



**HAL**  
open science

## Moment-SOS Hierarchies

Jean B Lasserre

► **To cite this version:**

Jean B Lasserre. Moment-SOS Hierarchies. Encyclopedia of Optimization, 2022, pp.1-7.  
10.1007/978-3-030-54621-2\_740-1 . hal-04054990

**HAL Id: hal-04054990**

**<https://laas.hal.science/hal-04054990>**

Submitted on 1 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.

# Moment-SOS hierarchies

Jean B. Lasserre

Received: date / Accepted: date

## Keywords

Non-Linear Programming; Global Optimization; Convex Relaxations, Semidefinite Programming;

## 1 Introduction

Consider the *Polynomial Optimization Problem*:

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

where  $\Omega \subset \mathbb{R}^n$  is a compact basic semi-algebraic set. For the hierarchy of upper bounds discussed below,  $\Omega$  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 ( $\Omega = \mathbb{R}_+^n$  or  $\Omega = \mathbb{R}^n$  is also permitted). Indeed, to define an SOS-hierarchy of upper bounds converging to the global minimum  $f^*$  as described in e.g. [5, 7, 12], one uses a measure  $\mu$  whose support is exactly  $\Omega$ , and for which all moments  $\mu_\alpha := \int_{\Omega} \mathbf{x}^\alpha d\mu$ ,  $\alpha \in \mathbb{N}^n$ , can be obtained numerically or in closed-form. For instance if  $\Omega$  is a box, an ellipsoid or a simplex,  $\mu$  can be chosen to be the Lebesgue measure restricted to  $\Omega$ . On the hypercube  $\{-1, 1\}^n$   $\mu$  one may choose for  $\mu$  the counting measure, etc.

---

Work partly funded by the AI Interdisciplinary Institute ANITI through the French “Investing for the Future PIA3” program under the Grant agreement 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.

J.B. Lasserre  
LAAS-CNRS & Institute of Mathematics  
University of Toulouse  
E-mail: lasserre@laas.fr LAAS-CNRS, 7 Avenue du Colonel Roche, BP 54200, 31031 Toulouse, Cedex 4, France

## 2 Moment-SOS hierarchies

### 2.1 Notation and definition

Let  $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \dots, x_n]$  be the ring of real polynomials in the variables  $x_1, \dots, x_n$  and let  $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$  be its subspace of polynomials of degree at most  $t$ . Let  $\mathbb{N}_t^n := \{\alpha \in \mathbb{N}^n : |\alpha| \leq t\}$  where  $|\alpha| = \sum_i \alpha_i$ . Denote by  $\Sigma[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]_{2t}$  the convex cone of sums-of-squares (SOS) polynomials of degree at most  $2t$ . Let  $\mathbf{v}_t(\mathbf{x}) := (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}_t^n}$  be the vector of the monomial basis, up to degree  $t$ . For an arbitrary Borel subset  $\mathcal{X}$  of  $\mathbb{R}^n$ , denote by  $\mathcal{M}(\mathcal{X})_+$  the convex cone of finite Borel measures on  $\mathcal{X} \subset \mathbb{R}^n$ , and by  $\mathcal{P}(\mathcal{X})$  is subset of probability measures on  $\mathcal{X}$ .

**Riesz functional.** Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , let  $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$  be the linear functional

$$f (= \sum_{\alpha} f_{\alpha} \mathbf{x}^{\alpha}) \mapsto L_{\mathbf{y}}(f) := \sum_{\alpha} f_{\alpha} y_{\alpha}. \quad (1)$$

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

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

$$\mathbf{M}_t(g\mathbf{y})(\alpha, \beta) := \sum_{\gamma} g_{\gamma} y_{\alpha+\beta+\gamma}, \quad \alpha, \beta \in \mathbb{N}_t^n. \quad (2)$$

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.

If  $\mathbf{y}$  has a representing measure supported on  $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$  then  $\mathbf{M}_t(\mathbf{y}) \geq 0$  and  $\mathbf{M}_t(g\mathbf{y}) \geq 0$  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 1** (Putinar [10]) *Let  $g_j \in \mathbb{R}[\mathbf{x}]$ ,  $j = 0, \dots, m$  with  $g_0(\mathbf{x}) = 1$  for all  $\mathbf{x}$ , and let  $G := \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}$  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  $\psi_0, \dots, \psi_m$ .*

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

**Orthonormal polynomials.** Let  $\Omega \subset \mathbb{R}^n$  be the compact basic semi-algebraic set defined in (4), assumed to have a nonempty interior. Let  $\mu$  be a finite Borel (reference) measure whose support is exactly  $\Omega$  and with associated sequence of orthonormal polynomials  $(T_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$ . That is:

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

For instance, if  $\Omega = [-1, 1]^n$  and  $\mu$  is the uniform probability distribution on  $\Omega$ , one may choose for the family  $(T_\alpha)$  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  $K_t : \Omega \times \Omega \rightarrow \mathbb{R}$ ,

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

is called the Cristoffel-Darboux kernel associated with  $\mu$ . An important property of  $K_t$  is to *reproduce* polynomials of degree at most  $t$ , that is, for all  $p \in \mathbb{R}[\mathbf{x}]_t$ :

$$p(\mathbf{x}) = \int_{\Omega} p(\mathbf{y}) K_t(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{y}) \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (3)$$

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  $L^2(\Omega, \mu)$ , 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. [1] and the many references therein.

Define the basic semi-algebraic set

$$\Omega := \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0 \quad j = 1, \dots, m\}, \quad (4)$$

for some polynomials  $g_j \in \mathbb{R}[\mathbf{x}]$ ,  $j = 1, \dots, m$ . Let  $g_0(\mathbf{x}) = 1$  for all  $\mathbf{x}$ , and let  $d_j := \lceil \deg(g_j)/2 \rceil$ ,  $j = 0, \dots, m$ . In the rest of the paper, and as  $\Omega \subset \mathbb{R}^n$  is compact, one also assumes for convenience that  $\Omega$  is contained in a Euclidean ball of radius  $M$ , and that  $g_1(\mathbf{x}) := M - \|\mathbf{x}\|^2$  (so that the constraint  $g_1(\mathbf{x}) \geq 0$  is redundant).

## 2.2 A hierarchy of lower bounds

Observe that the global minimum  $f^*$  also reads:

$$f^* = \inf_{\phi \in \mathcal{M}(\Omega)_+} \left\{ \int_{\Omega} f d\phi : \phi(\Omega) = 1 \right\}. \quad (5)$$

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

$$\rho_t = \inf_{\mathbf{y}} \left\{ L_{\mathbf{y}}(f) : y_0 = 1; \right. \\ \left. \mathbf{M}_{t-d_j}(g_j \mathbf{y}) \geq 0, \quad j = 0, \dots, m \right\}. \quad (6)$$

Of course  $\rho_t \leq f^*$  for all  $t$  because the constraints of (6) are only necessary conditions for  $\mathbf{y}$  to have a representing measure  $\phi$  supported on  $\Omega$ . Hence (6) is a semidefinite-relaxation of (5). Its dual reads:

$$\rho_t^* = \sup_{\lambda, \sigma_j} \left\{ \lambda : f - \lambda = \sum_{j=0}^m \psi_j g_j; \right. \\ \left. \psi_j \in \Sigma[\mathbf{x}]_{t-d_j}, \quad j = 0, \dots, m \right\}. \quad (7)$$

**Theorem 2** *Let  $\Omega \subset \mathbb{R}^n$  be as in (4) with nonempty interior. Then (6) has an optimal solution and  $\rho_t = \rho_t^*$ . In addition,  $\rho_t \uparrow f^*$  as  $t$  increases, and generically the convergence is finite, in which case a global minimizer can be extracted from an optimal solution of (6).*

The sequence of semidefinite programs (6) and their duals (7), 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 [2–4, 6, 9].

## 2.3 A hierarchy of upper bounds

Let  $\mu$  be a finite Borel measure whose support is exactly  $\Omega$ , where now  $\Omega$  is a “simple” set as mentioned earlier. (Hence all moments of  $\mu$  are available in closed form.) To approximate  $f^*$  from *above*, consider the hierarchy of semidefinite programs

$$\tau_t = \inf_{\sigma} \left\{ \int_{\Omega} f \sigma d\mu : \int_{\Omega} \sigma d\mu = 1; \sigma \in \Sigma[\mathbf{x}]_t \right\}. \quad (8)$$

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

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

for any feasible SOS  $\sigma$ . The dual of (8) is quite simple as it reads

$$\tau_t^* = \sup_{\lambda} \{ \lambda : \mathbf{M}_t(f\mu) \geq \lambda \mathbf{M}_t(\mu) \}, \quad (9)$$

where  $\mathbf{M}_t(\mu)$  (resp.  $\mathbf{M}_t(f\mu)$ ) is the moment matrix (resp. the localizing matrix) associated with  $\mu$  and  $f$ .

**Theorem 3** ([5]) *Let  $\Omega \subset \mathbb{R}^n$  be a simple set and  $\mu$  an arbitrary finite Borel measure whose support is  $\Omega$  and with all moments available (in closed-form or by simple computation). Then  $\tau_t^* = \tau_t \downarrow f^*$  as  $t$  increases.*

Observe that (9) is a standard generalized eigenvalue problem and an optimal SOS solution  $\sigma^*$  of (8) is the eigenvector associated with the eigenvalue  $\lambda^*$  (the optimal solution of (9)). In a series of papers, de Klerk, Laurent and co-workers have provided several rates of convergence of  $\tau_t \downarrow f^*$  for several examples of sets  $\Omega$ ; typically a  $O(1/t^2)$  rate of convergence takes place. For more details and results, the interested reader is referred to [11–13] and references therein.

The meaning of (8) is clear if one recalls that

$$f^* = \inf_{\phi} \left\{ \int_{\Omega} f d\phi : \phi(\Omega) = 1; \phi \in \mathcal{M}(\Omega)_+ \right\}, \quad (10)$$

where  $\mathcal{M}(\Omega)_+$  is the space of all finite Borel measures on  $\Omega$ . Indeed in (8) one only considers the (restricted) subset of probability measures on  $\Omega$  that have a density (an SOS of degree at most  $2t$ ) with respect to  $\mu$  whereas in (10) one considers all probability measures on  $\Omega$ . In particular, the Dirac measure  $\phi := \delta_{\xi}$  at any global minimiser  $\xi \in \Omega$  belongs to  $\mathcal{M}(\Omega)_+$  but does not have a density with respect to  $\mu$ , which explains why the convergence  $\tau_t \downarrow f^*$  as  $t$  increases, can be only asymptotic and not finite; an exception is when  $\Omega$  is a finite set (e.g.  $\Omega = \{-1, 1\}^n$  and  $\mu$  is the counting measure).

## 2.4 Main result

To compare the two hierarchies of upper and lower bounds one now expresses them in the same language of *polynomial densities* w.r.t.  $\mu$ , expressed in the orthonormal basis  $(T_{\alpha})$ , and described in [7] and [8].

**Lemma 1** *Let  $\Omega \subset \mathbb{R}^n$  be as in (4) and let  $\mu$  be a finite Borel (reference) measure whose support is exactly  $\Omega$  and with associated sequence of orthonormal polynomials  $(T_{\alpha})_{\alpha \in \mathbb{N}^n}$ . Let  $f^* = \min \{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$ . Then for every fixed  $t \geq \deg(f)$ :*

$$f^* = \min_{\sigma \in \mathbb{R}[\mathbf{x}]_t} \int_{\Omega} f(\mathbf{z}) \sigma(\mathbf{z}) d\mu(\mathbf{z}), \quad (11)$$

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

$$\sigma = \sum_{|\alpha| \leq t} \sigma_{\alpha} T_{\alpha}; \quad \sigma_{\alpha} = \int_{\Omega} T_{\alpha} d\phi, \quad \forall \alpha \in \mathbb{N}_t^n, \quad (12)$$

for some  $\phi \in \mathcal{P}(\Omega)$ .

In particular,  $f^* = \int_{\Omega} f(\mathbf{z}) \sigma^*(\mathbf{z}) d\mu(\mathbf{z})$  where  $\sigma^* \in \mathbb{R}[\mathbf{x}]_{2t}$  in (12) satisfies:  $\sigma_{\alpha}^* = T_{\alpha}(\xi)$  for all  $\alpha \in \mathbb{N}_t^n$ , and  $\xi$  is a global minimizer of  $f$  on  $\Omega$ .

*Proof* Recalling (5):

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

where the second equality follows from Fubini-Tonelli interchange theorem valid in this simple setting. Finally, the last statement follows from the reproducing property of the Christoffel-Darboux kernel.  $\square$

So solving (11) is equivalent to searching for a *signed* measure  $\sigma d\mu$  with polynomial (signed) density  $\sigma \in \mathbb{R}[\mathbf{x}]_t$  that satisfies (12). In particular the signed density  $\mathbf{z} \mapsto \sigma^* = \sum_{\alpha \in \mathbb{N}_t^n} T_{\alpha}(\xi) T_{\alpha}(\mathbf{z})$  is an optimal solution.

Then as done for (5), an obvious relaxation of the (difficult) moment constraint in (12) reads:

$$\begin{aligned} \rho_{2t} &= \inf_{\mathbf{y}} \left\{ \int_{\Omega} f(\mathbf{x}) \left( \sum_{\alpha \in \mathbb{N}_{2t}^n} L_{\mathbf{y}}(T_{\alpha}) T_{\alpha}(\mathbf{x}) \right) d\mu(\mathbf{x}) : \right. \\ &\quad \left. y_0 = 1; \mathbf{M}_{t-d_j}(g_j \mathbf{y}) \geq 0, \quad 0 \leq j \leq m \right\}, \quad (13) \end{aligned}$$

where  $L_{\mathbf{y}}$  is the Riesz functional in (1).

**Lemma 2** *The semidefinite relaxation (13) of (11) reads:*

$$\begin{aligned} \inf_{\mathbf{y}} \{ L_{\mathbf{y}}(f) : y_0 = 1; \\ \mathbf{M}_{t-d_j}(g_j \mathbf{y}) \geq 0, \quad 0 \leq j \leq m \}, \quad (14) \end{aligned}$$

which is exactly (6).

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

$$(T_\alpha)_{\alpha \in \mathbb{N}_{2t}^n} = \mathbf{D}_t \cdot (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}_{2t}^n} = \mathbf{D}_t \cdot \mathbf{v}_{2t}(\mathbf{x}) \quad (15)$$

and denote  $\mathbf{D}'_t$  the transpose of  $\mathbf{D}_t$ . The matrix  $\mathbf{D}_t$  is nonsingular with positive diagonal. Next, with  $f(\mathbf{x}) = \sum_\alpha f_\alpha \mathbf{x}^\alpha = \langle \mathbf{f}, \mathbf{v}_{2t}(\mathbf{x}) \rangle$ , write  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_{2t}^n} \tilde{f}_\alpha T_\alpha(\mathbf{x})$  in the basis  $(T_\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$ . Hence with  $\tilde{\mathbf{f}} = (\tilde{f}_\alpha)$  one obtains

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

Then

$$\begin{aligned} \int_{\Omega} f(\mathbf{x}) \left( \sum_{\alpha \in \mathbb{N}_{2t}^n} L_{\mathbf{y}}(T_\alpha) T_\alpha(\mathbf{x}) \right) d\mu(\mathbf{x}) &= \langle \tilde{\mathbf{f}}, (L_{\mathbf{y}}(T_\alpha))_\alpha \rangle \\ &= \langle (\mathbf{D}'_t)^{-1} \mathbf{f}, \mathbf{D}_t \cdot \mathbf{y} \rangle \\ &= \langle \mathbf{f}, \mathbf{y} \rangle = L_{\mathbf{y}}(f), \end{aligned}$$

which completes the proof.  $\square$

Of course by reverting the process of the above proof, the semidefinite program (14) can be transformed to (13) once a reference measure  $\mu$  with support exactly  $\Omega$  is defined with its associated orthonormal polynomials  $(T_\alpha)$ . Indeed, once  $\mu$  and the  $T_\alpha$ 's are defined, one may use the change of basis matrix  $\mathbf{D}$  in (15) to pass from (14) to (13).

Observe the difference between (8) which provides an upper bound on  $f^*$ , and (13) which provides a lower bound on  $f^*$ . In the former one searches for a *positive* polynomial density of degree  $2t$  (an SOS) while in the latter one searches for a *signed* polynomial density whose coefficients are pseudo-moments.

As the signed polynomial  $\mathbf{z} \mapsto \sum_\alpha T_\alpha(\boldsymbol{\xi}) T_\alpha(\mathbf{z})$  (with  $\boldsymbol{\xi} \in \Omega$  a global minimizer) is an optimal solution of (11), it is reasonable to expect that the hierarchy of lower bounds (13) (or, equivalently (6)) approximates  $f^*$  faster than the hierarchy of upper bounds (8).

**Corollary 1** *Let  $\Omega \subset \mathbb{R}^n$  be as in (4) and  $\mu$  be a finite Borel (reference) measure whose support is exactly  $\Omega$  and with associated sequence of orthonormal polynomials  $(T_\alpha)_{\alpha \in \mathbb{N}^n}$ . Let  $f^*$  be the global minimum on  $\Omega$ .*

*Let  $t$  be such that the semidefinite relaxation (13) (or equivalently (14)) is exact, i.e., if  $\rho_{2t} = f^*$ . If an optimal solution  $\mathbf{y}^*$  of (14) has a representing measure  $\phi^* \in \mathcal{M}(\Omega)_+$ , then an optimal polynomial density  $\sigma^* \in \mathbb{R}[\mathbf{x}]_{2t}$  of (13) satisfies:*

$$\sigma^*(\boldsymbol{\xi}) = \sum_{\alpha \in \mathbb{N}_{2t}^n} T_\alpha(\boldsymbol{\xi})^2 = K_{2t}(\boldsymbol{\xi}, \boldsymbol{\xi}),$$

*for all  $\boldsymbol{\xi} \in \text{supp}(\phi^*)$ , that is,  $\sigma^*(\boldsymbol{\xi})^{-1}$  is the Christoffel function evaluated at the global minimizer  $\boldsymbol{\xi} \in \Omega$ .*

*Proof* If  $\mathbf{y}^*$  has a representing measure  $\phi^* \in \mathcal{M}(\Omega)_+$  then necessarily  $f(\boldsymbol{\xi}) = f^*$  for all  $\boldsymbol{\xi} \in \text{supp}(\phi^*)$ ; see e.g. [3,6]. In particular, for every  $\boldsymbol{\xi} \in \text{supp}(\phi^*)$ , the vector  $\hat{\mathbf{y}} := (\boldsymbol{\xi}^\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$  is also an optimal solution of (14). Then

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

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

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

and so  $\sigma^*(\boldsymbol{\xi}) = \mathbf{K}_{2t}(\boldsymbol{\xi}, \boldsymbol{\xi})$ , i.e.,  $\sigma^*(\boldsymbol{\xi})^{-1}$  is the Christoffel function associated with  $\mu$ , evaluated at  $\boldsymbol{\xi} \in \Omega$ .  $\square$

**Discussion.** Observe that the formulation (13) does not require that the set  $\Omega$  is a “simple” set as it is required in §2.3. Indeed the orthonormal polynomials  $(T_\alpha)$  are only used to provide an *interpretation* of the hierarchy of lower bounds (14) (and its dual (7)). On the other hand, for the hierarchy of upper bounds (8),  $\Omega$  indeed needs to be a “simple” set for computational purposes. Indeed one needs the numerical value of the moments of  $\mu$  for practical implementation of (8).

Lemma 2 shows that the Moment-SOS hierarchy described in [3,6] amounts to compute a hierarchy of *signed* polynomial densities with respect to some reference measure  $\mu$  with support exactly  $\Omega$ . When the step- $t$  relaxation is exact (which takes place generically [9]), the resulting optimal density  $\sigma^*$  in Corollary 1 is nothing less than the polynomial  $\mathbf{x} \mapsto K_{2t}(\boldsymbol{\xi}, \mathbf{x})$  where  $\boldsymbol{\xi}$  is a global minimizer of  $f$  on  $\Omega$ , and  $K_{2t}(\boldsymbol{\xi}, \mathbf{x})$  is the celebrated Cristoffel-Darboux kernel in approximation theory, and  $\sigma^*(\boldsymbol{\xi}) = K_{2t}(\boldsymbol{\xi}, \boldsymbol{\xi})$  is the reciprocal of the Christoffel function evaluated at a global minimizer  $\boldsymbol{\xi}$ .

## 2.5 A variant hierarchy of upper bounds

Consider the optimization problems

$$\begin{aligned} \bar{u}_t &= \inf_{\sigma \in \mathbb{R}[\mathbf{x}]_{2t}} \left\{ \int_{\Omega} f(\mathbf{x}) \left( \sum_{\alpha \in \mathbb{N}_{2t}^n} \sigma_\alpha T_\alpha(\mathbf{x}) \right) d\mu(\mathbf{x}); \right. \\ &\sigma_0 = 1; \sum_{\alpha} \sigma_\alpha T_\alpha = \sum_{j=0}^m \psi_j g_j, \\ &\left. \psi_j \in \Sigma[\mathbf{x}]_{t-d_j}, \quad j = 0, \dots, m \right\}, \end{aligned} \quad (16)$$

indexed by  $t \in \mathbb{N}$ . Notice that the upper bound  $\bar{u}_t$  is *stronger* than the upper bound  $\tau_t$  in (8) (i.e.,  $\bar{u}_t \leq \tau_t$ ). Indeed the condition that  $\sigma$  is an SOS in (8) is now relaxed to the weaker condition in (16) that  $\sigma$  has a

Putinar's certificate of positivity on  $\Omega$ . The dual of (16) reads:

$$\bar{u}_t^* = \sup_{\lambda, \mathbf{y}} \left\{ \lambda : f - \lambda = \sum_{\alpha \in \mathbb{N}_{2t}^n} L_{\mathbf{y}}(T_{\alpha}) T_{\alpha} \right. \\ \left. \mathbf{M}_{t-d_j}(g_j \mathbf{y}) \geq 0, \quad j = 0, \dots, m \right\}. \quad (17)$$

Slater's condition holds for (16) so that  $\bar{u}_t = \bar{u}_t^*$  and (17) has an optimal solution  $(\bar{u}_t, \mathbf{y}^*)$ . Again  $\bar{u}_t > f^*$  and  $\mathbf{y}^*$  cannot have a representing measure  $\phi$  which is the Dirac  $\delta_{\xi}$  at a global minimizer  $\xi \in \Omega$ . Indeed suppose the contrary. Then

$$\underbrace{f(\xi) - \bar{u}_t}_{< 0} = \int_{\Omega} \sum_{\alpha \in \mathbb{N}_{2t}^n} T_{\alpha}(\mathbf{z}) T_{\alpha}(\xi) d\phi(\mathbf{z}) = K_{2t}(\xi, \xi) > 0.$$

In the Table 1 below, a final summary of our comparison displays a remarkable symmetry.

Lower bounds	Upper bounds
Primal	Primal
$\inf_{\mathbf{y}} \int_{\Omega} f \cdot \left( \sum_{\alpha} L_{\mathbf{y}}(T_{\alpha}) T_{\alpha} \right) d\mu$	$\inf_{\sigma, \psi_j} \int_{\Omega} f \cdot \left( \sum_{\alpha} \sigma_{\alpha} T_{\alpha} \right) d\mu$
s.t. $y_0 = 1;$	s.t. $\sigma_0 = 1;$
$\mathbf{M}_t(g_j \mathbf{y}) \geq 0, 0 \leq j \leq m.$	$\sum_{\sigma} \sigma_{\alpha} T_{\alpha} = \sum_{j=0}^m \psi_j g_j.$
	$\psi_j \in \Sigma[\mathbf{x}]_{t-d_j}$
Dual	Dual
$\sup_{\lambda, \psi_j} \lambda :$	$\sup_{\lambda, \mathbf{y}} \lambda :$
s.t. $f - \lambda = \sum_{j=0}^m \psi_j g_j$	s.t. $f - \lambda = \sum_{\alpha \in \mathbb{N}_{2t}^n} L_{\mathbf{y}}(T_{\alpha}) T_{\alpha}$
$\psi_j \in \Sigma[\mathbf{x}]_{t-d_j}$	$\mathbf{M}_t(g_j \mathbf{y}) \geq 0, 0 \leq j \leq m$

**Table 1** Comparing hierarchies of upper and lower bounds

### 3 Conclusions

The two Moment-SOS hierarchies of lower and upper bounds on the global minimum are both interpreted as searching for a polynomial density of increasing degree with respect to an arbitrary reference measure with support  $\Omega$ . In the former one searches for a *signed* polynomial density whose coefficients are moments of

the Dirac measure at some global minimizer, and the global optimum eventually is reached at some (in general early) step of the hierarchy. In contrast, in the latter one searches for a *positive* polynomial density and the convergence cannot be finite; surprisingly, good  $O(1/t^2)$  rates of convergence can be achieved but its practical implementation requires to know all moments of the reference measure, and therefore is limited to simple sets (e.g., box, ellipsoid, Euclidean sphere or unit ball, simplex, discrete hypercube, or their image by an affine map). On the other hand, if practice reveals fast (and even finite) convergence of the lower bound hierarchy, good rates of convergence are still not available (except in the special case of the unit Euclidean sphere).

### References

1. Dunkl, C., Xu, Y.: Orthogonal Polynomials of Several Variables. Cambridge University Press, Cambridge, UK (2001)
2. Lasserre, J.B.: Optimisation globale et théorie des moments. C.R. Acad. Sci. Sér. I-Math **331**(11), 929–934 (2000)
3. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001)
4. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press, London, UK (2009)
5. Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. SIAM J. Optim. **21**, 864–885 (2011)
6. Lasserre, J.B.: An Introduction to Polynomial and Semi-Algebraic Optimization. Cambridge University Press, Cambridge, UK (2015)
7. Lasserre, J.B.: Connecting optimization with spectral analysis of tri-diagonal matrices. Math. Program. **190**, 795–809 (2021). DOI: 10.1007/s10107-020-01549-3
8. Lasserre, J.B.: The Moment-SOS Hierarchy and the Christoffel-Darboux kernel. Optim. Letters **15**, 1835–1845 (2021). DOI: 10.1007/s11590-021-01713-4
9. Nie, J.: Optimality conditions and finite convergence of lasserre's hierarchy. Math. Program. Ser. A **146**(1-2), 97–121 (2014)
10. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Ind. Univ. Math. J. **42**, 969–984 (1993)
11. Slot, L., Laurent, M.: Near-optimal analysis of Lasserre's univariate measure-based bounds for multivariate polynomial optimization. Math. Program. **188**, 443–460 (2021). DOI: 10.1007/s10107-020-01586-y
12. Slot, L., Laurent, M.: Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. Math. Program. **193**, 831–871 (2022). DOI: 10.1007/s10107-020-01468-3
13. Slot, L., Laurent, M.: Sum-of-squares hierarchies for binary polynomial optimization. Math. Program. (2022). DOI: 10.1007/s10107-021-01745-9