

THE MOMENT-SOS HIERARCHY AND THE CHRISTOFFEL-DARBOUX KERNEL

Jean-Bernard Lasserre

▶ **To cite this version:**

Jean-Bernard Lasserre. THE MOMENT-SOS HIERARCHY AND THE CHRISTOFFEL-DARBOUX KERNEL. Optimization Letters, 2021, 15, pp.1835-1845. 10.1007/s11590-021-01713-4 . hal-03008801

HAL Id: hal-03008801

https://laas.hal.science/hal-03008801

Submitted on 16 Nov 2020

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

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

THE MOMENT-SOS HIERARCHY AND THE CHRISTOFFEL-DARBOUX KERNEL

JEAN B. LASSERRE

Abstract. We consider the global minimization of a polynomial on a compact set $\mathbf{B}$. We show that each step of the Moment-SOS hierarchy has a nice and simple interpretation that complements the usual one. Namely, it computes coefficients of a polynomial in an orthonormal basis of $L^2(\mathbf{B},\mu)$ where μ is an arbitrary reference measure whose support is exactly $\mathbf{B}$. The resulting polynomial is a certain density (with respect to μ) of some signed measure on $\mathbf{B}$. When some relaxation is exact (which generically takes place) the coefficients of the optimal polynomial density are values of orthonormal polynomials at the global minimizer and the optimal (signed) density is simply related to the Christoffel-Darboux (CD) kernel and the Christoffel function associated with μ. In contrast to the hierarchy of upper bounds which computes *positive* densities, the global optimum can be achieved exactly as integration against a polynomial (signed) density because the CD-kernel is a reproducing kernel, and so can mimic a Dirac measure (as long as finitely many moments are concerned).

1. Introduction

Consider the *Polynomial Optimization Problem* (POP):

$$f^* \,=\, \min_{\mathbf{x}}\{\, f(\mathbf{x}) :\, \mathbf{x}\in\mathbf{B}\,\},$$

where $\mathbf{B}\subset\mathbb{R}^n$ is a compact basic semi-algebraic set. For the hierarchy of upper bounds discussed below, $\mathbf{B}$ is restricted to be a "simple" set like e.g. a box, an ellipsoid, a simplex, a discrete-hypercube, or their image by an affine transformation. Indeed, to define an SOS-hierarchy of upper bounds converging to the global minimum f^* as described in e.g. [1, 4, 9], we use a measure μ whose support is exactly $\mathbf{B}$, and for which all moments

$$\mu_\alpha := \int_{\mathbf{B}} \mathbf{x}^\alpha\, d\mu\,, \quad \alpha\in\mathbb{N}^n\,,$$

can be obtained numerically or in closed-form. For instance if $\mathbf{B}$ is a box, an ellipsoid or a simplex, μ can chosen to be the Lebesgue measure restricted to $\mathbf{B}$. On the hypercube $\{-1,1\}^n$ μ one may choose for μ the counting measure, etc.

Work partly funded by the AI Interdisciplinary Institute ANITI through the French "Investing for the Future PI3A" program under the Grant agreement ANR-19-PI3A-0004.

1.1. **Background.** Let $\mathbf{B} \subset \mathbb{R}^n$ be the basic semi-algebraic set defined by

$$\mathbf{B} = \{\, \mathbf{x} \in \mathbb{R}^n \,:\, g_j(\mathbf{x}) \geq 0 \quad j = 1, \ldots, m \,\}, \tag{1.1}$$

for some polynomials $g_j \in \mathbb{R}[\mathbf{x}]$, $j = 1, \ldots, m$. Let $g_0(\mathbf{x}) = 1$ for all $\mathbf{x}$, and let $d_j := \lceil \deg(g_j)/2 \rceil$, $j = 0, \ldots, m$. Define $\Sigma[\mathbf{x}]_t$ be the set of sums-of-squares (SOS) polynomials of degree at most $2t$.

A hierarchy of lower bounds. To approximate f^* from below, consider the hierarchy of semidefinite programs indexed by $t \in \mathbb{N}$:

$$\rho_t = \sup_{\lambda\, \sigma_j} \{\, \lambda \,:\, f - \lambda = \sum_{j=0}^{m} \psi_j\, g_j\,; \quad \psi_j \in \Sigma[\mathbf{x}]_{t-d_j}\,,\ j = 0, \ldots, m \,\}, \tag{1.2}$$

where $\Sigma[\mathbf{x}]_t$ denotes the space of sum-of-squares (SOS) polynomials of degree at most $2t$. Under some Archimedean assumption on the g_j's, $\rho_t \leq f^*$ for all t and the sequence of lower bounds $(\rho_t)_{t\in\mathbb{N}}$ is monotone non decreasing and converges to f^* as t increases. Moreover, by a result of Nie [7], its convergence is finite generically, and global minimizers can be extracted from an optimal solution of the semidefinite program which is the dual of (1.2); see e.g. [5]. The sequence of semidefinite programs (1.2) and their duals, both indexed by t, forms what is called the Moment-SOS hierarchy initiated in the early 2000's. For more details on the Moment-SOS hierarchy and its numerous applications in and outside optimization, the interested reader is referred to [3, 5].

A hierarchy of upper bounds. Let μ be a finite Borel measure whose support is exactly $\mathbf{B}$, where now $\mathbf{B}$ is a "simple" set as mentioned earlier. (Hence all moments of μ are available in closed form.) To approximate f^* from above, consider the hierarchy of semidefinite programs

$$u_t = \inf_{\sigma} \{\, \int_{\mathbf{B}} f\, \sigma\, d\mu \,:\, \int_{\mathbf{B}} \sigma\, d\mu \,=\, 1\,;\ \sigma \in \Sigma[\mathbf{x}]_t \,\}. \tag{1.3}$$

That $u_t \geq f^*$ is straightforward since

$$f \geq f^* \text{ on } \mathbf{B} \;\Rightarrow \int_{\mathbf{B}} f\, \sigma\, d\mu \;\geq\; f^* \int_{\mathbf{B}} \sigma\, d\mu \;=\; f^*,$$

for any feasible SOS σ. In [4] it was proved that $u_t \downarrow f^*$ as t increases, and in fact solving the dual of (1.3) is solving a generalized eigenvalue problem for a certain pair of real symmetric matrices. In a series of papers, de Klerk, Laurent an co-workers have provided several rates of convergence of $u_t \downarrow f^*$ for several examples of sets $\mathbf{B}$. For more details and results, the interested reader is referred to [1, 9, 10, 11] and references therein.

The meaning of (1.3) is clear if one recalls that

$$f^* \;=\; \inf_{\phi} \{\, \int_{\mathbf{B}} f\, d\phi \,:\, \phi(\mathbf{B}) = 1\,;\ \phi \in \mathscr{M}(\mathbf{B})_+ \,\}, \tag{1.4}$$

where $\mathscr{M}(\mathbf{B})_+$ is the space of *all* finite Borel measures on $\mathbf{B}$. Indeed in (1.3) one only considers the (restricted) subset of probability measures on $\mathbf{B}$ that have a density (an SOS of degree at most $2t$) with respect to μ whereas in (1.4) one considers *all* probability measures on $\mathbf{B}$. In particular, the Dirac measure $\phi := \delta_{\boldsymbol{\xi}}$ at any global minimiser $\boldsymbol{\xi} \in \mathbf{B}$ belongs to $\mathscr{M}(\mathbf{B})_+$ but does *not* have a density with respect to μ, which explains why the convergence $u_t \downarrow f^*$ as t increases, can be only asymptotic and not finite; an exception is when $\mathbf{B}$ is a finite set (e.g. $\mathbf{B} = \{-1, 1\}^n$ and μ is the counting measure).

1.2. **Contribution.** Our contribution is to show that in fact the dual of the semidefinite program (1.2) for computing the lower bound ρ_t has also an interpretation of the same flavor as (1.3) where one now considers *signed* Borel measures ϕ_t with a distinguished *polynomial* density with respect to μ. Namely, the dual of (1.2) minimizes $\int_{\mathbf{B}} f d\phi_t$ over signed measures ϕ_t of the form:

$$d\phi_t(\mathbf{x}) \;=\; \sigma_t(\mathbf{x})\, d\mu(\mathbf{x}) \;=\; \left(\sum_{|\alpha| \le 2t} \sigma_\alpha\, T_\alpha(\mathbf{x}) \right) d\mu(\mathbf{x})\,, \tag{1.5}$$

where :

- $(T_\alpha) \subset \mathbb{R}[\mathbf{x}]$ is a family of polynomials that are orthonormal with respect to μ, and

- the coefficients $\boldsymbol{\sigma}_t = (\sigma_\alpha)_{\alpha \in \mathbb{N}^n_{2t}}$ of the polynomial $\sigma_t \in \mathbb{R}[\mathbf{x}]_{2t}$ satisfy the usual semidefinite constraints that are necessary for $\boldsymbol{\sigma}_t$ to be *moments* of a measure on $\mathbf{B}$.

Eventually for some $t \in \mathbb{N}$, $\boldsymbol{\sigma}_t$ satisfies:

$$\sigma_\alpha \;=\; T_\alpha(\boldsymbol{\xi}) \;=\; \int_{\mathbf{B}} T_\alpha(\mathbf{x})\, \delta_{\boldsymbol{\xi}}(d\mathbf{x})\,, \quad |\alpha| \;\le\; 2t\,, \tag{1.6}$$

where $\boldsymbol{\xi}$ is an arbitrary global minimizer and $\delta_{\boldsymbol{\xi}}$ is the Dirac measure at $\boldsymbol{\xi} \in \mathbf{B}$. Indeed then

$$\int_{\mathbf{B}} f(\mathbf{x})\, d\phi_t(\mathbf{x}) \;:=\; \int_{\mathbf{B}} f(\mathbf{x}) \sum_{|\alpha| \le 2t} T_\alpha(\boldsymbol{\xi})\, T_\alpha(\mathbf{x})\, d\mu(\mathbf{x}) \;=\; f(\boldsymbol{\xi})\,,$$

because the Christoffel-Darboux Kernel $K_t(\mathbf{x}, \mathbf{y}) := \sum_{|\alpha| \le 2t} T_\alpha(\mathbf{x})\, T_\alpha(\mathbf{y})$ is a *reproducing* kernel for $\mathbb{R}[\mathbf{x}]_{2t}$, considered to be a finite-dimensional subspace of the Hilbert space $L^2(\mathbf{B}, \mu)$. Moreover, $\sigma_t(\boldsymbol{\xi})^{-1}$ is nothing less than the Christoffel function evaluated at the global minimizer $\boldsymbol{\xi}$ of f on $\mathbf{B}$.

As a take home message and contribution of this paper, it turns out that the dual of the step-t semidefinite relaxation (1.2) *is a semidefinite program that computes the coefficients* $\boldsymbol{\sigma}_t = (\sigma_\alpha)$ *of the polynomial density* σ_t *in* (1.5)*. In addition, when the relaxation is exact then* $\sigma_t(\boldsymbol{\xi})^{-1}$ *is the Christoffel function of* μ*, evaluated at a global minimizer* $\boldsymbol{\xi}$ *of* f *on* $\mathbf{B}$*.*

Interestingly, in the dual of (1.2) there is no mention of the reference measure μ. Only *after* we fix some arbitrary reference measure μ on $\mathbf{B}$,

we can interpret an optimal solution as coefficients $\boldsymbol{\sigma}_t$ of an appropriate polynomial density with respect to μ.

So in both (1.3) and the dual of (1.2), one searches for a polynomial "density" with respect to μ. In (1.3) one searches for an SOS density (hence a positive density) whereas in the dual of (1.2) one searches for a *signed* polynomial density whose coefficients (in the basis of orthonormal polynomials) are moments of a measure on $\mathbf{B}$ (ideally the Dirac at a global minimizer).

The advantage of the (signed) polynomial density in (1.5) compared to the (positive) SOS density in (1.3), is to be able to obtain the global optimum f^* as the integral of f against this density, which is impossible with the SOS density of (1.3).

At last but not least, this interpretation establishes another (and rather surprising) simple link between polynomial optimization (here the Moment-SOS hierarchy), the Christoffel-Darboux kernel and the Christoffel function, fundamental tools in the theory of orthogonal polynomials and the theory of approximation. Previous contributions in this vein include [6] to characterize upper bounds (1.3), [1, 9, 10] to analyze their rate of convergence to f^*, and the more recent [11] for rate of convergence of both upper and lower bounds on $\mathbf{B} = \{0,1\}^n$.

2. Main result

2.1. **Notation and definition.** Let $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$ be the ring of real polynomials in the variables $x_1, \ldots, x_n$ and let $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$ be its subspace of polynomials of degree at most t. Let $\mathbb{N}^n_t := \{\alpha \in \mathbb{N}^n : |\alpha| \leq t\}$ where $|\alpha| = \sum_i \alpha_i$. For an arbitrary Borel subset $\mathcal{X}$ of $\mathbb{R}^n$, denote by $\mathscr{M}(\mathcal{X})_+$ the convex cone of finite Borel measures on $\mathcal{X} \subset \mathbb{R}^n$, and by $\mathscr{P}(\mathcal{X})$ is subset of probability measures on $\mathcal{X}$..

2.2. **Moment and localizing matrices.** Given an sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ and polynomial $g \in \mathbb{R}[\mathbf{x}]$, $\mathbf{x} \mapsto g(\mathbf{x}) := \sum_\gamma g_\gamma \, \mathbf{x}^\gamma$, the *localizing* matrix $\mathbf{M}_t(g\,\mathbf{y})$ associated with g and $\mathbf{y}$ is th real symmetric matrix with rows and columns indexed by $\alpha \in \mathbb{N}^n_t$ and with entries

$$\mathbf{M}_t(g\,\mathbf{y})(\alpha,\beta) := \sum_\gamma g_\gamma \, y_{\alpha+\beta+\gamma}\,, \quad \alpha,\beta \in \mathbb{N}^n_t\,. \tag{2.1}$$

If $g(\mathbf{x}) = 1$ for all $\mathbf{x}$ then $\mathbf{M}_t(g\,\mathbf{y})\,(= \mathbf{M}_t(\mathbf{y}))$ is called the *moment* matrix.

A sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ has a representing measure if there exists a (positive) finite Borel measure ϕ on $\mathbb{R}^n$ such that $y_\alpha = \int \mathbf{x}^\alpha \, d\phi$ for all $\alpha \in \mathbb{N}^n$.

If $\mathbf{y}$ has a representing measure supported on $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ then $\mathbf{M}_t(\mathbf{y}) \succeq 0$ and $\mathbf{M}_t(g\,\mathbf{y}) \succeq 0$ for all $t \in \mathbb{N}$. The converse is not true in general; however, the following important result is at the core of the Moment-SOS hierarchy.

Theorem 2.1. *(Putinar [8]) Let $g_j \in \mathbb{R}[\mathbf{x}]$, $j = 0, \ldots, m$ with $g_0(\mathbf{x}) = 1$ for all $\mathbf{x}$, and let $G := \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$ be compact. Moreover, assume that for some $M > 0$, the quadratic polynomial $\mathbf{x} \mapsto$*

$M - \|\mathbf{x}\|^2$ can be written in the form $\sum_{j=0}^m \psi_j\, g_j$, for some SOS polynomials $\psi_0, \ldots \psi_m$.

Then a sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n}$ has a representing measure on G if and only if $\mathbf{M}_t(g_j\,\mathbf{y}) \succeq 0$ for all $t \in \mathbb{N}$, and all $j = 0, \ldots, m$.

Orthonormal polynomials. Let $\mathbf{B} \subset \mathbb{R}^n$ be the compact basic semi-algebraic set defined in (1.1) assumed to have a nonempty interior. Let μ be a finite Borel (reference) measure whose support is exactly $\mathbf{B}$ and with associated sequence of orthonormal polynomials $(T_\alpha)_{\alpha\in\mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$. That is

$$\int_{\mathbf{B}} T_\alpha\, T_\beta\, d\mu \;=\; \delta_{\alpha=\beta}\,, \quad \forall \alpha, \beta \in \mathbb{N}^n\,.$$

For instance, if $\mathbf{B} = [-1, 1]^n$ and μ is the uniform probability distribution on $\mathbf{B}$, one may choose for the family (T_α) the tensorized *Legendre* polynomials. Namely if $(T_j) \subset \mathbb{R}[x]$ is the family of univariate Legendre polynomials, then

$$T_\alpha(\mathbf{x}) \;:=\; \prod_{j=1}^n T_{\alpha_j}(x_j)\,, \quad \alpha \in \mathbb{N}^n\,.$$

For every $t \in \mathbb{N}$, the mapping $K_t : \mathbf{B} \times \mathbf{B} \to \mathbb{R}$,

$$(\mathbf{x}, \mathbf{y}) \mapsto K_t(\mathbf{x}, \mathbf{y}) \;:=\; \sum_{|\alpha|\le t} T_\alpha(\mathbf{x})\, T_\alpha(\mathbf{y})\,, \quad \mathbf{x}, \mathbf{y} \in \mathbf{B}$$

is called the Cristoffel-Darboux kernel associated with μ. An important property of K_t is to *reproduce* polynomials of degree at most t, that is:

$$p(\mathbf{x}) \;=\; \int_{\mathbf{B}} p(\mathbf{y})\, K_t(\mathbf{x}, \mathbf{y})\, d\mu(\mathbf{y}) \quad \forall \mathbf{x} \in \mathbf{B}\,, \quad \forall p \in \mathbb{R}[\mathbf{x}]_t\,. \tag{2.2}$$

This is why K_t is called a reproducing kernel, and $\mathbb{R}[\mathbf{x}]_t$ viewed as a finite-dimensional vector subspace of the Hilbert space $L^2(\mathbf{B}, \mu)$, is called a Reproducing Kernel Hilbert Space (RKHS). For more details on the theory of orthogonal polynomials, the interested reader is referred to e.g. [2] and the many references therein.

2.3. **Main result.**

An observation. Let $f \in \mathbb{R}[\mathbf{x}]$ and let $t \geq \deg(f) = d_f$ be fixed. Let $\mathscr{P}(\mathbf{B}) \subset \mathscr{M}(\mathbf{B})_+$ be the space of probability measures on $\mathbf{B}$. Then

$$\begin{aligned}
f^* &= \min_{\phi \in \mathscr{P}(\mathbf{B})_+} \int_{\mathbf{B}} f \, d\phi \\
&= \min_{\phi \in \mathscr{P}(\mathbf{B})_+} \int_{\mathbf{B}} \int_{\mathbf{B}} f(\mathbf{y}) \, K_t(\mathbf{x},\mathbf{y}) \, d\mu(\mathbf{y}) \, d\phi(\mathbf{x}) \\
&= \min_{\phi \in \mathscr{P}(\mathbf{B})_+} \int_{\mathbf{B}} f(\mathbf{y}) \, (\int_{\mathbf{B}} K_t(\mathbf{x},\mathbf{y}) \, d\phi(\mathbf{x}) \,) \, d\mu(\mathbf{y}) \\
&= \min_{\phi \in \mathscr{P}(\mathbf{B})_+} \int_{\mathbf{B}} f(\mathbf{y}) \left[\sum_{|\alpha| \leq t} (\underbrace{\int_{\mathbf{B}} T_\alpha(\mathbf{x}) \, d\phi(\mathbf{x})}_{\sigma_\alpha}) \, T_\alpha(\mathbf{y}) \right] d\mu(\mathbf{y}) \\
&= \min_{\phi \in \mathscr{P}(\mathbf{B})_+} \int_{\mathbf{B}} f(\mathbf{y}) \, (\underbrace{\sum_{|\alpha| \leq t} \sigma_\alpha \, T_\alpha(\mathbf{y})}_{\sigma_t(\mathbf{y}) \in \mathbb{R}[\mathbf{x}]_t}) \, d\mu(\mathbf{y}) \,,
\end{aligned}$$

where the second equality follows from Fubini-Tonelli interchange theorem valid in this simple setting. In other words, we have proved the following:

Lemma 2.2. *Let $\mathbf{B} \subset \mathbb{R}^n$ be as in* (1.1) *and let μ be a finite Borel (reference) measure whose support is exactly $\mathbf{B}$ and with associated sequence of orthonormal polynomials $(T_\alpha)_{\alpha \in \mathbb{N}^n}$. Let $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{B}\}$. Then for every fixed $t \geq \deg(f)$:*

$$f^* = \inf_{\sigma \in \mathbb{R}[\mathbf{x}]_t} \int_{\mathbf{B}} f(\mathbf{y}) \, \sigma(\mathbf{y}) \, d\mu(\mathbf{y}) \,, \tag{2.3}$$

where the infimum is over all polynomials $\sigma \in \mathbb{R}[\mathbf{x}]_t$ of the form:

$$\sigma(\mathbf{x}) = \sum_{|\alpha| \leq t} \sigma_\alpha \, T_\alpha(\mathbf{x}) \,, \quad \forall \mathbf{x} \in \mathbf{B} \quad \textit{with} \tag{2.4}$$

$$\sigma_\alpha = \int_{\mathbf{B}} T_\alpha(\mathbf{x}) \, d\phi(\mathbf{x}) \,, \quad \forall \alpha \in \mathbb{N}^n_t \,, \quad \textit{for some } \phi \in \mathscr{P}(\mathbf{B}). \tag{2.5}$$

So solving (2.3) is equivalent to searching for a signed measure $\sigma \, d\mu$ with polynomial (signed) density $\sigma \in \mathbb{R}[\mathbf{x}]_t$ that satisfies (2.4)-(2.5).

2.4. **A hierarchy of relaxations of** (2.3)**.** In this section we show the SOS-hierarchy defined in (1.2) is the dual semidefinite program of a natural SDP-relaxation of (2.3). In fact the only difficult constraint in (2.3) is (2.5) which demands $\boldsymbol{\sigma}$ to admit a representing probability measure ϕ on $\mathbf{B}$.

Let $\mathbf{D}_t$ be the lower triangular matrix for the change of basis of $\mathbb{R}[\mathbf{x}]_{2t}$ from the monomial basis $\mathbf{v}_{2t}(\mathbf{x}) = (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}^n_{2t}}$ of $\mathbb{R}[\mathbf{x}]_{2t}$ to the basis $(T_\alpha)_\alpha$,

i.e.,

$$\left[\begin{array}{c} T_0 \\ \cdots \\ T_\alpha \\ \cdots \end{array}\right] = \mathbf{D}_t \cdot \left[\begin{array}{c} 1 \\ \cdots \\ \mathbf{x}^\alpha \\ \cdots \end{array}\right] = \mathbf{D}_t \cdot \mathbf{v}_{2t}(\mathbf{x}) \tag{2.6}$$

and denote $\mathbf{D}_t'$ the transpose of $\mathbf{D}_t$. The matrix $\mathbf{D}_t$ is nonsingular with positive diagonal. Then with $\boldsymbol{\sigma} = (\sigma_\alpha)_{\alpha\in\mathbb{N}^n_{2t}}$, (2.5) reads

$$\boldsymbol{\sigma} \,=\, \mathbf{D}_t \cdot \mathbf{y} \quad \text{with} \quad \mathbf{y} = \int_{\mathbf{B}} \mathbf{v}_{2t}(\mathbf{x})\, d\phi(\mathbf{x}). \tag{2.7}$$

That is, $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n_{2t}}$ is required to be a moment sequence as it has a representing probability measure $\phi \in \mathscr{P}(\mathbf{B})$. So in view of Theorem 2.1, the constraint (2.7) can be relaxed to

$$\boldsymbol{\sigma} = \mathbf{D}_t \cdot \mathbf{y} \quad \text{with } y_0 = 1 \text{ and } \quad \mathbf{M}_{t-d_j}(g_j\,\mathbf{y}) \,\succeq\, 0\,, \quad j = 0,\ldots,m\,.$$

Therefore, consider the following relaxation of (2.3)

$$\begin{aligned} \rho_{2t} = \inf_{\sigma\in\mathbb{R}[\mathbf{x}]_{2t}} \quad & \{ \int_{\mathbf{B}} f(\mathbf{x})\,(\sum_{\alpha\in\mathbb{N}^n_{2t}} \sigma_\alpha\, T_\alpha(\mathbf{x}))\, d\mu(\mathbf{x}) \;:\; \boldsymbol{\sigma} \,=\, \mathbf{D}_t \cdot \mathbf{y}\,; \\ & y_0 = 1\,;\, \mathbf{M}_{t-d_j}(g_j\,\mathbf{y}) \,\succeq\, 0\,, \quad j = 0,\ldots,m\,\}\,. \end{aligned} \tag{2.8}$$

Lemma 2.3. *Let $\mathbf{B} \subset \mathbb{R}^n$ be as in* (1.1) *and let μ be a finite Borel (reference) measure whose support is exactly $\mathbf{B}$ and with associated sequence of orthonormal polynomials $(T_\alpha)_{\alpha\in\mathbb{N}^n}$. The semidefinite relaxation* (2.8) *of* (2.3) *reads:*

$$\inf_{\mathbf{y}} \{\, \langle \mathbf{f}, \mathbf{y}\rangle \;:\; y_0 = 1\,;\, \mathbf{M}_{t-d_j}(g_j\,\mathbf{y}) \,\succeq\, 0\,, \quad j = 0,\ldots,m\,\}\,, \tag{2.9}$$

which is the dual of (1.2)

Proof. With $f(\mathbf{x}) = \sum_\alpha f_\alpha\, \mathbf{x}^\alpha = \langle \mathbf{f}, \mathbf{v}_{2t}(\mathbf{x})\rangle$, write $f(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n_{2t}} \tilde{f}_\alpha\, T_\alpha(\mathbf{x})$ in the basis $(T_\alpha)_{\alpha\in\mathbb{N}^n_{2t}}$. Then with $\tilde{\mathbf{f}} = (\tilde{f}_\alpha)$ one obtains

$$\langle \tilde{\mathbf{f}}, \mathbf{D}_t \cdot \mathbf{v}_{2t}(\mathbf{x})\rangle \,=\, \langle \mathbf{D}_t'\,\tilde{\mathbf{f}}, \mathbf{v}_{2t}(\mathbf{x})\rangle \,=\, \langle \mathbf{f}, \mathbf{v}_{2t}(\mathbf{x})\rangle \Rightarrow \tilde{\mathbf{f}} = (\mathbf{D}_t')^{-1}\mathbf{f}\,.$$

Finally, as the T_α's form an orthonormal basis, the criterion

$$\int_{\mathbf{B}} f(\mathbf{x})\,(\sum_{\alpha\in\mathbb{N}^n_{2t}} \sigma_\alpha\, T_\alpha(\mathbf{x}))\, d\mu(\mathbf{x})$$

to minimize in (2.8) reads:

$$\int_{\mathbf{B}} f(\mathbf{x})\,(\sum_{\alpha\in\mathbb{N}^n_{2t}} \sigma_\alpha\, T_\alpha(\mathbf{x}))\, d\mu(\mathbf{x}) \,=\, \langle \tilde{\mathbf{f}}, \boldsymbol{\sigma}\rangle \,=\, \langle (\mathbf{D}_t')^{-1}\mathbf{f}, \mathbf{D}_t\,\mathbf{y}\rangle \,=\, \langle \mathbf{f}, \mathbf{y}\rangle\,,$$

which yields that (2.8) is exactly (2.9). Next, that (2.9) is a dual of (1.2) is a standard result in polynomial optimization [3, 5]. □

Of course by reverting the process of the above proof, the semidefinite program (2.9) can be transformed to (2.8) once a reference measure μ with support exactly $\mathbf{B}$ is defined with its associated orthonormal polynomials (T_α). Indeed, once μ and the T_α's are defined, one may use the change of basis matrix $\mathbf{D}$ in (2.6) to pass from (2.9) to (2.8).

Corollary 2.4. *Let $\mathbf{B} \subset \mathbb{R}^n$ be as in (1.1) and let μ be a finite Borel (reference) measure whose support is exactly $\mathbf{B}$ and with associated sequence of orthonormal polynomials $(T_\alpha)_{\alpha\in\mathbb{N}^n}$. Let f^* be the global minimum of f on $\mathbf{B}$.*

Let t be such that the semidefinite relaxation (2.8) (or equivalently (2.9)) is exact, i.e., if $\rho_{2t} = f^*$. If an optimal solution $\mathbf{y}^*$ of (2.9) has a representing measure $\phi^* \in \mathscr{M}(\mathbf{B})_+$, then an optimal polynomial density $\sigma^* \in \mathbb{R}[\mathbf{x}]_{2t}$ of (2.8) satisfies:

$$\sigma^*(\boldsymbol{\xi}) \;=\; \sum_{\alpha\in\mathbb{N}^n_{2t}} T_\alpha(\boldsymbol{\xi})^2 \;=\; K_{2t}(\boldsymbol{\xi},\boldsymbol{\xi})\,, \quad \forall \boldsymbol{\xi} \in \mathrm{supp}(\phi^*)\,,$$

that is, $\sigma^*(\boldsymbol{\xi})^{-1}$ is the Christoffel function evaluated at the global minimizer $\boldsymbol{\xi} \in \mathbf{B}$.

Proof. If $\mathbf{y}^*$ has a representing measure $\phi^* \in \mathscr{M}(\mathbf{B})_+$ then necessarily $f(\boldsymbol{\xi}) = f^*$ for all $\boldsymbol{\xi} \in \mathrm{sup}(\phi^*)$; see e.g. [3, 5]. In particular, for every $\boldsymbol{\xi} \in \mathrm{sup}(\phi^*)$, the vector $\hat{\mathbf{y}} := (\boldsymbol{\xi}^\alpha)_{\alpha\in\mathbb{N}^n_{2t}}$ is also an optimal solution of (2.9). Then

$$\boldsymbol{\sigma}^* \;=\; \mathbf{D}_t \cdot \hat{\mathbf{y}} = \mathbf{D}_t \cdot \mathbf{v}_{2t}(\boldsymbol{\xi}) \;=\; \begin{bmatrix} T_0(\boldsymbol{\xi}) \\ \cdots \\ T_\alpha(\boldsymbol{\xi}) \\ \cdots \end{bmatrix},$$

i.e., $\sigma^*_\alpha = T_\alpha(\boldsymbol{\xi})$ for all $\alpha \in \mathbb{N}^n_{2t}$. Therefore,

$$\mathbf{x} \mapsto \sigma^*(\mathbf{x}) \;=\; \sum_{\alpha\in\mathbb{N}^n_{2t}} T_\alpha(\boldsymbol{\xi})\, T_\alpha(\mathbf{x}) \;=\; \mathbf{K}_{2t}(\boldsymbol{\xi},\mathbf{x})\,,$$

and so $\sigma^*(\boldsymbol{\xi}) = \mathbf{K}_{2t}(\boldsymbol{\xi},\boldsymbol{\xi})$. In other words, $\sigma^*(\boldsymbol{\xi})^{-1}$ is the Christoffel function associated with μ, evaluated at $\boldsymbol{\xi} \in \mathbf{B}$. □

Discussion. Observe that the formulation (2.8) does not require that the set $\mathbf{B}$ is a "simple" set as it is required in (1.3). Indeed the orthonormal polynomials (T_α) are only used to provide an *interpretation* of the hierarchy of lower bounds (2.9) (and its dual (1.2)). On the other hand, for the hierarchy of upper bounds (1.3), $\mathbf{B}$ indeed needs to be a "simple" set for computational purposes. This is because one needs the numerical value of the moments of μ for a practical implementation of (1.3).

Lemma 2.3 shows that the Moment-SOS hierarchy described in [3, 5] amounts to compute a hierarchy of signed polynomial densities with respect to some reference measure μ with support exactly $\mathbf{B}$. When the step-t relaxation is exact (which takes place generically [7]) the resulting optimal

density σ in (2.8) is nothing less than the polynomial $\mathbf{x} \mapsto K_t(\boldsymbol{\xi}, \mathbf{x})$ where $\boldsymbol{\xi}$ is a global minimizer of f on $\mathbf{B}$, $K_t(\boldsymbol{\xi}, \mathbf{x})$ is the celebrated Cristoffel-Darboux kernel in approximation theory, and $\sigma(\boldsymbol{\xi}, \boldsymbol{\xi})$ is the reciprocal of the Christoffel function evaluated at a global minimizer $\boldsymbol{\xi}$.

3. Conclusion

We have shown that the Moment-SOS hierarchy that provides an increasing sequence of lower bounds on the global minimum of a polynomial f on a compact set $\mathbf{B}$, has a simple interpretation related to orthogonal polynomials associated with an arbitrary reference measure whose support is exactly $\mathbf{B}$. This interpretation strongly relates polynomial optimization (here the Moment-SOS hierarchy) with the Christoffel-Darboux kernel and the Christoffel function, fundamental tools in the theory of orthogonal polynomials and the theory of of approximation.

It is another item in the list of previous contributions [6, 1, 9, 10] that also link some issues in polynomial optimization with orthogonal polynomials associated with appropriate measures. We hope that such connections will stimulate even further investigations in this direction.

References

[1] E. de Klerk, M. Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial optimization on the sphere. Math. Program. (2020). To appear.

[2] C. Dunkl, Y. Xu. *Orthogonal Polynomials of Several Variables*, Cambridge University Press, Cambridge, UK, 2001.

[3] J.B. Lasserre. Global optimization with polynomials and the problem of moments SIAM J. Optimization 11, pp 796–817, 2001.

[4] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization SIAM J. Optim. 21, pp. 864–885, 2011.

[5] J. B. Lasserre. *An introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press, Cambridge, 2015.

[6] J.B. Lasserre Connecting optimization with spectral analysis of tri-diagonal matrices. Math. Program. (2020). To appear.

[7] J. Nie. Optimality Conditions and Finite Convergence of Lasserre's Hierarchy Math. Program. Ser. A **146**(1-2), pp. 97-121, 2014.

[8] M. Putinar. Positive polynomial on compact semi-algebraic sets, Indiana Univ. Math. J. 42: 969–984, 1993

[9] L. Slot, M. Laurent. Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. Math. Program. (2020). To appear.

[10] L. Slot, M. Laurent. Near-optimal analysis of univariate moment bounds for polynomial optimization. Math. Program. (2020). To appear.

[11] L. Slot, M. Laurent. Sum-of-squares hierarchies for binary polynomial optimization. `arXiv:2011.04027`.

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4, France, email: lasserre@laas.fr