



HAL
open science

Polynomial Optimization, Certificates of Positivity, and Christoffel Function

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. Polynomial Optimization, Certificates of Positivity, and Christoffel Function. Michal Kočvara; Bernard Mourrain; Cordian Riener. Polynomial Optimization, Moments, and Applications, Springer, pp.1-20, 2023. hal-04076663

HAL Id: hal-04076663

<https://laas.hal.science/hal-04076663>

Submitted on 24 Apr 2023

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

POLYNOMIAL OPTIMIZATION, CERTIFICATES OF POSITIVITY, AND CHRISTOFFEL FUNCTION

JEAN B. LASSERRE

ABSTRACT. We briefly recall basics of the Moment-SOS hierarchy in polynomial optimization and the Christoffel-Darboux kernel (and the Christoffel function (CF)) in theory of approximation and orthogonal polynomials. We then (i) show a strong link between the CF and the SOS-based positive certificate at the core of the Moment-SOS hierarchy, and (ii) describe how the CD-kernel provides a simple interpretation of the SOS-hierarchy of lower bounds as searching for some signed polynomial density (while the SOS-hierarchy of upper bounds is searching for a positive (SOS) density). This link between the CF and positive certificates, in turn allows us (i) to establish a disintegration property of the CF much like for measures, and (ii) for certain sets, to relate the CF of their equilibrium measure with a certificate of positivity on the set, for constant polynomials.

CONTENTS

1. Introduction	2
2. Notation, definitions and preliminary results	3
3. The Moment-SOS hierarchy in polynomial optimization	5
3.1. A Moment-SOS hierarchy of lower bounds	5
3.2. A Moment-SOS hierarchy of upper bounds	7
4. The Christoffel-Darboux kernel and Christoffel functions	8
4.1. Christoffel-Darboux kernel	8
4.2. Christoffel function	9
4.3. Some distinguishing properties of the CF	10
5. CF, Optimization, and SOS-Certificates of Positivity	11
5.1. The CF to compare the hierarchies of upper and lower bounds	11
5.2. The CF and positive polynomials	12
5.3. A disintegration of the CF	14
5.4. Positive polynomials and equilibrium measure	14
6. Conclusion	16
Appendix	16
References	17

Research supported by the AI Interdisciplinary Institute ANITI funding through the french program "Investing for the Future PISA" under the grant agreement number ANR-19-PI3A-0004. It has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 813211 (POEMA). 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.

1. INTRODUCTION

In this chapter we describe (in our opinion, surprising) links between different fields, namely optimization – convex duality – certificates of positivity in real algebraic geometry on the one hand, and orthogonal polynomials – Christoffel function – approximation – equilibrium measures, on the other hand. More precisely, consider the polynomial optimization problem:

$$(1.1) \quad \mathbf{P} : f^* = \min \{f(\mathbf{x}) : \mathbf{x} \in \mathbf{S}\},$$

where f is a polynomial and \mathbf{S} is a basic semi-algebraic set¹. Importantly, f^* in (1.1) is understood as the *global* minimum of \mathbf{P} and *not* a local minimum. As a polynomial optimization problem, \mathbf{P} is NP-hard in general. However, in the early 2000 the Moment-SOS hierarchy (SOS stands for “sum-of-squares”) has emerged as a new methodology for solving \mathbf{P} . Its distinguishing feature is (i) to exploit powerful certificates of positivity from real algebraic geometry (and dual results on the \mathbf{S} -moment problem) and (ii) combine them with the computational power of semidefinite programming in conic optimization, to obtain a *hierarchy* of (convex) semidefinite relaxations of \mathbf{P} of increasing size.

The optimal values of such semidefinite relaxations provide a *monotone non decreasing sequence* of certified lower bounds which converges to the global minimum f^* . In addition, finite convergence is generic and when there are finitely many global minimizers, they can be obtained (also generically) from the optimal solutions of the exact semidefinite relaxation, via a simple linear algebra routine.

Moreover, this methodology is easily adapted to solve the *Generalized Moment Problem* (GMP) whose list of potential applications in mathematics, computer science, probability & statistics, quantum information, and many areas of engineering, is almost endless. For a detailed description of the methodology and an account of several of its applications, the interested reader is referred to e.g. the books [8, 11, 17] and the many references therein. Less known is another (still SOS-based) hierarchy but now with an associated monotone non increasing sequence of *upper bounds* which converges to f^* . While very general in its underlying principle, its practical implementation requires the feasible set \mathbf{S} to have a “simple” geometry like a box, a simplex, an ellipsoid, a hypercube, or their image by an affine mapping, and recently, rates of its asymptotic convergence have been obtained in e.g. [4, 28, 29, 30].

Crucial at each step t of the Moment-SOS hierarchy of lower bounds, is a dual pair of semidefinite programs associated with a dual pair (C_t, C_t^*) of convex cones. By a duality result of Nesterov [22], the respective interiors of C_t and C_t^* are in a simple one-to-one correspondence. In fact, its recent interpretation in [14, Lemma 3] states that every polynomial $p \in \text{int}(C_t)$ has a distinguished SOS-based representation in terms of Christoffel functions associated with some moment-sequence $\phi_p \in \text{int}(C_t^*)$. (In particular, every degree- $2t$ SOS p in the interior of the convex cone Σ_t of SOS of degree at most $2t$, is the reciprocal of the Christoffel function of some linear functional $\phi_p \in \Sigma_t^*$). In turn this duality result can be exploited to reveal additional properties of the CF. For instance we use it to obtain a *disintegration* property of the CF [14], very much in like for measures on a Cartesian product of Borel spaces. Also, for certain compact sets we can relate the CF of their

¹A basic semi-algebraic set in the intersection of finitely many sublevel sets of polynomials.

equilibrium measure with a certain SOS-based representation of constant polynomials. Finally, we reveal an interpretation of the latter representation [19] related to what we call a *generalized polynomial Pell equation* (an equation which originates in algebraic number theory).

So in this chapter we first briefly review basics of the moment-SOS hierarchies of lower and upper bounds. We next introduce the Christoffel-Darboux kernel (CD-kernel) and the Christoffel function (CF) and describe some of their basic properties, which in our opinion are interesting on their own and should deserve more attention from the optimization community. We then describe our interpretation of Nesterov's duality result to establish a strong link between the Christoffel functions and the SOS-based positivity certificate used in the Moment-SOS hierarchy. Conversely, we also describe how this duality result of convex analysis can be used to provide a disintegration property of the Christoffel function and a result on equilibrium measures of certain compact semi-algebraic sets.

We hope that this brief account on links between seemingly distinct disciplines will raise curiosity from the optimization community.

2. NOTATION, DEFINITIONS AND PRELIMINARY RESULTS

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

$$\mathbf{x} \mapsto p(\mathbf{x}) = \langle \mathbf{p}, \mathbf{v}_t(\mathbf{x}) \rangle, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

where $\mathbf{p} \in \mathbb{R}^{s(t)}$ is the vector of coefficients of p in the basis $(\mathbf{x}^\alpha)_{\boldsymbol{\alpha} \in \mathbb{N}_t^n}$.

Given a closed set $\mathcal{X} \subset \mathbb{R}^n$, denote by $\mathcal{M}(\mathcal{X})$ (resp. $\mathcal{M}(\mathcal{X})_+$) the space of finite signed Borel measures (resp. the convex cone of finite Borel measures) on \mathcal{X} . The support $\text{supp}(\mu)$ of a Borel measure μ 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.

Riesz linear functional With any real sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\boldsymbol{\alpha} \in \mathbb{N}^n}$ (in bold letter) is associated the Riesz linear functional $\phi \in \mathbb{R}[\mathbf{x}]^*$ (not in bold) defined by:

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

A sequence $\boldsymbol{\phi}$ has a representing measure if there exists a Borel measure $\phi \in \mathcal{M}(\mathbb{R}^n)_+$ such that $\phi_\alpha = \int \mathbf{x}^\alpha d\phi$ for all $\boldsymbol{\alpha} \in \mathbb{N}^n$, in which case

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

Given a sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\boldsymbol{\alpha} \in \mathbb{N}^n}$ and a polynomial $g \in \mathbb{R}[\mathbf{x}]$ ($\mathbf{x} \mapsto g(\mathbf{x}) := \sum_\gamma g_\gamma \mathbf{x}^\gamma$), denote by $g \cdot \boldsymbol{\phi}$ the new sequence $(g \cdot \boldsymbol{\phi})_\alpha := \sum_\gamma g_\gamma \phi_{\alpha+\gamma}$, $\boldsymbol{\alpha} \in \mathbb{N}^n$, with associated Riesz linear functional $g \cdot \phi \in \mathbb{R}[\mathbf{x}]^*$:

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

Moment matrix With $t \in \mathbb{N}$, the moment matrix $\mathbf{M}_t(\boldsymbol{\phi})$ associated with a real sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\boldsymbol{\alpha} \in \mathbb{N}^n}$ is the real symmetric matrix $\mathbf{M}_t(\boldsymbol{\phi})$ with rows and columns

indexed by \mathbb{N}_t^n , and with entries

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

If ϕ has a representing measure ϕ then necessarily $\mathbf{M}_t(\phi)$ is positive semidefinite (denoted $\mathbf{M}_t(\phi) \succeq 0$ or $\mathbf{M}_t(\phi) \succeq 0$) for all t . But the converse is not true in general.

Localizing matrix Similarly, with $t \in \mathbb{N}$, the localizing matrix $\mathbf{M}_t(g \cdot \phi)$ associated with a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ and a polynomial $\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{\gamma} g_\gamma \mathbf{x}^\gamma$, is the real symmetric matrix $\mathbf{M}_t(g \cdot \phi)$ with rows and columns indexed by \mathbb{N}_t^n , and with entries

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

Equivalently, $\mathbf{M}_t(g \cdot \phi)$ is the moment matrix of the sequence $g \cdot \phi$.

Orthonormal polynomials. Let $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ be a real sequence such that $\mathbf{M}_t(\phi)$ is positive definite (denoted $\mathbf{M}_t(\phi) \succ 0$) for all t . Then with ϕ one may associate a family of *orthonormal polynomials* $(P_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$, i.e., which satisfy:

$$(2.1) \quad \phi(P_\alpha \cdot P_\beta) = \delta_{\alpha=\beta}, \quad \forall \alpha, \beta \in \mathbb{N}^n,$$

where δ_\bullet is the Kronecker symbol. One way to obtain the P_α 's is via certain determinants formed from entries of $\mathbf{M}_t(\phi)$. For instance, in dimension $n = 1$, $P_0 = 1$ and

$$P_1 = \tau_1 \cdot \det \begin{bmatrix} \phi_0 & \phi_1 \\ 1 & x \end{bmatrix} \quad P_2 = \tau_2 \cdot \det \begin{bmatrix} \phi_0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_2 & \phi_3 \\ 1 & x & x^2 \end{bmatrix}, \quad \text{etc.,}$$

with τ_k being a scalar that ensures $\phi(P_k^2) = 1$, $k \in \mathbb{N}$. For more details the interested reader is referred to e.g. [5, 7].

Putinar's Positivstellensatz. Let $g_0 := 1$ and $G := \{g_0, g_1, \dots, g_m\} \subset \mathbb{R}[\mathbf{x}]$ with $t_g := \lceil \deg(g)/2 \rceil$ for all $g \in G$. Let

$$(2.2) \quad \mathbf{S} := \{ \mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \geq 0, \quad \forall g \in G \},$$

and define the sets

$$(2.3) \quad Q(G) = \left\{ \sum_{g \in G} \sigma_g g; \quad \sigma_g \in \Sigma[\mathbf{x}], \forall g \in G \right\}$$

$$(2.4) \quad Q_t(G) = \left\{ \sum_{g \in G} \sigma_g g; \quad \deg(\sigma_g g) \leq 2t, \forall g \in G \right\},$$

called respectively the *quadratic module* and the *t-truncated quadratic module* associated with G .

Remark 2.1. With $R > 0$, let $\mathbf{x} \mapsto \theta(\mathbf{x}) := R - \|\mathbf{x}\|^2$. The quadratic module $Q(G) \subset \mathbb{R}[\mathbf{x}]$ is said to be Archimedean if there exists $R > 0$ such that $\theta \in Q(G)$, in which case it provides an algebraic certificate that the set \mathbf{S} in (2.2) is compact. If one knows that $\mathbf{S} \subset \{ \mathbf{x} : \|\mathbf{x}\|^2 \leq R \}$ for some R then it is a good idea to include the additional (but redundant) constraint $R - \|\mathbf{x}\|^2 \geq 0$ in the definition (2.2) of \mathbf{S} , in which case the resulting associated quadratic module $Q(G)$ is Archimedean.

Theorem 2.2 (Putinar [27]). *Let \mathbf{S} be as in (2.2) and let $Q(G)$ be Archimedean.*

(i) *If $p \in \mathbb{R}[\mathbf{x}]$ is strictly positive on \mathbf{S} then $p \in Q(G)$.*

(ii) *A real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ has a representing Borel measure on \mathbf{S} if and only if $\mathbf{M}_t(g \cdot \phi) \succeq 0$ for all $t \in \mathbb{N}$, and all $g \in G$.*

Theorem 2.2 is central to prove convergence of the Moment-SOS hierarchy of lower bounds on f^* , described in Section 3.1.

Another Positivstellensatz We next provide an alternative Positivstellensatz where the compact set \mathbf{S} is not required to be semi-algebraic. Given a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$, define the convex cones

$$(2.5) \quad C_{t,s}^\phi := \{g \in \mathbb{R}[\mathbf{x}]_t : \mathbf{M}_s(g \cdot \phi) \succeq 0\}, \quad t, s \in \mathbb{N}.$$

Let t be fixed. Observe that for each s , the convex cone $C_{t,s}^\phi$ is defined in terms of the single linear matrix inequality $\mathbf{M}_s(g \cdot \phi) \succeq 0$, on the coefficients (g_α) of $g \in \mathbb{R}[\mathbf{x}]_t$. It defines a spectrahedron in the space $\mathbb{R}^{s(t)}$ of the coefficient vector $\mathbf{g} = (g_\alpha) \in \mathbb{R}^{s(t)}$ of $g \in \mathbb{R}[\mathbf{x}]_t$ (recall that $s(t) = \binom{n+t}{n}$). It is a closed convex cone.

Theorem 2.3 ([9]). *Let $\mathbf{S} \subset \mathbb{R}^n$ be a compact set and let ϕ be an arbitrary finite Borel measure on \mathbb{R}^n whose support is \mathbf{S} and with moments $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$. Then $g \in \mathbb{R}[\mathbf{x}]$ is nonnegative on \mathbf{S} if and only if $\mathbf{M}_s(g \cdot \phi) \succeq 0$ for all $s \in \mathbb{N}$.*

Theorem 2.3 is central to prove the convergence of the Moment-SOS hierarchy of upper bounds on f^* , described in Section 3.2. With t fixed, $(C_{t,s}^\phi)_{s \in \mathbb{N}}$ provides a monotone non increasing sequence of convex cones $C_{t,s}^\phi$, each being an outer approximation of the convex cone $\mathcal{C}_t(\mathbf{S})_+$ of polynomials of degree at most t , nonnegative on $\mathbf{S} = \text{supp}(\phi)$.

In addition, $\bigcap_{s=0}^\infty C_{t,s}^\phi = \mathcal{C}_t(\mathbf{S})_+$. Indeed if $g \in C_{t,s}^\phi$ for all s then by Theorem 2.3, $g \in \mathcal{C}_t(\mathbf{S})_+$. Conversely, if $g \in \mathcal{C}_t(\mathbf{S})_+$ then $\int_{\mathbf{S}} p^2 g d\phi \geq 0$, for all $p \in \mathbb{R}[\mathbf{x}]_s$, that is, $\mathbf{M}_s(g \cdot \phi) \succeq 0$, and as s was arbitrary, $g \in C_{t,s}^\phi$ for all s .

Notice that Theorem 2.3 is a Nichtnegativstellensatz and applies to sets with are not necessarily semi-algebraic. However, if on the one hand the set \mathbf{S} is *not* required to be semi-algebraic, on the other hand one needs to know the moment sequence ϕ to exploit numerically the convex cone $C_{t,s}^\phi$. In addition, the set \mathbf{S} may also be non-compact. It is then enough to take a reference measure ϕ on \mathbf{S} such that $\sup_i \int e^{|x_i|} d\phi < M$ for some $M > 0$; see e.g. [9, 10]. In particular, one may then approximate from above the convex cone $\mathcal{C}_t(\mathbb{R}^n)_+$ (resp. $\mathcal{C}(\mathbb{R}_+^n)_+$) of polynomials nonnegative on the whole \mathbb{R}^n (resp. \mathbb{R}_+^n). (Just take $\phi = \exp(-\|\mathbf{x}\|^2) d\mathbf{x}$ on \mathbb{R}^n (resp. $\phi = \exp(-2 \sum_i x_i) d\mathbf{x}$ on \mathbb{R}_+^n).

3. THE MOMENT-SOS HIERARCHY IN POLYNOMIAL OPTIMIZATION

Consider the optimization problem \mathbf{P} in (1.1) where $f \in \mathbb{R}[\mathbf{x}]$, \mathbf{S} is the basic semi-algebraic set described in (2.2), and f^* in (1.1) is the *global* minimum of \mathbf{P} .

3.1. A Moment-SOS hierarchy of lower bounds. Assumption 1: The set \mathbf{S} in (2.2) is compact and contained in the Euclidean ball of radius \sqrt{R} . Therefore with no loss of generality we may and will assume that the quadratic polynomial $\mathbf{x} \mapsto \theta(\mathbf{x}) := R - \|\mathbf{x}\|^2$ is in G . Technically this implies that the quadratic module $Q(G)$ is Archimedean; see Remark 2.1.

For every $g \in \mathbb{R}[\mathbf{x}]$, let $t_g := \lceil \deg(g)/2 \rceil$. Define $t_0 := \max[t_f, \max_{g \in G} t_g]$, and consider the sequence of semidefinite programs indexed by $t \in \mathbb{N}$:

$$(3.1) \quad \rho_t = \inf_{\phi \in \mathbb{R}^{s(2t)}} \{ \phi(f) : \phi(1) = 1; \mathbf{M}_{t-t_g}(g \cdot \phi) \succeq 0, \forall g \in G \}, \quad t \geq t_0.$$

For each $t \geq t_0$, (3.1) is a semidefinite program and a convex relaxation of (1.1) so that $\rho_t \leq f^*$ for all $t \geq t_0$. In addition, the sequence $(\rho_t)_{t \geq t_0}$ is monotone non decreasing. The dual of (3.1) reads:

$$(3.2) \quad \rho_t^* = \sup_{\sigma_g, \lambda} \{ \lambda : f - \lambda = \sum_{g \in G} \sigma_g g; \quad \sigma_g \in \Sigma[\mathbf{x}]_{t-t_g}, \forall g \in G \}.$$

By weak duality between (3.1) and (3.2), $\rho_t^* \leq \rho_t$ for all $t \geq t_0$ and in fact, under Assumption 1, there is no duality gap, i.e., $\rho_t^* = \rho_t$ for all $t \geq t_0$; see e.g. [11, 17].

KKT-optimality conditions. In the context of problem \mathbf{P} in (1.1) with feasible set \mathbf{S} as in (2.2), for $\mathbf{x} \in \mathbf{S}$, let $J(\mathbf{x}) := \{g \in G : g(\mathbf{x}) = 0\}$ identify the set of constraints that are *active* at \mathbf{x} . Let $\mathbf{x}^* \in \mathbf{S}$, and define

$$(3.3) \quad \text{CQ}(\mathbf{x}^*) : \quad \text{the vectors } (\nabla g(\mathbf{x}^*))_{g \in J(\mathbf{x}^*)} \text{ are linearly independent.}$$

In non linear programming, the celebrated first-order necessary Karush-Kuhn-Tucker (KKT) optimality conditions state that if $\mathbf{x}^* \in \mathbf{S}$ is a local minimizer for \mathbf{P} and $\text{CQ}(\mathbf{x}^*)$ holds, then there exists $\boldsymbol{\lambda}^* = (\lambda_g^*)_{g \in G} \subset \mathbb{R}_+$ such that

$$\nabla f(\mathbf{x}^*) - \sum_{g \in G} \lambda_g^* \nabla g(\mathbf{x}^*) = 0; \quad \lambda_g^* g(\mathbf{x}^*) = 0, \forall g \in G.$$

In addition if $\lambda_g^* > 0$ whenever $g(\mathbf{x}^*) = 0$, then *strict complementarity* is said to hold. Finally, the second-order sufficient optimality condition holds at \mathbf{x}^* if

$$\mathbf{u}^T \left(\nabla^2 f(\mathbf{x}^*) - \sum_{g \in G} \lambda_g^* \nabla^2 g(\mathbf{x}^*) \right) \mathbf{u} > 0, \quad \forall \mathbf{u} (\neq 0) \in \nabla G(\mathbf{x}^*)^\perp,$$

where $\nabla G(\mathbf{x}^*)^\perp := \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u}^T \nabla g(\mathbf{x}^*) = 0, \forall g \in J(\mathbf{x}^*)\}$, and $\nabla^2 h(\mathbf{x}^*)$ denotes the Hessian of h evaluated at \mathbf{x}^* .

Theorem 3.1. *Let Assumption 1 hold with \mathbf{S} as in (2.2), and consider the semidefinite program (3.1) and its dual (3.2).*

(i) $\rho_t^* = \rho_t$ for all $t \geq t_0$. Moreover (3.1) has an optimal solution ϕ^* for every $t \geq t_0$, and if \mathbf{S} has a nonempty interior then (3.2) also has an optimal solution $(\sigma_g^*)_{g \in G}$.

(ii) As t increases, $\rho_t \uparrow f^*$ and finite convergence takes place if (3.3), strict complementarity, and second-order sufficiency condition, hold at every global minimizer of \mathbf{P} (a condition that holds true generically).

(iii) Let $s := \max_{g \in G} t_g$. If $\text{rank}(\mathbf{M}_t(\phi^*)) = \text{rank}(\mathbf{M}_{t-s}(\phi^*))$ for some t , then $\rho_t = f^*$ (i.e., finite convergence takes place) and from ϕ^* one may extract $\text{rank}(\mathbf{M}_t(\phi^*))$ global minimizers of \mathbf{P} via a linear algebra subroutine.

In Theorem 3.1(iii), the (*flatness*) condition on the ranks of $\mathbf{M}_t(\phi^*)$ and $\mathbf{M}_{t-s}(\phi^*)$, also holds generically (e.g. if the second-order sufficiency condition holds at every global minimizer); see e.g. [1, 2, 25]. In the recent work [2], the authors have provided the first degree-bound on the SOS weights in Putinar's positivity certificate $f \in Q(G)$, with a *polynomial* dependence on the degree of f and a constant related

to how far is f from having a zero in \mathbf{S} . (The previous known bound of [24] has an exponential dependence.)

As stated in (3.1), the standard Moment-SOS hierarchy does not scale well with the dimension. This is because it involves $s(2t)$ moment variables ϕ_α and semidefinite matrices of size $s(t)$. Fortunately, for large-scale polynomial optimization problems, sparsity and/or symmetries are often encountered and can be exploited to obtain alternative hierarchies with much better scaling properties. The interested reader is referred to the recent book [20] and the many references therein where various such techniques are described and illustrated. Also in [23] are described first-order methods that exploit a *constant trace property* of matrices of the semidefinite program (3.1); they can provide an alternative to costly interior point methods for solving large-scale semidefinite relaxations.

3.2. A Moment-SOS hierarchy of upper bounds. In this section we now consider a hierarchy of upper bounds on the global minimum f^* of \mathbf{P} in (1.1) and where $\mathbf{S} \subset \mathbb{R}^n$ is a compact set with nonempty interior. Let μ be a probability measure with support \mathbf{S} and with associated sequence of moments $\boldsymbol{\mu} = (\mu_\alpha)_{\alpha \in \mathbb{N}^n}$. Consider the sequence of optimization problems indexed by $t \in \mathbb{N}$:

$$(3.4) \quad \tau_t = \min_{\sigma \in \Sigma[\mathbf{x}]_t} \left\{ \int_{\mathbf{S}} f \sigma d\mu : \int_{\mathbf{S}} \sigma d\mu = 1 \right\}$$

$$(3.5) \quad \tau_t^* = \sup_{\lambda} \{ \lambda : \mathbf{M}_t(f \cdot \boldsymbol{\mu}) \succeq \lambda \mathbf{M}_t(\boldsymbol{\mu}) \}$$

It is straightforward to see that $\tau_t \geq f^*$ for all t . Indeed let $\sigma \in \Sigma[\mathbf{x}]_t$ be a feasible solution of (3.4). Then as $f \geq f^*$ for all $\mathbf{x} \in \mathbf{S}$,

$$\int_{\mathbf{S}} f \sigma d\mu \geq f^* \int_{\mathbf{S}} \sigma d\mu = f^*.$$

Moreover, $\tau_t^* \leq \tau_t$ for every t because from the definition of the localizing and moment matrices associated with $\boldsymbol{\mu}$ and f ,

$$\mathbf{M}_t(f \cdot \boldsymbol{\mu}) \succeq \lambda \mathbf{M}_t(\boldsymbol{\mu}) \Rightarrow \int_{\mathbf{S}} f \sigma d\mu \geq \lambda \int_{\mathbf{S}} \sigma d\mu, \quad \forall \sigma \in \Sigma[\mathbf{x}]_t,$$

which in turn implies $\lambda \leq \int_{\mathbf{S}} f \sigma d\mu$ for all σ feasible in (3.4), and therefore $\lambda \leq \tau_t$.

Theorem 3.2 ([9]). *Let $S \subset \mathbb{R}^n$ be compact with nonempty interior and τ_t and τ_t^* be as in (3.4) and (3.5) respectively. Then $\tau_t = \tau_t^*$ for every t and $\tau_t \downarrow f^*$ as t increases. Moreover (3.4) (resp. (3.5)) has an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_t$ (resp. λ^*) and λ^* is the smallest generalized eigenvalue of the pair of matrices $(\mathbf{M}_t(f \cdot \boldsymbol{\mu}), \mathbf{M}_t(\boldsymbol{\mu}))$ with associated eigenvector σ^* .*

The proof of the convergence $\tau_t \downarrow f^*$ as t increases, is based on Theorem 2.3. The dual problem (3.5) has a single variable λ and is a generalized eigenvalue problem associated with the pair of matrices $(\mathbf{M}_t(f \cdot \boldsymbol{\mu}), \mathbf{M}_t(\boldsymbol{\mu}))$. Therefore τ_t can be computed by standard linear algebra routine with no optimization. See e.g. the discussion in [9, Section 4]. However the size of the involved matrices makes this technique quite difficult even for modest size problems. Nevertheless and fortunately, there is a variant [13] that reduces to computing generalized eigenvalues of related *univariate* Hankel moment matrices by using the pushforward (univariate) measure $\#\mu$ (on the real line) of μ by f . That is, $\#\mu(B) = \mu(f^{-1}(B))$ for all

$B \in \mathcal{B}(\mathbb{R})$, and therefore

$$f^* = \inf \{ z : z \in f(\mathcal{S}) \} = \inf \{ z : z \in \text{supp}(\#\mu) \}.$$

Then letting $\mathbf{H}_t(\#\mu)$ (resp. $\mathbf{H}_t(z \cdot \#\mu)$) be the (univariate) Hankel moment matrix associated with $\#\mu$ (resp. $z \cdot \#\mu$), the sequence of scalars $(\delta_t)_{t \in \mathbb{N}}$ defined by

$$(3.6) \quad \delta_t := \sup_{\lambda} \{ \lambda : \mathbf{H}_t(z \cdot \#\mu) \succeq \lambda \mathbf{H}_t(\#\mu) \}, \quad t \in \mathbb{N},$$

provides a monotone non-increasing sequence of upper bounds $(\delta_t)_{t \in \mathbb{N}}$ that converges to f^* . For more details, the interested reader is referred to [13, 29].

When comparing (3.6) with (3.5), the gain in the computational burden is striking. Indeed in (3.6) one has to compute generalized eigenvalues of Hankel matrices of size $t+1$ instead of size $\binom{n+t}{t}$ in (3.5). Recent works in [4, 28, 29, 30] have proven nice rates for the convergence $\delta_t \downarrow f^*$ and $\tau_t \downarrow f^*$, with an appropriate choice of the reference measure μ on specific sets \mathcal{S} (e.g., sphere, box, simplex, etc.). Interestingly, the analysis makes use of sophisticated results about zeros of orthogonal polynomials, and a clever perturbation of the Christoffel-Darboux kernel.

4. THE CHRISTOFFEL-DARBOUX KERNEL AND CHRISTOFFEL FUNCTIONS

In this section we briefly review basic properties of the Christoffel-Darboux (CD) kernel and Christoffel functions. For more details on these classical tools, the interested reader is referred to e.g. [15, 16] and the many references therein.

4.1. Christoffel-Darboux kernel. Let $\mathcal{S} \subset \mathbb{R}^n$ be compact with nonempty interior and let $\mu \in \mathcal{M}(\mathcal{S})_+$ be such that $\mathbf{M}_t(\mu) \succ 0$ for all $t \in \mathbb{N}$. Let $(P_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$ be a family of polynomials that are orthonormal with respect to μ , and view $\mathbb{R}[\mathbf{x}]_t$ as a finite-dimensional vector subspace of the Hilbert space $L^2(\mathcal{S}, \mu)$. Then the kernel

$$(4.1) \quad (\mathbf{x}, \mathbf{z}) \mapsto K_t^\mu(\mathbf{x}, \mathbf{z}) := \sum_{\alpha \in \mathbb{N}_t^n} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{z}), \quad \forall \mathbf{x}, \mathbf{z} \in \mathbb{R}^n, \quad t \in \mathbb{N},$$

is called the Christoffel-Darboux (CD) kernel associated with μ . It has an important property, namely it *reproduces* $\mathbb{R}[\mathbf{x}]_t$. Indeed, for every $p \in \mathbb{R}[\mathbf{x}]_t$,

$$(4.2) \quad p(\mathbf{x}) = \int_{\mathcal{S}} K_t^\mu(\mathbf{x}, \mathbf{z}) p(\mathbf{z}) d\mu(\mathbf{z}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

and for this reason, $(\mathbb{R}[\mathbf{x}]_t, K_t^\mu)$ is called a Reproducing Kernel Hilbert Space (RKHS). Then every $f \in L^2(\mathcal{S}, \mu)$ can be approximated by a sequence of polynomials $(\hat{f}_t)_{t \in \mathbb{N}}$, where $\hat{f}_t \in \mathbb{R}[\mathbf{x}]_t$ for every $t \in \mathbb{N}$, and

$$\mathbf{x} \mapsto \hat{f}_t(\mathbf{x}) := \int_{\mathcal{S}} f(\mathbf{z}) K_t^\mu(\mathbf{x}, \mathbf{z}) d\mu(\mathbf{z}) = \sum_{\alpha \in \mathbb{N}_t^n} \left(\int_{\mathcal{S}} f(\mathbf{z}) P_\alpha(\mathbf{z}) d\mu(\mathbf{z}) \right) P_\alpha(\mathbf{x}),$$

so that $\|f - \hat{f}_t\|_{L^2(\mathcal{S}, \mu)} \rightarrow 0$ as t increases; see e.g. [15, Section 2, p. 13].

Interpreting the reproducing property. Given $\mathbf{y} \in \mathbb{R}^n$ fixed, let $p \in \mathbb{R}[\mathbf{x}]_t$ be the polynomial defined by

$$(4.3) \quad \mathbf{x} \mapsto p(\mathbf{x}) := K_t^\mu(\mathbf{y}, \mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Then by the reproducing property (4.2), observe that

$$\int_{\mathcal{S}} \mathbf{x}^\alpha p(\mathbf{x}) d\mu(\mathbf{x}) = \mathbf{y}^\alpha = \int \mathbf{x}^\alpha \delta_{\{\mathbf{y}\}}(d\mathbf{x}), \quad \forall \alpha \in \mathbb{N}_t^n,$$

that is, viewing p as a *signed* density w.r.t. μ , the signed measure $p d\mu$ on \mathcal{S} , mimics the Dirac measure at \mathbf{y} , as long as only moments of order at most t are concerned. This is illustrated in Figure 1 where $\mathcal{S} = [-1, 1]$ and $d\mu = 1_{[-1,1]}(x)dx$, t varies between 1 and 10, and $y = 0, 1/2, 1$.

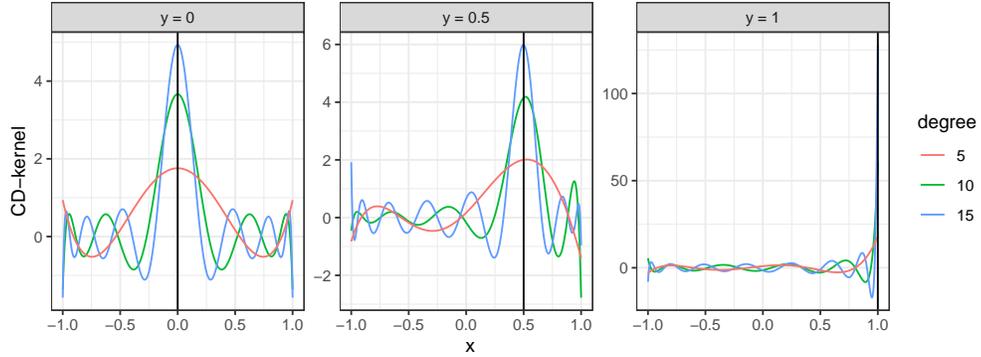


FIGURE 1. The signed measure $K_t^\mu(y, x)dx$ (with different values of t) mimics the Dirac measure at $y = 0$ (left) $y = 0.5$ (middle) and $y = 1$ (right); reprinted from [15, p. 45] with permission. ©Cambridge University Press

4.2. Christoffel function. With $t \in \mathbb{N}$, the function $\Lambda_t^\mu : \mathbb{R}^n \rightarrow \mathbb{R}_+$ associated with μ , and defined by

$$(4.4) \quad \mathbf{x} \mapsto \Lambda_t^\mu(\mathbf{x}) := K_t^\mu(\mathbf{x}, \mathbf{x})^{-1} = \left[\sum_{\alpha \in \mathbb{N}_t^n} P_\alpha(\mathbf{x})^2 \right]^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

is called the (degree- t) Christoffel function (CF), and recalling that $\mathbf{M}_t(\mu)$ is non-singular, it also turns out that

$$(4.5) \quad \Lambda_t^\mu(\mathbf{x}) = [\mathbf{v}_t(\mathbf{x})^T \mathbf{M}_t(\mu)^{-1} \mathbf{v}_t(\mathbf{x})]^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

The CF also has an equivalent and variational definition, namely:

$$(4.6) \quad \Lambda_t^\mu(\mathbf{x}) = \inf_{p \in \mathbb{R}[\mathbf{x}]_t} \left\{ \int_{\mathcal{S}} p^2 d\mu : p(\mathbf{x}) = 1 \right\}, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

$$(4.7) \quad = \inf_{\mathbf{p} \in \mathbb{R}^{s(t)}} \left\{ \langle \mathbf{p}, \mathbf{M}_t(\mu) \mathbf{p} \rangle : \langle \mathbf{p}, \mathbf{v}_t(\mathbf{x}) \rangle = 1 \right\}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

In (4.7) the reader can easily recognize a *convex quadratic* programming problem which can be solved efficiently even for large dimensions. However solving (4.7)

only provides the numerical value of Λ_t^μ at $\mathbf{x} \in \mathbb{R}^n$, whereas in (4.5) one obtains the coefficients of the polynomial $(\Lambda_t^\mu)^{-1}$ (but at the price of inverting $\mathbf{M}_t(\mu)$).

The reader will also notice that from its definitions (4.4) or (4.5), the CF depends only on the finite sequence $\boldsymbol{\mu}_{2t} = (\mu_\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$ of moments of μ , up to degree $2t$, and not on μ itself. Indeed there are potentially many measures on \mathbf{S} with same moments up to degree $2t$, and therefore indexing Λ_t^μ with μ is not totally correct; therefore a more correct labelling would be $\Lambda_t^{\boldsymbol{\mu}_{2t}}$. One reason for this labelling is that in theory of approximation, one is usually given a measure μ on a compact set \mathbf{S} and one is interested in the sequence $(\Lambda_t^\mu)_{t \in \mathbb{N}}$ and its asymptotic properties.

Remark 4.1. *In fact, one may also define the CD-kernel K_t^ϕ and the Christoffel function (CF) Λ_t^ϕ associated with a Riesz linear functional $\phi \in \mathbb{R}[\mathbf{x}]^*$ whose associated sequence ϕ is such that $\mathbf{M}_t(\phi) \succ 0$, no matter if ϕ is a measure on \mathbf{S} or not. Indeed for fixed t , and letting $(P_\alpha)_{\alpha \in \mathbb{N}_t^n}$ be orthonormal w.r.t. ϕ , the polynomial*

$$(\mathbf{x}, \mathbf{y}) \mapsto K_t^\phi(\mathbf{x}, \mathbf{y}) := \sum_{\alpha \in \mathbb{N}_t^n} P_\alpha(\mathbf{x}) P_\alpha(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

is well-defined, and all definitions (4.1)-(4.7) are still valid. But again, historically the CD-kernel was defined w.r.t. a given measure μ on \mathbf{S} . Finally, one may use interchangeably the notations K_t^ϕ (resp. Λ_t^ϕ) or K_t^μ (resp. Λ_t^μ), or $K_t^{\boldsymbol{\mu}_{2t}}$ (resp. $\Lambda_t^{\boldsymbol{\mu}_{2t}}$) as in all cases, the resulting mathematical object depends only on the finite moment sequence $\boldsymbol{\phi}_{2t} = (\phi_\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$ of ϕ .

4.3. Some distinguishing properties of the CF. The CF Λ_t^ϕ associated with a Borel measure ϕ on a compact $\mathbf{S} \subset \mathbb{R}^n$, has an interesting and distinguishing feature. As t increases, $\Lambda_t^\phi(\mathbf{x}) \downarrow 0$ exponentially fast for every $\mathbf{x} \notin \mathbf{S}$ whereas its decrease is at most polynomial in t whenever $\mathbf{x} \in \mathbf{S}$; see e.g. [15, Section 4.3, p. 50–51]. In other words, Λ_t^ϕ identifies the support of ϕ when t is sufficiently large. In addition, at least in dimension $n = 2$ or $n = 3$, one may visualize this property even for small t , as the resulting superlevel sets $\mathbf{S}_\gamma := \{\mathbf{x} : \Lambda_t^\phi(\mathbf{x}) \geq \gamma\}$, $\gamma \in \mathbb{R}$, capture the geometric shape of \mathbf{S} quite accurately; For instance in Figure 2 are displayed several level sets \mathbf{S}_γ associated with the empirical measure ϕ_N supported on a cloud of N points that approximates the geometric shape obtained with the letters “C” and “D” of Christoffel and Darboux. In [12], the interested reader can find many other examples of $2D$ -clouds with non-trivial geometric shapes which are captured quite well with levels set \mathbf{S}_γ associated with ϕ_N , even for relatively low degree t .

Another nice feature of the CF is its ability to approximate densities. Indeed let ϕ and μ be finite Borel measures on a compact set \mathbf{S} , and let μ be such that uniformly on compact subsets of $\text{int}(\mathbf{S})$, $\lim_{t \rightarrow \infty} s(t) \Lambda_t^\mu = h_\mu$, where h_μ is continuous and positive on $\text{int}(\mathbf{S})$ (and recall that $s(t)$ is the dimension of $\mathbb{R}[\mathbf{x}]_t$). In addition suppose that ϕ has continuous and positive density f_ϕ w.r.t. μ . Then uniformly on compact subsets of $\text{int}(\mathbf{S})$

$$\lim_{t \rightarrow \infty} s(t) \Lambda_t^\phi = f_\phi \cdot h_\mu$$

(see e.g. [15, Theorem 4.4.1]). So if the function h_μ is already known then one can approximate the density f_ϕ , uniformly on compact subsets of $\text{int}(\mathbf{S})$.

Finally, another distinguishing property of the CF is its link with the so-called *equilibrium measure* of the compact set \mathbf{S} . The latter is a measure on \mathbf{S} (let us

the reproducing property (4.2), observe that

$$\begin{aligned}
\phi(f) &= \phi \left(\int_{\mathcal{S}} \sum_{\alpha \in \mathbb{N}_{2t}^n} P_{\alpha}(\mathbf{x}) P_{\alpha}(\mathbf{y}) f(\mathbf{y}) d\mu(\mathbf{y}) \right) \\
&= \sum_{\alpha \in \mathbb{N}_{2t}^n} \phi(P_{\alpha}) \int_{\mathcal{S}} P_{\alpha}(\mathbf{y}) f(\mathbf{y}) d\mu(\mathbf{y}) \\
&= \int_{\mathcal{S}} f(\mathbf{y}) \sum_{\alpha \in \mathbb{N}_{2t}^n} \phi(P_{\alpha}) P_{\alpha}(\mathbf{y}) d\mu(\mathbf{y}) = \int_{\mathcal{S}} f(\mathbf{y}) \sigma_{\phi}(\mathbf{y}) d\mu(\mathbf{y})
\end{aligned}$$

where the degree- $2t$ polynomial $\mathbf{y} \mapsto \sigma_{\phi}(\mathbf{y}) := \sum_{\alpha \in \mathbb{N}_{2t}^n} \phi(P_{\alpha}) P_{\alpha}(\mathbf{y})$, is a *signed density* w.r.t. μ .

Therefore in the semidefinite relaxations (3.1) of lower bounds on f^* , one searches for a linear functional $\phi \in \mathbb{R}[\mathbf{x}]_{2t}^*$ which satisfies

$$\phi(1) = 1; \quad \mathbf{M}_t(g \cdot \phi) \succeq 0, \quad \forall g \in G,$$

and which minimizes $\phi(f) = \int_{\mathcal{S}} f \sigma_{\phi} d\mu$, where σ_{ϕ} is a degree- $2t$ polynomial signed density w.r.t. μ , with coefficients $(\sigma_{\phi, \alpha} := \phi(P_{\alpha}))_{\alpha \in \mathbb{N}_{2t}^n}$.

The reason why the semidefinite relaxations (3.1) can be exact (i.e., $\rho_t = f^*$ for some t), is that the signed probability measure $d\nu_t = \sigma_{\phi} d\mu$ can mimic the Dirac measure at a global minimizer $\boldsymbol{\xi} \in \mathcal{S}$ and so $\phi(f) = f(\boldsymbol{\xi})$; see Figure 1.

This is in contrast to the hierarchy of semidefinite relaxations (3.4) of upper bounds where one searches also for a polynomial density $\sigma d\mu$ w.r.t. μ , but as this density σ is an SOS (hence positive), it cannot be a Dirac measure, and therefore the resulting convergence $\tau_t \downarrow f^*$ is necessarily asymptotic and *not* finite. For more details on a comparison between the Moment-SOS hierarchies of upper and lower bounds, the interested reader is referred to [18].

5.2. The CF and positive polynomials. Of course, from its definition (4.5) the reciprocal $(\Lambda_t^{\mu})^{-1}$ of the CF is an SOS of degree $2t$. But we next reveal an even more interesting link with SOS polynomials. Observe that the $Q_t(G)$ in (2.4) is a convex cone and its dual reads

$$(5.1) \quad Q_t(G)^* = \{ \phi = (\phi_{\alpha})_{\alpha \in \mathbb{N}_{2t}^n} : \mathbf{M}_{t-t_g}(g \cdot \phi) \succeq 0, \quad \forall g \in G \}.$$

A duality result of Nesterov.

Lemma 5.1. *If $p \in \text{int}(Q_t(G))$ then there exists $\phi \in \text{int}(Q_t(G)^*)$ such that*

$$(5.2) \quad p = \sum_{g \in G} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{M}_{t-t_g}(g \cdot \phi)^{-1} \mathbf{v}_{t-t_g}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$$

$$(5.3) \quad = \sum_{g \in G} g \cdot (\Lambda_t^{g \cdot \phi})^{-1}.$$

In particular, for every SOS $p \in \text{int}(\Sigma_t[\mathbf{x}])$, $1/p$ is the CF of some linear functional $\phi \in \Sigma[\mathbf{x}]_{2t}^$, i.e., $1/p = \Lambda_t^\phi$ for some $\phi \in \mathbb{R}^{s(2t)}$ such that $\mathbf{M}_t(\phi) \succ 0$. In addition, in the univariate case, ϕ has a representing measure on \mathbb{R} .*

Equation (5.2) is from [22] while its interpretation (5.3) is from [14, Lemma 4]. Observe that (5.3) provides a distinguished representation of $p \in \text{int}(Q_t(G))$, and in view of its specific form, we propose to name (5.3) the *Christoffel representation* of $p \in \text{int}(Q_t(G))$, that is:

$$(5.4) \quad \text{int}(Q_t(G)) = \left\{ \sum_{g \in G} g \cdot (\Lambda_t^{g \cdot \phi})^{-1} : \phi \in \text{int}(Q_t(G)^*) \right\}.$$

Of course, an intriguing question is: *What is the link between $\phi \in \text{int}(Q_t(G)^*)$ in (5.3) and the polynomial $p \in \text{int}(Q_t(G))$?* A partial answer is provided in Section 5.4.

A numerical procedure to obtain the Christoffel representation. Consider the following optimization problems:

$$(5.5) \quad \mathbf{P} : \quad \inf_{\phi \in \mathbb{R}^{s(2t)}} \left\{ - \sum_{g \in G} \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi)) : \phi(p) = \sum_{g \in G} s(t-t_g) \right. \\ \left. \mathbf{M}_{t-t_g}(g \cdot \phi) \succeq 0, \quad \forall g \in G \right\}.$$

$$(5.6) \quad \mathbf{P}^* : \quad \sup_{\mathbf{Q}_g \succeq 0} \left\{ \sum_{g \in G} \log \det(\mathbf{Q}_g) : \right. \\ \left. p(\mathbf{x}) = \sum_{g \in G} g(\mathbf{x}) \mathbf{v}_{t-t_g}(\mathbf{x})^T \mathbf{Q}_g \mathbf{v}_{t-t_g}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \right\}.$$

Both \mathbf{P} and \mathbf{P}^* are convex optimization problems that can be solved by off-the-shelf software packages like e.g. CVX [6].

Theorem 5.2. *Let $p \in \text{int}(Q_t(G))$. Then \mathbf{P}^* is a dual of \mathbf{P} , that is, for every feasible solution $\phi \in \mathbb{R}^{s(2t)}$ of (5.5) and $(\mathbf{Q}_g)_{g \in G}$ of (5.6),*

$$(5.7) \quad \sum_{g \in G} \log \det(\mathbf{Q}_g) \leq - \sum_{g \in G} \log \det(\mathbf{M}_{t-t_g}(g \cdot \phi)).$$

Moreover, both \mathbf{P} and \mathbf{P}^ have a unique optimal solution ϕ^* and $(\mathbf{Q}_g^*)_{g \in G}$ respectively, which satisfy*

$$(5.8) \quad \mathbf{Q}_g^* = \mathbf{M}_{t-t_g}(g \cdot \phi^*)^{-1}, \quad \forall g \in G,$$

and which yields equality in (5.7).

The proof which uses Lemma 6.1 in Appendix, mimics that of [14, Theorem 3] (where p was a constant polynomial) and is omitted.

5.3. A disintegration of the CF. We next see how the above duality result, i.e., the Christoffel representation (5.3) of $\text{int}(Q_t(G))$, can be used to in turn infer a *disintegration property* of the CF. So let $\Lambda_t^\mu(\mathbf{x}, y)$ be the CF of a Borel probability measure μ on $\mathbf{S} \times Y$, where $\mathbf{S} \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}$ are compact. It is well-known that μ disintegrates into its marginal probability ϕ on \mathbf{S} , and a conditional measure $\hat{\mu}(dy|\mathbf{x})$ on Y , given $\mathbf{x} \in \mathbf{S}$, that is,

$$\mu(A \times B) = \int_{\mathbf{S} \cap A} \hat{\mu}(B|\mathbf{x}) \phi(d\mathbf{x}), \quad \forall A \in \mathcal{B}(\mathbf{S}), B \in \mathcal{B}(Y).$$

Theorem 5.3 ([14]). *Let $\mathbf{S} \subset \mathbb{R}^n$ (resp. $Y \subset \mathbb{R}$) be compact with nonempty interior, and let μ be a Borel probability measure on $\mathbf{S} \times Y$, with marginal ϕ on \mathbf{S} . Then for every $t \in \mathbb{N}$, and $\mathbf{x} \in \mathbf{S}$, there exists a probability measure $\nu_{\mathbf{x},t}$ on \mathbb{R} such that*

$$(5.9) \quad \Lambda_t^\mu(\mathbf{x}, y) = \Lambda_t^\phi(\mathbf{x}) \cdot \Lambda_t^{\nu_{\mathbf{x},t}}(y), \quad \forall \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R}.$$

The proof in [14, Theorem 5] heavily relies on the duality result of Lemma 5.1. In particular every degree- $2t$ univariate SOS p in the interior of $\Sigma[y]_t$ is the reciprocal of the Christoffel function of some Borel measure on \mathbb{R} .

5.4. Positive polynomials and equilibrium measure. This section is motivated by the following observation. Let $(T_n)_{n \in \mathbb{N}}$ (resp. $(U_n)_{n \in \mathbb{N}}$) be the family of Chebyshev polynomials of the first kind (resp. second kind). They are orthogonal w.r.t. measures $(1-x^2)^{-1/2}dx$ and $(1-x^2)^{1/2}dx$ on $[-1, 1]$, respectively. (The Chebyshev measure $(1-x^2)^{-1/2}dx/\pi$ is the *equilibrium measure* of the interval $[-1, 1]$.) They also satisfy the identity

$$T_n(x) + (1-x^2)U_n(x) = 1, \quad \forall x \in \mathbb{R}, \quad n = 1, \dots$$

Equivalently, it is said that the triple $(T_n, (1-x^2), U_n)$ is a solution to (polynomial) Pell's equation for every $n \geq 1$. For more details on polynomial Pell's equation (originally Pell's equation is a topic in algebraic number theory), the interested reader is referred to [21, 31]. Next, letting $x \mapsto g(x) := (1-x^2)$, and after normalization to pass to orthonormal polynomials, in summing up one obtains

$$(5.10) \quad \Lambda_t^\phi(x)^{-1} + (1-x^2)\Lambda^{g \cdot \phi}(x)^{-1} = 2t+1, \quad \forall x \in \mathbb{R}, \forall t = 0, 1, \dots$$

Now, invoking Lemma 5.1, observe that (5.10) also states that $p \in \text{int}(Q_t(G))$ where $G = \{g\}$ and p is the constant polynomial $x \mapsto p(x) = 2t+1$. In addition, it also means that if one solves \mathbf{P} in (5.5) with $p = s(t) + s(t-t_g) = 2t+1$ (recall that $t_g = 1$), then its unique optimal solution ϕ is just the vector of moments (up to degree $2t$) of the equilibrium measure $\phi = (1-x^2)^{-1/2}dx/\pi$ of the interval $\mathbf{S} = [-1, 1]$. So in Lemma 5.1 the linear functional ϕ_p associated with the constant polynomial $p = 2t+1$ is simply the equilibrium measure of \mathbf{S} (denote it $\lambda_{\mathbf{S}}$).

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) and some generalizations have been obtained in the multivariate case via pluripotential theory in \mathbb{C}^n . In particular if $\mathbf{S} \subset \mathbb{R}^n \subset \mathbb{C}^n$ is compact then the equilibrium measure $\lambda_{\mathbf{S}}$ is equivalent to Lebesgue measure on compact subsets of $\text{int}(\mathbf{S})$. See e.g. Bedford and Taylor [3, Theorem 1.1] and [3, Theorem 1.2].

The Bernstein-Markov property. A measure with compact support \mathcal{S} satisfies the Bernstein-Markov property if there exists a sequence of positive numbers $(M_t)_{t \in \mathbb{N}}$ such that for all $t \in \mathbb{N}$ and all $p \in \mathbb{R}[\mathbf{x}]_t$,

$$\sup_{\mathbf{x} \in \mathcal{S}} |p(\mathbf{x})| (= \|p\|_{\mathcal{S}}) \leq M_t \cdot \|p\|_{L^2(\mathcal{S}, \mu)},$$

and $\lim_{t \rightarrow \infty} \log(M_t)/t = 0$.

So when it holds, the Bernstein-Markov property describes how the sup-norm and the $L^2(\mathcal{S}, \mu)$ -norm of polynomials relate when the degree increases.

In [19] we have obtained the following result. Let $\mathbf{x} \mapsto \theta(\mathbf{x}) := 1 - \|\mathbf{x}\|^2$ and possibly after an appropriate scaling, let \mathcal{S} in (2.2) be such that $\theta \in Q_1(G)$ (so that $\mathcal{S} \subset [-1, 1]^n$); see Remark 2.1.

Theorem 5.4 ([19]). *Let $\phi \in \mathbb{R}[\mathbf{x}]^*$ (with $\phi_0 = 1$) be such that $\mathbf{M}_t(g \cdot \phi) \succ 0$ for all $t \in \mathbb{N}$ and all $g \in G$, so that the Christoffel functions $\Lambda_t^{g \cdot \phi}$ are all well defined. In addition, suppose that there exists $t_0 \in \mathbb{N}$ such that*

$$(5.11) \quad \sum_{g \in G_t} s(t - t_g) = \sum_{g \in G_t} g \cdot (\Lambda_{t-t_g}^{g \cdot \phi})^{-1}, \quad \forall t \geq t_0.$$

Then: (a) for every $t \geq t_0$, the finite moment sequence $\phi_t^ := (\phi_{\alpha})_{\alpha \in \mathbb{N}_{2t}^n}$ is the unique optimal solution of (5.5) (with p the constant polynomial $\mathbf{x} \mapsto \sum_{g \in G} s(t - t_g)$).*

(b) ϕ is a Borel measure on \mathcal{S} and the unique representing measure of ϕ . Moreover, if $(\mathcal{S}, g \cdot \phi)$ satisfies the Bernstein-Markov property for every $g \in G$, then ϕ is the equilibrium measure $\lambda_{\mathcal{S}}$ and therefore the Christoffel polynomials $(\Lambda_t^{g \cdot \lambda_{\mathcal{S}}})_{g \in G_t}^{-1}$ satisfy the generalized Pell's equations:

$$(5.12) \quad \sum_{g \in G_t} s(t - t_g) = \sum_{g \in G_t} g \cdot (\Lambda_{t-t_g}^{g \cdot \lambda_{\mathcal{S}}})^{-1}, \quad \forall t \geq t_0.$$

Importantly, the representation of \mathcal{S} in (2.2) depends on the chosen set G of generators, which is not unique. Therefore if (5.11) holds for some set G , it may not hold for another set G . The prototype of ϕ in Theorem 5.4 is the equilibrium measure of $\mathcal{S} = [-1, 1]$, i.e., the Chebyshev measure $(1 - x^2)^{-1/2} dx / \pi$ on $[-1, 1]$. So Theorem 5.4 is a strong result which is likely to hold only for quite specific sets \mathcal{S} (and provided that a good set of generators is used). In [19] the author could prove that (5.11) also holds for the equilibrium measure $\lambda_{\mathcal{S}}$ of the $2D$ -simplex, the $2D$ -unit unit box, the $2D$ -Euclidean unit ball, at least for $t = 1, 2, 3$.

However, if $\theta \in Q_1(G)$ then as proved in [19, 23], $1 \in \text{int}(Q_t(G))$ for all t , and therefore (5.5) has always a unique optimal solution ϕ_t^* . That is, for every $t \geq t_0$, (5.11) hold for some $\phi_t \in \mathbb{R}[\mathbf{x}]_{2t}^*$ which *depends on t* (whereas in (5.11) one considers moments up to degree $2t$ of the *same* ϕ). Moreover, every accumulation point ϕ of the sequence $(\phi_t^*)_{t \in \mathbb{N}}$ has a representing measure ϕ on \mathcal{S} . An interesting issue to investigate is the nature of ϕ , in particular its relationship with the equilibrium measure $\lambda_{\mathcal{S}}$ of \mathcal{S} .

Finally, for general compact sets \mathcal{S} with nonempty interior, to $\lambda_{\mathcal{S}}$ one may associate the polynomial

$$p_t^* := \frac{1}{\sum_{g \in G_t} s(t - t_g)} \sum_{g \in G_t} g \cdot (\Lambda_t^{g \cdot \lambda_{\mathcal{S}}})^{-1}$$

which is well-defined because the matrices $\mathbf{M}_{t-t_g}(g \cdot \lambda_{\mathcal{S}})$ are non singular. In Theorem 5.4 one has considered cases where p_t^* is exactly the constant (equal to 1) polynomial (like for the Chebyshev measure on $\mathcal{S} = [-1, 1]$). We now consider the measures $(\mu_t := p_t^* \lambda_{\mathcal{S}})_{t \in \mathbb{N}}$, with respective densities p_t^* w.r.t. $\lambda_{\mathcal{S}}$. Each μ_t is a probability measure on \mathcal{S} because

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

Moreover, preceding as in the proof of Theorem 5.4 in [19], it follows that

$$\lim_{t \rightarrow \infty} \int \mathbf{x}^\alpha p_t^* d\lambda_{\mathcal{S}} = \int \mathbf{x}^\alpha d\lambda_{\mathcal{S}}, \quad \forall \alpha \in \mathbb{N}^n.$$

As \mathcal{S} is compact it implies that the sequence of probability measures $(\mu_t)_{t \in \mathbb{N}} \subset \mathcal{M}(\mathcal{S})_+$ converges to $\lambda_{\mathcal{S}}$ for the weak- \star topology of $\mathcal{M}(\mathcal{S})$. In other words (and in an informal language), the density p_t^* of μ_t w.r.t. $\lambda_{\mathcal{S}}$ behaves like the constant (equal to 1) polynomial, which can be viewed as a weaker version of (5.12).

6. CONCLUSION

SOS polynomials play a crucial role in the Moment-SOS hierarchies of upper and lower bounds through their use in certificates of positivity of real algebraic geometry. We have shown that they are also related to the Christoffel function in theory of approximation. Interestingly, the link is provided by interpreting a duality result in convex optimization applied to a certain convex cone of polynomials and its dual cone of pseudo-moments. It also turns out that in this cone, the constant polynomial is strongly related to the equilibrium measure of the semi-algebraic set associated with the convex cone. We hope that these interactions between different and seemingly disconnected fields will raise the curiosity of the optimization community and yield further developments.

APPENDIX

Lemma 6.1. *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 $\mathbf{Q} \succ 0$). Then*

$$(6.1) \quad n + \log \det(\mathbf{M}) + \log \det(\mathbf{Q}) \leq \langle \mathbf{M}, \mathbf{Q} \rangle, \quad \forall \mathbf{M}, \mathbf{Q} \in \mathcal{S}_{++}^n.$$

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 (6.1). □

REFERENCES

- [1] Baldi, L.: Représentations Effectives en Géométrie Algébrique Réelle et Optimisation Polynomiale. Thèse de Doctorat, Université Côte d’Azur, Nice (2022).
- [2] Baldi, L., Mourrain, B.: On the effective Putinar’s Positivstellensatz and moment approximation. *Math. Program.* (2022). To appear.
- [3] Bedford, E., Taylor, B. A.: The complex equilibrium measure of a symmetric convex set in \mathbb{R}^n . *Trans. Amer. Math. Soc.* **294**, 705–717 (1986)
- [4] de Klerk, E., Laurent, M.: Convergence analysis of Lasserre hierarchy of upper bounds for polynomial optimization on the sphere. *Math. Program.* **193**, 665–685 (2022)
- [5] Dunkl, C. F., Xu, Yuan: *Orthogonal polynomials of several variables*, 2nd edition, Cambridge University Press, Cambridge, UK (2014)
- [6] Grant, M., Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 2.1. <http://cvxr.com/cvx> (2014)
- [7] Helton, J. W., Lasserre, J. B., Putinar, M.: Measures with zeros in the inverse of their moment matrix. *Annals Prob.* **36**, 1453–1471 (2008)
- [8] Henrion, D., Korda, M., Lasserre, J. B.: *The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs*. World Scientific, Singapore (2022)
- [9] Lasserre, J. B.: A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* **21**, 864–885 (2011)
- [10] Lasserre, J. B.: The K-Moment problem for continuous linear functionals. *Trans. Amer. Math. Soc.* **365**, 2489–2504 (2013)
- [11] Lasserre, J. B.: *Moments, Positive Polynomials and Their Applications*. Imperial College Press, London (2009)
- [12] Lasserre, J. B., Pauwels, E.: Sorting out typicality via the inverse moment matrix SOS polynomial. In: Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, pp. 190–198, Curran Associates, Inc., (2016)
- [13] Lasserre, J. B.: Connecting optimization with spectral analysis of tri-diagonal matrices. *Math. Program.* **190**, 795–809 (2021)
- [14] Lasserre, J. B.: A disintegration of the Christoffel function. *Comptes Rendus Mathématique* **360**, 1071–1079 (2022)
- [15] Lasserre, J. B., Pauwels, E., Putinar, M.: *The Christoffel-Darboux Kernel for Data Analysis*. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK (2022)
- [16] Lasserre, J. B., Pauwels, E.: The empirical Christoffel function with applications in data analysis. *Adv. Comput. Math.* **45**, 1439–1468 (2019)
- [17] Lasserre, J. B.: *Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press, Cambridge, UK (2015)
- [18] Lasserre, J. B.: The Moment-SOS hierarchy and the Christoffel-Darboux kernel. *Optim. Letters.* **15**, 1835–1845 (2021)
- [19] Lasserre, J. B.: Pell’s equation, sum-of-squares and equilibrium measures of compact sets. [arXiv:2210.07608](https://arxiv.org/abs/2210.07608) (2022). To appear in *Comptes Rendus Mathématique* (2023).
- [20] Magron, V., Wang, J.: *Sparse Polynomial Optimization: Theory and Practice*. World Scientific, Singapore (2023).
- [21] Mc Laughlin, J.: Multivariable-polynomial solutions to Pell’s equation and fundamental units in real quadratic fields. *Pacific J. Math.* **210**, 335–348 (2002)

- [22] Nesterov, Y.: Squared functional systems and optimization problems. In: Frenk, H., Roos, K., Terlaky, T., Zhang, Shuzong (eds.) High Performance Optimization, pp. 405–440 (2000) Applied Optimization Series vol 33, Springer, Boston MA (2000)
- [23] Ngoc Hoang Anh Mai, Lasserre, J. B., Magron, V., Wang, J.: Exploiting constant trace property in large scale polynomial optimization. ACM Trans. Math. Software (2022). To appear.
- [24] Nie, J., Schweighofer, M.: On the complexity of Putinar’s Positivstellensatz. J. Complexity **23**, 135–150 (2007)
- [25] Nie, J.: Certifying convergence of Lasserre’s hierarchy via flat truncation. Math. Program. **142**, 485–510 (2013)
- [26] Pauwels, E., Putinar, M., Lasserre, J. B.: Data analysis from empirical moments and the Christoffel function. Found. Comput. Math. **21**, 243–273 (2021)
- [27] Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**, 969–984 (1993)
- [28] Slot, L., Laurent, M.: Improved convergence analysis of Lasserre’s measure-based upper bounds for polynomial optimization on compact sets. Math. Program. **193**, 831–871 (2022)
- [29] Slot, L., Laurent, M.: Near-optimal analysis of Lasserre’s univariate measure-based bounds for multivariate polynomial optimization. Math. Program. **188**, 443–460 (2021)
- [30] Slot, L.: Sum-of-squares hierarchies for polynomial optimization and the Christoffel-Darboux kernel. SIAM J. Optim. **32**, 2612–2635 (2022)
- [31] Webb, W. A., Hisashi Yokota: Polynomial Pell’s equation. Proc. Amer. Math. Soc. **131**, 993–1006 (2002)

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE, LAAS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE, TEL: +33561336415

Email address: `lasserre@laas.fr`