



INSTITUT DE FRANCE  
Académie des sciences

# *Comptes Rendus*

---

# *Mathématique*

Jean B. Lasserre

**Pell's equation, sum-of-squares and equilibrium measures on a compact set**

Volume 361 (2023), p. 935-952

Published online: 18 July 2023

<https://doi.org/10.5802/crmath.465>



This article is licensed under the  
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.  
<http://creativecommons.org/licenses/by/4.0/>



*Les Comptes Rendus. Mathématique* sont membres du  
Centre Mersenne pour l'édition scientifique ouverte

[www.centre-mersenne.org](http://www.centre-mersenne.org)

e-ISSN : 1778-3569



Statistics / *Statistiques*

# Pell's equation, sum-of-squares and equilibrium measures on a compact set

## *Équation de Pell, sommes de carrés et mesure d'équilibre d'ensembles compacts*

Jean B. Lasserre<sup>®</sup> <sup>a</sup>

<sup>a</sup> LAAS-CNRS and Institute of Mathematics, BP 54200, 7 Avenue du Colonel Roche,  
31031 Toulouse cedex 4, France  
E-mail: [lasserre@laas.fr](mailto:lasserre@laas.fr)

**Abstract.** We first interpret Pell's equation satisfied by Chebyshev polynomials for each degree  $t$ , as a certain Positivstellensatz, which then yields for each integer  $t$ , what we call a generalized Pell's equation, satisfied by reciprocals of Christoffel functions of "degree"  $2t$ , associated with the equilibrium measure  $\mu$  of the interval  $[-1, 1]$  and the measure  $(1 - x^2)d\mu$ . We next extend this point of view to arbitrary compact basic semi-algebraic set  $S \subset \mathbb{R}^n$  and obtain a generalized Pell's equation (by analogy with the interval  $[-1, 1]$ ). Under some conditions, for each  $t$  the equation is satisfied by reciprocals of Christoffel functions of "degree"  $2t$  associated with (i) the equilibrium measure  $\mu$  of  $S$  and (ii), measures  $gd\mu$  for an appropriate set of generators  $g$  of  $S$ . These equations depend on the particular choice of generators that define the set  $S$ . In addition to the interval  $[-1, 1]$ , we show that for  $t = 1, 2, 3$ , the equations are indeed also satisfied for the equilibrium measures of the  $2D$ -simplex, the  $2D$ -Euclidean unit ball and unit box. Interestingly, this view point connects orthogonal polynomials, Christoffel functions and equilibrium measures on one side, with sum-of-squares, convex optimization and certificates of positivity in real algebraic geometry on another side.

**Résumé.** Nous fournissons d'abord une interprétation particulière de l'équation polynomiale de Pell satisfaite par les polynômes de Chebyshev. Pour chaque degré  $t$ , il en découle une équation similaire satisfaite par les fonctions de Christoffel de la mesure d'équilibre  $\mu$  de l'intervalle  $[-1, 1]$  et de la mesure  $(1 - x^2)d\mu$ . Nous généralisons ensuite ce point de vue à des ensembles semi-algébriques compacts, et vérifions le résultat pour  $t = 1, 2, 3$  sur la boule unité Euclidienne, la boîte unité, et le simplexe en dimension 2. Cette interprétation met en lumière une connexion plutôt inattendue entre d'un côté, polynômes orthogonaux, fonctions de Christoffel et mesure d'équilibre, et de l'autre, optimisation convexe et certificats de positivité en géométrie algébrique réelle.

**2020 Mathematics Subject Classification.** 42C05, 47B32, 33C47, 90C23, 90C46.

**Funding.** The author is supported by the AI Interdisciplinary Institute ANITI funding through the french program "Investing for the Future PI3A" under the grant agreement number ANR-19-PI3A-0004. This research is also part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

*Manuscript received 5 October 2022, revised 8 December 2022 and 12 January 2023, accepted 12 January 2023.*

### 1. Introduction

One goal of this paper is to introduce what we call a *generalized Pell's equation* which, under certain conditions, is satisfied by reciprocals of Christoffel functions associated with (i) the equilibrium measure  $\lambda_S$  of a compact basic semi-algebraic set  $S \subset \mathbb{R}^n$ , and (ii) associated measures  $g d\lambda_S$ ,  $g \in G$ , for an appropriate set  $G$  of generators of  $S$ . Moreover, checking whether a chosen set  $G$  of generators is appropriate, can be done by solving a sequence of convex optimization problems.

Another goal is to reveal via the path to obtain the result, strong links between orthogonal polynomials, Christoffel functions and equilibrium measures on one side, and certificates of positivity in real algebraic geometry, optimization and sum-of-squares, as well as a duality result on convex cones by Nesterov, on the other side.

#### 1.1. Initial and motivating example

The starting point is Pell's equation satisfied by Chebyshev polynomials. Pell's equation<sup>1</sup> is a topic in algebraic number theory and for more details (not needed here) the interested reader is referred to e.g. [7, 11]. When looking at this equation with special glasses, we can interpret this equation as a Putinar's certificate of positivity on the interval  $[-1, 1]$ , for the constant polynomial equal to 1. Then with  $g(x) = 1 - x^2$ , the reciprocal of the two Christoffel functions  $\Lambda_t^\mu$  and  $\Lambda_t^{g \cdot \mu}$  respectively associated with the equilibrium measure  $d\mu = dx/\pi\sqrt{1-x^2}$  and the measure  $g \cdot \mu := g d\mu$ , satisfy the same equation, for every  $t$ . Equivalently, for every integer  $t$ , the two polynomials  $1/(2t+1)\Lambda_t^\mu$  and  $(1-x^2)/(2t+1)\Lambda_t^{g \cdot \mu}$  form a *partition of unity* of  $[-1, 1]$ .

More precisely: In  $\mathbb{R}[x]$ , let  $(T_n)_{n \in \mathbb{N}} \subset \mathbb{R}[x]$  (resp.  $(U_n)_{n \in \mathbb{N}} \subset \mathbb{R}[x]$ ) be the Chebyshev polynomials of the first kind (resp. of the second kind). Then

$$T_n(x)^2 + (1-x^2)U_{n-1}(x)^2 = 1, \quad n = 1, \dots \tag{1}$$

In other words, for every integer  $n \geq 1$ , the triple  $(T_n, (x^2-1), U_{n-1})$  is a solution to the polynomial Pell's equation [7].

Next, let  $d\mu(x) = dx/\pi\sqrt{1-x^2}$ ,  $x \mapsto g(x) = 1-x^2$ , and denote by  $g \cdot \mu$  the measure  $g d\mu = \sqrt{1-x^2}dx/\pi$ . The family  $(\hat{T}_n)_{n \in \mathbb{N}}$  (resp.  $(\hat{U}_n)_{n \in \mathbb{N}}$ ) with

$$\hat{T}_0 = T_0; \hat{T}_n = \sqrt{2}T_n, \quad n \geq 1; \quad \hat{U}_n = \sqrt{2}U_n, \quad \forall n \in \mathbb{N},$$

is orthonormal w.r.t.  $\mu$  (resp.  $g \cdot \mu$ ). Then (1) yields

$$\hat{T}_n(x)^2 + (1-x^2)\hat{U}_{n-1}(x)^2 = 2, \quad n = 1, \dots \tag{2}$$

and consequently, summing up yields

$$\underbrace{\hat{T}_0^2 + \sum_{n=1}^t \hat{T}_n(x)^2}_{\Lambda_t^\mu(x)^{-1}} + (1-x^2) \underbrace{\sum_{n=0}^{t-1} \hat{U}_n(x)^2}_{\Lambda_{t-1}^{g \cdot \mu}(x)^{-1}} = 2t+1, \quad t = 0, 1, \dots \tag{3}$$

That is:

$$\Lambda_t^\mu(x)^{-1} + (1-x^2)\Lambda_{t-1}^{g \cdot \mu}(x)^{-1} = 2t+1, \quad t = 0, 1, \dots, \tag{4}$$

where  $\Lambda_t^\mu$  (resp.  $\Lambda_t^{g \cdot \mu}$ ) is the Christoffel function of "degree"  $2t$  associated with  $\mu$  (resp.  $g \cdot \mu$ ).

So for each integer  $t$ , the triple  $((\Lambda_t^\mu)^{-1}, (x^2-1), (\Lambda_{t-1}^{g \cdot \mu})^{-1})$  satisfies what we call a *generalized polynomial Pell's equation*; indeed (i)  $(\Lambda_t^\mu)^{-1} \in \mathbb{Z}[x]$  is a *sum* of  $t+1$  squares (in short, an SOS) and

<sup>1</sup>A multivariate polynomial  $F \in \mathbb{Z}[\mathbf{x}]$  is called a multi-variable Fermat–Pell polynomial if there exist polynomials  $C, H \in \mathbb{Z}[\mathbf{x}]$  such that  $C^2 - FH^2 = 1$  or  $C^2 - FH^2 = -1$  for all  $\mathbf{x}$ . Then the triple  $(C, H, F)$  is a multi-variable solution to Pell's equation; see [7].

not a *single* square, and (ii) after scaling,  $(\Lambda_t^\mu)^{-1}/(2t+1) \notin \mathbb{Z}[x]$  (and similarly for  $(\Lambda_{t-1}^{g \cdot \mu})^{-1}$ ). Also observe that the scaled polynomials  $\frac{1}{2t+1}(\Lambda_t^\mu)^{-1}$  and  $\frac{1}{2t+1}g \cdot (\Lambda_{t-1}^{g \cdot \mu})^{-1}$  form a partition of unity for the interval  $[-1, 1]$ .

The measure  $\mu$  is called the *equilibrium measure* associated with the interval  $[-1, 1]$ . Next, it turns out that (4) is in fact a particular case of [8, Theorem 17.7] which, rephrased later in the polynomial context by the author in [5, Lemma 4], states that every polynomial  $p \in \mathbb{R}[\mathbf{x}]$  (here the constant polynomial  $p = 2t + 1$ ) in the interior of a certain convex cone, has a distinguished representation in terms of certain SOS. Namely, such SOS are reciprocals of Christoffel functions associated with some rather “intriguing” linear functional  $\phi_p \in \mathbb{R}[\mathbf{x}]^*$  associated with  $p$  (see [5, Equation (10)]). However in [5, Lemma 4] we did not provide any clue on what is the link between  $p$  and  $\phi_p$ . So when  $S = [-1, 1]$ , (4) tells us that this intriguing linear functional  $\phi_p$  associated with *constant* polynomials  $p$ , is in fact proportional to the (Chebyshev) equilibrium measure  $dx/\pi\sqrt{1-x^2}$  of the interval  $[-1, 1]$ .

So the message of this introductory example is that we can view the polynomial Pell’s equation (1) as well as its generalization (4), as algebraic Putinar certificates of increasing degree  $t = 1, 2, \dots$ , that the constant polynomials ( $p = 1$  for (1) and  $p = 2t + 1$  for (4)) are positive on the interval  $[-1, 1]$ .

### 1.2. Contribution

The goal of this paper is (i) to define a framework that extends the above point of view to the broader context of compact basic semi-algebraic sets, (ii) to provide conditions under which a multivariate analogue of (4) holds, and (iii) to show that indeed (4) holds for  $t = 1, 2, 3$  for the  $2D$ -Euclidean ball, the  $2D$ -unit box, and the  $2D$ -simplex. As we next see, Equation (4) is particularly interesting as it links statistics, orthogonal polynomials and equilibrium measures on one side, with convex optimization and duality, sum-of-squares and algebraic certificates of positivity, on another side.

More precisely, with  $g_j \in \mathbb{R}[\mathbf{x}]$ ,  $j = 1, \dots, m$ , let

$$S := \{\mathbf{x} \in \mathbb{R}^n \mid g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}, \tag{5}$$

be compact with nonempty interior. Our contribution is to investigate an appropriate multivariate analogue for  $S$  in (5) and its equilibrium measure, of the SOS characterization (4) for the Chebyshev measure  $dx/\pi\sqrt{1-x^2}$  on  $[-1, 1]$ . Given  $g \in \mathbb{R}[\mathbf{x}]$ , let  $t_g := \lceil \deg(g)/2 \rceil$ , and let  $s(t) := \binom{n+t}{n}$ . With  $g_0 = 1$ , introduce  $G := \{g_0, g_1, \dots, g_m\}$  and for every  $t \in \mathbb{N}$ , let  $G_t := \{g \in G : t_g \leq t\}$  (when  $g \in \mathbb{R}[\mathbf{x}]_2$  for all  $g \in G$  then  $G_t = G$  for all  $t \geq 1$ ). For two polynomials  $g, h \in \mathbb{R}[\mathbf{x}]$ , we sometimes use the notation  $g \cdot h$  for their usual product, when needed to avoid ambiguity. Given a Borel measure  $\phi$  on  $S$ , denote by  $g \cdot \phi$ ,  $g \in G$ , the measure  $g d\phi$  on  $S$ . Then define the sets

$$Q(G) := \left\{ \sum_{g \in G} \sigma_g g \mid \sigma_g \in \Sigma[\mathbf{x}] \right\} \tag{6}$$

$$Q_t(G) := \left\{ \sum_{g \in G} \sigma_g g; \sigma_g \in \Sigma[\mathbf{x}] \mid \deg(\sigma_g g) \leq 2t \right\}, \quad t \in \mathbb{N}, \tag{7}$$

respectively called the quadratic module and  $2t$ -truncated quadratic module associated with  $\{g_1, \dots, g_m\} \subset \mathbb{R}[\mathbf{x}]$ .  $(\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$  is convex cone of sum-of-squares polynomials (SOS in short).)

(i). We first show that if a Borel probability measure  $\phi$  on  $S$  (with well-defined Christoffel functions  $\Lambda_t^{g \cdot \phi}$ ,  $g \in G$ ,  $t \in \mathbb{N}$ ) satisfies

$$\frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_{t-t_g}^{g \cdot \phi})^{-1} = 1, \quad \forall t \geq t_0, \tag{8}$$

for some  $t_0 \in \mathbb{N}$ , and  $(S, g \cdot \phi)$  satisfies the Bernstein–Markov property for every  $g \in G$ , then necessarily  $\phi$  is the equilibrium measure  $\lambda_S$  of  $S$  (as defined in e.g. [1]). Notice that (8) is the perfect multivariate analogue of the univariate (4) for  $S = [-1, 1]$  and its equilibrium measure  $\phi = dx/\pi\sqrt{1-x^2}$ ; therefore we propose to name (8) a *generalized Pell’s equation* as it is the analogue of (4) for several polynomials  $g$ , and the solutions  $(1/\Lambda_t^{g \cdot \phi})_{g \in G}$  are sums-of-squares (and not a single square as in the multivariate Pell’s equation [7].) So in this case, for every  $t \geq t_0$ , as an element of  $\text{int}(Q_t(G)^*)$ , the vector of degree- $2t$  moments of the equilibrium measure  $\lambda_S$ , is strongly related to the constant polynomial “1” in  $\text{int}(Q_t(G))$  (which can be viewed as the density of  $\lambda_S$  w.r.t.  $\lambda_S$ ). Such a situation is likely to hold only for specific cases of sets  $S$  (with  $S = [-1, 1]$  and  $\lambda_S = dx/\pi\sqrt{1-x^2}$  being the prototype example).

However we also show that in the general case, the vector of degree- $2t$  moments of the equilibrium measure  $\lambda_S$ , is still related to the constant polynomial “1” but in a weaker fashion. Namely, let  $\mu_t = p_t^* \lambda_S$  be the probability measure whose density  $p_t^*$  w.r.t.  $\lambda_S$  is the polynomial in the left-hand-side of (8) (with  $\phi = \lambda_S$ ). Then  $\lim_{t \rightarrow \infty} \mu_t = \lambda_S$  for the weak convergence of probability measures. That is, asymptotically as  $t$  grows, and as a density w.r.t.  $\lambda_S$ ,  $p_t^*$  behaves like the constant density “1” when integrating continuous functions against  $p_t^* \lambda_S$ .

(ii). We next provide an *if and only if* condition on  $S$  and its representation (5) so that indeed, for every  $t \geq t_0$ , there exists a distinguished linear functional  $\phi_{2t}^* \in \mathbb{R}[\mathbf{x}]_{2t}^*$ , positive on  $Q_t(G)$ , which satisfies

$$1 = \frac{1}{\sum_{g \in G_t} s(t-t_g)} \sum_{g \in G_t} g \cdot (\Lambda_t^{g \cdot \phi_{2t}^*})^{-1}, \tag{9}$$

an analogue of (8) with Christoffel functions  $\Lambda_t^{g \cdot \phi_{2t}^*}$  associated with  $\phi_{2t}^*$ . Interestingly, this condition which states that

$$1 \in \text{int}(Q_t(G)), \quad \forall t \in \mathbb{N}, \tag{10}$$

is a question of real algebraic geometry related to a (degree- $2t$  truncated) quadratic module associated with a set  $G$  of generators of  $S$ . Among all possible sets of generators for a given compact semi-algebraic set  $S$ , those  $G$  for which (10) holds, deserve to be distinguished.

(iii). Next, if condition (10) is satisfied then for every fixed  $t$ , the moment vector  $\phi_{2t}^*$  associated with the linear functional  $\phi_{2t}^*$  in (ii), is the *unique* optimal solution of a convex optimization problem (with a “logdet” criterion) which can be solved efficiently via off-the-shelf softwares like e.g. CVX [3] or Julia [2]. In fact, (9) is an algebraic “certificate” that condition (10) holds, and even more, (9) and (10) are equivalent. Of course, the larger  $t$  is, the larger is the size of the resulting convex optimization problem to solve.

Moreover, every (infinite sequence) accumulation point  $\phi^* = (\phi_\alpha^*)_{\alpha \in \mathbb{N}^n}$  of the sequence of finite moment-vectors  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  associated with the linear functional  $\phi_{2t}^*$ , is represented by a Borel measure  $\phi$  on  $S$ . Then  $\phi$  satisfies (8) if and only if the whole sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  converges to  $\phi^*$  and the convergence is *finite*. That is, there exists  $t_0 \in \mathbb{N}$  such that for every  $t \geq t_0$ ,  $\phi$  is a representing measure of  $\phi_{2t}^*$ . Equivalently, for every  $t \geq t_0$ ,  $\phi_{2(t+1)}^*$  is an *extension* of  $\phi_{2t}^*$ . In addition, if the measure  $\phi$  is such that  $(S, g \cdot \phi)$  satisfies the Bernstein–Markov property for all  $g \in G$ , then necessarily  $\phi$  is the equilibrium measure  $\lambda_S$  of  $S$  (by (i)).

Interestingly, this hierarchy of convex optimization problems provides a practical numerical scheme (at least for moderate values of  $t$ ) to check whether the (unique) optimal solution  $\phi_{2(t+1)}^*$  is an extension of  $\phi_{2t}^*$ , for an arbitrary fixed  $t \in \mathbb{N}$ , which should eventually happen if (8) has ever to hold for the limit measure  $\phi$  associated with the sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$ .

If  $\phi_{2(t+1)}^*$  is an extension of  $\phi_{2t}^*$  for some  $t$ , then it is a good indication that indeed (8) may hold with  $\phi_{2t}^*$  being moments of  $\phi$  (up to degree  $2t$ ). On the other hand, if  $\phi_{2(t+1)}^*$  is not an extension of  $\phi_{2t}^*$  then it may be because (i) there is no limit measure  $\phi$  that satisfies (8), or (ii) one must wait for a larger  $t$  ( $t \geq t_0$ ) to see a possible “extension”, or (iii) perhaps  $G$  is not an appropriate set of

generators of  $S$ . However, in that case it remains to check whether the limit measure  $\phi$  is still the equilibrium measure of  $S$ , and if not, to detect its distinguishing features.

(iv). Finally, in support that (8) may be valid for sets  $S$  other than  $[-1, 1]$ , we also show that for  $t = 1, 2, 3$ , (8) holds when  $S$  is the  $2D$ -Euclidean unit ball and unit box, as well as the  $2D$ -simplex, in which case  $\lambda_S$  is proportional to  $dx dy / \sqrt{1 - x^2 - y^2}$ ,  $dx dy / \sqrt{(1 - x^2)(1 - y^2)}$ , and  $dx dy / \sqrt{xy(1 - x - y)}$ , respectively.

## 2. Main result

### 2.1. Notation and definitions

Let  $\mathbb{R}[\mathbf{x}]$  denote the ring of real polynomials in the variables  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$  be its subset of polynomials of total degree at most  $t$ . Let  $\mathbb{N}_t^n := \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid |\boldsymbol{\alpha}| \leq t\}$  (where  $|\boldsymbol{\alpha}| = \sum_i \alpha_i$ ) with cardinal  $s(t) = \binom{n+t}{n}$ . Let  $\mathbf{v}_t(\mathbf{x}) = (\mathbf{x}^\boldsymbol{\alpha})_{\boldsymbol{\alpha} \in \mathbb{N}_t^n}$  be the vector of monomials up to degree  $t$ , and let  $\Sigma[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]_{2t}$  be the convex cone of polynomials of total degree at most  $2t$  which are sum-of-squares (in short SOS).

For a real symmetric matrix  $\mathbf{A} = \mathbf{A}^T$  the notation  $\mathbf{A} \geq 0$  (resp.  $\mathbf{A} > 0$ ) stands for  $\mathbf{A}$  is positive semidefinite (p.s.d.) (resp. positive definite (p.d.)). The support of a Borel measure  $\mu$  on  $\mathbb{R}^n$  is the smallest closed set  $A$  such that  $\mu(\mathbb{R}^n \setminus A) = 0$ , and such a set  $A$  is unique. Denote by  $\mathcal{C}(S)$  the space of real continuous functions on  $S$ .

**Riesz functional, moment and localizing matrix.** With a real sequence  $\boldsymbol{\phi} = (\phi_\boldsymbol{\alpha})_{\boldsymbol{\alpha} \in \mathbb{N}^n}$  (in bold) is associated the *Riesz* linear functional  $\phi \in \mathbb{R}[\mathbf{x}]^*$  (not in bold) defined by

$$p \left( = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \mathbf{x}^\boldsymbol{\alpha} \right) \mapsto \phi(p) = \langle \boldsymbol{\phi}, p \rangle = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \phi_{\boldsymbol{\alpha}}, \quad \forall p \in \mathbb{R}[\mathbf{x}],$$

and the moment matrix  $\mathbf{M}_t(\boldsymbol{\phi})$  with rows and columns indexed by  $\mathbb{N}_t^n$  (hence of size  $s(t) := \binom{n+t}{t}$ ), and with entries

$$\mathbf{M}_t(\boldsymbol{\phi})(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \phi(\mathbf{x}^{\boldsymbol{\alpha} + \boldsymbol{\beta}}) = \phi_{\boldsymbol{\alpha} + \boldsymbol{\beta}}, \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_t^n.$$

Similarly given  $g \in \mathbb{R}[\mathbf{x}]$  ( $\mathbf{x} \mapsto \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} \mathbf{x}^{\boldsymbol{\gamma}}$ ), define the new sequence

$$g \cdot \boldsymbol{\phi} := \left( \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} \phi_{\boldsymbol{\alpha} + \boldsymbol{\gamma}} \right)_{\boldsymbol{\alpha} \in \mathbb{N}^n},$$

and the localizing matrix associated with  $\boldsymbol{\phi}$  and  $g$ ,

$$\mathbf{M}_t(g \cdot \boldsymbol{\phi})(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} \phi_{\boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}}, \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_t^n.$$

Equivalently,  $\mathbf{M}_t(g \cdot \boldsymbol{\phi})$  is the moment matrix associated with the new sequence  $g \cdot \boldsymbol{\phi}$ . The Riesz linear functional  $g \cdot \phi$  associated with the sequence  $g \cdot \boldsymbol{\phi}$  satisfies

$$g \cdot \phi(p) = \phi(g \cdot p), \quad \forall p \in \mathbb{R}[\mathbf{x}].$$

In particular, for any real symmetric  $s(t) \times s(t)$  matrix  $\mathbf{Q}$

$$\phi(g(\mathbf{x}) \mathbf{v}_t(\mathbf{x})^T \mathbf{Q} \mathbf{v}_t(\mathbf{x})) = g \cdot \phi(\mathbf{v}_t(\mathbf{x})^T \mathbf{Q} \mathbf{v}_t(\mathbf{x})) = \langle \mathbf{Q}, \mathbf{M}_t(g \cdot \boldsymbol{\phi}) \rangle. \tag{11}$$

A real sequence  $\boldsymbol{\phi} = (\phi_\boldsymbol{\alpha})_{\boldsymbol{\alpha} \in \mathbb{N}^n}$  has a representing measure if its associated linear functional  $\phi$  is a Borel measure on  $\mathbb{R}^n$ . In this case  $\mathbf{M}_t(\boldsymbol{\phi}) \geq 0$  for all  $t$ ; the converse is not true in general. In addition, if  $\phi$  is supported on the set  $\{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \geq 0\}$  then necessarily  $\mathbf{M}_t(g \cdot \boldsymbol{\phi}) \geq 0$  for all  $t$ .

**Christoffel function.** Let  $\phi \in \mathbb{R}[\mathbf{x}]^*$  be a Riesz functional (not necessarily with a representing measure) such that  $\mathbf{M}_t(\phi) > 0$ . As for Borel measures, we may also define the (degree- $t$ ) Christoffel function

$$\mathbf{x} \mapsto \Lambda_t^\phi(\mathbf{x})^{-1} := \mathbf{v}_t(\mathbf{x})^T \mathbf{M}_t(\phi)^{-1} \mathbf{v}_t(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

associated with  $\phi$ . Alternatively, if  $(P_\alpha)_{\alpha \in \mathbb{N}_t^n} \subset \mathbb{R}[\mathbf{x}]$  is a family of polynomials which are orthonormal with respect to  $\phi$ , then

$$\Lambda_t^\phi(\mathbf{x})^{-1} = \sum_{\alpha \in \mathbb{N}_t^n} P_\alpha(\mathbf{x})^2, \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{12}$$

Similarly, if  $\mathbf{M}_t(g \cdot \phi) > 0$ , we may also define the (degree- $t$ ) Christoffel function

$$\mathbf{x} \mapsto \Lambda_t^{g \cdot \phi}(\mathbf{x})^{-1} := \mathbf{v}_t(\mathbf{x})^T \mathbf{M}_t(g \cdot \phi)^{-1} \mathbf{v}_t(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

associated with the Riesz functional  $g \cdot \phi$ .

All the above definitions also hold for finite sequences  $\phi_{2t} = (\phi_\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$  and associated Riesz linear functional  $\phi_{2t} \in \mathbb{R}[\mathbf{x}]_{2t}^*$ . Indeed when  $t$  is fixed,  $\Lambda_t^\phi$  and  $\mathbf{M}_t(\phi)$  only depend on the degree  $2t$ -truncation  $\phi_{2t}$  of the infinite sequence  $\phi$ . Then the notation  $\mathbf{M}_t(\phi)$  or  $\mathbf{M}_t(\phi_{2t})$  (and similarly  $\Lambda_t^{\phi_{2t}}$  or  $\Lambda_t^\phi$ ) can be used interchangeably. Finally, a sequence  $\phi_{2(t+1)}$  is an *extension* of  $\phi_{2t}$  if  $(\phi_{2(t+1)})_\alpha = (\phi_{2t})_\alpha$  for all  $\alpha \in \mathbb{N}_{2t}^n$ , i.e., if  $\phi_{2t}$  is the restriction of  $\phi_{2(t+1)}$  to all moments up to degree  $2t$ .

**Bernstein–Markov property.** A Borel measure  $\mu$  supported on a compact set  $S \subset \mathbb{R}^n$  satisfies the Bernstein–Markov property if there exists a sequence of positive numbers  $(M_t)_{t \in \mathbb{N}}$  such that for all  $t$  and  $p \in \mathbb{R}[\mathbf{x}]_t$ ,

$$\sup_{\mathbf{x} \in S} |p(\mathbf{x})| \leq M_t \cdot \left( \int_S p^2 d\mu \right)^{1/2}, \quad \text{and} \quad \lim_{t \rightarrow \infty} \log(M_t)/t = 0 \tag{13}$$

(see e.g. [6, Section 4.3.3]). The Bernstein–Markov property allows qualitative description for asymptotics of the Christoffel function as  $t$  grows. When it holds it permits to establish a strong link between the Christoffel function and Siciak’s extremal function

$$\mathbf{x} \mapsto V_S(\mathbf{x}) := \sup \left\{ \frac{\log |p(\mathbf{x})|}{\deg(p)} \mid \|p\|_S \leq 1, \deg(p) > 0 \right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

A compact set  $S$  is said to be *regular* if its associated Siciak’s function is continuous everywhere in  $\mathbb{R}^n$  (the same definition also extends to  $\mathbb{C}^n$ ; see [6, Definition 4.4.2, p. 53]). If  $S$  is regular and  $(S, \mu)$  satisfies the Bernstein–Markov property, then

$$\text{uniformly on compact subsets of } \mathbb{R}^n: \quad \lim_{t \rightarrow \infty} \frac{1}{2t} \log \Lambda_t^\mu(\mathbf{x}) = -V_S(\mathbf{x}).$$

**Equilibrium measure.** The notion of equilibrium measure associated to a given set, originates from logarithmic potential theory (working in  $\mathbb{C}$  in the univariate case) to minimize some energy functional. For instance, the equilibrium (Chebyshev) measure  $d\phi := dx/\pi\sqrt{1-x^2}$  minimizes the Riesz  $s$ -energy functional

$$\iint \frac{1}{|x-y|^s} d\mu(x) d\mu(y)$$

with  $s = 2$ , among all measure  $\mu$  equivalent to  $\phi$ . Some generalizations have been obtained in the multivariate case via pluripotential theory in  $\mathbb{C}^n$ . In particular if  $S \subset \mathbb{R}^n \subset \mathbb{C}^n$  is compact then the equilibrium measure (let us denote it by  $\lambda_S$ ) is equivalent to Lebesgue measure on compact subsets of  $\text{int}(S)$ . It has an even explicit expression if  $S$  is convex and symmetric about the origin; see e.g. Bedford and Taylor [1, Theorem 1.1] and [1, Theorem 1.2]. Moreover if  $\mu$  is a Borel

measure on  $S$  and  $(S, \mu)$  has the Bernstein–Markov property (13) then the sequence of measures  $dv_t = \frac{d\mu(\mathbf{x})}{s(t)\Lambda_t^\mu(\mathbf{x})}$ ,  $t \in \mathbb{N}$ , converges to  $\lambda_S$  for the weak- $\star$  topology and therefore in particular:

$$\lim_{t \rightarrow \infty} \int_S \mathbf{x}^\alpha dv_t = \lim_{t \rightarrow \infty} \int_S \frac{\mathbf{x}^\alpha d\mu(\mathbf{x})}{s(t)\Lambda_t^\mu(\mathbf{x})} = \int_S \mathbf{x}^\alpha d\lambda_S, \quad \forall \alpha \in \mathbb{N}^n \tag{14}$$

(see e.g. [6, Theorem 4.4.4]). In addition, if a compact  $S \subset \mathbb{R}^n$  is regular then  $(S, \lambda_S)$  has the Bernstein–Markov property; see [6, p. 59]. For a brief account on equilibrium measures see the discussion in [6, Section 4–5, pp. 56–60] while for more detailed expositions see some of the references indicated there.

### 2.2. Brief summary of main results

In Section 2.3, Theorem 2 shows that if a linear functional  $\phi \in \mathbb{R}[\mathbf{x}]^*$  satisfies the multivariate analogue (8) of (4) for  $S$  in (5), then under a certain technical assumption,  $\phi$  is necessarily the equilibrium measure  $\lambda_S$  of  $S$ . Corollary 4 shows that (8) is also a strong property of orthonormal polynomials associated with  $\lambda_S$ , the perfect analogue of (2) for Chebyshev polynomials on  $S = [-1, 1]$ . As this strong property is not expected to hold for general sets  $S$  in (5), we next show in Theorem 3 that in general, the polynomial

$$p_t^* := \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_t^{g \cdot \lambda_S})^{-1}, \quad t \in \mathbb{N},$$

associated with  $\lambda_S$  (now not necessarily constant (equal to 1) as in Theorem 2) has still a strong property related to the constant polynomial “1”. Namely asymptotically, the sequence of probability measures  $(\mu_t := p_t^* \lambda_S)_{t \in \mathbb{N}}$  (with densities  $p_t^*$  w.r.t.  $\lambda_S$ ) converges to  $\lambda_S$  for the weak- $\star$  topology of  $\mathcal{M}(S)$ . That is, informally, the polynomial density  $p_t^*$  “behaves” asymptotically like the constant (equal to 1) density when integrating continuous functions against  $p_t^* \lambda_S$ . Hence somehow, the vector of degree- $2t$  moments of  $\lambda_S$  in the convex cone  $(Q_t(G))^*$  are still intimately related to the constant polynomial 1 in  $Q_t(G)$  (but not as directly as in Theorem 2).

Next, in Section 2.4 we still consider again the constant polynomial 1 and in Theorem 6 we show that under a simple condition, indeed  $1 \in \text{int}(Q_t(G))$  for all  $t$ , and therefore there exists a sequence of linear functional  $(\phi_{2t})_{t \in \mathbb{N}}$  that satisfies (9) for all  $t$ . For each  $t$ , the linear functional  $\phi_{2t}$  is the unique optimal solution of a simple convex optimization problem with logdet criterion to maximize. (In addition, in the case when  $S = \{\mathbf{x} : g(\mathbf{x}) \geq 0\}$  for some  $g \in \mathbb{R}[\mathbf{x}]$ , Lemma 9 relates solutions to Pell’s equation with  $\phi_{2t}$  and  $S$ .)

In Section 2.5 one is concerned with the asymptotic behavior of the linear functionals  $(\phi_{2t})_{t \in \mathbb{N}}$  as  $t$  grows, and Theorem 10 shows that there exists a limit moment sequence  $\phi$  which has a representing probability measure  $\phi$  on  $S$ . Moreover  $\phi$  satisfies (8) and is the equilibrium measure  $\lambda_S$ , if and only if finite convergence takes place, that is, for every  $t \geq t_0$ ,  $\phi_{2t}$  is the vector of degree- $2t$  moments of  $\phi$ . So an interesting issue (not treated here) is to relate  $\phi$  and  $\lambda_S$  when the convergence is only asymptotic and not finite.

Finally in Section 3 we provide numerical examples of sets  $S$  where (8) holds at least for  $t = 1, 2, 3$ .

### 2.3. Two preliminary results

For simplicity of exposition, we will consider sets  $S$  in (5) for which the quadratic polynomial  $\mathbf{x} \mapsto R - \|\mathbf{x}\|^2$  belongs to  $Q_1(G)$ ; in particular,  $S$  is contained in the Euclidean ball of radius  $\sqrt{R}$  for some  $R > 0$ , and the quadratic module  $Q(G)$  is Archimedean; see e.g. [4]. Let  $\lambda_S$  be the equilibrium measure of  $S$  (as described in e.g. [1]) and recall that  $g_0 = 1$  (so that  $g_0 \cdot \lambda_S = \lambda_S$ ). Let  $\mathcal{C}(S)$  be the space of continuous functions on  $S$ .



**Assumption 1.** *The set  $S$  in (5) is compact with nonempty interior. Moreover, there exists  $R > 0$  such that the quadratic polynomial  $\mathbf{x} \mapsto \theta(\mathbf{x}) := R - \|\mathbf{x}\|^2$  is an element of  $Q_1(G)$ . In other words,  $h \in Q_1(G)$  is an “algebraic certificate” that  $S$  in (5) is compact.*

**Theorem 2.** *With  $S$  as in (5), let Assumption 1 hold. Let  $\boldsymbol{\phi} = (\phi_{\alpha})_{\alpha \in \mathbb{N}^n}$  (with  $\phi_0 = 1$ ) be such that  $\mathbf{M}_t(\mathbf{g} \cdot \boldsymbol{\phi}) > 0$  for all  $t \in \mathbb{N}$  and all  $g \in G$ , so that the Christoffel functions  $\Lambda_t^{\mathbf{g} \cdot \boldsymbol{\phi}}$  are all well defined (recall that  $\phi \in \mathbb{R}[\mathbf{x}]^*$  is the Riesz linear functional associated with the moment sequence  $\boldsymbol{\phi}$ ). In addition, suppose that there exists  $t_0 \in \mathbb{N}$  such that*

$$1 = \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_{t-t_g}^{\mathbf{g} \cdot \boldsymbol{\phi}})^{-1}, \quad \forall t \geq t_0. \quad (15)$$

*Then  $\phi$  is a Borel measure on  $S$  and the unique representing measure of  $\boldsymbol{\phi}$ . Moreover, if  $(S, \mathbf{g} \cdot \boldsymbol{\phi})$  satisfies the Bernstein–Markov property for every  $g \in G$ , then  $\phi = \lambda_S$  and therefore the Christoffel polynomials  $(\Lambda_t^{\mathbf{g} \cdot \lambda_S})_{g \in G_t}^{-1}$  satisfy the generalized Pell’s equations:*

$$1 = \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_{t-t_g}^{\mathbf{g} \cdot \lambda_S})^{-1}, \quad \forall t \geq t_0. \quad (16)$$

**Proof.** In view of Assumption 1, the quadratic module  $Q(G)$  is Archimedean. Next, as  $\mathbf{M}_t(\mathbf{g} \cdot \boldsymbol{\phi}) > 0$  for all  $t \in \mathbb{N}$  and all  $g \in G$ , then by Putinar’s Positivstellensatz [10],  $\boldsymbol{\phi}$  has a unique representing measure on  $S$ ; that is, the Riesz linear functional  $\phi$  associated with  $\boldsymbol{\phi}$  is a Borel measure on  $S$ . Next, write (15) as

$$1 = \sum_{g \in G_t} g \cdot \frac{(\Lambda_{t-t_g}^{\mathbf{g} \cdot \boldsymbol{\phi}})^{-1}}{s(t - t_g)} \cdot \frac{s(t - t_g)}{\sum_{g \in G} s(t - t_g)}, \quad \forall t \geq t_0, \quad (17)$$

and let  $\boldsymbol{\alpha} \in \mathbb{N}^n$  be fixed arbitrary. As  $(S, \mathbf{g} \cdot \boldsymbol{\phi})$  satisfies the Bernstein–Markov property for every  $g \in G$ , then by [6, Theorem 4.4.4],

$$\lim_{t \rightarrow \infty} \int_S \mathbf{x}^{\boldsymbol{\alpha}} \frac{(\Lambda_t^{\mathbf{g} \cdot \boldsymbol{\phi}})^{-1}}{s(t)} g d\phi = \int_S \mathbf{x}^{\boldsymbol{\alpha}} d\lambda_S, \quad \forall g \in G,$$

where  $\lambda_S$  is the equilibrium measure of  $S$ ; see [1, 6]. Hence multiplying (17) by  $\mathbf{x}^{\boldsymbol{\alpha}}$  and integrating w.r.t.  $\phi$  yields

$$\int_S \mathbf{x}^{\boldsymbol{\alpha}} d\phi = \sum_{g \in G_t} \frac{s(t - t_g)}{\sum_{g \in G_t} s(t - t_g)} \cdot \int_S \frac{\mathbf{x}^{\boldsymbol{\alpha}} \cdot (\Lambda_{t-t_g}^{\mathbf{g} \cdot \boldsymbol{\phi}})^{-1}}{s(t - t_g)} g d\phi, \quad \forall t \geq t_0.$$

Each term of the product in the above sum of the right-hand-side has a limit as  $t$  grows. Moreover  $G_t = G$  for  $t$  sufficiently large. Therefore taking limit as  $t$  increases yields

$$\begin{aligned} \int_S \mathbf{x}^{\boldsymbol{\alpha}} d\phi &= \sum_{g \in G} \lim_{t \rightarrow \infty} \frac{s(t - t_g)}{\sum_{g \in G} s(t - t_g)} \cdot \lim_{t \rightarrow \infty} \int_S \frac{\mathbf{x}^{\boldsymbol{\alpha}} \cdot (\Lambda_{t-t_g}^{\mathbf{g} \cdot \boldsymbol{\phi}})^{-1}}{s(t - t_g)} g d\phi \\ &= \int_S \mathbf{x}^{\boldsymbol{\alpha}} d\lambda_S \cdot \sum_{g \in G} \lim_{t \rightarrow \infty} \sum_{g \in G} \frac{s(t - t_g)}{\sum_{g \in G} s(t - t_g)} \\ &= \int_S \mathbf{x}^{\boldsymbol{\alpha}} d\lambda_S \cdot \sum_{g \in G} (\#G)^{-1} = \int_S \mathbf{x}^{\boldsymbol{\alpha}} d\lambda_S. \end{aligned}$$

As  $\boldsymbol{\alpha} \in \mathbb{N}^n$  was arbitrary and  $S$  is compact, then necessarily  $\phi = \lambda_S$ . □

Theorem 2 reveals a strong property of Christoffel functions  $\Lambda_t^{\mathbf{g} \cdot \lambda_S}$  and is likely to hold only in some specific cases. The prototype example is  $S = [-1, 1] = \{x : g(x) \geq 0\}$  with  $x \mapsto g(x) = 1 - x^2$ . Then indeed (4) is exactly (16), and by analogy with the Chebyshev univariate case, we propose to call Equation (16) a *generalized Pell’s (polynomial) equation* of degree  $2t$ . It is satisfied by the

polynomials  $(g \cdot (\Lambda_{t-t_g}^{g \cdot \lambda_S})^{-1})_{g \in G_t}$ , all of degree less than  $2t$ . If true for all  $t$ , then  $(S, \lambda_S)$  satisfies the generalized Pell's equations for all degrees.

Of course, to be valid (16) requires conditions on  $S$  and its representation (5) by the polynomials  $g \in G$ . For instance, as shown in Section 3 below, if  $S$  is the  $2D$ -Euclidean unit ball with  $g = 1 - \|\mathbf{x}\|^2$ , (in which case  $G_t = G_1$  for all  $t \geq 1$ ), then  $\lambda_S = d\mathbf{x}/(\pi\sqrt{1 - \|\mathbf{x}\|^2})$  and we can show that (16) holds for  $t = 1, 2, 3$ . Similarly, if  $S$  is the  $2D$ -simplex  $\{\mathbf{x} : x_1, x_2 \geq 0; x_1 + x_2 \leq 1\}$ , then  $\lambda_S = d\mathbf{x}/(\pi\sqrt{x_1 \cdot x_2 \cdot (1 - x_1 - x_2)})$  and we can show that (16) holds for  $t = 1, 2, 3$ , for the quadratic generators in  $G = \{g_0, g_1, g_2, g_3\}$  with  $g_1(\mathbf{x}) = x_1 \cdot (1 - x_1 - x_2)$ ,  $g_2(\mathbf{x}) = x_2 \cdot (1 - x_1 - x_2)$ , and  $g_3(\mathbf{x}) = x_1 \cdot x_2$ .

However in the general case we have the following weaker result, still related to Theorem 2.

**Theorem 3.** *Let  $\lambda_S$  be the equilibrium measure of  $S$  and assume that for every  $g \in G$ ,  $(S, g \cdot \lambda_S)$  satisfies the Bernstein–Markov property. For every  $t$ , define the polynomial*

$$p_t^* := \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_t^{g \cdot \lambda_S})^{-1}, \quad t \in \mathbb{N}. \tag{18}$$

*Then the sequence of probability measures  $(\mu_t := p_t^* \lambda_S)_{t \geq t_0}$  converges to  $\lambda_S$  for the weak- $\star$  topology of  $\mathcal{M}(S)$ , i.e.,*

$$\lim_{t \rightarrow \infty} \int_S f p_t^* d\lambda_S = \int_S f d\lambda_S, \quad \forall f \in \mathcal{C}(S). \tag{19}$$

**Proof.** The polynomial  $p_t^*$  in (18) is well defined because the matrices  $\mathbf{M}_{t-t_g}(g \cdot \lambda_S)$  are non singular. Each  $\mu_t$  is a probability measure on  $S$  because

$$\begin{aligned} \int p_t^* d\lambda_S &= \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} \int g \cdot (\Lambda_t^{g \cdot \lambda_S})^{-1} d\lambda_S \\ &= \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} \langle \mathbf{M}_{t-t_g}(g \cdot \lambda_S), \mathbf{M}_{t-t_g}(g \cdot \lambda_S)^{-1} \rangle \\ &= \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} s(t - t_g) = 1. \end{aligned}$$

As  $(S, g \cdot \lambda_S)$  satisfies the Bernstein–Markov property for every  $g \in G$ , then by [6, Theorem 4.4.4],

$$\lim_{t \rightarrow \infty} \int_S f \frac{(\Lambda_t^{g \cdot \lambda_S})^{-1}}{s(t)} g d\lambda_S = \int_S f d\lambda_S, \quad \forall f \in \mathcal{C}(S), \forall g \in G.$$

Hence multiplying (18) by  $f \in \mathcal{C}(S)$  and integrating w.r.t.  $\lambda_S$ , yields

$$\int_S f d\mu_t = \int_S f p_t^* d\lambda_S = \sum_{g \in G_t} \frac{s(t - t_g)}{\sum_{g \in G_t} s(t - t_g)} \cdot \int_S \frac{f \cdot (\Lambda_{t-t_g}^{g \cdot \lambda_S})^{-1}}{s(t - t_g)} g d\lambda_S, \quad \forall t \geq t_0.$$

Each term of the product in the above sum of the right-hand-side has a limit as  $t$  grows. Moreover  $G_t = G$  for  $t$  sufficiently large. Therefore taking limit as  $t$  increases, yields

$$\begin{aligned} \lim_{t \rightarrow \infty} \int_S f d\mu_t &= \lim_{t \rightarrow \infty} \int_S f p_t^* d\lambda_S = \sum_{g \in G} \lim_{t \rightarrow \infty} \frac{s(t - t_g)}{\sum_{g \in G} s(t - t_g)} \cdot \lim_{t \rightarrow \infty} \int_S \frac{f \cdot (\Lambda_{t-t_g}^{g \cdot \lambda_S})^{-1}}{s(t - t_g)} g d\lambda_S \\ &= \int_S f d\lambda_S \cdot \sum_{g \in G} \lim_{t \rightarrow \infty} \sum_{g \in G} \frac{s(t - t_g)}{\sum_{g \in G} s(t - t_g)} \\ &= \int_S f d\lambda_S \cdot \sum_{g \in G} (\#G)^{-1} = \int_S f d\lambda_S, \quad \forall f \in \mathcal{C}(S). \end{aligned}$$

As  $S$  is compact it implies that the sequence of probability measures  $(\mu_t)_{t \in \mathbb{N}} \subset \mathcal{M}(S)_+$  converges to  $\lambda_S$  for the weak- $\star$  topology  $\sigma(\mathcal{M}(S), \mathcal{C}(S))$  of  $\mathcal{M}(S)$ . □

In other words (and in an informal language), when integrating continuous functions against  $\mu_t$ , the density  $p_t^*$  of  $\mu_t$  w.r.t.  $\lambda_S$  behaves asymptotically like the constant (equal to 1) density. That is, Theorem 3 is a more general (but weaker) version of Theorem 2.

**Corollary 4.** *Let  $\phi$  be the Borel measure on  $S$  in Theorem 2, and for each  $g \in G$ , let  $(P_\alpha^{g \cdot \phi})_{\alpha \in \mathbb{N}^n}$  be a family of polynomials, orthonormal with respect to the measure  $g \cdot \phi$ . Then for every  $t \geq t_0 + 1$ :*

$$\begin{aligned} \sum_{g \in G_t} \sum_{|\alpha|=t-t_g} g \cdot (P_\alpha^{g \cdot \phi})^2 &= \sum_{g \in G_t} s(t-t_g) - \sum_{g \in G_t} s(t-t_g-1) \\ &= \sum_{g \in G_t} \binom{n-1+t-t_g}{n-1}. \end{aligned} \tag{20}$$

**Proof.** Recalling (12), for each  $g \in G_t$  with  $t \geq t_0 + 1$ :

$$(\Lambda_{t-t_g}^{g \cdot \phi})^{-1} = \sum_{\alpha \in \mathbb{N}_{t-t_g}^n} (P_\alpha^{g \cdot \phi})^2 = \sum_{|\alpha| < t-t_g} (P_\alpha^{g \cdot \phi})^2 + \sum_{|\alpha|=t-t_g} (P_\alpha^{g \cdot \phi})^2 = (\Lambda_{t-t_g-1}^{g \cdot \phi})^{-1} + \sum_{|\alpha|=t-t_g} (P_\alpha^{g \cdot \phi})^2,$$

which combined with (15) yields (20). □

**Remark 5.** Observe that (20) which states a property satisfied by orthonormal polynomials associated with  $g \cdot \phi$ ,  $g \in G_t$ , is a multivariate and multi-generator analogue of (2), the polynomial Pell’s equation satisfied by normalized Chebyshev polynomials. However there are several differences between (20) and (2).

In (2), where  $G = \{g\}$  with  $g = (1-x^2)$  (and so with  $t_g = 1$ ), the triplet  $(\widehat{T}_t, -g, \widehat{U}_{t-t_g})$  is a solution to the polynomial Pell equation  $C^2 - FH^2 = 1$  which involves single squares  $C^2$  and  $H^2$  and a single generator  $F$ . On the other hand, (20) addresses the *multivariate* case with possibly *several* generators  $g \in G_t$  and in compact form reads  $\sum_{g \in G_t} g C_g = 1$  which now involves SOS polynomials  $(C_g)_{g \in G_t}$  and several generators  $g \in G_t$ .

For instance, in case of a single generator  $G = \{g\}$ , in compact form (20) reads  $C - FH = 1$  with now *SOS polynomials*  $C, H$  and generator  $F = -g$ . Writing  $C = \sum_{i=1}^r C_i^2$  and  $H = \sum_{j=1}^q H_j^2$ ,

$$1 = C - FH = \sum_{i=1}^r C_i^2 - \sum_{j=1}^q FH_j^2.$$

This is why we think that it is fair to call (20) (as well as (8)) a *generalized* polynomial Pell equation where SOS (rather than just single squares) are allowed. In fact, even in the univariate Chebyshev case, when summing over  $n$ , (3) is a generalized Pell equation in the form  $C - FH = 1$  with SOS polynomials  $C, H$ .

### 2.4. A convex optimization problem and its dual

In Theorem 2 we have taken for granted existence of a linear functional  $\phi$  such that its moment sequence  $\phi$  satisfies (15). The next issue is:

*Given a compact set  $S$  as in (5), can we provide such a moment sequence  $\phi$ ? At least, can we define a numerical scheme which provides finite sequences  $(\phi_{2t})_{t \in \mathbb{N}}$  which “converge” to such a desirable  $\phi$  as  $t$  grows?*

As we next see, this issue essentially translates to the following simple issue in real algebraic geometry. Do we have  $1 \in \text{int}(Q_t(G))$  for every  $t \in \mathbb{N}$ ? If the answer is yes then indeed such a  $\phi$  exists. But then the associated linear functional  $\phi$  will satisfy (15) only if the convergence is *finite*. Moreover the conditions can be checked by solving a sequence of convex optimization problems described in the next section.

With  $t_g := \lceil \deg(g)/2 \rceil$ , for every  $t \in \mathbb{N}$ , consider the two convex optimization problems:

$$\rho_t = \inf_{\phi_{2t}} \left\{ - \sum_{g \in G_t} \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi_{2t})) \mid \phi_{2t}(1) = 1; \text{ s.t. } \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}) \geq 0, \forall g \in G_t \right\}, \tag{21}$$

and:

$$\rho_t^* = \sup_{\mathbf{Q}_g} \left\{ \sum_{g \in G_t} \log \det(\mathbf{Q}_g) \mid \mathbf{Q}_g \geq 0, \forall g \in G_t; \text{ s.t. } \sum_{g \in G_t} s(t-t_g) = \sum_{g \in G_t} g(\mathbf{x}) \cdot \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{Q}_g \mathbf{v}_{t-t_g}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n \right\}. \tag{22}$$

Problem (21) and (22) are convex optimization problems.

**Theorem 6.** *With  $t \in \mathbb{N}$  fixed, Problems (21) and (22) have same finite optimal value  $\rho_t = \rho_t^*$  if and only if  $1 \in \text{int}(Q_t(G))$ . Then both have a unique optimal solution  $\phi_{2t}^* \in \mathbb{R}^{s(2t)}$  and  $(\mathbf{Q}_g^*)_{g \in G_t}$  respectively, which satisfy  $\mathbf{Q}_g^* = \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}^*)^{-1}$  for all  $g \in G_t$ . Therefore*

$$\begin{aligned} 1 &= \frac{1}{\sum_{g \in G_t} s(t-t_g)} \sum_{g \in G_t} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}^*)^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}) \\ &= \frac{1}{\sum_{g \in G_t} s(t-t_g)} \sum_{g \in G_t} g(\mathbf{x}) \Lambda_{t-t_g}^{g \cdot \phi_{2t}^*}(\mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n. \end{aligned} \tag{23}$$

**Proof.** For every fixed  $t$ , the convex cone  $Q_t(G)$  is a particular case of the convex cone  $K(\bar{q})$  investigated in Nesterov [8, p. 415, Section 2.2] when the functional system  $\{v(\mathbf{x})\}$  in [8] is the set of monomials  $(\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$  and the functions  $(\bar{q}_1, \dots, \bar{q}_l)$  are our polynomials  $g$  in  $G_t$ . Then

$$K(\bar{q}_1, \dots, \bar{q}_l)^* = Q_t(G)^* = \{ \phi_{2t} : \mathbf{M}_{t-t_g}(g \cdot \phi) \geq 0, g \in G_t \}.$$

By [8, Theorem 17.7]

$$p \in \text{int}(K(\bar{q}_1, \dots, \bar{q}_l)) \quad \text{if and only if} \quad p = \sum_{g \in G_t} g \cdot \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{M}_{t-t_g}(g \cdot \phi_p)^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}),$$

for some unique  $\phi_p \in K(\bar{q}_1, \dots, \bar{q}_l)^*$ . In addition, letting  $\mathbf{Q}_g := \mathbf{M}_{t-t_g}(g \cdot \phi_p)^{-1}$ ,  $g \in G_t$ , the sequence  $(\mathbf{Q}_g)_{g \in G_t}$  is the unique solution of (22), with  $p$  instead of  $\sum_{g \in G_t} s(t-t_g)$  in the left-hand-side of the constraint. Therefore, by [8, Theorem 17.7] for the constant polynomial  $p = 1$ ,

$$1 \in \text{int}(Q_t(G)) \iff 1 = \sum_{g \in G_t} g \cdot \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{M}_{t-t_g}(g \cdot \phi)^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}),$$

for some distinguished  $\phi \in Q_t(G)^*$ . Then as  $1 \in \text{int}(Q_t(G))$  for every  $t$ , letting  $p$  be the constant polynomial  $\sum_{g \in G_t} s(t-t_g)$ , one obtains

$$\sum_{g \in G_t} s(t-t_g) = \sum_{g \in G_t} g \cdot \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}^*)^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}),$$

for some unique  $\phi_{2t}^* \in Q_t(G)^*$ , and  $\mathbf{Q}_g^* := \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}^*)^{-1}$ ,  $g \in G_t$ , is the unique optimal solution of (22). Next,  $\phi_{2t}^*$  is a feasible solution of (21), and

$$\sum_{g \in G_t} \log \det(\mathbf{Q}_g^*) = - \sum_{g \in G_t} \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi_{2t}^*)) \geq \rho_t.$$

We next prove weak duality, i.e.,  $\rho_t^* \leq \rho_t$ , so that  $\phi_{2t}^*$  (resp.  $(\mathbf{Q}_g^*)_{g \in G_t}$ ) is the unique optimal solution of (21) (resp. (22)) and  $\rho_t = \rho_t^*$ . So let  $\phi_{2t}$  (resp.  $(\mathbf{Q}_g)_{g \in G_t}$ ) be an arbitrary feasible solution of (21) (resp. (22)). Then by Lemma 12, for every  $g \in G_t$ ,

$$s(t-t_g) + \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi_{2t})) + \log \det(\mathbf{Q}_g) \leq \langle \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}), \mathbf{Q}_g \rangle.$$

In addition, as  $\phi_{2t}(1) = 1$

$$\begin{aligned} \sum_{g \in G_t} s(t - t_g) &= \phi_{2t} \left( \sum_{g \in G_t} s(t - t_g) \right) = \sum_{g \in G_t} \phi_{2t}(g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{Q}_g \mathbf{v}_{t-t_g}(\mathbf{x})) \\ &= \sum_{g \in G_t} g \cdot \phi_{2t}(\mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{Q}_g \mathbf{v}_{t-t_g}(\mathbf{x})) \\ &= \sum_{g \in G_t} \langle \mathbf{Q}_g, \mathbf{M}_{t-t_g}(g \cdot \phi_{2t}) \rangle \quad [\text{by (11)}] \\ &\geq \sum_{g \in G_t} [s(t - t_g) + \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi_{2t})) + \log \det(\mathbf{Q}_g)], \end{aligned}$$

from which we deduce weak duality, that is,

$$\sum_{g \in G_t} \log \det(\mathbf{Q}_g) \leq - \sum_{g \in G_t} \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi_{2t})). \quad \square$$

So as one can see, (23) is a multivariate analogue of (4). Crucial in Theorem 6 is the condition  $1 \in \text{int}(Q_t(G))$  for all  $t$ . Below is a simple sufficient condition.

**Lemma 7.** *Let  $S$  be as in (5) with  $G = \{g_0, g_1, \dots, g_m\}$ , and let Assumption 1 hold. Then  $1 \in \text{int}(Q_t(G))$  for every  $t$ .*

For clarity of exposition the proof is postponed to Section 3.1.

**Remark 8.** Let  $n = 1$  and  $S = [-1, 1] = \{x \in \mathbb{R} : g(x) \geq 0\}$  with  $x \mapsto g(x) = 1 - x^2$ . Then  $G = \{g\}$ , the unique optimal solution  $\phi_{2t}^*$  of (21) is the vector of moments up to degree  $2t$  of the Chebyshev measure  $dx/\pi\sqrt{1-x^2}$  on  $[-1, 1]$ , and (23) is exactly (4).

**Lemma 9.** *Let  $g \in \mathbb{R}[\mathbf{x}]$  of even degree be fixed,  $G := \{g\}$ , and suppose that there are two polynomials of even degree  $p \in \text{int}(\Sigma_t)$ , and  $q \in \text{int}(\Sigma_{t-t_g})$  such that  $p + gq = 1$ . Then there exists a linear functional  $\phi \in \mathbb{R}[\mathbf{x}]_{2t}^*$  with  $\phi \in \text{int}(Q_t(G)^*)$  such that*

$$1 = \mathbf{v}_t(\mathbf{x})^T \Lambda_t^\phi(\mathbf{x})^{-1} \mathbf{v}_t(\mathbf{x}) + g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x})^T \Lambda_{t-t_g}^{g \cdot \phi}(\mathbf{x})^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (24)$$

*In particular with  $g \in \mathbb{R}[\mathbf{x}]$  fixed: If there exist polynomials  $(C_i, H_i)_{i \in I} \subset \mathbb{Z}[\mathbf{x}]$  that solve Pell's polynomial equation  $C_i^2 + g H_i^2 = 1$ ,  $i \in I$ , and if  $\sum_{i \in I} C_i^2 \in \text{int}(\Sigma[\mathbf{x}]_t)$ ,  $\sum_{i \in I} H_i^2 \in \text{int}(\Sigma[\mathbf{x}]_{t-t_g})$ , then (24) holds for some  $\phi \in \text{int}(Q_t(G)^*)$ .*

**Proof.** Let  $G := \{g\}$  and let  $Q_t(G)$  be as in (7). As  $p \in \text{int}(\Sigma_t)$ , and  $q \in \text{int}(\Sigma_{t-t_g})$ ,  $1 = p + gq \in \text{int}(Q_t(G))$  and by [5, Lemma 4], (24) holds. The second statement is a direct consequence by taking  $p = \sum_{i \in I} C_i^2$  and  $q = \sum_{i \in I} H_i^2$ .  $\square$

So Lemma 9 states that if the triple  $(p, g, q)$  solve the generalized Pell's equation  $p + gq = 1$ , with  $p \in \text{int}(\Sigma_t)$  and  $q \in \text{int}(\Sigma_{t-t_g})$ , then  $p$  (resp.  $q$ ) is the Christoffel polynomial  $(\Lambda_t^\phi)^{-1}$  (resp.  $(\Lambda_{t-t_g}^{g \cdot \phi})^{-1}$ ) associated with some linear functional  $\phi \in \mathbb{R}[\mathbf{x}]_{2t}^*$  such that  $\phi \in \text{int}(Q_t(G)^*)$ .

### 2.5. An asymptotic result

We now consider asymptotics for the sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  obtained in Theorem 6, as  $t$  grows.

**Theorem 10.** *Under Assumption 1, let  $\phi_{2t}^*$  be an optimal solution of (21),  $t \in \mathbb{N}$ , guaranteed to exist by Theorem 6. Then:*

- (i) *The sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  has accumulation points, and for each converging subsequence  $(t_k)_{k \in \mathbb{N}}$ ,  $(\phi_{2t_k}^*)_{k \in \mathbb{N}}$  converges pointwise to the vector  $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$  of moments of some probability measure  $\phi$  on  $S$ , that is,*

$$\lim_{k \rightarrow \infty} (\phi_{2t_k}^*)_\alpha = \phi_\alpha = \phi(\mathbf{x}^\alpha) = \int_S \mathbf{x}^\alpha d\phi, \quad \forall \alpha \in \mathbb{N}^n. \quad (25)$$

(ii) A limit probability measure  $\phi$  as in (i) satisfies (15) if and only if the whole sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  converges to  $\phi$  and finite convergence takes place. That is, there exists  $t_0$  such that for all  $t \geq t_0$ ,

$$(\phi_{2t}^*)_{\alpha} = \phi_{\alpha} = \phi(\mathbf{x}^{\alpha}) = \int_S \mathbf{x}^{\alpha} d\phi, \quad \forall \alpha \in \mathbb{N}_{2t}^n, \tag{26}$$

and so  $\phi$  is a representing measure of  $\phi_{2t}^*$  for all  $t \geq t_0$ . In addition, under the condition of Theorem 2,  $\phi$  is the equilibrium measure  $\lambda_S$  of  $S$ .

**Proof.**

(i). As  $R - \|\mathbf{x}\|^2 \in Q_1(G)$ , the set of feasible solutions of (21) is compact. Indeed, let  $\phi_{2t}$  be feasible for (21). Then as  $\phi_{2t}(1) = 1$ ,  $R \geq \phi_{2t}(x_i^2)$  for all  $i = 1, \dots, n$ . Next in multiplying by  $x_i$  (with  $i$  arbitrary),  $Rx_i^2 - x_i^2 \cdot \|\mathbf{x}\|^2 \in Q_2(G)$  and so

$$R^2 \geq R\phi_{2t}(x_i^2) \geq \phi_{2t}(x_i^2 \cdot \|\mathbf{x}\|^2) \geq \phi_{2t}(x_i^4) \Rightarrow R^2 \geq \phi_{2t}(x_i^4).$$

Iterating yields  $R^t \geq \phi_{2t}(x_i^{2t})$  for every  $i = 1, \dots, n$ . Then by [4, Proposition 2.38, p. 41] one obtains  $|(\phi_{2t})_{\alpha}| \leq \max[1, R^t]$  for all  $\alpha \in \mathbb{N}_{2t}^n$ , and all  $\phi_{2t} \in Q_t(G)^*$ . In fact (and assuming  $R \geq 1$ ) we even have  $|(\phi_{2t})_{\alpha}| \leq R^{|\alpha|/2}$ , for all  $|\alpha| \leq 2t$ , and all  $\phi_{2t} \in Q_t(G)^*$ .

By completing with zeros, the finite sequence  $\phi_{2t}^*$  is viewed as an infinite sequence indexed by  $\mathbb{N}^n$ . Then by a standard argument involving scaling and the  $\sigma(\ell_{\infty}, \ell_1)$  weak- $\star$  topology, the sequence  $(\phi_{2t}^*)_{t \in \mathbb{N}}$  has accumulation points and for each subsequence  $(t_k)_{k \in \mathbb{N}}$  converging to some  $\phi \in \mathbb{N}^n$ , one obtains the pointwise convergence  $\lim_{k \rightarrow \infty} (\phi_{2t_k}^*)_{\alpha} = \phi_{\alpha}$ , for every  $\alpha \in \mathbb{N}^n$ . Next, let  $d \in \mathbb{N}$  and  $g \in G$  be fixed, arbitrary. Observe that  $\mathbf{M}_d(\phi_{2t_k}^*) \geq 0$  as a principal submatrix of  $\mathbf{M}_{t_k}(\phi_{2t_k}^*) \geq 0$ , and similarly  $\mathbf{M}_d(g \cdot \phi_{2t_k}^*) \geq 0$  as a principal submatrix of  $\mathbf{M}_{t_k}(g \cdot \phi_{2t_k}^*) \geq 0$  (when  $k$  is sufficiently large so that  $G_{t_k} = G$ ). Therefore by the above pointwise convergence,  $\mathbf{M}_d(g \cdot \phi_{2t_k}^*) \rightarrow \mathbf{M}_d(g \cdot \phi) \geq 0$  as  $k$  increases. As  $Q(G)$  is Archimedean, then by Putinar’s Positivstellensatz [10],  $\phi$  is a Borel probability measure on  $S$  (as  $\phi_{2t_k}^*(1) = 1$  for all  $k$ ).

(ii). Let  $\phi$  be as in (i) and suppose that  $\phi$  satisfies (15). Then for each  $t \geq t_0$ , the vector  $\phi_{2t} = (\phi_{\alpha})_{\alpha \in \mathbb{N}_{2t}^n}$  is an optimal solution of (21), and by uniqueness,  $\phi_{2t} = \phi_{2t}^*$ . That is,  $\phi$  is a representing measure for  $\phi_{2t}^*$  for all  $t \geq t_0$ . But this implies that  $\phi_{2(t+1)}^*$  is an extension of  $\phi_{2t}^*$  for all  $t \geq t_0$ , and therefore the whole sequence converges to  $\phi$ , and the convergence is finite.

Conversely, if finite convergence takes place, that is, if  $\phi_{2(t+1)}^*$  is an extension of  $\phi_{2t}^*$  for all  $t \geq t_0$ , then  $\phi$  in (i) is the unique accumulation point and its associated measure  $\phi$  satisfies (15).

Finally, if  $(S, g \cdot \phi)$  satisfies the Bernstein–Markov property for all  $g \in G$ , then by Theorem 2,  $\phi = \lambda_S$ , which concludes the proof. □

**Remark 11.** Theorem 10 provides a simple test to detect whether the set  $G$  of generators of  $S$  is a good one, and if so, a numerical scheme to compute moments of the equilibrium measure  $\lambda_S$  of  $S$ . Indeed if (26) has to hold for the equilibrium measure  $\lambda_S$ , then necessarily, the unique optimal solution  $\phi_{2(t+1)}^*$  of (21) for  $t + 1$  must be an extension of the unique optimal solution  $\phi_{2t}^*$  of (21) for  $t$ , whenever  $t$  is sufficiently large. So for instance, if one observes that  $\phi_2^*$  is an extension of  $\phi_1^*$  after solving (21) for  $t = 1$  and  $t = 2$ , then it already provides a good indication that finite convergence may indeed take place.

**3. Examples on some particular sets  $S$**

We know that Theorem 10 holds for the equilibrium measure  $\lambda_S = 1_{[-1,1]}(x) \frac{dx}{\pi\sqrt{1-x^2}}$  of the interval  $S = [-1, 1]$ . Next, we first show how (26) holds at least for  $t = 1, 2, 3$  in the bivariate case with  $S$  being the Euclidean unit box and unit ball, or the canonical simplex, indeed  $\phi_2^*$ ,  $\phi_4^*$ , and  $\phi_6^*$ , are moment vectors up to degree 2 and 4 and 6, of the equilibrium measures  $dxdy/2\pi\sqrt{1-x^2-y^2}$ ,  $dxdy/\pi^2\sqrt{(1-x^2)(1-y^2)}$ , and  $dxdy/\pi\sqrt{x \cdot y \cdot (1-x-y)}$ , respectively.

**On the Euclidean unit box.** With  $S := [-1, 1]^2$  and  $\lambda_S = dx dy / \pi^2 \sqrt{(1-x^2)(1-y^2)}$ . With the univariate Chebyshev polynomials  $T_n$  of first kind and  $U_n$  of second kind, and letting

- $g_1(x, y) := (1 - x^2)$ ;  $g_2(x, y) := (1 - y^2)$ ;  $g_3(x, y) := (1 - x^2)(1 - y^2)$ ;  $G = \{g_1, g_2, g_3\}$ ,
- $(P_{ij}^1(x, y) := \widehat{T}_i(x)\widehat{T}_j(y))_{i,j \in \mathbb{N}}$  form an orthonormal family with respect to  $d\lambda_S$ ,
- $(P_{ij}^{g_1}(x, y) := \widehat{U}_i(x)\widehat{T}_j(y))_{i,j \in \mathbb{N}}$  form an orthonormal family with respect to  $(1 - x^2)d\lambda_S$ ,
- $(P_{ij}^{g_2}(x, y) := \widehat{T}_i(x)\widehat{U}_j(y))_{i,j \in \mathbb{N}}$  form an orthonormal family with respect to  $(1 - y^2)d\lambda_S$ ,
- $(P_{ij}^{g_3}(x, y) := \widehat{U}_i(x)\widehat{U}_j(y))_{i,j \in \mathbb{N}}$  form an orthonormal family with respect to  $(1 - x^2)(1 - y^2)d\lambda_S$ .

Then

$$P_{10}(x, y)^2 + P_{01}^2(x, y) + (1 - x^2) P_{00}^{g_1}(x, y)^2 + (1 - y^2) P_{00}^{g_2}(x, y)^2 = 2,$$

from which we obtain:

$$(\Lambda_1^{\lambda_S})^{-1} + g_1 \cdot (\Lambda_0^{g_1 \cdot \lambda_S})^{-1} + g_2 \cdot (\Lambda_0^{g_2 \cdot \lambda_S})^{-1} = 5 = s(1) + 2s(0).$$

Next,

$$\begin{aligned} P_{20}(x, y)^2 + P_{11}(x, y)^2 + P_{02}(x, y)^2 &= 8x^4 + 8y^4 - 8x^2 - 8y^2 + 4x^2y^2 + 4 \\ (1 - x^2) ((P_{10}^{g_1})^2 + (P_{10}^{g_2})^2) &= 8x^2 - 8x^4 + 4y^2 - 4x^2y^2 \\ (1 - y^2) ((P_{10}^{g_2})^2 + (P_{10}^{g_1})^2) &= 8y^2 - 8y^4 + 4x^2 - 4x^2y^2 \\ (1 - x^2)(1 - y^2) (P_{00}^{g_3})^2 &= 4 - 4x^2 - 4y^2 + 4x^2y^2. \end{aligned}$$

Again after scaling (to get borthonormal polynomials) and summing up, one obtains

$$(\Lambda_2^{\lambda_S})^{-1} + g_1 \cdot (\Lambda_1^{g_1 \cdot \lambda_S})^{-1} + g_2 \cdot (\Lambda_1^{g_2 \cdot \lambda_S})^{-1} + g_3 \cdot (\Lambda_0^{g_3 \cdot \lambda_S})^{-1} = 13 = s(2) + 2s(1) + s(0).$$

**On the 2D-Euclidean ball.** With  $\lambda_S = \frac{dx dy}{2\pi\sqrt{1-x^2-y^2}}$  and  $\mu_{ij} := \int x^i y^j d\mu(x, y)$ , for all  $i, j \in \mathbb{N}$ , one obtains:

$$\mathbf{M}_2(\mu) = \begin{bmatrix} 1 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/5 & 0 & 1/15 \\ 0 & 0 & 0 & 0 & 1/15 & 0 \\ 1/3 & 0 & 0 & 1/15 & 0 & 1/5 \end{bmatrix}.$$

Similarly, with  $g \cdot \mu = (1 - x^2 - y^2) dx dy = \sqrt{1 - x^2 - y^2} dx dy / 2\pi$ ,

$$\mathbf{M}_1(g \cdot \mu) = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/15 & 0 \\ 0 & 0 & 1/15 \end{bmatrix},$$

which yields

$$\Lambda_1^\mu(x, y)^{-1} + (1 - x^2 - y^2) \cdot \Lambda_0^{g \cdot \mu}(x, y)^{-1} = 3 + 1 = s(1) + s(0),$$

as well as

$$\Lambda_2^\mu(x, y)^{-1} + (1 - x^2 - y^2) \cdot \Lambda_1^{g \cdot \mu}(x, y)^{-1} = 9 = s(2) + s(1),$$

and similarly

$$\Lambda_3^\mu(x, y)^{-1} + (1 - x^2 - y^2) \cdot \Lambda_2^{g \cdot \mu}(x, y)^{-1} = 16 = s(3) + s(2) + s(1),$$

as predicted by Theorem 10.

**On the simplex.** Consider the canonical simplex  $S = \{(x, y) \in \mathbb{R}^2 : x + y \leq 1; x, y \geq 0\}$  with equilibrium measure reads:

$$\lambda_S = \frac{1_S(x, y) dx dy}{2\pi \sqrt{x} \sqrt{y} \sqrt{1-x-y}}.$$

There are several ways to represent  $S$  and in particular, consider  $G = \{g_1, g_2, g_3\}$  with

$$(x, y) \mapsto g_1(x, y) := x \cdot (1 - x - y); \quad g_2(x, y) := y \cdot (1 - x - y),$$

and  $(x, y) \mapsto g_3(x, y) := x \cdot y$ . Do we have

$$\Lambda_t^{\lambda_S}(x)^{-1} + g_1 \cdot \Lambda_{t-1}^{g_1 \cdot \lambda_S}(x)^{-1} + g_2 \cdot \Lambda_{t-1}^{g_2 \cdot \lambda_S}(x)^{-1} + g_3 \cdot \Lambda_{t-1}^{g_3 \cdot \lambda_S}(x, y)^{-1} = 1?$$

With  $t = 1$ , the moment matrix of  $\lambda_S$  reads:

$$\mathbf{M}_1(\mu) = \frac{1}{15} \begin{bmatrix} 15 & 5 & 5 \\ 5 & 3 & 1 \\ 5 & 1 & 3 \end{bmatrix}.$$

Next, we obtain

$$\int_S x \cdot (1 - x - y) d\lambda_S = \int_S y \cdot (1 - x - y) d\lambda_S = \frac{1}{15}; \quad \int_S xy \cdot (1 - x - y) d\lambda_S = \frac{1}{15}.$$

Hence

$$\begin{aligned} \Lambda_1^\mu(x, y)^{-1} &= 15 \left( \frac{2}{5} - x - y + x^2 + y^2 + xy \right) \\ x \cdot (1 - x - y) \Lambda_0^{v_1}(x, y)^{-1} &= 15(x - x^2 - xy) \\ y \cdot (1 - x - y) \Lambda_0^{v_2}(x, y)^{-1} &= 15(y - y^2 - xy) \\ x \cdot y \Lambda_0^{v_3}(x, y)^{-1} &= 15xy, \end{aligned}$$

and therefore

$$\Lambda_1^\mu(x, y)^{-1} + x \cdot (1 - x - y) \Lambda_0^{v_1}(x, y)^{-1} + y \cdot (1 - x - y) \Lambda_0^{v_2}(x, y)^{-1} + x \cdot y \Lambda_0^{v_3}(x, y)^{-1} = 6 = s(1) + 3s(0),$$

as predicted by Theorem 10. Similarly,

$$\mathbf{M}_2(\lambda_S) = \begin{bmatrix} 1.0000 & 0.3333 & 0.3333 & 0.2000 & 0.0667 & 0.2000 \\ 0.3333 & 0.2000 & 0.0667 & 0.1429 & 0.0286 & 0.0286 \\ 0.3333 & 0.0667 & 0.2000 & 0.0286 & 0.0286 & 0.1429 \\ 0.2000 & 0.1429 & 0.0286 & 0.1111 & 0.0159 & 0.0095 \\ 0.0667 & 0.0286 & 0.0286 & 0.0159 & 0.0095 & 0.0159 \\ 0.2000 & 0.0286 & 0.1429 & 0.0095 & 0.0159 & 0.1111 \end{bmatrix}$$

and

$$\mathbf{M}_1(g_1 \cdot \lambda_S) = \begin{bmatrix} 0.0667 & 0.0286 & 0.0095 \\ 0.0286 & 0.0159 & 0.0032 \\ 0.0095 & 0.0032 & 0.0032 \end{bmatrix}; \quad \mathbf{M}_1(g_2 \cdot \lambda_S) = \begin{bmatrix} 0.0667 & 0.0095 & 0.0286 \\ 0.0095 & 0.0032 & 0.0032 \\ 0.0286 & 0.0032 & 0.0159 \end{bmatrix}$$

and

$$\mathbf{M}_1(g_3 \cdot \lambda_S) = \begin{bmatrix} 0.0667 & 0.0286 & 0.0286 \\ 0.0286 & 0.0159 & 0.0095 \\ 0.0286 & 0.0095 & 0.0159 \end{bmatrix}.$$

Then this implies

$$(\Lambda_2^{\lambda_S})^{-1} + \sum_{i=1}^3 g_i \cdot (\Lambda_1^{g_i \cdot \lambda_S})^{-1} = 15 = s(2) + 3s(1).$$

In continuing with  $t = 3$  we also obtain

$$(\Lambda_3^{\lambda_S})^{-1} + \sum_{i=1}^3 g_i \cdot (\Lambda_2^{g_i \cdot \lambda_S})^{-1} = 28 = s(3) + 3s(2).$$



**Intersection of two ellipsoids.** Here we consider  $S \subset \mathbb{R}^2$ ,  $G = \{g_0, g_1, g_2\}$ , with

$$g_1(x, y) := 1 - 2x^2 - 3y^2; \quad g_2(x, y) := 1 - 3x^2 - 2y^2.$$

so that  $S$  is the intersection of two ellipsoids. With  $t = 1, 2, 3$ , the respective optimal solutions  $\phi_2^*$ ,  $\phi_4^*$  and  $\phi_6^*$  of (21) are such that  $\phi_4^*$  seems to be an extension of  $\phi_2^*$ , and  $\phi_6^*$  seems to be an extension of  $\phi_4^*$ , up to some numerical imprecision due to the solver; indeed the norm of the difference between  $\phi_4^*$  and the restriction of  $\phi_6^*$  is about 0.006, and for instance

$$\begin{aligned} \phi_2^* &= (1, 0, 0, 0.00999961, 0, 0.00999962) & \phi_4^* &= (1, 0, 0, 0.0117564, 0, 0.01175, \dots). \\ \phi_4^* &= (1, 0, 0, 0.0117564, 0, 0.01175, \dots); & \phi_6^* &= (1, 0, 0, 0.011506, 0, 0.011425, \dots). \end{aligned}$$

However, as the solver is not very accurate, it is difficult to conclude whether or not the differences between  $\phi_2$ ,  $\phi_4$  and  $\phi_6$  are due to numerical inaccuracies.

**The TV-screen.** Let  $S := \{(x, y) \in \mathbb{R}^2 : x^4 + y^4 \leq 1\}$ . By solving numerically (21)-(22) with  $t = 2$  and  $t = 3$ , we find

$$\begin{aligned} \Lambda_2^{\phi_4^*}(x, y)^{-1} + (1 - x^4 - y^4) \Lambda_0^{g \cdot \phi_4^*}(x, y)^{-1} &= 6 + 1 = s(2) + s(0) \\ \Lambda_3^{\phi_6^*}(x, y)^{-1} + (1 - x^4 - y^4) \Lambda_1^{g \cdot \phi_6^*}(x, y)^{-1} &= 10 + 3 = s(3) + s(1) \end{aligned}$$

respectively, as predicted by Theorem 6. However, we observed that  $\phi_6^*$  is not an extension of  $\phi_4^*$ .

**The Gaussian case.** Finally, in the same spirit but not in the preceding context of a compact set  $S \subset \mathbb{R}^n$ , we consider the case of  $\mathbb{R}^n$  where by [5, Lemma 3] any polynomial  $p \in \text{int}(\Sigma[\mathbf{x}]_t)$  is the Christoffel function of some linear functional  $\phi_p \in \Sigma[\mathbf{x}]_t^*$ . Let  $\Sigma > 0$  be a real symmetric  $n \times n$  matrix, and associated with  $\Sigma$ , let  $p$  be the quadratic polynomial

$$\mathbf{x} \mapsto p(\mathbf{x}) := 1 + \langle \mathbf{x}, \Sigma^{-1} \mathbf{x} \rangle, \quad \mathbf{x} \in \mathbb{R}^n,$$

which is in the interior of the convex cone  $\Sigma[\mathbf{x}]_1$ . Therefore, by [8] and [5, Lemma 3]

$$p(\mathbf{x}) = \mathbf{v}_1(\mathbf{x})^T \mathbf{M}_1(\phi)^{-1} \mathbf{v}_1(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some unique  $\phi \in \Sigma[\mathbf{x}]_1^*$ . It is straightforward to check that

$$p(\mathbf{x}) = \mathbf{v}_1(\mathbf{x})^T \begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}^{-1} \mathbf{v}_1(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

and  $\begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}^{-1}$  is the moment matrix of the Gaussian measure

$$d\phi_p = (\det(2\pi \Sigma))^{-1/2} \exp(-\mathbf{x}^T \Sigma^{-1} \mathbf{x} / 2) d\mathbf{x}.$$

That is,  $\phi$  is represented by the Gaussian measure  $\phi_p$  and  $\Lambda_1^\phi$  is the Christoffel function of degree 2 of the Gaussian measure  $\phi_p$ .

$$\Lambda_1^{\phi_p}(\mathbf{x})^{-1} = (1, \mathbf{x})^T \begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}^{-1} (1, \mathbf{x}) = 1 + \mathbf{x}^T \Sigma^{-1} \mathbf{x} = p(\mathbf{x}).$$

### 3.1. Discussion

There are several issues that are worth investigating. The first one is to completely validate our result for  $t > 3$ , for the cases where  $S$  is the unit box, the Euclidean unit ball, and the simplex. One possibility is to use Corollary 4 for each degree  $t$ , which only requires to show (20) (a property of orthonormal polynomials associated with the measures  $(g \cdot \lambda_S)_{g \in G}$ ) as we did on some of the above examples.

Another issue is to investigate what is a distinguishing feature of the limit measure  $\phi$  in Theorem 10 when  $\phi$  does not satisfy the generalized Pell's equation (8). Could  $\phi$  still be the equilibrium measure of  $S$ ?

Finally, one would like to extend the present framework and characterize the linear functional  $\phi_p \in Q_t(G)^*$  in [5, Lemma 4], associated with a non constant polynomial  $p \in \text{int}(Q_t(G))$ . Then with arguments that mimic those used for the constant polynomial  $p = 1$ , if  $p \in \text{int}(Q_t(G))$  for every  $t \in \mathbb{N}$ , then a natural candidate seems to be  $\phi_p := \lambda_S/p$ , that is,  $\phi_p$  is the measure with density  $1/p$  with respect to the equilibrium measure  $\lambda_S$  of  $S$ .

### Appendix

**Lemma 12.** *Let  $\mathcal{S}^n$  be the space of real symmetric  $n \times n$  matrices and let  $\mathcal{S}_{++}^n \subset \mathcal{S}^n$  be the convex cone of real  $n \times n$  positive definite matrices  $\mathbf{Q}$  (denoted  $\mathbf{Q} > 0$ ). Then*

$$n + \log \det(\mathbf{M}) + \log \det(\mathbf{Q}) \leq \langle \mathbf{M}, \mathbf{Q} \rangle, \quad \forall \mathbf{M}, \mathbf{Q} \in \mathcal{S}_{++}^n. \tag{27}$$

with equality if and only if  $\mathbf{Q} = \mathbf{M}^{-1}$ .

**Proof.** Consider the concave function

$$f : \mathcal{S}^n \rightarrow \mathbb{R} \cup \{-\infty\} \quad \mathbf{Q} \mapsto f(\mathbf{Q}) = \begin{cases} \log \det(\mathbf{Q}) & \text{if } \mathbf{Q} \in \mathcal{S}_{++}^n, \\ -\infty & \text{otherwise,} \end{cases}$$

and let  $f^*$  be its (concave analogue) of Legendre–Fenchel conjugate, i.e.,

$$\mathbf{M} \mapsto f^*(\mathbf{M}) := \inf_{\mathbf{Q} \in \mathcal{S}^n} \langle \mathbf{M}, \mathbf{Q} \rangle - f(\mathbf{Q}).$$

It turns out that

$$f^*(\mathbf{M}) = \begin{cases} n + \log \det(\mathbf{M}) (= n + f(\mathbf{M})) & \text{if } \mathbf{M} \in \mathcal{S}_{++}^n, \\ -\infty & \text{otherwise.} \end{cases}$$

Hence the concave analogue of Legendre–Fenchel inequality states that

$$f^*(\mathbf{M}) + f(\mathbf{Q}) \leq \langle \mathbf{M}, \mathbf{Q} \rangle, \quad \forall \mathbf{M}, \mathbf{Q} \in \mathcal{S}^n,$$

and yields (27). □

**Proof of Lemma 7.** Recall that  $\theta(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$ . By [9, Lemma 3.4]

$$(R + 1)^t = \underbrace{(1 + \|\mathbf{x}\|^2)^t}_{\Delta} + \theta(\mathbf{x}) \underbrace{\sum_{j=0}^{t-1} (R + 1)^j (1 + \|\mathbf{x}\|^2)^{t-j-1}}_{\Gamma}.$$

Note that

$$\Delta(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_t^n} \Theta_{\alpha} \mathbf{x}^{2\alpha} = \mathbf{v}_t(\mathbf{x})^T \mathbf{G}_0 \mathbf{v}_t(\mathbf{x}), \quad \forall \mathbf{x},$$

$$\Gamma(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_{t-1}^n} \Gamma_{\alpha} \mathbf{x}^{2\alpha} = \mathbf{v}_{t-1}(\mathbf{x})^T \mathbf{G}_1 \mathbf{v}_{t-1}(\mathbf{x}), \quad \forall \mathbf{x},$$

where  $\mathbf{G}_0, \mathbf{G}_1$  are diagonal positive definite matrices (i.e.  $\mathbf{G}_0, \mathbf{G}_1 > 0$ ), and so  $(R + 1)^t = \Delta + \theta \Gamma$ . Next, let  $\mathbf{W}$  be a real symmetric matrix such that

$$\mathbf{v}_t(\mathbf{x})^T \mathbf{W} \mathbf{v}_t(\mathbf{x}) = \sum_{g \in G_t} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x}) \mathbf{I}_{t-t_g} \mathbf{v}_{t-t_g}(\mathbf{x}),$$

where  $\mathbf{I}_{t-t_g}$  is the  $s(t - t_g)$ -identity matrix. As  $\mathbf{G}_0 > 0$  there exists  $\delta > 0$  such that  $\mathbf{G}_0 - \delta \mathbf{W} > 0$ . But then

$$(R + 1)^t = \mathbf{v}_t(\mathbf{x})^T (\mathbf{G}_0 - \delta \mathbf{W}) \mathbf{v}_t(\mathbf{x}) + \delta \sum_{g \in G_t} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x}) \mathbf{I}_{t-t_g} \mathbf{v}_{t-t_g}(\mathbf{x}) + \theta(\mathbf{x}) \mathbf{v}_{t-1}(\mathbf{x})^T \mathbf{G}_1 \mathbf{v}_{t-1}(\mathbf{x}).$$

Finally, as  $\theta \in Q_1(G)$  and  $\deg(g) \leq 2$  for all  $g \in G_1$ , write

$$\theta(\mathbf{x}) = \mathbf{v}_1(\mathbf{x})^T \mathbf{A}_0 \mathbf{v}_1(\mathbf{x}) + \sum_{g_0 \neq g \in G_1} g(\mathbf{x}) a_g$$

with  $a_g \geq 0$  and  $\mathbf{A}_0 \geq 0$ , to obtain

$$\begin{aligned} (R+1)^t = & \mathbf{v}_t(\mathbf{x})^T (\mathbf{G}_0 - \delta \mathbf{W}) \mathbf{v}_t(\mathbf{x}) + \delta \sum_{g \in G_t \setminus G_1} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x}) \mathbf{I}_{t-t_g} \mathbf{v}_{t-t_g}(\mathbf{x}) \\ & + \sum_{g \in G_1; g \neq g_0} g(\mathbf{x}) [\delta \mathbf{v}_{t-1}(\mathbf{x}) \mathbf{I}_{t-1} \mathbf{v}_{t-1}(\mathbf{x}) + a_g \mathbf{v}_{t-1}(\mathbf{x})^T \mathbf{G}_1 \mathbf{v}_{t-1}(\mathbf{x})], \\ & + \delta \mathbf{v}_t(\mathbf{x}) \mathbf{I}_t \mathbf{v}_t(\mathbf{x}) + \underbrace{\mathbf{v}_1(\mathbf{x})^T \mathbf{A}_0 \mathbf{v}_1(\mathbf{x}) \mathbf{v}_{t-1}(\mathbf{x})^T \mathbf{G}_1 \mathbf{v}_{t-1}(\mathbf{x})}_{\mathbf{v}_t(\mathbf{x})^T \mathbf{V} \mathbf{v}_t(\mathbf{x}); \mathbf{V} \geq 0}, \end{aligned}$$

and therefore  $(R+1)^t \in \text{int}(Q_t(G))$ . □

## References

- [1] E. Bedford, B. A. Taylor, "The complex equilibrium measure of a symmetric convex set in  $\mathbb{R}^n$ ", *Trans. Am. Math. Soc.* **294** (1986), no. 2, p. 705-717.
- [2] I. Dunning, J. Huchette, M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization", *SIAM Rev.* **59** (2017), no. 2, p. 295-320.
- [3] M. Grant, S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1", <http://cvxr.com/cvx>, 2014.
- [4] J. B. Lasserre, *Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2015.
- [5] ———, "A disintegration of the Christoffel function", *C. R. Math. Acad. Sci. Paris* **360** (2022), p. 1071-1079.
- [6] J. B. Lasserre, E. Pauwels, M. Putinar, *The Christoffel–Darboux Kernel for Data Analysis*, Cambridge Monographs on Applied and Computational Mathematics, vol. 38, Cambridge University Press, 2022.
- [7] J. Mc Laughlin, "Multivariable-polynomial solutions to Pell's equation and fundamental units in real quadratic fields", *Pac. J. Math.* **210** (2002), no. 2, p. 335-348.
- [8] Y. Nesterov, "Squared functional systems and optimization problems", in *High Performance Optimization* (H. Frenk, K. Roos, T. Terlaky, S. Zhang, eds.), Applied Optimization, vol. 33, Springer, 2000, p. 405-440.
- [9] Ngoc Hoang Anh Mai, J. B. Lasserre, V. Magron, J. Wang, "Exploiting constant trace property in large scale polynomial optimization", <https://arxiv.org/abs/2012.08873>, to appear in *ACM Trans. Math. Softw.*, 2020.
- [10] M. Putinar, "Positive polynomials on compact semi-algebraic sets", *Indiana Univ. Math. J.* **42** (1993), no. 3, p. 969-984.
- [11] W. A. Webb, H. Yokota, "Polynomial Pell's equation", *Proc. Am. Math. Soc.* **131** (2003), no. 4, p. 993-1006.