ON SOME GATEWAYS BETWEEN SUM RULES

Fabrice Gamboa, Jan Nagel, Alain Rouault

ON SOME GATEWAYS BETWEEN SUM RULES

FABRICE GAMBOA, JAN NAGEL, AND ALAIN ROUAULT

Abstract. We present correspondences induced by some classical mappings between measures on an interval and measures on the unit circle. More precisely, we link their sequences of orthogonal polynomial and their recursion coefficients. We also deduce some correspondences between particular equilibrium measures of random matrix ensembles. Additionally, we show that these mappings open up gateways between the sum rules associated with some classical models, leading to new formulations of several sum rules.

1. Introduction

The relation between orthogonal polynomials on the unit circle (OPUC) and orthogonal polynomials on the line (OPRL) is a longstanding problem. When a measure on the unit circle is mapped to a measure on the real line, what is the relation between the orthogonal polynomials related to these measures or their recursion coefficients? First results in this direction go back to Szegő, who found a relation between the orthogonal polynomials when the mapping on the real line is the pushforward under $z \mapsto z + z^{-1}$, now called the Szegő mapping, see [40], p. 880 for a historical account. The relation between the recursion coefficients was found by Geronimus: surprisingly, the so-called Verblunsky coefficients $(\alpha_k)_{k\geq 0}$ of the recursion on the unit circle appear in a decomposition of the Jacobi coefficients on the real line, forming an identity now known as the Geronimus relations. Since then, a variety of mappings have been studied, motivated from applications for orthogonal polynomials [5, 16] operator theory [28, 20, 13] or signal processing [17, 19, 9].

Let us highlight the implications of such mappings and relations in particular on important identities in spectral theory called sum rules. Sum rules are identities between two nonnegative functionals of a probability measure μ compactly supported on $\mathbb{R}$ (resp. ν supported on $\mathbb{T}$). On the one hand, the first functional is an entropy-like functional with respect to some reference measure. On the other hand, the second functional is built from Jacobi coefficients (resp. Verblunsky coefficients) of μ (resp. ν) and vanishes only for the reference measure. We call the left hand side (LHS) the spectral side and the right hand side (RHS) the coefficient side.

Date: January 25, 2023.

2020 *Mathematics Subject Classification.* 42C05, 47B36, 15B52, 34L05.

Key words and phrases. Sum rules, Szegő mapping, Verblunsky coefficients, Delsarte-Genin mapping, Jacobi coefficients, relative entropy.

The first historical example of such a sum rule is the classical Szegő-Verblunsky identity,

$$\frac{1}{2\pi}\int_0^{2\pi} \log g_\nu(\theta)d\theta = \sum_{k=0}^{\infty} \log(1-|\alpha_k|^2)\,, \tag{1.1}$$

where ν is a measure on the unit circle with Lebesgue decomposition

$$d\nu(\theta) = g_\nu(\theta)\tfrac{d\theta}{2\pi} + d\nu_s(\theta)$$

having Verblunsky coefficients $(\alpha_k)_{k\geq 0}$. Both sides of (1.1) vanish if, and only if, ν is the uniform measure on the circle (the reference measure in this case). We refer to Chapter 1 of [40] for a discussion of the origin of this sum rule.

The most famous sum rule for measures on the line is the Killip-Simon sum rule [32]. An exhaustive discussion and history of this sum rule can be found in Section 1.10 of the book [40] and a deep analytical proof in Chapter 3 therein. The reference measure for this sum rule is the semicircle law (SC).

An important consequence of these two sum rules is the equivalence of two conditions for the finiteness of both sides, one formulated in terms of Verblunsky or Jacobi coefficients and the other as a spectral condition. In the words of Simon [40], these are the *gems* of spectral theory. In [26] and [23], we revisited these results from a probabilistic point of view and gave a new proof based on large deviations. We also refer to the work of Breuer et al. [10] which enlightens non-probabilists about [23], [26]. The method was robust enough to prove new sum rules with reference measures such as Marchenko-Pastur (MP), Kesten-McKay (KMK) on the real line and Gross-Witten (GW), Hua-Pickrell (HP) on the unit circle.

The main contribution of this paper is two-fold. On the one hand, we gather a series of results on relations between measures on the unit circle and measures on the real line and their orthogonal polynomials under several well known mappings: Szegő, Delsarte-Genin (DG), Derevyagin-Vinet-Zhedanov (DVZ) and Möbius. On the other hand, we show how these relations allow to catch a –potentially new– sum rule from an existing one. The main idea is easy: we transform both sides according to the mapping. While our first contribution is merely expository in nature, we believe the second contribution can be of great interest, either to find new sum rules or to highlight connections, or "gateways", between existing identities.

As an easy example for such a gateway, the Szegő-Verblunsky sum rule (1.1) leads to an identity for measures on $[-2,2]$, when both sides are transformed according to the Szegő mapping. The LHS may be written as an integral with respect to the Arcsine law while the Geronimus relations allow to rewrite the RHS (see Section 6.1). To give an overview of further results (we refer to Section 3 for the statement of the sum rules and Section 4 for the definition of the mappings):

- Particular cases of the KMK-sum rule can be obtained from the HP-sum rule by the Szegő mapping or by the DG mapping (Section 6.2).
- The GW-sum rule implies the new sum rules (6.7) and (6.13) by the Szegő mapping.
- The GW-sum rule leads to the reformulations (6.19) and (6.18) under the DG mapping, with a new formula for the RHS in Theorem 6.1.
- Under the DVZ mapping, the GW-sum rule leads to a variant of the Killip-Simon sum rule (6.22).
- We prove a new sum rule with reference to the Poisson measure Pois(ζ) in Theorem 6.2.

- Another new Poisson sum rule is obtained from a recent result of [6] in Proposition 6.3.
- A new analytical proof of a weak version of the HP-sum rule is in Proposition 8.1.

The paper is organized as follows: In Section 2 we recap the required background on orthogonal polynomials on the real line and on the unit circle with corresponding recursion formulas. In Section 3 we recall the main known sum rules with their reference measures. Section 4 discusses the mappings used in our work. Then, in Section 5 we apply these mappings to our reference measures. Section 6 presents the gateways between OPUC sum rules and OPRL sum rules. In Section 7 are the proofs of some new sum rules and some auxiliary results are in Section 8.

2. Orthogonal polynomials

Let $\mathcal{M}_1(\mathbb{R})$ (resp. $\mathcal{M}_1(\mathbb{T})$) denote the set of all probability measures on $\mathbb{R}$ (resp. on the unit circle $\mathbb{T} = \partial\mathbb{D}$, where $\mathbb{D}$ is the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$). Additionally, we write $\mathcal{M}_{1,s}(\mathbb{T})$ for the set of all symmetric probability measures on $\mathbb{T}$, invariant under the transformation $z \mapsto \bar{z}$.

2.1. **OPRL.**

The sequence of orthogonal polynomials on the real line (OPRL) is well defined for a probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$ with a compact support consisting of infinitely many points, a.k.a. nontrivial case (in constrast to a finite support consisting of n points, a.k.a. trivial case). They are obtained by applying to the sequence $1, x, x^2, \dots$ the orthonormalizing Gram-Schmidt procedure. The resulting polynomials $p_0, p_1, \dots$, with p_k of degree k, obey the recursion relation

$$xp_k(x) = a_{k+1}p_{k+1}(x) + b_{k+1}p_k(x) + a_k p_{k-1}(x) \tag{2.1}$$

for $k \geq 0$, with $p_{-1} = 0$. The recursion or Jacobi coefficients (or short "J-coefficients") satisfy that for all k, $b_k \in \mathbb{R}$ and $a_k > 0$. Notice that here the orthogonal polynomials are not monic but normalized in $L^2(\mu)$. The monic polynomials satisfy the recursion

$$xP_k(x) = P_{k+1}(x) + b_{k+1}P_k(x) + a_k^2 P_{k-1}(x)\,. \tag{2.2}$$

When the support of μ consists of n points, the orthogonal polynomials might be defined up to degree $n-1$ and J-coefficients $b_1, a_1, \dots, a_{n-1}, b_n$ are well defined.

For a non-trivial measure μ let us equip the vector space $L^2(\mu)$ with the basis $(p_k)_{k\geq 0}$. Then the linear map $f \mapsto xf$, multiplication by the identity, is represented by the tridiagonal matrix

$$J_\mu = \begin{pmatrix} b_1 & a_1 & 0 & 0 & \cdots \\ a_1 & b_2 & a_2 & 0 & \cdots \\ 0 & a_2 & b_3 & a_3 & \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{2.3}$$

Conversely, if H is a bounded Hermitian operator on an infinitely dimensional Hilbert space $\mathcal{H}$, and e is a cyclic vector, then we can define the spectral measure μ of the pair (H, e) and then $(\mathcal{H}, H, e)$ is isomorphic to (ℓ^2, J_μ, e_1) where $e_1 = (1, 0, 0, \dots)^t$. Such a correspondence still holds between Hermitian operators on an n-dimensional space and measures supported by n points and $n \times n$ tridiagonal matrices.

If the support of μ is contained in $[0,\infty)$, there is a decomposition of J-coefficients,

$$\begin{aligned} b_k &= z_{2k-2} + z_{2k-1}, \\ a_k^2 &= z_{2k-1} z_{2k}, \end{aligned} \tag{2.4}$$

with $z_0 = 0$ and $z_k \geq 0$ for $k \geq 1$ (see [15] p.47). The z_k will be called the canonical coefficients and they are uniquely determined by the J-coefficients.

If μ is nontrivial with support contained in the interval $[-2,2]$, there exists a decomposition of J-coefficients,

$$\begin{aligned} b_{k+1} &= (1-u_{2k})u_{2k+1} - (1+u_{2k})u_{2k-1}, \\ a_{k+1}^2 &= (1-u_{2k})(1-u_{2k+1}^2)(1+u_{2k+2}), \end{aligned} \tag{2.5}$$

with $u_0 = -1$ and $u_k \in (-1,1)$ for all $k \geq 1$. The u_k will be called canonical moments (although more classically, the $\frac{1}{2}(u_k+1)$ are called canonical moments [21]) and they are uniquely determined by the J-coefficients. If μ is nontrivial, $u_k \in (-1,1)$ for all $k \geq 1$, while if μ is supported by n points, we can still define $u_1, \dots, u_{2n-2} \in (-1,1)$ and $u_{2n-1} \in \{-1,1\}$. Let us notice that if μ is symmetric, then $u_{2k+1} = 0$ for all k, all the diagonal coefficients b_k vanish and

$$a_{k+1}^2 = (1-u_{2k})(1+u_{2k+2}) \quad (k \geq 0)\,. \tag{2.6}$$

2.2. **OPUC.** For a probability measure $\nu \in \mathcal{M}_1(\mathbb{T})$ supported by at least $k+1$ points, the inductive relation between two successive monic polynomials Φ_{k+1} and Φ_k, where Φ_k has degree k, orthogonal with respect to ν involves a complex number α_k and may be written as

$$\Phi_{k+1}(z) = z\Phi_k(z) - \overline{\alpha}_k \Phi_k^*(z) \text{ where } \Phi_k^*(z) := z^k \overline{\Phi_k(1/\bar{z})}. \tag{2.7}$$

The complex numbers $\alpha_k = -\overline{\Phi_{k+1}(0)}$, $k \geq 0$ are the so-called Verblunsky coefficients (in short V-coefficients). They are also called Schur, Levinson, Szegő coefficients in other contexts or canonical moment as well [21]. We also set $\alpha_{-1} = -1$. The V-coefficients satisfy $|\alpha_{k-1}| < 1$ if $k \geq 1$ and the support of ν contains at least $k+1$ points and $|\alpha_{k-1}| = 1$ if the support consists of exactly k points. For a symmetric measure $\nu \in \mathcal{M}_{1,s}(\mathbb{T})$, the V-coefficients are real. We will denote by $(\varphi_k)_{k\geq 0}$ the sequence of orthonormal polynomials on the unit circle (OPUC).

In the basis $(\chi_k)_{k\geq 0}$ obtained by orthonormalizing $1, z, z^{-1}, z^2, z^{-2}, \dots$, the linear transformation $f \to zf$ in $L^2(\nu)$ is represented by the so-called CMV-matrix

$$\mathcal{C}_\mu = \begin{pmatrix} \bar{\alpha}_0 & \bar{\alpha}_1\rho_0 & \rho_1\rho_0 & 0 & 0 & \dots \\ \rho_0 & -\bar{\alpha}_1\alpha_0 & -\rho_1\alpha_0 & 0 & 0 & \dots \\ 0 & \bar{\alpha}_2\rho_1 & -\bar{\alpha}_2\alpha_1 & \bar{\alpha}_3\rho_2 & \rho_3\rho_2 & \dots \\ 0 & \rho_2\rho_1 & -\rho_2\alpha_1 & -\bar{\alpha}_3\alpha_2 & -\rho_3\alpha_2 & \dots \\ 0 & 0 & 0 & \bar{\alpha}_4\rho_3 & -\bar{\alpha}_4\alpha_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \tag{2.8}$$

with $\rho_k = \sqrt{1-|\alpha_k|^2}$ for every $k \geq 0$ in the non-trivial case.

Conversely, if U is a unitary operator on an infinite dimensional Hilbert space $\mathcal{H}$ and e is a cyclic vector, then we can define the spectral measure ν of the pair (U,e) and then $(\mathcal{H},U,e)$ is isomorphic to $(\ell^2, \mathcal{C}_\nu, e_1)$. Let

$$\Theta_k = \begin{pmatrix} \alpha_k & \rho_k \\ \rho_k & -\alpha_k \end{pmatrix} \tag{2.9}$$

and

$$\mathcal{L} = \Theta_0 \oplus \Theta_2 \oplus \cdots , \qquad \mathcal{M} = \mathbf{1} \oplus \Theta_1 \oplus \Theta_3 \oplus \cdots , \tag{2.10}$$

where $\mathbf{1}$ denotes the 1×1 identity matrix and $\oplus$ is the direct sum operator. The unitary operators $\mathcal{L}$ and $\mathcal{M}$ satisfy

$$\mathcal{C}_\nu = \mathcal{L}\mathcal{M} . \tag{2.11}$$

For probability measures ν, μ both on $\mathbb{R}$ or on $\mathbb{T}$, let $\mathcal{K}(\nu|\mu)$ denote the Kullback-Leibler divergence or relative entropy of ν with respect to μ:

$$\mathcal{K}(\nu \,|\, \mu) = \begin{cases} \displaystyle\int \log \frac{d\nu}{d\mu}\, d\nu & \text{if } \nu \ll \mu \text{ and } \log \frac{d\nu}{d\mu} \in L^1(\nu), \\ \infty & \text{otherwise.} \end{cases} \tag{2.12}$$

3. Reference measures and sum rules

3.1. **Measures on $\mathbb{R}$.** We start with measures on the real line and state sum rules relative to these measures. In order to formulate the spectral side, we need some support conditions. For $c^- < c^+$ we define the set $\mathcal{S}_1(c^-, c^+)$ as the set of probability measures μ on $\mathbb{R}$ whose support satisfies

$$\operatorname{supp}(\mu) = I \cup E,$$

where $I \subset [c^-, c^+]$ and $E = E(\mu)$ is an at most countable subset of $[c^-, c^+]^c$.

3.1.1. *Semicircle distribution.* The semicircle law is

$$\mathrm{SC}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbb{1}_{\{-2 \le x \le 2\}}\, dx. \tag{3.1}$$

It is the central probability measure in classical random matrix theory. Indeed, it is the equilibrium measure for a large class of random matrix models (the limit of their empirical eigenvalue distribution). The Jacobi matrix of SC is also called the free Jacobi matrix with J-coefficients

$$a_k^{\mathrm{SC}} = 1,\ b_k^{\mathrm{SC}} = 0 \ \text{ for all } k \ge 1 . \tag{3.2}$$

We start by stating the classical sum rule of [33] (and explained in [41] p.37), the new probabilistic proof using large deviations might be found in [23]. For a probability measure μ on $\mathbb{R}$ with recursion coefficients a_k, b_k as in (2.1), define the sum

$$\mathcal{I}_H(\mu) = \frac{1}{2} \sum_{k=1}^{\infty} b_k^2 + G(a_k^2), \tag{3.3}$$

where

$$G(x) = x - 1 - \log x . \tag{3.4}$$

Furthermore, define

$$\mathcal{F}_{\mathrm{SC}}(x) := \int_2^{|x|} \sqrt{t^2 - 4}\, dt = \frac{|x|}{2} \sqrt{x^2 - 4} - 2 \log \left(\frac{|x| + \sqrt{x^2 - 4}}{2} \right)$$

if $|x| \ge 2$ and $\mathcal{F}_H(x) = \infty$ otherwise. Then the following remarkable identity holds.

Theorem 3.1 ([33])**.** *Let J be a Jacobi matrix with diagonal entries $b_1, b_2, \dots \in \mathbb{R}$ and subdiagonal entries $a_1, a_2, \dots > 0$ satisfying $\sup_k(a_k + |b_k|) < \infty$ and let μ be the associated spectral measure. Then $\mathcal{I}_H(\mu)$ is infinite if $\mu \notin \mathcal{S}_1(-2,2)$ and for $\mu \in \mathcal{S}_1(-2,2)$,*

$$\mathcal{K}(\mathrm{SC} \mid \mu) + \sum_{\lambda \in E(\mu)} \mathcal{F}_{\mathrm{SC}}(\lambda) = \mathcal{I}_H(\mu),$$

where both sides may be infinite simultaneously.

Let us emphazise that for a sum rule as in Theorem 3.1, both sides are nonnegative and vanish if and only if μ is equal to the reference measure, which is the semicircle law SC in this case.

3.1.2. *Marchenko-Pastur distribution.* The Marchenko-Pastur distribution with parameter $\tau \in (0,1]$ is

$$\mathrm{MP}_\tau(dx) = \frac{\sqrt{(x-\tau_-)(\tau_+ - x)}}{2\pi\tau x} \mathbb{1}_{[\tau_-,\tau_+]}(x)dx\,. \tag{3.5}$$

where $\tau_\pm = \sqrt{1 \pm \tau}$. In random matrix theory, it is the equilibrium measure of the Laguerre ensemble. Its canonical coefficients (see (2.4)) are

$$z_{2k-1} = 1, \quad z_{2k} = \tau \ (k \geq 1)$$

with $z_0 = 0$, which correspond to the J-coefficients

$$a_k = \sqrt{\tau}, \quad b_{k+1} = 1 + \tau \ (k \geq 1)$$

with $b_1 = 1$. Notice that the MP distribution is not symmetric. Let $\mathcal{I}_L$ be defined by

$$\mathcal{I}_L(\mu) = \sum_{k=1}^{\infty} \tau^{-1} G(z_{2k-1}) + G(\tau^{-1} z_{2k})\,. \tag{3.6}$$

Furthermore, define for $x \notin (\tau^-, \tau^+)$

$$\mathcal{F}_{\mathrm{MP}}(x) := \int_{I(x)} \frac{\sqrt{(t-\tau^+)(t-\tau^-)}}{\tau t}\, dt$$

where $I(x) = [\tau^+, x]$ if $x \geq \tau^+$ and $I(x) = [x, \tau^-]$ if $0 < x \leq \tau^-$. Then we have the following theorem.

Theorem 3.2 ([23] Theorem 2.2)**.** *Let $\mu \in \mathcal{M}_1([0,\infty))$ be a non-trivial measure with compact support and $0 < \tau \leq 1$. Then $\mathcal{I}_L(\mu) = \infty$ if $\mu \notin \mathcal{S}_1(\tau^-, \tau^+)$ and if $\mu \in \mathcal{S}_1(\tau^-, \tau^+)$ we have*

$$\mathcal{K}(\mathrm{MP}(\tau)\,|\mu) + \sum_{\lambda \in E(\mu)} \mathcal{F}_{\mathrm{MP}}(\lambda) = \mathcal{I}_L(\mu), \tag{3.7}$$

where both sides may be infinite simultaneously.

3.1.3. *The KMK distribution.* The Kesten-McKay law with parameters $\kappa_1, \kappa_2 \geq 0$ is denoted by $\mathrm{KMK}(\kappa_1, \kappa_2)$ and has the density

$$\mathrm{KMK}(\kappa_1, \kappa_2)(dx) = \frac{(2+\kappa_1+\kappa_2)}{2\pi} \frac{\sqrt{(x-u_-)(u_+-x)}}{4-x^2} \mathbb{1}_{\{u^- < x < u^+\}} \, dx \tag{3.8}$$

where

$$u^{\pm} = \frac{2\left(\kappa_2^2 - \kappa_1^2 \pm 4\sqrt{(1+\kappa_1)(1+\kappa_2)(1+\kappa_1+\kappa_2)}\right)}{(2+\kappa_1+\kappa_2)^2} . \tag{3.9}$$

It is the equilibrium measure of the Jacobi ensemble. The canonical moments of $\mathrm{KMK}(\kappa_1, \kappa_2)$ of even and odd index are, respectively:

$$u_{2k}^{\kappa_1,\kappa_2} \equiv u_e^{\kappa_1,\kappa_2} := -\frac{\kappa_1+\kappa_2}{2+\kappa_1+\kappa_2} \ , \ u_{2k+1}^{\kappa_1,\kappa_2} \equiv u_o^{\kappa_1,\kappa_2} := \frac{\kappa_2-\kappa_1}{2+\kappa_1+\kappa_2} . \tag{3.10}$$

We will consider also a symmetric version with $\kappa_1 = \kappa_2 = \kappa$, we will denote it $\mathrm{KMK}(\kappa) := \mathrm{KMK}(\kappa, \kappa)$:

$$\mathrm{KMK}(\kappa)(dx) = \frac{(1+\kappa)}{\pi} \frac{\sqrt{u^2-x^2}}{4-x^2} \mathbb{1}_{\{|x| \leq u\}} \, dx, \tag{3.11}$$

where

$$u = 2\frac{\sqrt{1+2\kappa}}{1+\kappa} .$$

The canonical moments of $\mathrm{KMK}(\kappa)$ are

$$u_{2k}^{\kappa} = u^{(\kappa)} := \frac{-\kappa}{1+\kappa}, \quad u_{2k-1}^{(\kappa)} = 0 , \tag{3.12}$$

see [27, Section 6] for the linearly transformed canonical moments. For $\kappa = 0$, the Kesten-McKay law is the Arcsine distribution

$$\mathrm{Arcsine}(dx) := \frac{1}{\pi\sqrt{4-x^2}} \mathbb{1}_{\{-2 < x < 2\}} \, dx , \tag{3.13}$$

whose canonical coefficients are all zero.

To state the sum rule, we need some more notation. Set for $u \in (-1, 1)$

$$\begin{aligned} \mathcal{H}_e^{\kappa_1,\kappa_2}(u) &:= -(1+\kappa_1+\kappa_2)\log\frac{1-u}{1-u_e^{\kappa_1,\kappa_2}} - \log\frac{1+u}{1+u_e^{\kappa_1,\kappa_2}} , \\ \mathcal{H}_o^{\kappa_1,\kappa_2}(u) &:= -(1+\kappa_1)\log\frac{1-u}{1-u_o^{\kappa_1,\kappa_2}} - (1+\kappa_2)\log\frac{1+u}{1+u_o^{\kappa_1,\kappa_2}} . \end{aligned} \tag{3.14}$$

For a non-trivial measure $\mu \in \mathcal{M}_1([-2, 2])$ with canonical moments $u_k \in (-1, 1)$, define

$$\mathcal{I}_J(\mu) = \sum_{k=1}^{\infty} \mathcal{H}_o^{\kappa_1,\kappa_2}(u_{2k-1}) + \mathcal{H}_e^{\kappa_1,\kappa_2}(u_{2k}) . \tag{3.15}$$

Finally, for the contribution of the outlying support points, we define for $x \notin (u^-, u^+)$

$$\mathcal{F}_{\mathrm{KMK}(\kappa_1,\kappa_2)}(x) = \int_{I(x)} (2+\kappa_1+\kappa_2) \frac{\sqrt{(t-u^+)(t-u^-)}}{4-t^2} \, dt \tag{3.16}$$

where $I(x) = [u^+, x]$ if $x \in [u^+, 2]$ and $I(x) = [x, u^-]$ if $x \in [-2, u^-]$.

We are now able to give the sum rule relative to the KMK measure. It is Theorem 2.3 in [23], where it is formulated for linearly transformed measures on $[0, 1]$.

Theorem 3.3. *Let $\mu \in \mathcal{M}_1([-2,2])$ be a nontrivial measure and $\kappa_1, \kappa_2 \geq 0$. Then $\mathcal{I}_J(\mu) = \infty$ if $\mu \notin \mathcal{S}_1(u^-, u^+)$, and if $\mu \in \mathcal{S}_1(u^-, u^+)$ we have*

$$\mathcal{K}(\mathrm{KMK}(\kappa_1, \kappa_2) \,|\, \mu) + \sum_{\lambda \in E(\mu)} \mathcal{F}_{\mathrm{KMK}(\kappa_1,\kappa_2)}(\lambda) = \mathcal{I}_J(\mu), \tag{3.17}$$

where both sides may be infinite simultaneously.

Remark 1. In the particular case $\kappa_1 = \kappa_2 = 0$, we obtain the identity

$$\mathcal{K}(\mathrm{Arcsine} \,|\, \mu) = -\sum_{k=1}^{\infty} \log(1 - u_k^2)\,. \tag{3.18}$$

It is very close to the results of Gamboa and Lozada [22] and equivalent to the so-called C_0 sum rule of Simon and Zlatos [40, Theorem 13.8.8]:

$$\mathcal{K}(\mathrm{Arcsine} \,|\, \mu) = \log 2 - \sum_{k=1}^{\infty} \log a_k^2\,. \tag{3.19}$$

Indeed, using (2.5) we can write

$$-\sum_{k=1}^{n} \log a_k^2 = -\log 2 - \log(1 - u_{2n}) + \sum_{k=1}^{2n} -\log(1 - u_k^2)\,. \tag{3.20}$$

Suppose that the last sum is bounded, then $\lim_{n\to\infty} u_{2n} = 0$ and hence

$$-\sum_{k=1}^{\infty} \log a_k^2 = \log 2 - \sum_{k=1}^{\infty} \log(1 - u_k^2)\,. \tag{3.21}$$

On the other hand, since $\log(1 - u_{2n+2}) \leq \log 2$, (3.20) implies

$$\sum_{k=1}^{2n} -\log(1 - u_k^2) \leq 2\log 2 - \sum_{k=1}^{n} \log a_k^2$$

and when the sum on the LHS diverges, $-\sum_{k=1}^{\infty} \log a_k^2$ does as well, so that (3.21) holds true in any case.

3.2. **Measures on $\mathbb{T}$.** In analogy to the real case we introduce for $0 \leq \theta^- < \theta^+ \leq 2\pi$ the set $\mathcal{S}_1^{\mathbb{T}}(\theta^-, \theta^+)$ of probability measures $\nu \in \mathcal{M}_1(\mathbb{T})$ supported on $I \cup E$, where I is a subset of the arc

$$\{z = e^{\mathrm{i}\theta} \in \mathbb{T} \,|\, \theta \in [\theta^-, \theta^+]\} \tag{3.22}$$

and where $E = E(\nu)$ is an at most countable subset of the complement of the set (3.22).

3.2.1. *Uniform distribution.* We write UNIF for the normalized Lebesgue measure on $\mathbb{T}$

$$\mathrm{UNIF}(d\theta) = \frac{d\theta}{2\pi}\,.$$

Its V-coeffcients are

$$\alpha_k = 0, \quad k \geq 0\,.$$

The classical Szegő-Verblunsky theorem (see [41], Theorem 1.8.6) is the identity

$$\frac{1}{2\pi} \int_{\mathbb{T}} \log g_\nu(e^{\mathrm{i}\theta}) d\theta = \sum_{k=0}^{\infty} \log(1 - |\alpha_k|^2)\,, \tag{3.23}$$

where $\nu \in \mathcal{M}_1(\mathbb{T})$ is nontrivial with V-coefficients α_k and with Lebesgue decomposition

$$d\nu = g_\nu d \,\mathrm{UNIF} + d\nu_s$$

with respect to UNIF. Changing signs in both sides of this equation leads to

$$\mathcal{K}(\mathrm{UNIF}\,|\nu) = -\sum_{k=0}^{\infty} \log(1-|\alpha_k|^2). \tag{3.24}$$

3.2.2. *Gross-Witten.* The Gross-Witten measures are a class of equilibrium measures for random matrix distributions with potential

$$\mathcal{V}_{\mathtt{g}}(z) = -\mathtt{g}\,\mathrm{Re}(z)\,, \tag{3.25}$$

with parameter $\mathtt{g} \in \mathbb{R}$. For details and applications of this distribution we refer to [30] p. 203, [29], and [42].

If $-1 \le \mathtt{g} \le 1$ (ungapped or strongly coupled phase), the Gross-Witten measure $\mathrm{GW}(\mathtt{g})$ is supported by $\mathbb{T}$ and is given by :

$$\mathrm{GW}(\mathtt{g})(dz) = \frac{1}{2\pi}(1+\mathtt{g}\cos\theta)\,d\theta, \tag{3.26}$$

with $z = e^{\mathrm{i}\theta}, \theta \in [-\pi,\pi)$. Note that $\tau_\pi(\mathrm{GW}(\mathtt{g})) = \mathrm{GW}(-\mathtt{g})$, where

$$\int f(\theta) d\tau_\pi(\mu)(\theta) = \int f(\theta+\pi) d\mu(\theta)\,. \tag{3.27}$$

Since[1]

$$\alpha_k(\tau_\pi(\mu)) = (-1)^{k+1}\alpha_k(\mu), \tag{3.28}$$

see [39], we state the V-coefficients only for the case $\mathtt{g} < 0$.

For $-1 \le \mathtt{g} < 0$, the measure $\mathrm{GW}(\mathtt{g})$ has V-coefficients

$$\alpha_n^{\mathtt{g}} = \alpha_n(\mathrm{GW}(\mathtt{g})) = \begin{cases} -\dfrac{x_+ - x_-}{x_+^{n+2} - x_-^{n+2}} & \text{if } -1 < \mathtt{g} < 0, \\ -\dfrac{1}{n+2} & \text{if } \mathtt{g} = -1\,, \end{cases} \tag{3.29}$$

(see Simon [39], p. 86), where $x_\pm = -\mathtt{g}^{-1} \pm \sqrt{\mathtt{g}^{-2}-1}$ are roots of the equation

$$x + \frac{1}{x} = -\frac{2}{\mathtt{g}}\,.$$

We remark that the measure $\mathrm{GW}(\mathtt{g})$ has only nontrivial moments of order ± 1.

For $|\mathtt{g}| \ge 1$ (gapped or weakly coupled phase), let $\theta_g \in [0,\pi]$ be the solution of

$$\sin^2(\theta_{\mathtt{g}}/2) = |\mathtt{g}|^{-1}\,. \tag{3.30}$$

When $\mathtt{g} \le -1$, the Gross-Witten measure is for $z = e^{\mathrm{i}\theta}, \theta \in [0, 2\pi)$

$$\mathrm{GW}(\mathtt{g})(dz) = \frac{|\mathtt{g}|}{\pi}\sin(\theta/2)\sqrt{\sin^2(\theta/2) - \cos^2(\theta_{\mathtt{g}}/2)}\,1_{[\pi-\theta_{\mathtt{g}},\pi+\theta_{\mathtt{g}}]}\,d\theta\,. \tag{3.31}$$

Summarizing formula (7.22) in Zhedanov [43], we have that in the case $\mathtt{g} < -1$ the V-coefficients are

$$\alpha_{n-1}^{\mathtt{g}} = \alpha_{n-1}(\mathrm{GW}(\mathtt{g})) = 1 - \frac{2}{1+q}\frac{1-q^{n+2}}{1-q^{n+1}}\,, \tag{3.32}$$

[1] In the sequel, we will use the notation $\alpha_k(\nu)$ or $u_k(\mu)$ when the context needs the name of the measure we work with.

where

$$q = \left(\sqrt{|\mathsf{g}|} - \sqrt{|\mathsf{g}|-1}\right)^2 . \tag{3.33}$$

Since $0 < q < 1$, it holds that

$$\lim_{n\to\infty} \alpha_n^{\mathsf{g}} = -\sqrt{1-|\mathsf{g}|^{-1}} = -\cos(\theta_{\mathsf{g}}/2) . \tag{3.34}$$

When $g \geq 1$, the equilibrium measure is

$$\mathrm{GW}(\mathsf{g})(dz) = \frac{|\mathsf{g}|}{\pi}\cos(\theta/2)\sqrt{\sin^2(\theta_{\mathsf{g}}/2) - \sin^2(\theta/2)}\, 1_{[-\theta_{\mathsf{g}},\theta_{\mathsf{g}}]}\, d\theta . \tag{3.35}$$

Note that again $\tau_\pi(\mathrm{GW}(\mathsf{g})) = \mathrm{GW}(-\mathsf{g})$, so that by (3.28) the V-coefficients in this case can be obtained from (3.32).

Remark 2. We may rotate GW as in [38] and consider the equilibrium measure obtained by pushing forward $\mathrm{GW}(\mathsf{g})$ by a rotation of angle η instead of π in (3.27).

The first sum rule relative to the Gross-Witten equilibrium measure $\mathrm{GW}(\mathsf{g})$ was discovered by Simon for $\mathsf{g} = -1$ (see [39, Theorem 2.8.1]), proved later with probabilistic methods by Breuer, Simon and Zeitouni [11]. It is easily extended to $|\mathsf{g}| \leq 1$ ([24, Corollary 5.4]). For $-1 \leq \mathsf{g} \leq 0$ and $\nu \in \mathcal{M}_1(\mathbb{T})$ nontrivial, it is the identity

$$\begin{aligned}\mathcal{K}(\mathrm{GW}(\mathsf{g}) \mid \nu) = H(\mathsf{g}) + \frac{\mathsf{g}}{2} - \frac{\mathsf{g}}{2}\sum_{k=0}^{\infty} |\alpha_k - \alpha_{k-1}|^2 \\ + \sum_{k=0}^{\infty} -\log(1-|\alpha_k|^2) + \mathsf{g}|\alpha_k|^2 ,\end{aligned} \tag{3.36}$$

where

$$H(\mathsf{g}) = \mathcal{K}(\mathrm{GW}(\mathsf{g}) \mid \mathrm{UNIF}) = 1 - \sqrt{1-\mathsf{g}^2} + \log\frac{1+\sqrt{1-\mathsf{g}^2}}{2}. \tag{3.37}$$

We recall that in (3.36), $\alpha_{-1} = -1$. The sum rule (3.36) implies the following *gem*, conditions for finiteness of the Kullback-Leibler divergence. The RHS of (3.36) is finite if and only if

$$\sum_{k=0}^{\infty} |\alpha_k|^2 < \infty \text{ if } -1 < \mathsf{g} \leq 0, \tag{3.38}$$

$$\sum_{k=0}^{\infty} |\alpha_k|^4 < \infty \text{ and } \sum_{k=1}^{\infty} |\alpha_k - \alpha_{k-1}|^2 < \infty \text{ if } \mathsf{g} = -1 . \tag{3.39}$$

Remark 3. Provided that $\sum_k |\alpha_k|^2 < \infty$, we may rewrite the sum rule as

$$\mathcal{K}(\mathrm{GW}(\mathsf{g}) \mid \nu) = H(\mathsf{g}) + \mathsf{g}\,\mathrm{Re}\sum_{k=0}^{\infty} \alpha_k\bar{\alpha}_{k-1} - \sum_{k=0}^{\infty}\log(1-|\alpha_k|^2), \tag{3.40}$$

[39, p. 174], for the case $\mathsf{g} = -1$, which is extended to $-1 < \mathsf{g} \leq 0$ [24, Corollary 5.4]. Actually, since the LHS vanishes for $\nu = \mathrm{GW}(\mathsf{g})$, we can also rewrite the sum rule (3.36) as

$$\mathcal{K}(\mathrm{GW}(\mathsf{g}) \mid \nu) = \mathsf{g}\,\mathrm{Re}\sum_{k=0}^{\infty}(\alpha_k\bar{\alpha}_{k-1} - \alpha_k^{\mathsf{g}}\bar{\alpha}_{k-1}^{\mathsf{g}}) - \sum_{k=0}^{\infty}\log\frac{1-|\alpha_k|^2}{1-|\alpha_k^{\mathsf{g}}|^2} , \tag{3.41}$$

where $\alpha_k^{\mathtt{g}}$ is in (3.29). This RHS is also the RHS of a sum rule for GW($\mathtt{g}$) with $|\mathtt{g}| > 1$ (see [25]).

3.2.3. *Hua-Pickrell.* The Hua-Pickrell distribution appears in the study of random matrices corresponding to the potential

$$\mathcal{V}_{\mathtt{d}}(z) = -2\mathtt{d} \log |1 - z| \,, \tag{3.42}$$

which is invariant by $z \mapsto \bar{z}$. Here, $\mathtt{d}$ is a complex parameter. It has been introduced in [31] and appeared later in [36]. We also refer to [34], [7] and [8]. We will consider here only the case of real parameter $\mathtt{d} > 0$.

The equilibrium measure is the measure

$$\mathrm{HP}(\mathtt{d})(dz) = (1+\mathtt{d}) \frac{\sqrt{\sin^2(\theta/2) - \sin^2(\theta_{\mathtt{d}}/2)}}{2\pi \, \sin(\theta/2)} \mathbb{1}_{(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})}(\theta) d\theta \,, \tag{3.43}$$

with $z = e^{\mathrm{i}\theta}, \theta \in [0, 2\pi]$ and where $\theta_{\mathtt{d}} \in (0, \pi)$ is such that

$$\sin(\theta_{\mathtt{d}}/2) = \frac{\mathtt{d}}{1+\mathtt{d}} \,. \tag{3.44}$$

The orthogonal polynomials with respect to HP($\mathtt{d}$) are the Geronimus polynomials with constant V-coefficients

$$\alpha_k \equiv \gamma_{\mathtt{d}} := -\frac{\mathtt{d}}{1+\mathtt{d}} \ , \ k \geq 0 \,. \tag{3.45}$$

For $\gamma \in \mathbb{D}$, let

$$H_{\mathtt{d}}(\gamma) = -\log \frac{1 - |\gamma|^2}{1 - \gamma_{\mathtt{d}}^2} - 2\mathtt{d} \log \frac{|1 - \gamma|}{1 - \gamma_{\mathtt{d}}} \,. \tag{3.46}$$

The arguments of the functions $H_{\mathtt{d}}$ in the sum rule are the deformed V-coefficients (see [8, Section 2.2]). For a nontrivial measure $\nu \in \mathcal{M}_1(\mathbb{T})$ they form a sequence of parameters $\gamma_k \in \mathbb{D}$, $k \geq 0$ defined by

$$\gamma_k = \bar{\alpha}_k \frac{\Phi_k^*(1)}{\Phi_k(1)}, \quad (k \geq 0). \tag{3.47}$$

and can be computed via the recursive definition

$$\gamma_0 = \bar{\alpha}_0, \quad \gamma_k = \bar{\alpha}_k \prod_{j=0}^{k-1} \frac{1 - \bar{\gamma}_j}{1 - \gamma_j}, \quad (k \geq 1). \tag{3.48}$$

Of course, when ν is symmetric, then $\Phi_k^*(1) = \Phi_k(1)$ and α_k is real, so that the deformed V-coefficients are the genuine V-coefficients.

Furthermore, define the function $\mathcal{F}_{\mathrm{HP}}$ for $\theta \notin (\theta_{\mathtt{d}}, 2\pi - \theta_d)$:

$$\mathcal{F}_{\mathrm{HP}}(e^{\mathrm{i}\theta}) := \int_{I(\theta)} (1+\mathtt{d}) \frac{\sqrt{\sin^2\left(\theta_{\mathtt{d}}/2\right) - \sin^2(\varphi/2)}}{\sin(\varphi/2)} \, d\varphi \tag{3.49}$$

where $I(\theta) = [\theta, \theta_{\mathtt{d}}]$ if $\theta \in (0, \theta_{\mathtt{d}}]$ and $I(\theta) = [2\pi - \theta_{\mathtt{d}}, \theta]$ if $\theta \in [2\pi - \theta_{\mathtt{d}}, 2\pi)$.

Then the following sum rule holds.

Theorem 3.4. [24, Theorem 5.1] *Let* $\mathtt{d} \geq 0$ *and* $\nu \in \mathcal{M}_1(\mathbb{T})$ *be nontrivial with* $(\gamma_k)_{k\geq 0} \in \mathbb{D}^{\mathbb{N}}$ *the sequence of its deformed V-coefficients. Then, if* $\nu \in \mathcal{S}_1^{\mathbb{T}}(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$,

$$\mathcal{K}(\mathrm{HP}(\mathtt{d})|\nu) + \sum_{\lambda \in E(\nu)} \mathcal{F}_{\mathrm{HP}}(\lambda) = \sum_{k=0}^{\infty} H_{\mathtt{d}}(\gamma_k)\,, \tag{3.50}$$

where both sides may be infinite simultaneously. If $\mu \notin \mathcal{S}_1^{\mathbb{T}}(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$*, the RHS equals* $+\infty$.

3.2.4. *Poisson.* The Poisson kernel is the probability measure on $\mathbb{T}$ given by

$$\mathrm{Pois}(\zeta)(dz) = \frac{1 - |\zeta|^2}{2\pi|z - \zeta|^2} dz\,. \tag{3.51}$$

It is the equilibrium measure of random matrices with potential

$$\mathcal{V}_\zeta(z) = \log|z\bar{\zeta} - 1|^2, \tag{3.52}$$

see [30, Proposition 5.3.9], or [31] and [2] for the study of the random matrix ensembles. Note that $\mathrm{Pois}(0) = \mathrm{UNIF}$.

The V-coefficients of $\mathrm{Pois}(\zeta)$ are

$$\alpha_0 = \zeta, \quad \alpha_k = 0 \ (k \geq 1)\,. \tag{3.53}$$

We are aware of two sum rules relative to the Poisson measure $\mathrm{Pois}(\zeta)$. The first one (Theorem 2.5.1 and formula (2.2.77) in [39]) is

$$\mathcal{K}(\mathrm{Pois}(\zeta)|\nu) = -\log \lambda_\infty(\zeta),$$

where, with φ_n the n-th orthonormal polynomial with respect to ν,

$$\lambda_\infty(\zeta) = (1 - |\zeta|^2) \lim_{n\to\infty} |\varphi_n^*(\zeta)|^{-2}\,.$$

The second one is quoted in Proposition 6.3. Its statement needs some notations given later in the paper. We state a third new sum rule in Theorem 6.2.

4. Mappings

Apart from the last one, all the mappings presented here are from $\mathbb{T}$ to $\mathbb{R}$. They push forward a probability measure ν on the circle to a probability measure μ on the real line, which implies a possible connection of the J-coefficients of μ in terms of the V-coefficients of ν. This may induce a connection between J_μ and $\mathcal{C}_\nu$ and also a correspondence, or "gateway" between sum rules.

4.1. Szegő.

The Szegő mapping from $\mathbb{T}$ to $[-2, 2]$ is defined by

$$z \mapsto \mathrm{Sz}(z) = z + z^{-1}\,, \tag{4.1}$$

or in angular coordinates,

$$\mathrm{Sz}(e^{\mathrm{i}\theta}) = 2\cos\theta\,. \tag{4.2}$$

The mapping Sz is two-to-one from $\mathbb{T}$ to $[-2, 2]$. For a symmetric $\nu \in \mathcal{M}_{1,s}(\mathbb{T})$, we let $\mathrm{Sz}(\nu) = \nu \circ \mathrm{Sz}^{-1}$ be the pushforward of ν by the Szegő mapping, which induces a bijection between $\mathcal{M}_{1,s}(\mathbb{T})$ and $\mathcal{M}_1([-2, 2])$, the set of probability measures on

$[-2, 2]$. The OPUC $(\varphi_n)_{n\geq 0}$ with respect to ν and the OPRL $(p_n)_{n\geq 0}$ with respect to $\mathrm{Sz}(\nu)$ are related by

$$p_n(z) = z^{-n} \frac{\varphi_{2n}(z) + \varphi_{2n}^*(z)}{\sqrt{2(1-\alpha_{2n-1})}}, \tag{4.3}$$

where α_{2n-1} are the real V-coefficients of ν. The Geronimus relations [40, Theorem 13.1.7] and equation (2.5) give the remarkable identity

$$u_k(\mathrm{Sz}(\nu)) = \alpha_{k-1}(\nu) \tag{4.4}$$

for $k \geq 1$ between the canonical moments of $\mathrm{Sz}(\nu)$ and the V-coefficients of ν.

4.2. **Delsarte-Genin (DG).** For $\mathfrak{d} > 0$ we consider the following relation between a point $z \in \mathbb{T}$ and $x \in [-2\mathfrak{d}, 2\mathfrak{d}]$ given by

$$x = \mathfrak{d}\left(z^{1/2} + z^{-1/2}\right) \quad \text{or} \quad x = 2\mathfrak{d}\cos(\theta/2)\,. \tag{4.5}$$

The following computations mainly come from [18], [19], [17]. Therein, the parameter $\mathfrak{d}$ is fixed to $\frac{1}{2}$. Notice that the concern about branches of the square-root is addressed in [20, p. 518] . With the right choice, this mapping is a bijection from $\mathbb{T}\setminus\{1\}$ to $(-2\mathfrak{d}, 2\mathfrak{d})$, which we denote by $\mathrm{DG}_\mathfrak{d}$, the point $1 \in \mathbb{T}$ corresponds to both $-2\mathfrak{d}$ and $2\mathfrak{d}$.

Let $\nu \in \mathcal{M}_{s,1}(\mathbb{T})$ and fix $\mathfrak{d} = 1$. We let $\mathrm{DG}_1(\nu)$ be the pushforward of ν by DG_1, with the convention that $\mathrm{DG}_1(\nu)(\{-2\}) = \mathrm{DG}_1(\nu)(\{2\}) = \frac{1}{2}\nu(\{1\})$. It is a symmetric measure on $[-2, 2]$. The monic orthogonal polynomials $(\Phi_n)_{n\geq 0}$ with respect to ν and the monic orthogonal polynomials $(P_n)_{n\geq 0}$ with respect to $\mu = \mathrm{DG}_1(\nu)$ are related by

$$P_n(x) = \frac{z^{-n/2}(\Phi_n(z) + \Phi_n^*(z))}{\sqrt{2(1-\alpha_{n-1})}}\,, \tag{4.6}$$

where α_{n-1} are the real V-coefficients of ν. The J-coefficients of μ are

$$a_n^2 = (1 + \alpha_{n-1})(1 - \alpha_{n-2}), \quad b_n = 0 \qquad (n \geq 1). \tag{4.7}$$

The canonical coefficients of μ of odd index are zero by symmetry, so comparing (2.5) and (4.7) we conclude

$$u_{2n}(\mathrm{DG}_1(\nu)) = \alpha_{n-1}(\nu)\ (n \geq 1), \qquad u_0 = \alpha_{-1} = -1\,. \tag{4.8}$$

The inverse relation between Φ_n and P_n is

$$\Phi_n(z) = \frac{z^{n/2}\left(z^{1/2}P_{n+1}(x) - \sigma_n P_n(x)\right)}{z-1}, \tag{4.9}$$

with

$$\sigma_n = \frac{P_{n+1}(2)}{P_n(2)} = 1 - \alpha_{n-1}. \tag{4.10}$$

An easy rescaling is helpful when considering the general case $\mathfrak{d} \neq 1$. Indeed, write the polynomials orthogonal to $\mathrm{DG}_\mathfrak{d}(\nu)$ as $P_n(x;\mathfrak{d}) = \mathfrak{d}^n P_n(x/\mathfrak{d})$. Then their V-coefficients satisfy

$$a_n^2 = \mathfrak{d}^2(1 + \alpha_{n-1})(1 - \alpha_{n-2})\,. \tag{4.11}$$

Sometimes it is more convenient to use the mapping

$$x = -i\mathfrak{d}(z^{1/2} - z^{-1/2})\,, \tag{4.12}$$

or $x = 2\mathfrak{d}\sin\theta$. In this case we denote this mapping by $\mathrm{DG}_{\mathfrak{d}}^{-}$ and the classical mapping by $\mathrm{DG}_{\mathfrak{d}}^{+}$.

4.3. **Derevyagin-Vinet-Zhedanov (DVZ).** This map was intruduced in [20] and generalized in [14]. It gives a remarkable relation between symmetric measures on $\mathbb{T}$ and measures on $\mathbb{R}$, induced by a algebraic relation between the CMV matrix and the Jacobi matrix, also called the Schur-Delsarte-Genin (SDG) map by [14].

Let $\nu \in \mathcal{M}_{s,1}(\mathbb{T})$ be a symmetric measure. Its V-coefficients are real, and when the associated CMV matrix is written as in (2.11) in the form $\mathcal{C} = \mathcal{L}\mathcal{M}$, we have in this case $\Theta_k^2 = I_2$ (the identity in $\mathbb{R}^2$) for all k. This implies $\mathcal{L}^2 = \mathcal{M}^2 = I$ and the matrix $J_+ := \mathcal{L} + \mathcal{M}$ satisfies the following properties:

(1) J_+ is real tridiagonal symmetric.
(2) $J_+^2 - 2I = \mathcal{C} + \mathcal{C}^t$
(3) The J-coefficients in J_+ are

$$a_k = \rho_{k-1}, \quad b_{k+1} = \alpha_k - \alpha_{k-1} \ \ (k \geq 1) \tag{4.13}$$

and $b_1 = \alpha_0 + 1$.
(4) Its spectral measure is given by

$$d\mu(x) = \frac{1}{2}(2 + x)\, d\,\mathrm{DG}_1(\nu), \tag{4.14}$$

supported on $[-2, 2]$. Let us notice that this measure is not symmetric.

The measure μ defined by (4.14) will be denoted by $\mathrm{DVZ}^+(\nu)$.

If we consider $J_- = \mathcal{L} - \mathcal{M}$ then the spectral measure satisfies

$$d\mu(x) = \frac{1}{2}(2 - x)\, d\,\mathrm{DG}_1(\nu)\,, \tag{4.15}$$

and it is denoted by $\mathrm{DVZ}^-(\nu)$.

4.4. **Möbius.** The Möbius transform m_{z_0} for $z_0 \in \mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ is defined by

$$m_{z_0}(z) = \frac{z - z_0}{1 - \bar{z}_0 z}\,. \tag{4.16}$$

It is an automorphism of $\mathbb{D}$, sending z_0 to 0, or of $\mathbb{T}$. Its inverse is m_{-z_0}.

5. Pushing forward measures

5.1. **Transformation of** UNIF**.** From the definitions we see easily that

$$\mathrm{Sz}(\mathrm{UNIF}) = \mathrm{DG}_1(\mathrm{UNIF}) = \mathrm{Arcsine} \tag{5.1}$$

To compute the DVZ transform of UNIF, let us introduce the following notation. For $a < b$ the measure $\mathcal{D}(a, b)$ (resp. $\mathcal{D}(b, a)$) is supported by (a, b) with density

$$\frac{2}{\pi(b-a)}\sqrt{\frac{x-a}{b-x}} \quad \left(\text{resp.} \quad \frac{2}{\pi(b-a)}\sqrt{\frac{b-x}{x-a}}\right)\,. \tag{5.2}$$

These measures are affine pushforwards of the beta-distribution with parameter $\frac{1}{2}, -\frac{1}{2}$ (resp. $-\frac{1}{2}, \frac{1}{2}$). The measure $\mathcal{D}(2, -2)$ is also a shift of the Marchenko-Pastur distribution, in the hard edge case.

The associated orthonormal polynomials are (up to an affine change) Chebyshev of the third type (resp. fourth type). We then have

$$\mathrm{DVZ}^+(\mathrm{UNIF}) = \mathcal{D}(-2,2) \quad \text{and} \quad \mathrm{DVZ}^-(\mathrm{UNIF}) = \mathcal{D}(2,-2).$$

The J-coefficients of $\mathcal{D}(-2,2)$ are by (4.13)

$$a_k = 1, \quad b_{k+1} = 0 \ (k \geq 1), \tag{5.3}$$

and $b_1 = 1$.

5.2. **Transformation of** GW**.** From the density (3.31), we deduce for $\mathsf{g} \leq -1$

$$\mathrm{Sz}(\mathrm{GW}(\mathsf{g}))(dx) = \frac{|\mathsf{g}|}{2\pi}\sqrt{\frac{4|\mathsf{g}|^{-1} - 2 - x}{x+2}}\,\mathbb{1}_{(-2,4|\mathsf{g}|^{-1}-2)}(x)\ dx$$

or in other words

$$\mathrm{Sz}(\mathrm{GW}(\mathsf{g})) = \mathcal{D}(4|\mathsf{g}|^{-1} - 1, -2)\,. \tag{5.4}$$

For $|\mathsf{g}| \leq 1$ let us notice that the Gross-Witten density (3.26) is the mixture:

$$\mathrm{GW}(\mathsf{g}) = |\mathsf{g}|\,\mathrm{GW}(\epsilon(g)) + (1 - |\mathsf{g}|)\,\mathrm{UNIF}\,, \tag{5.5}$$

where $\epsilon(\mathsf{g})$ is the sign of g. Since the Szegő mapping acts linearly on measures, we obtain the complete picture

$$\mathrm{Sz}(\mathrm{GW}(\mathsf{g})) = \mu_{\mathsf{g}} =: \begin{cases} \mathcal{D}(-2 + 4|\mathsf{g}|^{-1}, -2) & \text{if } \mathsf{g} \leq -1, \\ |\mathsf{g}|\mathcal{D}(2,-2) + (1 - |\mathsf{g}|)\,\mathrm{Arcsine} & \text{if } -1 \leq \mathsf{g} \leq 0, \\ \mathsf{g}\mathcal{D}(-2,2) + (1-\mathsf{g})\,\mathrm{Arcsine} & \text{if } 0 \leq \mathsf{g} \leq 1, \\ \mathcal{D}(2 - 4\mathsf{g}^{-1}, 2) & \text{if } \mathsf{g} \geq 1. \end{cases} \tag{5.6}$$

5.2.1. *With* DG *when* $|\mathsf{g}| \geq 1$*.* For $\mathsf{g} < -1$, the change of variable

$$x = 2\sqrt{|\mathsf{g}|}\cos(\theta/2)$$

gives

$$\mathrm{DG}_{\sqrt{|\mathsf{g}|}}(\mathrm{GW}(\mathsf{g})) = \mathrm{SC} \tag{5.7}$$

When $\mathsf{g} > 1$, the change of variable

$$x = 2\sqrt{\mathsf{g}}\sin(\theta/2)$$

gives also $\mathrm{DG}^-_{\sqrt{\mathsf{g}}}(\mathrm{GW}(\mathsf{g})) = \mathrm{SC}$.

5.2.2. *With* DG *when* $|\mathsf{g}| \leq 1$*.* Starting from (5.5) and since DG is linear, we get for $-1 \leq \mathsf{g} < 0$

$$\mathrm{DG}_1(\mathrm{GW}(\mathsf{g})) = \rho_{\mathsf{g}} := |\mathsf{g}|\,\mathrm{SC} + (1 - |\mathsf{g}|)\,\mathrm{Arcsine}\,. \tag{5.8}$$

The corresponding canonical moments are

$$u_{2k} = \alpha^{\mathsf{g}}_{k-1}, \quad u_{2k+1} = 0 \quad (k \geq 0)\,, \tag{5.9}$$

where α^{g}_k is in (3.29).

Similarly the application of DG^-_1 leads, for $0 < \mathsf{g} \leq 1$, to the mixture

$$\mathrm{DG}^-_1(\mathrm{GW}(\mathsf{g})) = \mathsf{g}\,\mathrm{SC} + (1-\mathsf{g})\,\mathrm{Arcsine}\,. \tag{5.10}$$

5.3. **Transformation of** HP**.** The change of variable $x = 2\cos\theta$ in (3.43) gives

$$\frac{\mathrm{Sz}(\mathrm{HP}(\mathtt{d}))(dx)}{dx} = \frac{2(1+\mathtt{d})}{2\pi}\frac{\sqrt{x_{\mathtt{d}}-x}}{(2-x)\sqrt{2+x}} = \frac{1+\mathtt{d}}{\pi}\frac{\sqrt{(x_{\mathtt{d}}-x)(2+x)}}{(4-x^2)}\,.$$

where

$$x_{\mathtt{d}} = \frac{2(1+2\mathtt{d}-\mathtt{d}^2)}{1+2\mathtt{d}+\mathtt{d}^2}\,. \tag{5.11}$$

We conclude that

$$\mathrm{Sz}(\mathrm{HP}(\mathtt{d})) = \mathrm{KMK}(2\mathtt{d}, 0)\,, \tag{5.12}$$

(recall that $\mathrm{KMK}(2\mathtt{d}, 0)$ is supported on $[-2, x_{\mathtt{d}}]$).

The change of variable $x = 2\cos(\theta/2)$ in (3.43) gives for the density of $\mathrm{DG}_1(\mathrm{HP}(\mathtt{d}))$

$$(1+\mathtt{d})\frac{\sqrt{\cos^2(\theta_{\mathtt{d}}/2)-\cos^2(\theta/2)}}{2\pi\sin^2(\theta/2)} = (1+\mathtt{d})\frac{\sqrt{4\cos^2(\theta_{\mathtt{d}}/2)-x^2}}{\pi(4-x^2)}\,,$$

so that we conclude

$$\mathrm{DG}_1(\mathrm{HP}(\mathtt{d})) = \mathrm{KMK}(\mathtt{d})\,. \tag{5.13}$$

which is supported by $[-\hat{x}_{\mathtt{d}}, \hat{x}_{\mathtt{d}}]$, with

$$\hat{x}_{\mathtt{d}} = 2\frac{\sqrt{1+2\mathtt{d}}}{1+\mathtt{d}}. \tag{5.14}$$

Note that the V-coefficients of $\mathrm{HP}(\mathtt{d})$ are constant equal to $\gamma_{\mathtt{d}}$ and then by (4.7) the J-coefficients of $\mathrm{KMK}(\mathtt{d})$ are

$$a_1^2 = 2(1+\gamma_{\mathtt{d}}),\quad a_n^2 = (1-\gamma_{\mathtt{d}}^2)\ \ (n\geq 2),\quad b_n = 0\ \ (n\geq 1)\,,$$

which agrees with (3.12).

Let us summarize the above results by two tables :

	Sz
UNIF	Arcsine
$\mathrm{GW}(\mathtt{g}), \lvert\mathtt{g}\rvert \leq 1$	$\lvert\mathtt{g}\rvert\mathcal{D}(-2\epsilon(\mathtt{g}), 2\epsilon(g)) + (1-\lvert\mathtt{g}\rvert)$ Arcsine
$\mathrm{GW}(\mathtt{g}), \lvert\mathtt{g}\rvert > 1$	$\mathcal{D}(2\epsilon(\mathtt{g}) - 4\mathtt{g}^{-1}, 2\epsilon(\mathtt{g}))$
$\mathrm{HP}(\mathtt{d})$	$\mathrm{KMK}(2\mathtt{d}, 0)$

	DG_1^+	DG_1^-
UNIF	Arcsine	Arcsine
$\mathrm{GW}(\mathtt{g}), -1 \leq \lvert\mathtt{g}\rvert \leq 0$	$\lvert\mathtt{g}\rvert$ SC $+(1-\lvert\mathtt{g}\rvert)$ Arcsine	
$\mathrm{GW}(\mathtt{g}), 0 \leq \lvert\mathtt{g}\rvert \leq 1$		$\lvert\mathtt{g}\rvert$ SC $+(1-\lvert\mathtt{g}\rvert)$ Arcsine
$\mathrm{GW}(\mathtt{g}), \mathtt{g} \leq -1$	SC $(*)$	
$\mathrm{GW}(\mathtt{g}), \mathtt{g} > 1$		SC $(*)$
$\mathrm{HP}(\mathtt{d})$	$\mathrm{KMK}(\mathtt{d})$	

Here $(*)$ means that $\mathrm{DG}^{\pm}_{\sqrt{\lvert\mathtt{g}\rvert}}$ is applied.

6. Gateways

In this section, we highlight connections between different sum rules arising when measures are transformed by the mappings of Section 4. Unlike the large deviation technique developed in [23], this method designing new sum rules is purely analytical. Nevertheless, it requires an existing sum rule to run. In most cases, the aim is to obtain an OPRL sum rule from an OPUC one.

A measurable mapping $\varphi : X \to Y$ between metric spaces induces a mapping from $\mathcal{M}_1(X)$ to $\mathcal{M}_1(Y)$ by $\mu \mapsto \varphi(\mu) = \mu \circ \varphi^{-1}$. A sum rule for measures in $\mathcal{M}_1(Y)$ may lead to an identity for measures $\mu \in \mathcal{M}_1(X)$ (or vice versa) by evaluating both sides of the sum rule for $\varphi(\mu)$.

Suppose the mapping $\varphi : X \to Y$ is a bijection. Then $\mu \mapsto \varphi(\mu)$ is a bijection from $\mathcal{M}_1(X)$ to $\mathcal{M}_1(Y)$. The entropy part of a sum rule can then be obtained directly by the reversible entropy principle:

$$\mathcal{K}(\mu_0 \mid \mu) = \mathcal{K}(\varphi(\mu_0) \mid \varphi(\mu)). \tag{6.1}$$

Among the mappings introduced in Section 4, only the Möbius mapping is one to one. However, the Szegő mapping is a bijection between symmetric measures on $\mathbb{T}$ and measures on $[-2, 2]$ and (6.1) still holds for $\varphi = \mathrm{Sz}$ and $\mu, \mu_0 \in \mathcal{M}_{1,s}(\mathbb{T})$. The mappings $\mathrm{DG}^{\pm}_{\mathfrak{d}}$ are bijective when restricted to $\mathbb{T} \setminus \{1\}$. With the convention on mapping the mass at 1, (6.1) also holds for $\varphi = \mathrm{DG}^{\pm}_{\mathfrak{d}}$. Then (6.1) also holds for $\varphi = \mathrm{DVZ}^{\pm}$ and therefore for all mappings considered in this paper.

Transforming the RHS of a sum rule is less straightforward, but may be simplified if the coefficients of $\varphi(\mu)$ are connected with those of μ in a convenient way.

6.1. **From** UNIF **to** Arcsine.

Using (6.1) we get, with $\nu \in \mathcal{M}_{1,s}(\mathbb{T})$ such that $\mathrm{Sz}(\nu) = \mu$,

$$\mathcal{K}(\mathrm{Arcsine} \mid \mu) = \mathcal{K}(\mathrm{UNIF} \mid \nu).$$

The Szegő formula (3.24), jointly with (4.4), gives

$$\mathcal{K}(\mathrm{UNIF} \mid \nu) = -\sum_{k=0}^{\infty} \log(1 - |\alpha_k|^2(\nu)) = -\sum_{k=1}^{\infty} \log(1 - u_k^2(\mu)) \,,$$

and we recover the sum rule relative to the arcsine law (3.18). Notice that UNIF is a particular case of the following distributions,

$$\mathrm{UNIF} = \mathrm{HP}(0) = \mathrm{GW}(0).$$

So that, the sum rule relative to Arcsine can be recovered from any sum rule relative to one of these distributions.

6.2. **From** HP **to** KMK.

The Kesten-McKay laws can be obtained from the Hua-Pickrell distribution as

$$\mathrm{Sz}(\mathrm{HP}(\mathtt{d})) = \mathrm{KMK}(2\mathtt{d}, 0), \quad \mathrm{DG}_1(\mathrm{HP}(\mathtt{d})) = \mathrm{KMK}(\mathtt{d}).$$

These distributional identities allow to recover the sum rules with reference measure $\mathrm{KMK}(2\mathtt{d}, 0)$ or $\mathrm{KMK}(\mathtt{d})$. If ν is a symmetric distribution in $\mathcal{S}_1^{\mathbb{T}}(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$ then $\mathrm{Sz}(\mathrm{HP}(\mathtt{d}))$ (resp. $\mathrm{DG}_1(\nu)$) is supported on $[-2, 2]$ and belongs to $\mathcal{S}_1^{\mathbb{R}}(-2, x_{\mathtt{d}})$ (resp. $\mathcal{S}_1^{\mathbb{R}}(-\hat{x}_{\mathtt{d}}, \hat{x}_{\mathtt{d}})$).

Let us consider the Szegő mapping of the sum rule (3.50). Let μ be supported on $[-2,2]$. If $\mu \in \mathcal{S}_1^{\mathbb{R}}(-2, x_{\mathtt{d}})$, then a symmetric measure ν with $\mathrm{Sz}(\nu) = \mu$ belongs to $\mathcal{S}_1^{\mathbb{T}}(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$. The LHS of the sum rule (3.50) evaluated at ν is

$$\mathcal{K}(\mathrm{HP}(\mathtt{d})|\nu) + \sum_{\lambda\in E(\nu)} \mathcal{F}_{\mathrm{HP}(\mathtt{d})}(\lambda) = \mathcal{K}(\mathrm{KMK}(2\mathtt{d},0)|\mu) + 2\sum_{\lambda\in E(\mu)} \mathcal{F}(\lambda) \tag{6.2}$$

where the factor 2 comes from the two support points in $E(\nu)$ corresponding to one support point in $E(\mu)$, and where

$$2\mathcal{F}(x) = 2\mathcal{F}_{\mathrm{HP}(\mathtt{d})}(e^{\mathrm{i}\arccos(x/2)}) = 2\int_{x_{\mathtt{d}}}^{x} (1+\mathtt{d})\frac{\sqrt{t-x_{\mathtt{d}}}}{(2-t)\sqrt{2+t}}\, dt \ = \mathcal{F}_{\mathrm{KMK}(2\mathtt{d},0)}(x). \tag{6.3}$$

We therefore obtain (a particular case of) the LHS of (3.17). When looking at the RHS of (3.50) evaluated at the symmetric measure ν, we see that the deformed V-coefficients γ_k are the regular V-coefficients α_k. But by the relation (4.4), they are the canonical moments u_k of $\mu = \mathrm{Sz}(\nu)$. Consequently, the RHS becomes

$$-\sum_{k=1}^{\infty}\log\frac{1-u_k^2}{1-\gamma_{\mathtt{d}}^2} - 2\mathtt{d}\sum_{k=1}^{\infty}\log\frac{1-u_k}{1-\gamma_{\mathtt{d}}} = -(1+2\mathtt{d})\sum_{k=1}^{\infty}\log\frac{1-u_k}{1-\gamma_{\mathtt{d}}} - \sum_{k=1}^{\infty}\log\frac{1+u_k}{1+\gamma_{\mathtt{d}}}\,. \tag{6.4}$$

The measure $\mathrm{KMK}(2\mathtt{d},0)$ has all canonical moments equal to $\gamma_{\mathtt{d}}$, and so (6.4) is equal to $I_J(\mu)$ as given in (3.15). We have recovered the sum rule corresponding to $\mathrm{KMK}(2\mathtt{d},0)$.

Let us consider the DG mapping and let μ supported on $[-2,2]$ and symmetric. If ν is a symmetric measure on $\mathbb{T}$ such that $\mathrm{DG}_1(\nu) = \mu$ and $\nu \in \mathcal{S}_1^{\mathbb{T}}(\theta_{\mathtt{d}}, 2\pi-\theta_{\mathtt{d}})$, then $\mu \in \mathcal{S}_1^{\mathbb{R}}(-\hat{x}_{\mathtt{d}}, \hat{x}_{\mathtt{d}})$. The LHS of (3.50) can then be transformed as

$$\begin{aligned}&\mathcal{K}(\mathrm{HP}(\mathtt{d})|\nu) + \sum_{\lambda\in E(\nu)} \mathcal{F}_{\mathrm{HP}(\mathtt{d})}(\lambda)\\ &= \mathcal{K}(\mathrm{KMK}(\mathtt{d})|\mu) + \sum_{\lambda\in E(\mu)} \mathcal{F}(\lambda).\end{aligned} \tag{6.5}$$

where for $x > \hat{x}_{\mathtt{d}}$

$$\mathcal{F}(x) = \mathcal{F}_{\mathrm{HP}(\mathtt{d})}(e^{2\mathrm{i}\arccos(x/2)}) = \mathcal{F}_{\mathrm{KMK}(\mathtt{d})}(x),$$

and we obtain the LHS of the KMK sum rule (3.17). Turning to the RHS, we first observe that the canonical moments of μ satisfy by (4.8)

$$u_{2k+1}(\mu) = 0\ ,\ \ u_{2k}(\mu) = \alpha_{k-1}(\nu)\,,$$

hence

$$\begin{aligned}H_{\mathtt{d}}(\gamma_{k-1}) = H_{\mathtt{d}}(\alpha_{k-1}) &= -\log\frac{1-\alpha_{k-1}^2}{1-\gamma_{\mathtt{d}}^2} - 2\mathtt{d}\log\frac{1-\alpha_{k-1}}{1-\gamma_{\mathtt{d}}}\\ &= -(1+2\mathtt{d})\log\frac{1-u_{2k}}{1-\gamma_{\mathtt{d}}} - \log\frac{1+u_{2k}}{1+\gamma_{\mathtt{d}}}\\ &= \mathcal{H}_e^{\kappa,\kappa}(u_{2k})\end{aligned}$$

since $\gamma_{\mathtt{d}} = u_e^{\mathtt{d},\mathtt{d}}$. For odd index, we have $\mathcal{H}_o^{\mathtt{d},\mathtt{d}}(u_{2k-1}) = \mathcal{H}_o^{\mathtt{d},\mathtt{d}}(0) = 0$, since both μ and the reference measure are symmetric. We conclude that the RHS of the DG sum rule transforms exactly to the RHS of the KMK sum rule for symmetric measures.

6.3. **From** GW**.** We now discuss how one may obtain new sum rules starting from the sum rule (3.41) relative to $\mathrm{GW}(\mathsf{g})$, for $-1 \le \mathsf{g} \le 0$. Applying the Szegő mapping leads to a sum rule relative to μ_{g} in (5.6), a mixture of beta distributions. On the other hand, the mapping DG_1 leads to a sum rule relative to a mixture of SC and Arcsine.

6.3.1. *With* Sz. Let μ by a measure on $[-2, 2]$ and $\nu \in \mathcal{M}_{1,s}(\mathbb{T})$ with $\mathrm{Sz}(\nu) = \mu$. Then the LHS of (3.36) applied to ν gives by (6.1)

$$\mathcal{K}(\mathrm{GW}(\mathsf{g})|\nu) = \mathcal{K}(\mu_{\mathsf{g}}|\mu). \tag{6.6}$$

In the RHS of (3.36) evaluated at ν only real V-coefficients appear. By the relation (4.4) we obtain the sum rule for μ with support $[-2, 2]$:

$$\begin{aligned}\mathcal{K}(\mu_{\mathsf{g}}|\mu) = H(\mathsf{g}) + \frac{\mathsf{g}}{2} - \frac{\mathsf{g}}{2}\sum_{k=1}^{\infty}(u_k - u_{k-1})^2 \\ + \sum_{k=1}^{\infty} - \log(1 - u_k^2) + \mathsf{g}u_k^2,\end{aligned} \tag{6.7}$$

where u_k are the canonical moments of μ.

Remark 4. When $\mathsf{g} = -1$, (6.7) becomes

$$\mathcal{K}(\mu_{-1}|\mu) = \frac{1}{2} - \log 2 + \frac{1}{2}\sum_{k=0}^{\infty}(u_{k+1} - u_k)^2 + \sum_{k=1}^{\infty} - \log(1 - u_k^2) - u_k^2, \tag{6.8}$$

which is a version of [39, formula (2.8.6)]. But

$$\mu_{-1} = \mathcal{D}(2, -2) = T(\mathrm{MP}_1)\,, \tag{6.9}$$

and where $T : \xi \mapsto \xi - 2$, so that, by (6.1)

$$\mathcal{K}(\mu_{-1}|\mu) = \mathcal{K}(\mathrm{MP}_1\,|T^{-1}(\mu)) \tag{6.10}$$

The RHS of the sum rule corresponding to $\mathcal{K}(\mathrm{MP}_1\,|T^{-1}(\mu))$ uses coefficients (z_k) associated to $T^{-1}(\mu)$. To get an expression in terms of the (u_k), we notice that if a_k, b_k are the J-coefficients of μ, the J-coefficients of $T^{-1}(\mu)$ are

$$\tilde{a}_k = a_k, \quad \tilde{b}_k = b_k + 2. \tag{6.11}$$

Applying to a_k and b_k the decomposition into the canonical moment u_k of μ according to (2.5), and using the parameters (z_k) defined in (2.4) we obtain the relations

$$z_1(T^{-1}(\mu)) = 2(1 + u_1)(\mu), \quad z_k(T^{-1}(\mu)) = (1 - u_{k-1}(\mu))(1 + u_k(\mu)), \quad (k \ge 1). \tag{6.12}$$

Combining the sum rule (3.7) relative to MP_1 and (6.12), we obtain the identity

$$\mathcal{K}(\mathcal{D}(2, -2)|\mu) = \sum_{k=1}^{\infty}\big((1 - u_{k-1})(1 + u_k) - \log[(1 - u_{k-1})(1 + u_k)] - 1\big) \tag{6.13}$$

for μ with support $[-2, 2]$.

Let us compare the RHS of (6.13) and (6.8) by direct calculation. Denote by S_N the partial sum, up to N, of the RHS in (6.13). Obviously, we may write

$$S_N = \frac{1}{2} - \log 2 + u_N - \log(1+u_N) - \frac{u_N^2}{2} \tag{6.14}$$

$$+ \frac{1}{2}\sum_{k=1}^{N}(u_k - u_{k-1})^2 + \sum_{k=1}^{N-1}(-\log(1-u_k^2) - u_k^2) \tag{6.15}$$

(recall that $u_0 = -1$). If the RHS of (6.7) is finite, the gem (3.39) warrants that

$$\sum_{k=1}^{\infty} u_k^4 < \infty \ , \ \sum_{k=1}^{\infty}(u_{k+1} - u_k)^2 < \infty \ . \tag{6.16}$$

Hence, in particular $u_N \to 0$ and the two sums in (6.15) converge. We therefore recover (6.7). Conversely, if one of the conditions in (6.16) is not satisfied, we have $S_N \to \infty$ since the RHS of (6.14) is bounded below by $-2\log 2 - 1$

6.3.2. *With* DG. Recall that $\mathrm{DG}_1(\mathrm{GW}(\mathtt{g})) = \rho_{\mathtt{g}}$ as in (5.8). If μ is symmetric and supported on $[-2,2]$ and ν is symmetric on $\mathbb{T}$ such that $\mathrm{DG}_1(\nu) = \mu$, then from (6.1)

$$\mathcal{K}(\rho_{\mathtt{g}}|\mu) = \mathcal{K}(\mathrm{GW}(\mathtt{g})|\nu) \tag{6.17}$$

Now, in the sum rule (3.36) the V-coefficients of ν are real and using (4.8), we may rewrite the last identity as

$$\mathcal{K}(\rho_{\mathtt{g}}|\mu) = H(\mathtt{g}) + \frac{\mathtt{g}}{2} - \frac{\mathtt{g}}{2}\sum_{k=1}^{\infty}(u_{2k} - u_{2k-2})^2 + \sum_{k=1}^{\infty} - \log(1-u_{2k}^2) + \mathtt{g}u_{2k}^2. \tag{6.18}$$

Here, u_k is the k-th canonical moment of μ. The following theorem gives an alternative form of the RHS, obtained by combining the two sum rules relative to SC and Arcsine.

Theorem 6.1. *For μ symmetric and supported on $[-2,2]$ and $-1 \le \mathtt{g} \le 0$,*

$$\mathcal{K}(\rho_{\mathtt{g}}|\mu) = C_{\mathtt{g}} + |\mathtt{g}|\sum_{k=1}^{\infty}(a_k^2 - 1 - \log a_k^2) + (1-|\mathtt{g}|)\sum_{k=1}^{\infty}\log a_k^2 \tag{6.19}$$

where

$$C_{\mathtt{g}} = -|\mathtt{g}|(1-\log 2) + 1 - \sqrt{1-\mathtt{g}^2} + \log\frac{1+\sqrt{1-\mathtt{g}^2}}{2}\ . \tag{6.20}$$

Proof. From (5.8), $\rho_{\mathtt{g}} = |\mathtt{g}|\,\mathrm{SC} + (1-|\mathtt{g}|)\,\mathrm{Arcsine}$, and applying Proposition 8.2 we thus obtain

$$\begin{aligned}\mathcal{K}(\rho_{\mathtt{g}} \mid \mu) &= |\mathtt{g}|\big(\mathcal{K}(\mathrm{SC}\,|\mu) - \mathcal{K}(\mathrm{SC}\,|\rho_{\mathtt{g}})\big)\\ &\quad + (1-|\mathtt{g}|)\big(\mathcal{K}(\mathrm{Arcsine}\,|\mu) - \mathcal{K}(\mathrm{Arcsine}\,|\rho_{\mathtt{g}})\big)\end{aligned} \tag{6.21}$$

From the Killip-Simon sum rule (Theorem 3.1) and from (3.19) we know $\mathcal{K}(\mathrm{SC}\,|\mu)$ and $\mathcal{K}(\mathrm{Arcsine}\,|\mu)$ respectively as functions of the J-coefficients. This gives the coefficient dependent part of the RHS of (6.19). To compute the constant $C_{\mathtt{g}}$ we use (6.1), so that

$$\begin{aligned}C_{\mathtt{g}} &= -|\mathtt{g}|\mathcal{K}\left(\mathrm{SC}\,|\rho_{\mathtt{g}}\right) - (1-|\mathtt{g}|)\mathcal{K}\left(\mathrm{Arcsine}\,|\rho_{\mathtt{g}}\right)\\ &= -|\mathtt{g}|\mathcal{K}(\mathrm{GW}(-1)|\,\mathrm{GW}(\mathtt{g})) - (1-|\mathtt{g}|)\mathcal{K}(\mathrm{UNIF}\,|\,\mathrm{GW}(\mathtt{g})).\end{aligned}$$

The final value of $C_{\mathtt{g}}$ is then calculated with the help of [24, formula (7.5)]. $\square$

Remark 5. When $\mathsf{g} \in (-1, 1]$, we may use the alternative formulation (3.41) and compute the RHS using (4.7) and (4.8).

6.3.3. *With* DVZ. Let us restrict again to the case $\mathsf{g} = -1$.

Since $\mathrm{DG}_1(\mathrm{GW}(-1)) = \mathrm{SC}$, the measure $\hat{\mu} = \mathrm{DVZ}(\mathrm{GW}(-1))$ on $[-2, 2]$ is by (4.14)

$$d\hat{\mu}(x) = \frac{2+x}{2} d\,\mathrm{SC}(x) = \frac{1}{2\pi}(2+x)^{3/2}(2-x)^{1/2}dx\,.$$

We can then transform the sum rule (3.36) as follows. Assume that μ is a nontrivial measure supported by $[-2, 2]$ and is such that there exists $\nu \in \mathcal{M}_{s,1}(\mathbb{T})$ such that $\mu = \mathrm{DVZ}(\nu)$. We have by (6.1)

$$\mathcal{K}(\hat{\mu}|\mu) = \mathcal{K}(\mathrm{GW}(-1)|\nu) \tag{6.22}$$

As before the V-coefficients of ν are real and related to the J-coefficients a_k, b_k of μ by (4.13). The sum rule relative to $\mathcal{K}(\mathrm{GW}(-1)|\nu)$ can be rewritten as

$$\mathcal{K}(\hat{\mu}|\mu) = H(-1) - \frac{1}{2} + \frac{1}{2}\sum_{k=0}^{\infty} b_k^2 + \sum_{k=1}^{\infty} a_k^2 - \log(a_k^2) - 1, \tag{6.23}$$

The RHS is therefore $\mathcal{I}_H(\mu)$ as in the Killip-Simon sum rule, (Theorem 3.1), plus the negative constant $H(-1) - \frac{1}{2} = 1/2 - \log 2$. Notice that this does not mean that the RHS may be negative, actually this formula holds for μ in a restricted class.

Alternatively, the sum rule relative to $\hat{\mu}$ may be obtained directly from Theorem 3.1, since

$$\begin{aligned}
\mathcal{K}(\hat{\mu}|\mu) &= \mathcal{K}(\mathrm{GW}(-1)|\nu) = \mathcal{K}(\mathrm{DG}_1(\mathrm{GW}(-1))|\,\mathrm{DG}(\nu)) \\
&= \int \log \frac{d\,\mathrm{SC}}{d\,\mathrm{DG}(\nu)} d\,\mathrm{SC} \\
&= \int \log \frac{d\,\mathrm{SC}}{d\,\mathrm{DVZ}(\nu)} d\,\mathrm{SC} + \int \log\left(1 + \frac{x}{2}\right) d\,\mathrm{SC}(x) \\
&= \mathcal{K}(\mathrm{SC}\,|\,\mathrm{DVZ}(\nu)) + \int \log\left(1 + \frac{x}{2}\right) d\,\mathrm{SC}(x)
\end{aligned}$$

and, as in [1, Exercise 2.6.4],

$$\int \log\left(1 + \frac{x}{2}\right) \mathrm{SC}(dx) = -\log 2 + \frac{1}{2}.$$

6.4. **From** UNIF **to** Pois**.** In this section we investigate sum rules relative to the measure $\mathrm{Pois}(\zeta)$ with $\zeta \in \mathbb{D}$ as given in (3.51). The (reverse) entropy with respect to the Poisson measure is called the Arov-Krein entropy (see [37]). The Möbius transform m_ζ defined in (4.16) maps $\mathrm{Pois}(\zeta)$ to the uniform measure UNIF. Using (6.1) and the Szegő sum rule (3.24) shows that

$$\mathcal{K}(\mathrm{Pois}(\zeta)|\nu) = \sum_{k=0}^{\infty} -\log(1 - |\alpha_k(m_\zeta(\nu))|^2)\,. \tag{6.24}$$

In the following, we analyze this sum rule and obtain alternative expressions for the RHS.

Let us recall the connection between the V-coefficients and the Schur function of a measure [39, Chapter 1]. First, the Caratheodory function of a probability measure ν on $\mathbb{T}$ is defined as

$$F(z) = \int \frac{e^{\mathrm{i}\theta}+z}{e^{\mathrm{i}\theta}-z} d\nu(\theta) = 1 + 2z \int \frac{d\nu(\theta)}{e^{\mathrm{i}\theta}-z}\,.$$

It is analytic on $\mathbb{D}$. The Schur function is then defined from the Caratheodory function as

$$f(z) = \frac{1}{z}\frac{F(z)-1}{F(z)+1} = \frac{1}{z} - \frac{2}{z(F(z)+1)} \tag{6.25}$$

and conversely we have

$$F(z) = \frac{1+zf(z)}{1-zf(z)}, \quad \frac{F(z)-1}{z} = \frac{2f(z)}{1-zf(z)}\,. \tag{6.26}$$

The V-coefficients can be obtained from f by the classical Schur algorithm:

$$S(g)(z) = \frac{1}{z}\frac{g(z)-g(0)}{1-\overline{g(0)}g(z)} \tag{6.27}$$

$$S^{[0]}(g) = g\ ,\ S^{[k+1]}(g) = S \circ S^{[k]} \tag{6.28}$$

$$\alpha_k = S^{[k]}f(0)\,. \tag{6.29}$$

To tackle the Poisson case, we use an extension of the Schur algorithm named the Nevanlinna-Pick algorithm that is defined as follows.

For $\rho \in \mathbb{D}\setminus\{0\}$ set, as in [35]

$$S_\rho(g) = \frac{\omega(\rho)}{m_\rho}\frac{g-g(\rho)}{1-\overline{g(\rho)}g}\ ,\ \omega(\rho) = -\frac{\rho}{|\rho|} \tag{6.30}$$

$$S_\rho^{[0]}(g) = g\ ,\ S_\rho^{[k+1]} = S_\rho \circ S_\rho^{[k]} \quad (k \geq 0)\,, \tag{6.31}$$

where we recall that m_ρ is defined in (4.16). To simplify we set $S_0 = S$.

We have then the following new sum rule, whose proof is postponed to Section 7.

Theorem 6.2. *For ν a nontrivial measure on $\mathbb{T}$ with Schur function f,*

$$\mathcal{K}(\mathrm{Pois}(\zeta)|\nu) = -\log(1-|m_{\bar\zeta}\circ f(\zeta)|^2) + \sum_{k=1}^{\infty} -\log(1-|S_\zeta^{[k]}(f)(\zeta)|^2)\,. \tag{6.32}$$

Remark 6. The prefactor ω introduced in Nevanlinna-Pick theory for technical reasons of infinite product convergence can be omitted here. Noticing that $S_\zeta(\omega g) = \omega S_\zeta(g)$ we can set

$$\hat S_\zeta = \omega^{-1} S_\zeta$$

and get recursively

$$\hat S_\zeta^{[2k]} = S_\zeta^{[2k]}\ , \hat S_\zeta^{[2k+1]} = \omega S_\zeta^{[2k+1]}$$

so that the sum in (6.32) also holds for $\hat S_\zeta^{[k]}$ instead of $S_\zeta^{[k]}$.

There is another Poisson sum rule which is a direct consequence of a recent formula of Bessonov [6]. We neither are able to give any probabilistic interpretation nor to recover it from a pushforward of some other sum rule. The proof of this Poisson sum rule is also postponed to Section 7.

Proposition 6.3. *For ν a nontrivial measure on $\mathbb{T}$ with Schur function f,*

$$\mathcal{K}(\mathrm{Pois}(\zeta)|\nu) = \log \frac{|1-\zeta f(\zeta)|^2}{(1-|\zeta|^2)(1-|f(\zeta)|^2)} + \sum_{k=1}^{\infty} \log \frac{1-|\zeta S_0^{[k]}(f)(\zeta)|^2}{1-|S_0^{[k]}(f)(\zeta)|^2}\,. \tag{6.33}$$

The above series has positive terms since

$$|1-\zeta f|^2 - (1-|\zeta|^2)(1-|f|^2) = |\bar{\zeta} - f|^2 \geq 0,$$

and $1-|\zeta f|^2 > 1-|f|^2$, so that all the terms in the RHS of (6.33) are positive. We conclude with the *gems* corresponding to the above sum rules.

Remark 7. From (6.32) and (6.33) we deduce that the Kullback-Leibler divergence $\mathcal{K}(\mathrm{Pois}(\zeta)|\nu)$ is finite if and only if

$$\sum_{k=1}^{\infty} |S_\zeta^{[k]}(f)(\zeta)|^2 < \infty,$$

or equivalently,

$$\sum_{k=1}^{\infty} |S_0^{[k]}(f)(\zeta)|^2 < \infty\,.$$

7. Proofs of Theorem 6.2 and Prop. 6.3

7.1. **Proof of Theorem 6.2.** The OPUC theory is in many ways an approximation theory. The information carried by the V-coefficients is the same as the one carried by the iterated Schur functions evaluated in 0. This relies on the Schur function and its derivatives at 0. If we are interested in the Schur function and its derivatives at another point $\zeta \in \mathbb{D}$ we fall into the extension of the OPUC theory called Orthogonal Rational Functions (ORF) theory. Our main source for the following developments is [35] (see also [12]).

We start with the sequence of rational functions

$$1, m_\zeta, (m_\zeta)^2, ..., (m_\zeta)^n, ...$$

and we apply the Gram-Schmidt orthonormalization in $L^2(\nu)$ to get

$$1, \varphi_1^o, \varphi_2^o, ...$$

we put the superscript o to stress on the ORF aspect. Let (Φ_n^o) be the corresponding monic ORF's.

Actually, we have

$$\int \overline{\Phi_j^o(z)} \Phi_k^o(z) d\nu(z) = \kappa_k \delta_{jk}$$

and if we set $z = m_{-\zeta}(\tau)$ we get

$$\int \overline{\Phi_j^o \circ m_{-\zeta}(\tau)} \Phi_k^o \circ m_{-\zeta}(\tau) d\nu^o(\tau) = \kappa_k \delta_{jk},,$$

where $\nu^o = \nu \circ m_{-\zeta} = m_\zeta(\nu)$ is the pushforward of ν by m_ζ. Of course the $\Phi_k^o \circ m_{-\zeta}$'s are the monic OPUC with respect to ν^o.

At the level of V-coefficients we have

$$\Phi_k^o \circ m_{-\zeta}(0) = \Phi_k^o(\zeta), \quad \alpha_{k-1}(m_\zeta(\nu)) = -\overline{\Phi_k^o(\zeta)}\,. \tag{7.1}$$

Now, let us study the relation between the Schur functions. We write F and f for the Caratheodory and Schur function of ν and F^o and f^o for the functions of $\nu^o = m_\zeta(\nu)$. With $\tau = m_{-\zeta}(z)$, we have then

$$\begin{aligned} F^o(z) &= 1 + 2z \int \frac{d\nu^o(\theta)}{e^{\mathrm{i}\theta} - z} = 1 + 2z \int \frac{1 - \bar\zeta e^{\mathrm{i}\theta}}{e^{\mathrm{i}\theta}(1 - \bar\zeta z) - \zeta - z} d\nu(\theta) \\ &= \frac{1 - z\bar\zeta}{1 + z\bar\zeta} + \frac{2z(1 - \tau\bar\zeta)}{(1 + z\bar\zeta)} \int \frac{d\nu(\theta)}{e^{\mathrm{i}\theta} - \tau} . \end{aligned} \tag{7.2}$$

Since

$$\int \frac{d\nu(\theta)}{e^{\mathrm{i}\theta} - \tau} = \frac{F(\tau) - 1}{2\tau} , \tag{7.3}$$

we obtain the relation

$$F^o(z) = \frac{1 - z\bar\zeta}{1 + z\bar\zeta} + \frac{z(1 - \tau\bar\zeta)}{\tau(1 + z\bar\zeta)}(F(\tau) - 1). \tag{7.4}$$

By (6.25), this implies

$$f^o(z) = \frac{-\bar\zeta + (1 - \tau\bar\zeta)\frac{f(\tau)}{1 - \tau f(\tau)}}{1 + z(1 - \tau\bar\zeta)\frac{f(\tau)}{1 - \tau f(\tau)}} = \frac{f(\tau) - \bar\zeta}{1 + f(\tau)(z - \tau z\bar\zeta - \tau)} = \frac{f(\tau) - \bar\zeta}{1 - \bar\zeta f(\tau)} \tag{7.5}$$

or in other words

$$f^o = m_{\bar\zeta} \circ f \circ m_{-\zeta} . \tag{7.6}$$

The first V-coefficient is then

$$\alpha_0^o = f^o(0) = \left(m_{\bar\zeta} \circ f\right)(\zeta) = \frac{f(\zeta) - \bar\zeta}{1 - \bar\zeta f(\zeta)} .$$

To compute the higher order coefficients, let us begin with two auxiliary results. Observing that

$$\begin{aligned} S_\alpha(m_\beta \circ h)(z) &= \frac{\omega(\alpha)}{m_\alpha(z)} \frac{m_\beta \circ h(z) - m_\beta \circ h(\alpha)}{1 - \overline{m_\beta \circ h(\alpha)} m_\beta \circ h(z)} = \frac{\omega(\alpha)}{m_\alpha(z)} \frac{\frac{h(z) - \beta}{1 - \beta h(z)} - \frac{h(\alpha) - \beta}{1 - \beta h(\alpha)}}{1 - \overline{\frac{h(\alpha) - \bar\beta}{1 - \beta h(\alpha)}} \frac{h(z) - \beta}{1 - \beta h(z)}} \\ &= \varepsilon(\alpha, \beta, h) \frac{\omega(\alpha)}{m_\alpha(z)} \frac{h(z) - h(\alpha)}{1 - \overline{h(\alpha)} h(z)} \end{aligned} \tag{7.7}$$

with

$$\varepsilon(\alpha, \beta, h) = \frac{1 - \beta\overline{h(\alpha)}}{1 - \bar\beta h(\alpha)} \in \mathbb{T}.$$

So that, we have obtained the first auxiliary result

$$S_\alpha(m_\beta \circ h) = \varepsilon(\alpha, \beta, h) S_\alpha(h) . \tag{7.8}$$

The second one is the following

$$S_0(h \circ m_\gamma)(z) = \frac{-1}{z} \frac{h \circ m_\gamma(z) - h(-\gamma)}{1 - \overline{h(-\gamma)} h \circ m_\gamma(z)} = -\overline{\omega(\gamma)} S_{-\gamma}(h)(m_\gamma(z)) . \tag{7.9}$$

We have therefore

$$\begin{aligned} S_0(f^o) = S(m_{\bar\zeta} \circ f \circ m_{-\zeta}) &\overset{(7.9)}{=} -\overline{\omega(-\zeta)} \left[S_\zeta(m_{\bar\zeta} \circ f)\right] \circ m_{-\zeta} \\ &\overset{(7.8)}{=} \varepsilon_1 S_\zeta(f) \circ m_{-\zeta}, \end{aligned} \tag{7.10}$$

with $\varepsilon_1 = \overline{\omega(\zeta)}\varepsilon(\zeta, \bar\zeta, f) \in \mathbb{T}$, hence

$$\alpha_1^o = S_0(f^o)(0) = \varepsilon_1 S_\zeta(f)(\zeta)\,. \tag{7.11}$$

This representation can be iterated. Assuming that

$$S_0^{[k]}(f^o) = \varepsilon_k \left[S_\zeta^{[k]}(f)\right] \circ m_{-\zeta}, \quad \text{with } |\varepsilon_k| = 1\,, \tag{7.12}$$

we have since $S_\alpha(\varepsilon h) = \varepsilon S_\alpha(h)$ when $\varepsilon \in \mathbb{T}$

$$\begin{aligned} S_0^{[k+1]}(f^o) &= S_0\left[S_0^{[k]}(f^o)\right] = \varepsilon_k S\left[\left[S_\zeta^{[k]}(f)\right] \circ m_{-\zeta}\right] \\ &\overset{(7.9)}{=} -\overline{\omega(-\zeta)}\varepsilon_k \left[S_\zeta\left[S_\zeta^{[k]}(f)\right]\right] \circ m_{-\zeta} = \varepsilon_{k+1}\left[S_\zeta^{[k+1]}(f)\right] \circ m_{-\zeta}. \end{aligned} \tag{7.13}$$

Inductively, (7.12) holds for every $k \geq 0$ and

$$\alpha_k(m_\zeta(\nu)) = \alpha_k^o = S_0^{[k]}(f^0)(0) = \varepsilon_k S_\zeta^{[k]}(f)(\zeta)\,. \tag{7.14}$$

This finished the proof, since $|\varepsilon_k| = 1$. □

7.2. **Proof of Proposition 6.3.** Let us recall the Bessonov formula of [6, Theorem 1]. Let ν be a probability measure on $\mathbb{T}$ with Lebesgue decomposition

$$d\nu = g_\nu dz + d\nu_s$$

(with respect to the uniform measure). The Bessonov formula is

$$\log \int \frac{d\,\mathrm{Pois}(\zeta)}{dz}\, d\nu - \int (\log g_\nu)\, d\,\mathrm{Pois}(\zeta) = \sum_{k=0}^\infty \log \frac{1 - |\zeta f_k(\zeta)|^2}{1 - |f_k(\zeta)|^2}\,.$$

Here, we set $f_k = S_0^{[k]}(f)$. It allows the following slight transformation. Since UNIF and $\mathrm{Pois}(\zeta)$ are mutually absolute continuous, it follows that $g_\nu \frac{dz}{d\,\mathrm{Pois}(\zeta)}$ is the density of the absolutely continuous part of ν with respect to $\mathrm{Pois}(\zeta)$. Hence we get,

$$\begin{aligned} -\int (\log g_\nu)\, d\,\mathrm{Pois}(\zeta) &= -\int (\log g_\nu) \frac{dz}{d\,\mathrm{Pois}(\zeta)} d\,\mathrm{Pois}(\zeta) - \int \log \frac{d\,\mathrm{Pois}(\zeta)}{dz} d\,\mathrm{Pois}(\zeta) \\ &= \mathcal{K}(\mathrm{Pois}(\zeta)|\nu) - \mathcal{K}(\mathrm{Pois}(\zeta)|\,\mathrm{UNIF})\,. \end{aligned}$$

Besides, using the beginning of the proof of Lemma 1 in [6], we have

$$\int \frac{d\,\mathrm{Pois}(\zeta)}{dz}\, d\nu = \frac{1 - |\zeta f(\zeta)|^2}{|1 - \zeta f(\zeta)|^2},$$

Transforming the Kullback-Leibler distance according to (6.1) with the Möbius mapping m_ζ, using (3.24) and (3.53), we also have

$$\mathcal{K}(\mathrm{Pois}(\zeta)|\,\mathrm{UNIF}) = \mathcal{K}(\mathrm{UNIF}\,|\,\mathrm{Pois}(\zeta)) = -\log(1 - |\zeta|^2).$$

Consequently, we can write

$$\begin{aligned} \mathcal{K}(\mathrm{Pois}(\zeta) \mid \nu) &= -\log(1 - |\zeta|^2) - \log \frac{1 - |\zeta f(\zeta)|^2}{|1 - \zeta f(\zeta)|^2} + \sum_{k=0}^\infty \log \frac{1 - |\zeta f_k(\zeta)|^2}{1 - |f_k(\zeta)|^2} \\ &= \log \frac{|1 - \zeta f(\zeta)|^2}{(1 - |\zeta|^2)(1 - |f(\zeta)|^2)} + \sum_{k=1}^\infty \log \frac{1 - |\zeta f_k(\zeta)|^2}{1 - |f_k(\zeta)|^2}\,, \end{aligned}$$

which is the claimed sum rule. □

8. Appendix

8.1. **Analytical proof of a weak version of the HP sum rule.** Up to our knowledge, no analytical proof of the sum rule (3.4) is known. Nevertheless, we can express the coefficient side in terms of the limiting orthogonal polynomials and then use some limit theorems in the OP literature to try to recover the entropy of the spectral side, at least when there are no outliers.

Proposition 8.1. *If the probability measure $\mu = h\,\mathrm{HP}_{\mathtt{d}}$ is such that there exists a polynomial Q such that Qh and Qh^{-1} are bounded on the arc $(\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$, then*

$$\mathcal{K}(\mathrm{HP}(\mathtt{d}) \mid \mu) = \sum_{k=0}^{\infty} H_{\mathtt{d}}(\gamma_k) < \infty\,. \tag{8.1}$$

It is a weaker form of (3.4) since we impose stronger conditions on μ.

Proof. We will put a superscript $\mathtt{d}$ to all quantities relative to the reference measure.

Step 1: Rewriting the coefficient side. The Szegő recursion (2.7) with (3.47) implies:

$$\Phi_n(1) = \prod_{k=0}^{n-1} (1 - \gamma_k)\,, \tag{8.2}$$

so that

$$\sum_{k=0}^{n-1} \log|1 - \gamma_j| = \log|\Phi_n(1)|\,,$$

and then

$$\sum_{k=0}^{n-1} \log \frac{|1 - \gamma_j|}{1 - \gamma_{\mathtt{d}}} = \log \frac{|\Phi_n(1)|}{|\Phi_n^{\mathtt{d}}(1)|}\,. \tag{8.3}$$

If we go back to orthonormal polynomials, we have

$$\Phi_n(t) = \kappa_n^{-1}\varphi_n(t), \quad \kappa_n^{-2} = \prod_{k=0}^{n-1}(1 - |\alpha_k|^2)$$

and since $|\gamma_k| = |\alpha_k|$,

$$\sum_{k=0}^{n-1} - \log \frac{1 - |\gamma_k|^2}{1 - \gamma_{\mathtt{d}}^2} = 2\log\frac{\kappa_n}{\kappa_n^{\mathtt{d}}}\,. \tag{8.4}$$

So that

$$\mathcal{S}_n := \sum_{k=0}^{n-1} H_{\mathtt{d}}(\gamma_k) = 2\log\frac{\kappa_n}{\kappa_n^{\mathtt{d}}} - 2\mathtt{d}\log\frac{|\Phi_n(1)|}{|\Phi_n^{\mathtt{d}}(1)|}\,. \tag{8.5}$$

Coming back to the normalized polynomials, we thus obtain,

$$\mathcal{S}_n = \sum_{k=0}^{n-1} H_{\mathtt{d}}(\gamma_k) = 2(1+\mathtt{d})\log\frac{\kappa_n}{\kappa_n^{\mathtt{d}}} - 2\mathtt{d}\log\frac{|\varphi_n(1)|}{|\varphi_n^{\mathtt{d}}(1)|}. \tag{8.6}$$

Step 2: Computation of the limit. We use an extension to measures supported on an arc of the classical Maté-Nevai-Totik result on the full unit circle [40, Theorem

9.4.1]. The result for an arc is due to Bello Hernandez and Lopez Lagomasino [4]. Theorem 2 therein shows, that if

$$\mu = h\,\mathrm{HP}_{\mathtt{d}}$$

is such that there exists a polynomial Q such that Qh and Qh^{-1} are bounded on the arc $a = (\theta_{\mathtt{d}}, 2\pi - \theta_{\mathtt{d}})$, then

$$\lim_{n\to\infty} \frac{\varphi_n(\zeta)}{\varphi_n^{\mathtt{d}}(\zeta)} = D_a(h,\zeta), \quad \text{and} \quad \lim_{n\to\infty} \frac{\kappa_n}{\kappa_n^{\mathtt{d}}} = D_a(h,\infty)\,,$$

uniformly on compact subsets of $\bar{\mathbb{C}} \setminus a$. Here, the subscript a stands for "the arc". To understand the limit, we need some more notations (well detailed in [3, Section 2.2]).

Let

$$\eta(\tau) = \tau + \sqrt{\tau^2 - 1}$$

(with root such that $|\eta(\tau)| > 1$) be the conformal mapping of $\bar{\mathbb{C}} \setminus [-1,1]$ onto $\bar{\mathbb{C}} \setminus \{z : |z| \le 1\}$ such that $\eta(\infty) = \infty$ and $\eta'(\infty) > 0$. Set

$$\mathtt{c} = \cot(\theta_{\mathtt{d}}/2)\,.$$

Let also

$$\nu(\zeta) = \eta\left(\frac{i}{\mathtt{c}}\frac{\zeta+1}{\zeta-1}\right)\,,$$

be the conformal mapping from $\bar{\mathbb{C}} \setminus a$ onto $\bar{\mathbb{C}} \setminus \{z : |z| \le 1\}$. In particular

$$\nu(1) = \infty\ ,\ \nu(\infty) = \eta(i/\mathtt{c}) = i\sqrt{1+2\mathtt{d}}\,.$$

Following [3, formula (10)], or [4, Lemma 9], we have the indentity

$$D_a(h,\zeta) = \frac{D(h,\nu(\zeta))|D(h,\eta(i/\mathtt{c}))|}{D(h,\eta(i/\mathtt{c}))}\,,$$

where D is a variant of the famous Szegő function:

$$D(h,z) = \exp\left\{\frac{1}{4\pi}\int_0^{2\pi} \log[h(\tau)]\frac{e^{i\theta}+z}{e^{i\theta}-z}d\theta\right\}$$

with

$$\cot\frac{\tau}{2} = \mathtt{c}\cos\theta\,. \tag{8.7}$$

This yields, respectively

$$\begin{aligned} |D_a(h,1)| &= \exp\left\{-\frac{1}{4\pi}\int_0^{2\pi} \log h(\tau) d\theta\right\}\,, \\ |D_a(h,\infty)| &= \exp\left\{\frac{1}{4\pi}\int_0^{2\pi} \log[h(\tau)]\,\mathrm{Re}\,\frac{e^{i\theta}+i\sqrt{1+2\mathtt{d}}}{e^{i\theta}-i\sqrt{1+2\mathtt{d}}}d\theta\right\} \\ &= \exp\left\{-\frac{1}{4\pi}\int_0^{2\pi} \log[h(\tau)]\frac{\mathtt{d}}{1+\mathtt{d}-\sqrt{1+2\mathtt{d}}\sin\theta}d\theta\right\}. \end{aligned}$$

Going back to (8.6), we see that the limit as $n \to \infty$ exists and is given by

$$\begin{aligned}\mathcal{S}_\infty = \lim_{n\to\infty} \mathcal{S}_n &= 2(1+\mathtt{d})\log|D_a(h,\infty)| - 2\mathtt{d}\log|D_a(h,1)| \\ &= -\frac{1}{2\pi}\int_0^{2\pi} \log[h(\tau)]\frac{\mathtt{d}\sqrt{1+2\mathtt{d}}\sin\theta}{1+\mathtt{d}-\sqrt{1+2\mathtt{d}}\sin\theta}\, d\theta \\ &= -\frac{1}{2\pi}\int_0^{2\pi} \log[h(\tau)]\frac{\mathtt{d}\cos(\theta_\mathtt{d}/2)\sin\theta}{1-\cos(\theta_\mathtt{d}/2)\sin\theta}\, d\theta\,, \end{aligned} \tag{8.8}$$

where τ and θ are connected by (8.7). Actually, splitting this integral in two parts and using $\sin(-\theta) = -\sin\theta$ leads to,

$$\begin{aligned}\mathcal{S}_\infty &= -\frac{1}{2\pi}\int_0^{\pi} \log[h(\tau)]\mathtt{d}\cos(\theta_\mathtt{d}/2)\left(\frac{\sin\theta}{1-\cos(\theta_\mathtt{d}/2)\sin\theta} - \frac{\sin\theta}{1+\cos(\theta_\mathtt{d}/2)\sin\theta}\right) d\theta \\ &= -\frac{1}{2\pi}\int_0^{\pi} \log[h(\tau)]\frac{2\mathtt{d}\cos^2(\theta_\mathtt{d}/2)\sin^2\theta}{1-\cos^2(\theta_\mathtt{d}/2)\sin^2\theta}\, d\theta\,.\end{aligned}$$

Now, we have successively

$$\begin{aligned} d\theta &= \left(2\mathtt{c}\sin^2(\tau/2)\sin\theta\right)^{-1} d\tau\,, \\ \sin\theta &= \frac{\sqrt{\sin^2(\tau/2)-\sin^2(\theta_\mathtt{d}/2)}}{\cos(\theta_\mathtt{d}/2)\sin(\tau/2)}\,, \\ 1-\cos^2(\theta_\mathtt{d}/2)\sin^2\theta &= \frac{\sin^2(\theta_\mathtt{d}/2)}{\sin^2(\tau/2)}\,, \end{aligned}$$

so that, using the values of $\mathtt{c}$ and $\sin(\theta_\mathtt{d}/2)$:

$$\begin{aligned}\mathcal{S}_\infty &= -\int_{\theta_\mathtt{d}}^{2\pi-\theta_\mathtt{d}} \log[h(\tau)](1+\mathtt{d})\frac{\sqrt{\sin^2(\tau/2)-\sin^2(\theta_\mathtt{d}/2)}}{2\pi\sin(\tau/2)}\, d\tau \\ &= -\int \log h(\tau)\,\mathrm{HP}(d\tau) = \mathcal{K}(\mathrm{HP}(\mathtt{d})\mid\mu)\,.\end{aligned}$$

This ends the proof. □

8.2. **Kullback-Leibler distances for mixtures.** Suppose μ_1 and μ_2 are probability measures on some measurable space S. The following proposition is useful in the study of sum rules relative to a mixture of μ_1 and μ_2.

Proposition 8.2. *Let $\tau_1, \tau_2 > 0$ with $\tau_1 + \tau_2 = 1$. Then,*

$$\mathcal{K}(\mu_i \mid \tau_1\mu_1 + \tau_2\mu_2) < \infty, \quad (i = 1,2). \tag{8.9}$$

Moreover, for any probability measure μ on S,

$$\begin{aligned}\mathcal{K}(\tau_1\mu_1+\tau_2\mu_2|\mu) &= \tau_1\mathcal{K}(\mu_1|\mu) + \tau_2\mathcal{K}(\mu_2|\mu) \\ &\quad - \tau_1\mathcal{K}(\mu_1|\tau_1\mu_1+\tau_2\mu_2) - \tau_2\mathcal{K}(\mu_2|\tau_1\mu_1+\tau_2\mu_2),\end{aligned} \tag{8.10}$$

where both sides are simultaneously finite or infinite.

Proof. Since, for $i = 1,2$, $\mu_i \ll \tau_1\mu_1 + \tau_2\mu_2$ and

$$\frac{d\mu_i}{d(\tau_1\mu_1+\tau_2\mu_2)} \le \frac{1}{\tau_i}\,,$$

we obtain (8.9).

For the proof of (8.10) let us begin with a useful (but obvious) remark. If ν_1 and ν_2 are two probability measures, such that $\nu_1 \ll \nu_2$, then

$$K(\nu_1 \mid \nu_2) < \infty \Longleftrightarrow \int \left|\log \frac{d\nu_1}{d\nu_2}\right| d\nu_1 < \infty.$$

This follows from the inequality $u(\log u)_- \leq 1/e$ for $u > 0$.
Now, if $\mathcal{K}(\tau_1\mu_1 + \tau_2\mu_2 \mid \mu) < \infty$, then

$$\int \left|\log \frac{d(\tau_1\mu_1 + \tau_2\mu_2)}{d\mu}\right| d(\tau_1\mu_1 + \tau_2\mu_2) < \infty$$

hence for $i = 1, 2$

$$\int \left|\log \frac{d(\tau_1\mu_1 + \tau_2\mu_2)}{d\mu}\right| d\mu_i < \infty$$

and

$$\int \log \frac{d(\tau_1\mu_1 + \tau_2\mu_2)}{d\mu} d\mu_i < \infty$$

and eventually

$$\mathcal{K}(\tau_1\mu_1 + \tau_2\mu_2 \mid \mu) = \sum_{i=1}^{2} \tau_i \int \log \frac{d(\tau_1\mu_1 + \tau_2\mu_2)}{d\mu} d\mu_i$$

Adding $\sum_{i=1}^{2} \tau_i \mathcal{K}(\mu_i \mid \tau_1\mu_1 + \tau_2\mu_2)$ we get

$$\begin{aligned}\mathcal{K}(\tau_1\mu_1 + \tau_2\mu_2 \mid \mu) &+ \sum_{i=1}^{2} \tau_i \mathcal{K}(\mu_i \mid \tau_1\mu_1 + \tau_2\mu_2) \\ &= \sum_{i=1}^{2} \tau_i \int \log \left(\frac{d(\tau_1\mu_1 + \tau_2\mu_2)}{d\mu} \times \frac{d\mu_i}{d(\tau_1\mu_1 + \tau_2\mu_2)}\right) d\mu_i \\ &= \sum_{i=1}^{2} \tau_i \mathcal{K}(\mu_i \mid \mu)\,.\end{aligned}$$

Conversely, if $\mathcal{K}(\mu_i \mid \mu)$ for $i = 1, 2$ are finite, then $\mathcal{K}(\tau_1\mu_1 + \tau_2\mu_2 \mid \mu)$ is finite by convexity. □

Acknowledgement. Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is gratefully acknowledged.

References

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

[2] T.H. Baker, P.J. Forrester, and P.A. Pearce. Random matrix ensembles with an effective extensive external charge. *J. Phys. A*, 31(29):6087, 1998.

[3] M. Bello Hernandez and E. Miña Diaz. Strong asymptotic behavior and weak convergence of polynomials orthogonal on an arc of the unit circle. *J. Approx. Th.*, 111(2):233–255, 2001.

[4] M. Bello Hernandez and G. Lopez Lagomasino. Ratio and relative asymptotics of polynomials orthogonal on an arc of the unit circle. *J. Approx. Th.*, 92(2):216–244, 1998.

[5] D. Bessis and P. Moussa. Orthogonality properties of iterated polynomial mappings. *Comm. Math. Phys.*, 88(4):503–529, 1983.

[6] R. Bessonov and S. Denisov. Zero sets, entropy, and pointwise asymptotics of orthogonal polynomials. *J. Funct. Anal.*, 280(12):109002, 2021.
[7] A. Borodin and G. Olshanski. Infinite random matrices and ergodic measures. *Comm. Math. Phys.*, 223(1):87–123, 2001.
[8] P. Bourgade, A. Nikeghbali, and A. Rouault. Circular Jacobi ensembles and deformed Verblunsky coefficients. *Int. Math. Res. Not.*, 23:4357–4394, 2009.
[9] C. F. Bracciali, X. Li, and A. Sri Ranga. Real orthogonal polynomials in frequency analysis. *Mathematics of computation*, 74(249):341–362, 2005.
[10] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and sum rules for spectral theory - A pedagogical approach. *J. Spectr. Theory*, 8(4):1551–1581, 2018.
[11] J. Breuer, B. Simon, and O. Zeitouni. Large deviations and the Lukic conjecture. *Duke Math. J.*, 167(15):2857–2902, 2018.
[12] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njastad. *Orthogonal rational functions*. Cambridge University Press, 1999.
[13] M. J. Cantero, F. Marcellán, L. Moral, and L. Velázquez. Darboux transformations for CMV matrices. *Adv. Math.*, 298:122–206, 2016.
[14] M.J. Cantero, F. Marcellán, L. Moral, and L. Velázquez. A CMV connection between orthogonal polynomials on the unit circle and the real line. *J. Approx. Th.*, 266:105579, 2021.
[15] T. S. Chihara. *An introduction to orthogonal polynomials*. Mathematics and its Applications, Vol. 13. Gordon and Breach Science Publishers, New York-London-Paris, 1978.
[16] M. S. Costa, H. M. Felix, and A. Sri Ranga. Orthogonal polynomials on the unit circle and chain sequences. *J. Approx. Th.*, 173:14–32, 2013.
[17] P. Delsarte and Y. Genin. The split Levinson algorithm. *IEEE Trans. Acoust. Speech Signal Process.*, 34(3):470–478, 1986.
[18] Ph. Delsarte and Y. Genin. Tridiagonal approach to the algebraic environment of Toeplitz matrices, part I: Basic results. *SIAM J. Matrix Anal. Appl.*, 12(2):220–238, 1991.
[19] Ph. Delsarte and Y. Genin. Tridiagonal approach to the algebraic environment of Toeplitz matrices, part II: Zero and eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 12(3):432–448, 1991.
[20] M. Derevyagin, L. Vinet, and A. Zhedanov. CMV matrices and little and big- 1 Jacobi polynomials. *Constr. Approx.*, 36(3):513–535, 2012.
[21] H. Dette and W. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics, 1997.
[22] F. Gamboa and L.-V. Lozada-Chang. Large deviations for random power moment problem. *Ann. Probab.*, 32(3B):2819–2837, 2004.
[23] F. Gamboa, J. Nagel, and A. Rouault. Sum rules via large deviations. *J. Funct. Anal.*, 270:509–559, 2016.
[24] 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.
[25] F. Gamboa, J. Nagel, and A. Rouault. A higher-order Szegő theorem in the gapped case. in preparation, 2023.
[26] F. Gamboa and A. Rouault. Canonical moments and random spectral measures. *J. Theoret. Probab.*, 23:1015–1038, 2010. Erratum in the same journal (2015) doi 10.1007/s10959-015-0653-5.
[27] F. Gamboa and A. Rouault. Large deviations for random spectral measures and sum rules. *Appl. Math. Res. Express*, 2011(2):281–307, 2011.
[28] L. Golinskii, A. Kheifets, F. Peherstorfer, and P. Yudistkii. On a class of Verblunsky parameters that corresponds to Guseinov's class of Jacobi parameters. *Zh. Mat. Fiz. Anal. Geom.*,, 6(3):277–290, 338, 341, 2010.
[29] 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.
[30] F. Hiai and D. Petz. *The Semicircle Law, Free Random Variables and Entropy*, volume 77 of *Mathematical Surveys and Monographs*. Amer. Math. Soc., Providence, 2000.
[31] L.K. Hua. *Harmonic analysis of functions of several complex variables in the classical domains*, volume 6 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, R.I., 1963.
[32] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math. (2)*, 158(1):253–321, 2003.

[33] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. of Math.*, 158(1):253–321, 2003.
[34] Y.A. Neretin. Hua-type integrals over unitary groups and over projective limits of unitary groups. *Duke Math. J.*, 114(2), 2002.
[35] O. Njåstad and L. Velázquez. Wall rational functions and Khrushchev's formula for orthogonal rational functions. *Constr. Approx.*, 30(2):277–297, 2009.
[36] D. Pickrell. Measures on infinite dimensional Grassmann manifolds. *J. Funct. Anal.*, 70(2):323–356, 1987.
[37] I. Roitberg and A.L. Sakhnovich. Arov–Krein entropy functionals and indefinite interpolation problems. *Integr. Equ. Oper. Theory*, 91(6):50, 2019.
[38] A. Rouault. A matrix version of a higher-order Szegő theorem. *J. Approx. Theory*, 266:105578, 2021.
[39] 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.
[40] 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.
[41] B. Simon. *Szegő's theorem and its descendants.* M. B. Porter Lectures. Princeton University Press, Princeton, NJ, 2011.
[42] S. Wadia. A study of U(N) lattice gauge theory in 2-dimensions. *arXiv preprint arXiv:1212.2906*, 2012.
[43] A. Zhedanov. On some classes of polynomials orthogonal on arcs of the unit circle connected with symmetric orthogonal polynomials on an interval. *J. Approx. Theory*, 94(1):73–106, 1998.

Université Paul Sabatier, Institut de Mathématiques de Toulouse, 31062-Toulouse Cedex 9, France and ANITI,
Email address: `fabrice.gamboa@@math.univ-toulouse.fr`

Technische Universität Dortmund, Fakultät für Mathematik, 44227 Dortmund, Germany
Email address: `jan.nagel@tu-dortmund.de`

Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035-Versailles Cedex France
Email address, corresponding author: `alain.rouault@uvsq.fr`