w$ is a $p \times p$ matrix), then for a.e. $\theta$

$$w(\theta) = \text{Re } F(e^{i\theta}),$$

and for a.e. $\theta$, $\det w(\theta) = 0$ if and only if $f(e^{i\theta})^\dag f(e^{i\theta}) < 1$.

**2.2. Main result**

When $\Sigma = 1 \cdot \sigma$ is a pseudo-scalar measure and $d\mu(\theta) = h(\theta)d\sigma(\theta) + d\mu_s(\theta)$, we define the relative entropy

$$K(\Sigma | \mu) = -\int \log \det h(\theta)d\sigma(\theta). \quad (2.9)$$

We will consider two reference measures:

$$d\Lambda_0(\theta) = 1 \cdot d\lambda_0(\theta), \quad d\Lambda_g(\theta) = 1 \cdot d\lambda_g(\theta). \quad (2.10)$$

Our main result is the following.

**Theorem 2.2.** For $|g| \leq 1$, let $d\mu(\theta) = w(\theta)d\lambda_0(\theta) + d\mu_s(\theta)$ be a non-trivial matrix measure, then

$$\int_0^{2\pi} (1 - g \cos \theta) \log \det w(\theta)d\lambda_0(\theta) = \sum_{0}^{\infty} \log \det(1 - \alpha_k\alpha_k^\dag) - T(\alpha_0, \alpha_1, \cdots) \quad (2.11)$$
with
\[ T(\alpha_0, \alpha_1, \cdots) := \text{Re tr} (\alpha_0 - \sum_0^\infty \alpha_k \alpha_{k+1}^\dagger), \] (2.12)

or in an equivalent form
\[ K(\Lambda_0 \mid \mu) = K(\Lambda_g \mid \lambda_0) - \sum_0^\infty \log \det (1 - \alpha_k \alpha_k^\dagger) + g T(\alpha_0, \alpha_1, \cdots). \] (2.13)

In (2.13), both sides, which are nonnegative, may be simultaneously infinite.

It is exactly Conjecture 6.11 1. in [5]. For \( g = 0 \), we recover of course the matrix Szegő formula.

The right hand side may also be written
\[
\begin{align*}
T(\alpha_0, \alpha_1, \cdots) &= \text{Re tr} \alpha_0 + \frac{1}{2} \text{tr} \alpha_0 \alpha_0^\dagger \\
&\quad + \frac{1}{2} \sum_0^\infty \text{tr} (\alpha_k - \alpha_{k+1})(\alpha_k^\dagger - \alpha_{k+1}^\dagger) - \sum_0^\infty \text{tr} \alpha_k \alpha_k^\dagger.
\end{align*}
\] (2.14)

According to the definition of B. Simon [9], the gems are equivalent conditions for the finiteness of entropies. Like in Corollary 5.4 in [5], we have the following result.

**Proposition 2.3.**

1. If \(|g| < 1\),
\[
K(\Lambda_g \mid \mu) < \infty \iff \sum_k \text{tr} \alpha_k \alpha_k^\dagger < \infty
\] (2.15)

2. \[
K(\Lambda_1 \mid \mu) < \infty \iff \sum_k \text{tr} (\alpha_k \alpha_k^\dagger)^2 + \sum_k \text{tr} (\alpha_{k+1} - \alpha_k)(\alpha_{k+1}^\dagger - \alpha_k^\dagger) < \infty
\] (2.16)

\[
K(\Lambda_{-1} \mid \mu) < \infty \iff \sum_k \text{tr} (\alpha_k \alpha_k^\dagger)^2 + \sum_k \text{tr} (\alpha_{k+1} + \alpha_k)(\alpha_{k+1}^\dagger + \alpha_k^\dagger) < \infty.
\] (2.17)

**3. Proof of Theorem 2.2**

We need a preliminary remark to reduce the case \( g < 0 \) to the case \( g > 0 \).

**Lemma 3.1** (Simon [7] 3.2.6 and [8] 9.5.28). If \( \mu \) is a non-trivial matrix measure and \( \tilde{\mu} \) is defined by
\[
d\tilde{\mu}(\theta) = \begin{cases} 
\frac{d\mu(\pi + \theta)}{d\mu} & \text{if } \theta \in [0, \pi] \\
\frac{d\mu(\theta - \pi)}{d\mu} & \text{if } \theta \in [\pi, 2\pi]
\end{cases}
\]
then
\[ \alpha_k(\hat{\mu}) = (-1)^{k+1} \alpha_k(\mu), \quad (k \geq 0). \] (3.1)

If \( g = -\gamma \) with \( \gamma > 0 \), we have,
\[ \int (1 - g \cos \theta) \log \det w(\theta) d\lambda_0(\theta) = \int (1 - \gamma \cos \theta) \log \det \hat{w}(\theta) d\lambda_0(\theta), \]
where \( w \) (resp. \( \hat{w} \)) is the a.c. part of \( \mu \) (resp. \( \hat{\mu} \)).

If we take for granted the result for \( \gamma \), we get
\[ \int (1 - \gamma \cos \theta) \log \det \hat{w}(\theta) d\lambda_0(\theta) = \sum_0^{\infty} \log \det(1 - \alpha_k(\hat{\mu})\alpha_k^{\dagger}(\hat{\mu})) - \gamma T(\alpha_0(\hat{\mu}), \alpha_1(\hat{\mu}), \cdots) \]
but, it is straightforward to see that from (2.12) and (3.1)
\[ T(\alpha_0(\mu), \alpha_1(\mu), \cdots) = -T(\alpha_0(\hat{\mu}), \alpha_1(\hat{\mu}), \cdots) \] (3.2)
so that (2.11) holds true.

From now on, in this section we assume \( 0 \leq g \leq 1 \).

If \( \mu \) is a probability measure on \( \mathbb{T} \) with V-coefficients \( (\alpha_j(\mu))_{j \geq 0} \) and if \( N \) is some positive integer, we denote by \( \mu_N \) the measure whose V-coefficients are shifted:
\[ \alpha_j(\mu_N) = \alpha_{j+N}(\mu), \quad j \geq 0. \]
When \( \mu \) has a density \( w \) with respect to \( \Lambda_0 \), we denote by \( w_N \) the density of \( \mu_N \).

The key point is the following "recursion" theorem, matrix version of Theorem 2.8.2 in \cite{9}, whose proof is postpone to Sect. 5.

**Theorem 3.2.** If \( \det w \neq 0 \) a.e., we have
\[ \int \log \det \left( w(\theta)w_1(\theta)^{-1} \right) d\lambda_\theta(\theta) = \log \det(1 - \alpha_0^0\alpha_0^{\dagger}) - g \Re \text{tr} (\alpha_0 - \alpha_1 - \alpha_1\alpha_0^{\dagger}). \] (3.3)

This implies that \( \det w_1 \neq 0 \) a.e. and then we may iterate. We get, for \( N > 1 \)
\[ \int \log \det \left( w(\theta)w_N(\theta)^{-1} \right) d\lambda_\theta(\theta) = G_N(\mu) \] (3.4)
where
\[ G_N(\mu) = -g \Re \text{tr} (\alpha_N - \alpha_0) + g \sum_0^{N-1} \Re \text{tr} \alpha_k\alpha_k^{\dagger} + \sum_0^{N-1} \log \det(1 - \alpha_k\alpha_k^{\dagger}) \] (3.5)
In terms of entropy, we have the equivalent form of (3.3):

\[ K(\Lambda_g | \mu_N) - K(\Lambda_g | \mu) = G_N(\mu). \] (3.6)

To look for a limit when \( N \to \infty \), we need a careful study of \( G_N(\mu) \). We have

\[ G_N(\mu) = -g \text{Re} \text{tr} (\alpha_N - \alpha_0) + \frac{g}{2} \text{tr} (\alpha_N \alpha_N^\dagger - \alpha_0 \alpha_0^\dagger) - \sum_{k=0}^{N-1} A_k, \] (3.7)

with

\[ A_k := -\log \det (1 - \alpha_k \alpha_k^\dagger) - g \text{tr} \alpha_k \alpha_k^\dagger + \frac{g}{2} \text{tr} (\alpha_{k+1} - \alpha_k) (\alpha_{k+1}^\dagger - \alpha_k^\dagger). \] (3.8)

For \( \alpha \alpha^\dagger < 1 \), we have

\[ -\log \det (1 - \alpha \alpha^\dagger) = \text{tr} \alpha \alpha^\dagger + \frac{1}{2} \text{tr}(\alpha \alpha^\dagger)^2 + R(\alpha), \]

with

\[ R(\alpha) > 0, \quad R(\alpha) = o(\text{tr}(\alpha \alpha^\dagger)^2). \] (3.9)

This yields

\[ A_k \geq (1 - g) \text{tr} \alpha_k \alpha_k^\dagger + \frac{1}{2} \text{tr}(\alpha_k \alpha_k^\dagger)^2 + \frac{g}{2} \text{tr} (\alpha_{k+1} - \alpha_k) (\alpha_{k+1}^\dagger - \alpha_k^\dagger). \] (3.10)

In particular, \( A_k \geq 0 \) for every \( k \) (remind that we have assumed \( g \geq 0 \)), which gives

\[ S_N(\mu) := \sum_{k=0}^{N-1} A_k \uparrow S_\infty(\mu) = \sum_{k=0}^\infty A_k \leq \infty, \]

(this argument of monotonicity is like in Simon [9] Prop. 2.8.6.

The identity (2.13) will be the result of two inequalities.

A) The first one uses the Bernstein-Szegő approximation of \( \mu \). We know, from Theorem 3.9 in [2], for every \( \theta \) and every integer \( k \), \( \varphi^R_k(e^{i\theta}) \) is invertible and from Theorem 3.11 of the same article that the measure

\[ d\mu^{(N)}(\theta) = \left[ \varphi_{N-1}(e^{i\theta}) \varphi_{N-1}(e^{i\theta})^\dagger \right]^{-1} d\lambda_0(\theta) \] (3.11)

satisfies

\[ \alpha_j(\mu^{(N)}) = \begin{cases} \alpha_j(\mu) & \text{if } 0 \leq j \leq N - 1 \\
0 & \text{if } j \geq N. \end{cases} \] (3.12)

We have \( (\mu^{(N)})_N = \Lambda_0 \). We may apply (3.6) with \( \mu = \mu^{(N)} \), which gives

\[ K(\Lambda_g | \Lambda_0) - K(\Lambda_g | \mu^{(N)}) = G_N(\mu^{(N)}) = g \text{Re} \text{tr} \alpha_0 - \frac{g}{2} \text{tr} \alpha_0 \alpha_0^\dagger - S_N(\mu). \]
Since \( \mu^{(N)} \) converges weakly to \( \mu \), the lower semicontinuity of \( \mathcal{K}(\Lambda \mid \cdot) \) gives
\[
\mathcal{K}(\Lambda \mid \Lambda_0) - \mathcal{K}(\Lambda \mid \mu) \geq \mathcal{K}(\Lambda \mid \Lambda_0) - \liminf_N \mathcal{K}(\Lambda \mid \mu^{(N)}) \geq \text{Re } \alpha_0 - \frac{g}{2} \text{ tr } \alpha_0 \alpha_0^\dagger - S_\infty(\mu) \geq -\infty .
\] (3.13)

**B)** If \( \mathcal{K}(\Lambda \mid \mu) = \infty \) the inequality
\[
\mathcal{K}(\Lambda \mid \Lambda_0) - \mathcal{K}(\Lambda \mid \mu) \leq \text{Re } \alpha_0 - \frac{g}{2} \text{ tr } \alpha_0 \alpha_0^\dagger - S_\infty(\mu)
\] (3.14)
is trivial.

If \( \mathcal{K}(\Lambda \mid \mu) < \infty \), then \( \det w(\theta) > 0 \) a.e. and then from (3.6) we have \( \det w_N(\theta) > 0 \) a.s. too. We want to let \( N \to \infty \) in (3.6) in order to get (3.14). To begin with, let us prove that
\[
\lim_N \alpha_N(\mu) = 0 .
\] (3.15)

From (3.6) we deduce
\[
G_N(\mu) \leq \mathcal{K}(\Lambda \mid \mu) < \infty ,
\]
and then, since
\[
-p \leq -\text{Re } \alpha_N + \frac{1}{2} \text{ tr } \alpha_N \alpha_N^\dagger \leq \frac{3p}{2}
\]
(\( p \) is the dimension) we have \( S_\infty(\mu) < \infty \).

Let us split the study into two cases:

1. if \( 0 \leq g < 1 \), \( S_\infty(\mu) < \infty \) implies
\[
\sum_k \text{ tr } \alpha_k \alpha_k^\dagger < \infty
\] (3.16)
hence (3.15) holds true.

2. if \( g = 1 \), we have
\[
\sum_k (\alpha_k \alpha_k^\dagger)^2 + \text{ tr } (\alpha_{k+1} - \alpha_k)(\alpha_{k+1}^\dagger - \alpha_k)^\dagger \leq \infty
\] (3.17)
which in particular implies that (3.15) holds true.

This result has consequences for both sides of (3.6). On the one hand, since for every \( j \)
\[
\lim_N \alpha_j(\mu_N) = \lim_N \alpha_{N+j}(\mu) \to 0 ,
\]
the sequence \( (\mu_N) \) converges weakly to \( \Lambda_0 \), so using again the semicontinuity, we get
\[
\mathcal{K}(\Lambda \mid \Lambda_0) - \mathcal{K}(\Lambda \mid \mu) \leq \liminf_N \mathcal{K}(\Lambda \mid \mu_N) - \mathcal{K}(\Lambda \mid \mu) .
\]
On the other hand, from (3.7)
\[ \lim G_N(\mu) = g \text{Re} \, \text{tr} \alpha_0 - \frac{g}{2} \text{tr} \alpha_0\alpha_0^\dagger - gS_\infty(\mu). \] (3.18)
and then (3.14) holds true also in this case.

Gathering (3.13) and (3.14) ends the proof of (2.13) hence (2.11) when \( 0 \leq g \leq 1 \).

4. Proof of Proposition 2.3

We consider only the case \( 0 \leq g \leq 1 \), since for \( -1 < g < 0 \) the reduction from \( g < 0 \) to \( \gamma > 0 \) as in the beginning of Sect. 3 leads directly to the result.

We already saw in the above section, that when \( K(\Lambda_g | \mu) < \infty \) and \( 0 \leq g \leq 1 \), the good conditions are fulfilled.

Conversely, we consider three cases.

If \( 0 \leq g < 1 \) and (3.16) is fulfilled, then
\[ -\sum_k \log \det \alpha_k\alpha_k^\dagger < \infty \]
and since
\[ \text{tr} (\alpha_{k+1} - \alpha_k)(\alpha_{k+1}^\dagger - \alpha_k^\dagger) \leq 2(\text{tr} \alpha_k\alpha_k^\dagger + \text{tr} \alpha_{k+1}\alpha_{k+1}^\dagger), \]
the expression \( T(\alpha_0, \alpha_1, \cdots) \) in (2.14) is well defined and finite, so is the left hand side of (2.13) and then \( K(\Lambda_g | \mu) \) is finite.

If \( g = 1 \), condition (3.17), jointly with (3.9) entails that
\[ \sum_0^\infty -\log \det (1 - \alpha_k\alpha_k^\dagger) - \text{tr} \alpha_k\alpha_k^\dagger + \frac{1}{2} \text{tr} (\alpha_k - \alpha_{k+1})(\alpha_k^\dagger - \alpha_{k+1}^\dagger) < \infty \]
and then gathering (2.13) and (2.14) show that \( K(\Lambda_g | \mu) \) is finite.

5. Proofs of intermediate results

5.1. Proof of Theorem 3.2

To compute the LHS of (3.3) we need the values of the Fourier coefficients:
\[ \int e^{ik\theta} \log \det \left( w(\theta)w_1(\theta)^{-1} \right) \frac{d\theta}{2\pi} \text{ for } k = -1, 0, 1. \]

The strategy is to approach \( \log \det (w(\theta)w_1(\theta)^{-1}) \) by a function of \( z = re^{i\theta} \), sufficiently smooth to apply Cauchy's formula.

In view of Theorem 2.1, it is natural to approximate \( w(\theta)(w_1(\theta))^{-1} \) by \( \text{Re} F(z)(\text{Re} F_1(z))^{-1} \) with \( z = re^{i\theta} \). We define the auxiliary matrix function:
\[ D_0(z) := (1 - zf(z))^{-1}(1 - zf_1(z)) \left( \rho_0^\dagger \right)^{-1} \left( 1 - f(z)\alpha_0^\dagger \right). \] (5.1)

We need the following formula whose proof is postponed in Sect. 5.2
Lemma 5.1.
\[
\det \left( \text{Re} F(z) (\text{Re} F_1(z))^{-1} \right) = \det(D_0(z)D_0(z)\dagger) \frac{\det(1 - |z|^2 f(z)^\dagger f(z))}{\det(1 - f(z)^\dagger f(z))}.
\] (5.2)

From Theorem 2.1 for a.e. \( \theta \) we have
\[
\lim_{r \to 1} \det \left( \text{Re} F(re^{i\theta}) (\text{Re} F_1(re^{i\theta}))^{-1} \right) = \det \left( w(\theta)w_1(\theta)^{-1} \right)
\]
\[
\lim_{r \to 1} \frac{\det(1 - |r|^2 f(re^{i\theta})^\dagger f(re^{i\theta}))}{\det(1 - f(re^{i\theta})^\dagger f(re^{i\theta}))} = 1,
\]
so that,
\[
\det(w(\theta)w_1(\theta)^{-1}) = \lim_{r \to 1} \det(D_0(re^{i\theta})D_0(re^{i\theta})\dagger),
\]
and the remaining part of the proof is based on the study of \( \det(D_0(z)D_0(z)\dagger) \).

Some properties of \( D_0 \) are collected in the following lemma, whose proof is also in Sect. 5.2.

Lemma 5.2. The function \( \det D_0 \) is analytic in \( \mathbb{D} \) and non-vanishing. Moreover
\[
h := 2 \log \det D_0 \in H^2(\mathbb{D}).
\] (5.3)

Since \( h \in H^2(\mathbb{D}) \subset H^1(\mathbb{D}) \), we have
\[
\int e^{-i\theta} h(e^{i\theta}) \frac{d\theta}{2\pi} = h'(0), \quad \int h(e^{i\theta}) \frac{d\theta}{2\pi} = h(0), \quad \int e^{i\theta} h(e^{i\theta}) \frac{d\theta}{2\pi} = 0,
\] (5.4)
and then
\[
\int (1 - g \cos \theta) \text{Re } h(e^{i\theta})d\lambda_0(\theta) = \text{Re } h(0) - \frac{g}{2} \text{Re } h'(0).
\] (5.5)

Let us compute \( h(0) \) and \( h'(0) \). As \( |z| \to 0 \),
\[
det(1 - zf(z)) = 1 - z(\text{tr } \alpha_0) + O(z^2), \quad \det(1 - zf_1(z)) = 1 - z(\text{tr } \alpha_1) + O(z^2)
\] (5.6)

Now, formula (2.7) can be inverted into
\[
f(z) = (\rho_0^R)^{-1}(\alpha_0 + zf_1(z)) \left(1 + z\alpha_0^\dagger f_1(z)\right)^{-1} \rho_0^L,
\] (5.7)
which gives the expansion
\[
f(z) = (\rho_0^R)^{-1} \left(\alpha_0 + z(1 - \alpha_0\alpha_0^\dagger)f_1(z) + O(z^2)\right) \rho_0^L

\]
\[
= \alpha_0 + z(\rho_0^R\alpha_1\rho_0^L) + O(z^2),
\]
so that
\[
1 - f(z)\alpha_0^\dagger = 1 - \alpha_0\alpha_0^\dagger - z(\rho_0^R\alpha_1\rho_0^R\alpha_0^\dagger) + O(z^2) = (\rho_0^R)^2 - z(\rho_0^R\alpha_1\rho_0^R) + O(z^2)
\]
\[
= \rho_0^R \left(1 - z(\alpha_1\alpha_0^\dagger) + O(z^2)\right) \rho_0^R
\]
and
\[
\det \left(1 - f(z)\alpha_0^\dagger\right) = \det(\rho_0^R)^2 \det \left(1 - z(\alpha_1\alpha_0^\dagger) + O(z^2)\right)
\]
\[
= \det(\rho_0^R)^2 \left(1 - z \text{tr} (\alpha_1\alpha_0^\dagger) + O(z^2)\right). \tag{5.8}
\]
Gathering (5.1), (5.6) and (5.8) and using \(\det \rho_0^R = \det \rho_0^L\) we get
\[
\det D_0(z) = (\det \rho_0^R) \left(1 - z \text{tr} (\alpha_0 - \alpha_1 - \alpha_1\alpha_0^\dagger) + O(z^2)\right)
\]
Coming back to the definition of \(h\), we get
\[
h(z) = \log \det(1 - \alpha_0\alpha_0^\dagger) - 2z \text{tr} \left(\alpha_0 - \alpha_1 - \alpha_1\alpha_0^\dagger\right) + O(z^2)
\]
and from (5.5)
\[
\int (1 - g \cos \theta) \text{Re} \ h(e^{i\theta}) d\lambda_0(\theta) = \log \det(1 - \alpha_0\alpha_0^\dagger) + g \text{Re} \left(\alpha_0 - \alpha_1 - \alpha_1\alpha_0^\dagger\right).
\]

5.2. Proof of Lemma 5.1

To simplify, we omit the variable \(z\) if unnecessary. Applying (2.8) to \(F_1\)
\[
\text{Re} \ F_1 = (1 - \bar{z}f_1^\dagger)^{-1}(1 - |z|^2 f_1^\dagger f_1)(1 - z)^{-1} \tag{5.9}
\]
so we need an expression of \(1 - |z|^2 f_1^\dagger f_1\) as a function of \(f\). From (2.7) we get
\[
|z|^2 f_1(z)^\dagger f_1(z) = \rho_0^L \left(1 - f(z)^\dagger\alpha_0\right)^{-1}(f(z)^\dagger - \alpha_0^\dagger)(\rho_0^R)^{-2}(f(z) - \alpha_0) \left(1 - \alpha_0^\dagger f(z)\right)^{-1} \rho_0^L
\]
which, with the help of the trivial identity
\[
1 = \rho_0^L \left(1 - f(z)^\dagger\alpha_0\right)^{-1} \left(1 - f(z)^\dagger\alpha_0\right) (\rho_0^L)^{-2} \left(1 - \alpha_0^\dagger f(z)\right) \left(1 - \alpha_0^\dagger f(z)\right)^{-1} \rho_0^L,
\]
yields
\[
\left(1 - f^\dagger\alpha_0\right) (\rho_0^L)^{-1} \left(1 - |z|^2 f_1^\dagger f_1\right) (\rho_0^L)^{-1} \left(1 - \alpha_0^\dagger f\right) = \left(1 - f^\dagger\alpha_0\right) (\rho_0^L)^{-2} \left(1 - \alpha_0^\dagger f\right) - (f^\dagger - \alpha_0^\dagger)(\rho_0^R)^{-2}(f - \alpha_0). \tag{5.10}
\]
Now, we use (2.6) and
\[ (\rho_0^R)^{-2} = \sum_{n \geq 0} (\alpha_0 \alpha_0^\dagger)^n, \quad (\rho_0^L)^{-2} = \sum_{n \geq 0} (\alpha_0^\dagger \alpha_0)^n \]
(\alpha_0 is a contraction). Expanding the RHS of (5.10) and cancelling terms gives
\[ (1 - f^\dagger \alpha_0) (\rho_0^L)^{-2} \left(1 - \alpha_0^\dagger f\right) - (f^\dagger - \alpha_0^\dagger)(\rho_0^R)^{-2}(f - \alpha_0) = 1 - f^\dagger f \]
so that
\[ 1 - |z|^2 f_1^\dagger f_1 = \rho_0^L (1 - f^\dagger \alpha_0)^{-1} (1 - f^\dagger f) \left(1 - f \alpha_0^\dagger\right)^{-1} \rho_0^R. \quad (5.11) \]
Plugging into (5.9) yields
\[
\text{Re } F_1 = (1 - \bar{z} f_1^\dagger)^{-1} \rho_0^L (1 - f(z)^\dagger \alpha_0)^{-1} (1 - f(z) f(z)) \left(1 - f(z) \alpha_0^\dagger\right)^{-1} \rho_0^R (1 - zf_1)^{-1}
\]
and
\[
(\text{Re } F)(\text{Re } F_1)^{-1} = (1 - \bar{z} f^\dagger)^{-1} (1 - |z|^2 f^\dagger f)(1 - zf)^{-1}
\times (1 - zf_1) (\rho_0^L)^{-1} (1 - f \alpha_0^\dagger)^{-1} (1 - f^\dagger \alpha_0) (\rho_0^R)^{-1} (1 - \bar{z} f_1^\dagger)
= (1 - \bar{z} f^\dagger)^{-1} (1 - |z|^2 f^\dagger f) D_0 (1 - f^\dagger f)^{-1} D_0^\dagger (1 - \bar{z} f^\dagger).
\]
Then, taking determinants
\[
\det \left( (\text{Re } F)(\text{Re } F_1)^{-1} \right) = \det(D_0 D_0^\dagger) \frac{\det(1 - |z|^2 f^\dagger f)}{\det(1 - f^\dagger f)}
\]
ends the proof.

5.3. Proof of Lemma 5.2

We repeat here the argument of Theorem 2.6.2 in [9] for the sake of completeness. For \( z \in \mathbb{D} \) we have \( f(z) f^\dagger(z) < 1 \), hence analyticity and non-vanishing are straightforward. Moreover, since \( |\zeta| < 1 \) implies \( |\text{arg}(1 - \zeta)| < \pi/2 \), we conclude from (5.1) and (5.3) that
\[ |\text{Im } h| < 3\pi/2. \]
Since \( |h|^2 - 2(\text{Im } h)^2 \) is harmonic we have
\[
\int |h|^2 d\lambda_0 - 2 \int (\text{Im } h)^2 d\lambda_0 = |h(0)|^2 - 2 (\text{Im } h(0))^2,
\]
and since \( h(0) = \log \det(1 - \alpha_0 \alpha_0^\dagger) < 0 \), we get
\[
\int |h(e^{i\theta})|^2 d\lambda_0(\theta) \leq \frac{9\pi^2}{2} + \left( \log \det(1 - \alpha_0 \alpha_0^\dagger) \right)^2,
\]
which yields
\[
\sup_{r < 1} \int |h(e^{i\theta})|^2 d\lambda_0(\theta) < \infty.
\]
[1] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and the Lukic conjecture. *Duke Math. J.*, 167(15):2857–2902, 2018.

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

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

[4] M. Derevyagin, O. Holtz, S. Khrushchev, and M. Tyaglov. Szegő’s theorem for matrix orthogonal polynomials. *J. Approx. Theory*, 164(9):1238–1261, 2012.

[5] F. Gamboa, J. Nagel, and A. Rouault. Sum rules and large deviations for spectral measures on the unit circle. *Random Matrices Theory Appl.*, 6(1):1750005, 49, 2017.

[6] D.J. Gross and E. Witten. Possible third-order phase transition in the large-N lattice gauge theory. *Phys. Rev. D*, 21(2):446–453, 1980.

[7] B. Simon. *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. Colloquium Publications. American Mathematical Society 54, Part 1. Providence, RI: American Mathematical Society (AMS), 2005.

[8] B. Simon. *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*. Colloquium Publications. American Mathematical Society 51, Part 2. Providence, RI: American Mathematical Society, 2005.

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