



HAL
open science

A new look at nonnegativity on closed sets and polynomial optimization

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. 2010. hal-00512560v1

HAL Id: hal-00512560

<https://hal.science/hal-00512560v1>

Preprint submitted on 30 Aug 2010 (v1), last revised 11 May 2011 (v3)

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.

A NEW LOOK AT NONNEGATIVITY ON CLOSED SETS AND POLYNOMIAL OPTIMIZATION

JEAN B. LASSERRE

ABSTRACT. We first show that a continuous function f is nonnegative on a closed set $\mathbf{K} \subseteq \mathbb{R}^n$ if and only if (countably many) moment matrices of some signed measure $d\nu = fd\mu$ with $\text{supp } \mu = \mathbf{K}$, are all positive semidefinite (if \mathbf{K} is compact μ is an arbitrary finite Borel measure with $\text{supp } \mu = \mathbf{K}$). In particular, we obtain a convergent explicit hierarchy of semidefinite (outer) approximations with *no* lifting, of the cone of nonnegative polynomials of degree at most d . When used in polynomial optimization on certain simple closed sets \mathbf{K} (like e.g., the whole space \mathbb{R}^n , the positive orthant, a box, a simplex, or the vertices of the hypercube), it provides a nonincreasing sequence of upper bounds which converges to the global minimum by solving a hierarchy of semidefinite programs with only one variable. This convergent sequence of upper bounds complements the convergent sequence of lower bounds obtained by solving a hierarchy of semidefinite relaxations as in e.g. [7].

1. INTRODUCTION

This paper is concerned with a concrete characterization of continuous functions that are nonnegative on a closed set $\mathbf{K} \subseteq \mathbb{R}^n$ and its application for optimization purposes. By concrete we mean a systematic procedure, e.g. a numerical test that can be implemented by an algorithm, at least in some interesting cases. For polynomials, Stengle's Nichtnegativstellensatz [17] provides a certificate of nonnegativity (or absence of nonnegativity) on a semi-algebraic set. Moreover, in principle, this certificate can be obtained by solving a single semidefinite program (although the size of this semidefinite program is far beyond the capabilities of today's computers). Similarly, for compact basic semi-algebraic sets, Schmüdgen's and Putinar's Positivstellensätze [15, 13] provide certificates of strict positivity that can be obtained by solving finitely many semidefinite programs (of increasing size). Extensions of those certificates to some algebras of non-polynomial functions have been recently proposed in Lasserre and Putinar [9] and in Marshall and Netzer [11]. However, and to the best of our knowledge, there is still no hierarchy of explicit (outer or inner) semidefinite approximations (with or without lifting) of the cone of polynomials nonnegative on a closed set \mathbf{K} , even if \mathbf{K} is compact and basic semi-algebraic.

Contribution: In this paper, we present a different approach based on a new (at least to the best of our knowledge) and simple characterization of continuous functions that are nonnegative on a closed set $\mathbf{K} \subseteq \mathbb{R}^n$. This characterization

1991 *Mathematics Subject Classification.* 90C25 28C15.

Key words and phrases. closed sets; nonnegative functions; nonnegative polynomials; semidefinite approximations; moments.

involves a *single* (but known) measure μ with $\text{supp } \mu = \mathbf{K}$, and sums of squares of polynomials. Namely, our contribution is twofold:

(a) We first show that a continuous function f is nonnegative on a closed set $\mathbf{K} \subseteq \mathbb{R}^n$ if and only if $\int h^2 f d\mu$ is nonnegative for all polynomials $h \in \mathbb{R}[\mathbf{x}]$, where μ is a finite Borel measure with $\text{supp } \mu = \mathbf{K}$. The measure μ is arbitrary if \mathbf{K} is compact. If \mathbf{K} is not compact, all moments of the signed measure $f d\mu$ must be finite. For instance, if f is a polynomial one may choose μ to be the restriction to \mathbf{K} of the product of normal distributions (so that all polynomials are integrable) and consider those closed sets \mathbf{K} such that $\text{supp } \mu = \mathbf{K}$ (for instance \mathbf{K} is the closure of some open set).

Equivalently, f is nonnegative on \mathbf{K} if and only if every element of the countable family \mathcal{T} of moment matrices associated with the signed Borel measure $f d\mu$, is positive semidefinite. The absence of nonnegativity on \mathbf{K} can be *certified* by exhibiting a polynomial $h \in \mathbb{R}[\mathbf{x}]$ such that $\int h^2 f d\mu < 0$, or equivalently, when some moment matrix in the family \mathcal{T} is not positive semidefinite. And so, interestingly, as for nonnegativity or positivity, our certificate for absence of nonnegativity is also in terms of sums of squares. When f is a polynomial, these moment matrices are easily obtained from the moments of μ and this criterion for absence of nonnegativity complements Stengle's Nichtnegativstellensatz [17] (which provides a certificate of nonnegativity on a semi-algebraic set \mathbf{K}) or Schmüdgen and Putinar's Positivstellensätze [15, 13] (for certificates of strict positivity on compact basic semi-algebraic sets). At last but not least, we obtain a convergent *explicit* hierarchy of semidefinite (outer) approximations with *no* lifting, of the cone \mathcal{C}_d of nonnegative polynomials of degree at most $2d$. That is, we obtain a nested sequence $\mathcal{C}_d^0 \supset \dots \mathcal{C}_d^k \supset \dots \supset \mathcal{C}_d$ such that each \mathcal{C}_d^k is a spectrahedron defined solely in terms of the vector of coefficients of the polynomial, with no additional variable (i.e., no projection is needed). Similar explicit hierarchies can be obtained for the cone of polynomials nonnegative on a closed set \mathbf{K} , provided all moments of an appropriate measure μ can be obtained. To the best of our knowledge, this is first result of this kind.

(b) As a potential application, we consider the problem of minimizing a polynomial f on some unbounded closed sets (e.g., $\mathbf{K} = \mathbb{R}^n$ or $\mathbf{K} = \mathbb{R}_+^n$), or on any compact set \mathbf{K} for which one may compute all moments of a measure μ with $\text{supp } \mu = \mathbf{K}$. Typical examples of such sets \mathbf{K} are a box, a ball, a simplex, or the vertices of the hypercube. We then provide a hierarchy of semidefinite programs (with only one variable!) whose optimal values form a monotone sequence of *upper* bounds which converges to the global minimum f^* . This is in contrast with the hierarchy of semidefinite relaxations defined in Lasserre [7, 8] which provide a nondecreasing sequence of *lower* bounds that also converges to f^* . Hence these two convergent hierarchies of upper and lower bounds complement each other and permit to locate the global minimum f^* in smaller and smaller intervals. Notice that convergence of the hierarchy of convex relaxations in [7] is guaranteed only for compact basic semi-algebraic sets, whereas for the new hierarchy of upper bounds, the only requirement is to know all moments of a measure μ with $\text{supp } \mu = \mathbf{K}$. On the other hand, computing such moments is possible only for relatively simple (but not necessarily compact) sets.

When $\mathbf{K} = \mathbb{R}^n$, we also obtain a nonincreasing sequence of upper bounds that converges to f^* even if f^* is not attained, an alternative and/or complement to the hierarchy of convex relaxations provided in Schweighofer [16] (based on gradient

tentacles) and in Hà and So'n [5] (based on the truncated tangency variety), which both provide again a monotone sequence of lower bounds.

Finally, we also give a very simple interpretation of the hierarchy of dual semidefinite programs, which provide some information on the location of global minimizers.

2. NOTATION AND PRELIMINARY RESULTS

Let $\mathbb{R}[\mathbf{x}]$ be the ring of polynomials in the variables $\mathbf{x} = (x_1, \dots, x_n)$. Denote by $\mathbb{R}[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]$ the vector space of polynomials of degree at most d , which forms a vector space of dimension $s(d) = \binom{n+d}{d}$, with e.g., the usual canonical basis (\mathbf{x}^α) of monomials. Also, denote by $\Sigma[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_{2d}$ the space of sums of squares (s.o.s.) polynomials of degree at most $2d$. If $f \in \mathbb{R}[\mathbf{x}]_d$, write $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \mathbf{x}^\alpha$ in the canonical basis and denote by $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(d)}$ its vector of coefficients.

Moment matrix. With $\mathbf{y} = (y_\alpha)$ being a sequence indexed in the canonical basis (\mathbf{x}^α) of $\mathbb{R}[\mathbf{x}]$, let $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$ be the linear functional

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

and let $\mathbf{M}_d(\mathbf{y})$ be the symmetric matrix with rows and columns indexed in the canonical basis (\mathbf{x}^α) , and defined by:

$$(2.1) \quad \mathbf{M}_d(\mathbf{y})(\alpha, \beta) := L_{\mathbf{y}}(\mathbf{x}^{\alpha+\beta}) = y_{\alpha+\beta}, \quad \alpha, \beta \in \mathbb{N}_d^n$$

with $\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| (= \sum_i \alpha_i) \leq d\}$.

If \mathbf{y} has a representing measure μ , i.e., if $y_\alpha = \int \mathbf{x}^\alpha d\mu$ for every $\alpha \in \mathbb{N}^n$, then

$$\langle \mathbf{f}, \mathbf{M}_d(\mathbf{y})\mathbf{f} \rangle = \int f(\mathbf{x})^2 d\mu(\mathbf{x}) \geq 0, \quad \forall f \in \mathbb{R}[\mathbf{x}]_d,$$

and so $\mathbf{M}_d(\mathbf{y}) \succeq 0$, where for a real symmetric matrix \mathbf{A} , the notation $\mathbf{A} \succeq 0$ (resp. $\mathbf{A} \succ 0$) stands for \mathbf{A} is positive semidefinite (resp. positive definite). A measure μ is said to be moment determinate if there is no other measure with same moments. In particular, and as an easy consequence of the Stone-Weierstrass theorem, every measure with compact support is determinate.

Not every sequence \mathbf{y} satisfying $\mathbf{M}_d(\mathbf{y}) \succeq 0$, $d \in \mathbb{N}$, has a representing measure. However:

Proposition 2.1 (Berg [2]). *Let $\mathbf{y} = (y_\alpha)$ be such that $\mathbf{M}_d(\mathbf{y}) \succeq 0$, for every $d \in \mathbb{N}$. Then:*

(a) *The sequence \mathbf{y} has a representing measure whose support is contained in the ball $[-a, a]^n$ if there exists $a, c > 0$ such that $|y_\alpha| \leq ca^{|\alpha|}$ for every $\alpha \in \mathbb{N}^n$.*

(b) *The sequence \mathbf{y} has a unique representing measure μ on \mathbb{R}^n if*

$$(2.2) \quad \sum_{t=1}^{\infty} L_{\mathbf{y}}(x_i^{2t})^{-1/2t} = +\infty$$

In addition, the measure μ is moment determinate.

Condition (b) is an extension to the multivariate case of Carleman's condition in the univariate case and is due to Nussbaum [12]. For more details see e.g. Berg [2] and/or Maserick and Berg [6].

Localizing matrix. Similarly, with $\mathbf{y} = (y_\alpha)$ and $f \in \mathbb{R}[\mathbf{x}]$ written

$$\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\gamma \in \mathbb{N}^n} f_\gamma \mathbf{x}^\gamma,$$

let $\mathbf{M}_d(f \mathbf{y})$ be the symmetric matrix with rows and columns indexed in the canonical basis (\mathbf{x}^α) , and defined by:

$$(2.3) \quad \mathbf{M}_d(f \mathbf{y})(\alpha, \beta) := L_{\mathbf{y}}(f(\mathbf{x}) \mathbf{x}^{\alpha+\beta}) = \sum_{\gamma} f_\gamma y_{\alpha+\beta+\gamma}, \quad \forall \alpha, \beta \in \mathbb{N}_d^n.$$

If \mathbf{y} has a representing measure μ , then $\langle \mathbf{g}, \mathbf{M}_d(f \mathbf{y}) \mathbf{g} \rangle = \int g^2 f d\mu$, and so if μ is supported on the set $\{\mathbf{x} : f(\mathbf{x}) \geq 0\}$, then $\mathbf{M}_d(f \mathbf{y}) \succeq 0$ for all $d = 0, 1, \dots$ because

$$(2.4) \quad \langle \mathbf{g}, \mathbf{M}_d(f \mathbf{y}) \mathbf{g} \rangle = \int g(\mathbf{x})^2 f(\mathbf{x}) d\mu(\mathbf{x}) \geq 0, \quad \forall g \in \mathbb{R}[\mathbf{x}]_d.$$

3. NONNEGATIVITY ON CLOSED SETS

Recall that if \mathbf{X} is a separable metric space with Borel σ -field \mathcal{B} , the support $\text{supp } \mu$ of a Borel measure μ on \mathbf{X} is the (unique) smallest closed set $B \in \mathcal{B}$ such that $\mu(\mathbf{X} \setminus B) = 0$. Also, if f is a nonnegative measurable function, then the set function $B \mapsto \nu(B) := \int_B f d\mu$, $B \in \mathcal{B}$, defines a measure (which is finite if f is μ -integrable); see e.g. Royden [14, p. 276 and p. 408]. If f is not nonnegative then setting $B_1 := \{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ and $B_2 := \{\mathbf{x} : f(\mathbf{x}) < 0\}$, $d\nu = f d\mu$ is a *signed* measure that can be written as the difference $\nu_1 - \nu_2$ of two (positive) measures ν_1, ν_2 , where:

$$\nu_1(B) = \int_{B_1 \cap B} f d\mu, \quad \nu_2(B) = - \int_{B_2 \cap B} f d\mu, \quad \forall B \in \mathcal{B}.$$

We first provide the following auxiliary result which is also of self-interest.

Lemma 3.1. *Let \mathbf{X} be a separable metric space, $\mathbf{K} \subseteq \mathbf{X}$ a closed set, and μ a Borel measure on \mathbf{X} with $\text{supp } \mu = \mathbf{K}$. A continuous function $f : \mathbf{X} \rightarrow \mathbb{R}$ is nonnegative on \mathbf{K} if and only if the signed Borel measure $\nu(B) = \int_{\mathbf{K} \cap B} f d\mu$, $B \in \mathcal{B}$, is a positive measure.*

Proof. The *only if part* is straightforward. For the *if part*, if ν is a positive measure then f is the Radon-Nikodym derivative of ν with respect to μ and $f(\mathbf{x}) \geq 0$ for μ -almost all $\mathbf{x} \in \mathbf{K}$. That is, there is a Borel set $\mathbf{G} \subset \mathbf{K}$ such that $\mu(\mathbf{G}) = 0$ and $f(\mathbf{x}) \geq 0$ on $\mathbf{K} \setminus \mathbf{G}$. Next, observe that $\overline{\mathbf{K} \setminus \mathbf{G}} \subset \mathbf{K}$ and $\mu(\overline{\mathbf{K} \setminus \mathbf{G}}) = \mu(\mathbf{K})$. Therefore $\overline{\mathbf{K} \setminus \mathbf{G}} = \mathbf{K}$ because $\text{supp } \mu (= \mathbf{K})$ is the unique smallest closed set such that $\mu(\mathbf{X} \setminus \mathbf{K}) = 0$. Hence, let $\mathbf{x} \in \mathbf{K}$ be fixed, arbitrary. As $\mathbf{K} = \overline{\mathbf{K} \setminus \mathbf{G}}$, there is a sequence $(\mathbf{x}_k) \subset \mathbf{K} \setminus \mathbf{G}$, $k \in \mathbb{N}$, with $\mathbf{x}_k \rightarrow \mathbf{x}$ as $k \rightarrow \infty$. But since f is continuous and $f(\mathbf{x}_k) \geq 0$ for every $k \in \mathbb{N}$, we obtain the desired result $f(\mathbf{x}) \geq 0$. \square

Lemma 3.1 itself (of which we have not been able to find a trace in the literature) is a characterization of nonnegativity on \mathbf{K} for a continuous function f on \mathbf{X} . However, one goal of this paper is to provide a more concrete characterization. To do so we first consider the case of a compact set $\mathbf{K} \subset \mathbb{R}^n$.

3.1. The compact case. Let \mathbf{K} be a compact subset of \mathbb{R}^n . For simplicity, and with no loss of generality, we may and will assume that $\mathbf{K} \subseteq [-1, 1]^n$.

Theorem 3.2. *Let $\mathbf{K} \subseteq [-1, 1]^n$ be compact and let μ be an arbitrary, fixed, finite Borel measure on \mathbf{K} with $\text{supp } \mu = \mathbf{K}$. Let f be a continuous function on \mathbb{R}^n . Then:*

(a) *f is nonnegative on \mathbf{K} if and only if*

$$(3.1) \quad \int_{\mathbf{K}} g^2 f d\mu \geq 0, \quad \forall g \in \mathbb{R}[\mathbf{x}],$$

or, equivalently, if and only if

$$(3.2) \quad \mathbf{M}_d(\mathbf{z}) \succeq 0, \quad d = 0, 1, \dots$$

where $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, with $z_\alpha = \int \mathbf{x}^\alpha f(\mathbf{x}) d\mu(\mathbf{x})$, and with $\mathbf{M}_d(\mathbf{z})$ as in (2.1).

If in addition $f \in \mathbb{R}[\mathbf{x}]$ then (3.2) reads $\mathbf{M}_d(f\mathbf{y}) \succeq 0$, $d = 0, 1, \dots$, where $\mathbf{M}_d(f\mathbf{y})$ is the localizing matrix defined in (2.3).

(b) *If in addition to be continuous, f is also concave on \mathbf{K} , then f is nonnegative on $\text{co}(\mathbf{K})$ if and only if (3.1) holds.*

Proof. The *only if* part is straightforward. Indeed, if $f \geq 0$ on \mathbf{K} then $\mathbf{K} \subseteq \{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ and so for any finite Borel measure μ on \mathbf{K} , $\int_{\mathbf{K}} g^2 f d\mu \geq 0$ for every $g \in \mathbb{R}[\mathbf{x}]$. Next, if f is concave and $f \geq 0$ on $\text{co}(\mathbf{K})$ then $f \geq 0$ on \mathbf{K} and so the “only if” part of (b) also follows.

If part. Let ν be the finite signed Borel measure $\nu(B) = \int_B f d\mu$, $B \in \mathcal{B}$, and let $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, be the sequence defined by:

$$(3.3) \quad z_\alpha := \int_{\mathbf{K}} \mathbf{x}^\alpha d\nu(\mathbf{x}) = \int_{\mathbf{K}} \mathbf{x}^\alpha f(\mathbf{x}) d\mu(\mathbf{x}), \quad \forall \alpha \in \mathbb{N}^n.$$

Then the condition

$$\int_{\mathbf{K}} g(\mathbf{x})^2 f(\mathbf{x}) d\mu(\mathbf{x}) \geq 0, \quad \forall f \in \mathbb{R}[\mathbf{x}]_d,$$

reads $\langle \mathbf{g}, \mathbf{M}_d(\mathbf{z})\mathbf{g} \rangle \geq 0$ for all $\mathbf{g} \in \mathbb{R}^{s(d)}$, that is, $\mathbf{M}_d(\mathbf{z}) \succeq 0$, where $\mathbf{M}_d(\mathbf{z})$ is the moment matrix defined in (2.1). And so (3.1) implies $\mathbf{M}_d(\mathbf{z}) \succeq 0$ for every $d \in \mathbb{N}$. Moreover, as $\mathbf{K} \subseteq [-1, 1]^n$,

$$|z_\alpha| \leq c := \int_{\mathbf{K}} |f| d\mu, \quad \forall \alpha \in \mathbb{N}^n.$$

Hence, by Proposition 2.1, \mathbf{z} is the moment sequence of a finite (positive) Borel measure ψ on $[-1, 1]^n$, that is, as $\text{supp } \nu \subseteq \mathbf{K} \subseteq [-1, 1]^n$,

$$(3.4) \quad \int_{[-1, 1]^n} \mathbf{x}^\alpha d\nu(\mathbf{x}) = \int_{[-1, 1]^n} \mathbf{x}^\alpha d\psi(\mathbf{x}), \quad \forall \alpha \in \mathbb{N}^n.$$

Define the two Borel sets $B_1 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \geq 0\}$ and $B_2 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) < 0\}$. The signed Borel measure ν can be written as

$$\nu(B) = \nu_1(B) - \nu_2(B), \quad B \in \mathcal{B},$$

where the two positive measures ν_1, ν_2 are defined by

$$\nu_1(B) := \int_{B_1 \cap B} f(\mathbf{x}) d\mu(\mathbf{x}); \quad \nu_2(B) := - \int_{B_2 \cap B} f(\mathbf{x}) d\mu(\mathbf{x}), \quad B \in \mathcal{B}.$$

But then (3.4) yields

$$\int_{[-1,1]^n} \mathbf{x}^\alpha d\nu_1(\mathbf{x}) = \int_{[-1,1]^n} \mathbf{x}^\alpha d(\nu_2 + \psi)(\mathbf{x}), \quad \forall \alpha \in \mathbb{N}^n,$$

which in turn implies $\nu_1 = \nu_2 + \psi$ because measures on a compact set are determinate. Next, this implies $0 = \nu_1(B_2) = \nu_2(B_2) + \psi(B_2) \geq \nu_2(B_2)$, and so $\nu = \nu_1$, that is, ν is a positive Borel measure on \mathbf{K} . Hence by Lemma 3.1, $f(\mathbf{x}) \geq 0$ on \mathbf{K} .

If in addition $f \in \mathbb{R}[\mathbf{x}]$, the sequence $\mathbf{z} = (z_\alpha)$ is obtained as a linear combination of (y_α) . Indeed if $f(\mathbf{x}) = \sum_\beta f_\beta \mathbf{x}^\beta$ then

$$z_\alpha = \sum_{\beta \in \mathbb{N}^n} f_\beta y_{\alpha+\beta}, \quad \forall \alpha \in \mathbb{N}^n,$$

and so in (3.2), $\mathbf{M}_d(\mathbf{z})$ is nothing less than the *localizing* matrix $\mathbf{M}_d(f \mathbf{y})$ associated with $\mathbf{y} = (y_\alpha)$ and $f \in \mathbb{R}[\mathbf{x}]$, defined in (2.3), and (3.2) reads $\mathbf{M}_d(f \mathbf{y}) \succeq 0$ for all $d = 0, 1, \dots$

Finally, if f is concave then $f \geq 0$ on \mathbf{K} implies $f \geq 0$ on $\text{co}(\mathbf{K})$, and so the *only* if part of (b) also follows. \square

Therefore, to check whether a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is nonnegative on \mathbf{K} , it suffices to check if every element of the countable family of real symmetric matrices $(\mathbf{M}_d(f \mathbf{y}))$, $d \in \mathbb{N}$, is positive semidefinite.

3.2. The non-compact case. We now consider the more delicate case where \mathbf{K} is a closed set of \mathbb{R}^n , not necessarily compact.

Given a continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we need choose a measure μ on \mathbb{R}^n such that

- $\mathbf{x} \mapsto \mathbf{x}^\alpha f$ is μ -integrable, for all $\alpha \in \mathbb{N}^n$, and
- $\text{supp } \mu = \mathbf{K}$.

So consider measures μ of the form:

$$(3.5) \quad \mu(B) := \int_{\mathbf{K} \cap B} d\nu_1(x_1) \cdots d\nu_n(x_n), \quad \forall B \in \mathcal{B},$$

where (ν_i) , $i = 1, \dots, n$, is a family of finite Borel measures on \mathbb{R} such that for every $i = 1, \dots, n$, and every fixed $s \in \mathbb{N}$, the measure $\psi_s(B) = \int_B x^{2s} d\nu_i(x)$, $B \in \mathcal{B}(\mathbb{R})$, satisfy Carleman's condition (2.2). That is, if $\mathbf{z}^i = (z_k^i)$, $k \in \mathbb{N}$, is the moment sequence of ν_i , then

$$(3.6) \quad \sum_{k=1}^{\infty} L_{\mathbf{z}^i}(x^{2(k+s)})^{-1/2k} = +\infty, \quad i = 1, \dots, n.$$

For instance, if f is a polynomial and $\mathbf{K} \subseteq \mathbb{R}^n$ is the closure of an open set, then for every $i = 1, \dots, n$, one may choose the normal distribution $\nu_i(B) = (2\pi)^{-1/2} \int_B e^{-x^2/2} dx$, $B \in \mathcal{B}(\mathbb{R})$, so that,

$$(3.7) \quad \mu(B) := \frac{1}{(\sqrt{2\pi})^n} \int_{\mathbf{K} \cap B} e^{-\|\mathbf{x}\|^2/2} d\mathbf{x}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Notice that with such a μ , every polynomial is μ -integrable.

Theorem 3.3. *Let f be a continuous function on \mathbb{R}^n and let $\mathbf{K} \subseteq \mathbb{R}^n$ be closed. Choose a measure μ of the form (3.5) and such that (3.6) holds, $\text{supp } \mu = \mathbf{K}$, and $\mathbf{x}^\alpha f$ is μ -integrable, for every $\alpha \in \mathbb{N}^n$. Then (a) and (b) of Theorem 3.2 hold.*

Proof. The *only if* part is exactly the same as in the proof of Theorem 3.2. For the *if* part, let ν and $\mathbf{z} = (z_\alpha)$ be the finite signed Borel measure and the sequence as in (3.3). The sequence \mathbf{z} is well defined because $\mathbf{x} \mapsto \mathbf{x}^\alpha f(\mathbf{x})$ is μ -integrable. The same arguments show that $\mathbf{M}_d(\mathbf{z}) \succeq 0$ for every $d \in \mathbb{N}$. Next by Lemma 6.1, the sequence \mathbf{z} satisfies the generalized Carleman's condition (2.2) and so, by Proposition 2.1, \mathbf{z} is the moment sequence of a measure ψ on \mathbb{R}^n and ψ is determinate. In addition from the definition (3.3) of \mathbf{z} we have

$$(3.8) \quad \int_{\mathbb{R}^n} \mathbf{x}^\alpha d\psi(\mathbf{x}) = z_\alpha = \int_{\mathbf{K}} \mathbf{x}^\alpha d\nu(\mathbf{x}), \quad \forall \alpha \in \mathbb{N}^n.$$

Define the two Borel sets $B_1 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \geq 0\}$ and $B_2 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) < 0\}$. The signed Borel measure ν can be written as

$$\nu(B) = \varphi_1(B) - \varphi_2(B), \quad B \in \mathcal{B},$$

where the two positive measures φ_1, φ_2 on \mathbf{K} are defined by

$$\varphi_1(B) := \int_{B_1 \cap B} f(\mathbf{x}) d\mu(\mathbf{x}); \quad \varphi_2(B) := - \int_{B_2 \cap B} f(\mathbf{x}) d\mu(\mathbf{x}), \quad B \in \mathcal{B}.$$

But then (3.8) reads

$$(3.9) \quad \int_{\mathbb{R}^n} \mathbf{x}^\alpha d(\psi + \varphi_2)(\mathbf{x}) = \int_{\mathbf{K}} \mathbf{x}^\alpha d\varphi_1(\mathbf{x}), \quad \forall \alpha \in \mathbb{N}^n.$$

Let $\mathbf{v} = (v_\alpha)$, $\alpha \in \mathbb{N}^n$, be the sequence of moments associated with φ_1 . Of course, $\mathbf{M}_d(\mathbf{v}) \succeq 0$ for all $d \in \mathbb{N}$. Next, by inspection of its proof, one may see that Lemma 6.1 also holds for the measure φ_1 , the sequence \mathbf{v} and $\mathbf{K} = B_1$. Hence by Proposition 2.1, φ_1 is moment determinate, so that (3.9) yields $\varphi_1 = \psi + \varphi_2$. But this implies

$$0 = \varphi_1(B_2) = \psi(B_2) + \varphi_2(B_2) \geq \varphi_2(B_2),$$

so that $\varphi_2(B_2) = 0$. But this in turn implies that $\nu = \varphi_1 = \psi$, and so ν is a positive measure, which in turn implies that f is the Radon-Nikodym derivative of ν with respect to μ . The rest of the proof is exactly the same as for proof of Theorem 3.2. \square

3.3. The cone of nonnegative polynomials. The convex cone $\mathcal{C}_d \subset \mathbb{R}[\mathbf{x}]_{2d}$ of nonnegative polynomials of degree at most $2d$ (a nonnegative polynomial has necessarily even degree) is much harder to characterize than its subcone $\Sigma[\mathbf{x}]_d$ of sums of squares. Indeed, while the latter has a simple semidefinite representation with lifting (i.e. $\Sigma[\mathbf{x}]_d$ is the projection in $\mathbb{R}^{s(2d)}$ of a spectrahedron¹ in a higher dimensional space), so far there is no such simple representation for the former. In addition, when d is fixed, Blekherman [3] has shown that after proper normalization, the ‘‘gap’’ between \mathcal{C}_d and $\Sigma[\mathbf{x}]_d$ increases unboundedly with the number of variables.

We next provide a convergent hierarchy of (outer) semidefinite approximations (\mathcal{C}_d^k), $k \in \mathbb{N}$, of \mathcal{C}_d where each \mathcal{C}_d^k has a semidefinite representation with *no* lifting (i.e., no projection is needed and \mathcal{C}_d^k is a spectrahedron). To the best of our knowledge, this is the first result of this kind.

Recall that with every $f \in \mathbb{R}[\mathbf{x}]_d$ is associated its vector of coefficients $\mathbf{f} = (f_\alpha)$, $\alpha \in \mathbb{N}_d^n$, in the canonical basis of monomials, and conversely, with every $\mathbf{f} \in \mathbb{R}^{s(d)}$

¹A spectrahedron is the intersection of the cone of positive semidefinite matrices with an affine-linear space. Its algebraic representation is called a Linear Matrix Inequality (LMI).

is associated a polynomial $f \in \mathbb{R}[\mathbf{x}]_d$ with vector of coefficients $\mathbf{f} = (f_\alpha)$ in the canonical basis. Recall that for every $k = 1, \dots$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k+1} e^{-x^2/2} dx = 0; \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-x^2/2} dx = \prod_{j=1}^k (2j-1).$$

Corollary 3.4. *Let μ be the probability measure on \mathbb{R}^n which is the n -times product of the normal distribution on \mathbb{R} , and so with moments $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$,*

$$(3.10) \quad y_\alpha = \int_{\mathbb{R}^n} \mathbf{x}^\alpha d\mu = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{\alpha_i} e^{-x^2/2} dx \right), \quad \forall \alpha \in \mathbb{N}^n.$$

For every $k \in \mathbb{N}$, let $\mathcal{C}_d^k := \{\mathbf{f} \in \mathbb{R}^{s(2d)} : \mathbf{M}_k(f, \mathbf{y}) \succeq 0\}$, where $\mathbf{M}_k(f, \mathbf{y})$ is the localizing matrix in (2.3) associated with \mathbf{y} and $f \in \mathbb{R}[\mathbf{x}]_{2d}$. Each \mathcal{C}_d^k is a closed convex cone and a spectrahedron.

Then $\mathcal{C}_d^0 \supset \mathcal{C}_d^1 \cdots \supset \mathcal{C}_d^k \cdots \supset \mathcal{C}_d$ and $f \in \mathcal{C}_d$ if and only if its vector of coefficients $\mathbf{f} \in \mathbb{R}^{s(2d)}$ satisfies $\mathbf{f} \in \mathcal{C}_d^k$, for every $k = 0, 1, \dots$

Proof. Following its definition (2.3), all entries of the localizing matrix $\mathbf{M}_k(f, \mathbf{y})$ are linear in $\mathbf{f} \in \mathbb{R}^{s(2d)}$, and so $\mathbf{M}_k(f, \mathbf{y}) \succeq 0$ is an LMI. Therefore \mathcal{C}_d^k is a spectrahedron and a closed convex cone. Next, let $\mathbf{K} := \mathbb{R}^n$ and let μ be as in Corollary 3.4 and so of the form (3.5). In addition, the normal distribution on \mathbb{R} satisfies Carleman's condition (3.6). Hence by Theorem 3.3 with $\mathbf{K} = \mathbb{R}^n$, f is nonnegative on \mathbf{K} if and only if (3.2) holds, which is equivalent to stating that $\mathbf{M}_k(f, \mathbf{y}) \succeq 0$, $k = 0, 1, \dots$, which in turn is equivalent to stating that $\mathbf{f} \in \mathcal{C}_d^k$, $k = 0, 1, \dots$ \square

So the nested sequence of convex cones $\mathcal{C}_d^0 \supset \mathcal{C}_d^1 \cdots \supset \mathcal{C}_d$ defines arbitrary close outer approximations of \mathcal{C}_d . In fact $\bigcap_{k=0}^{\infty} \mathcal{C}_d^k$ is closed and $\mathcal{C}_d = \bigcap_{k=0}^{\infty} \mathcal{C}_d^k$. It is worth emphasizing that each \mathcal{C}_d^k is a spectrahedron with *no lifting*, that is, \mathcal{C}_d^k is defined solely in terms of the vector of coefficients \mathbf{f} with no additional variable (i.e., no projection is needed).

For instance, the first approximation \mathcal{C}_d^0 is just the set $\{\mathbf{f} \in \mathbb{R}^{s(2d)} : \int f d\mu \geq 0\}$, which is a half-space of $\mathbb{R}^{s(2d)}$. And with $n = 2$,

$$\mathcal{C}_d^1 = \left\{ \mathbf{f} \in \mathbb{R}^{s(d)} : \begin{bmatrix} \int f d\mu & \int x_1 f d\mu & \int x_2 f d\mu \\ \int x_1 f d\mu & \int x_1^2 f d\mu & \int x_1 x_2 f d\mu \\ \int x_2 f d\mu & \int x_1 x_2 f d\mu & \int x_2^2 f d\mu \end{bmatrix} \succeq 0 \right\},$$

or, equivalently, \mathcal{C}_d^1 is the convex basic semi-algebraic set:

$$\begin{aligned} \left\{ \mathbf{f} \in \mathbb{R}^{s(2d)} : \int f d\mu \geq 0 \right. \\ \left. \left(\int x_i^2 f d\mu \right) \left(\int f d\mu \right) \geq \left(\int x_i f d\mu \right)^2, \quad i = 1, 2 \right. \\ \left. \left(\int x_1^2 f d\mu \right) \left(\int x_2^2 f d\mu \right) \geq \left(\int x_1 x_2 f d\mu \right)^2 \right. \\ \left. \left(\int f d\mu \right) \left[\left(\int x_1^2 f d\mu \right) \left(\int x_2^2 f d\mu \right) - \left(\int x_1 x_2 f d\mu \right)^2 \right] \right\} \end{aligned}$$

$$\left(\int x_1 f d\mu \right)^2 \left(\int x_2^2 f d\mu \right) - \left(\int x_2 f d\mu \right)^2 \left(\int x_1^2 f d\mu \right) + 2 \left(\int x_1 f d\mu \right) \left(\int x_2 f d\mu \right) \left(\int x_1 x_2 f d\mu \right) \geq 0 \Big\},$$

where we have just expressed the nonnegativity of all principal minors of $\mathbf{M}_1(f \mathbf{y})$.

A very similar result holds for the convex cone $\mathcal{C}_d(\mathbf{K})$ of polynomials of degree at most d , nonnegative on a closed set $\mathbf{K} \subset \mathbb{R}^n$, where \mathbf{K} has the property that the product measure μ defined in (3.5) (and such that (3.6) holds), satisfies $\text{supp } \mu = \mathbf{K}$.

Corollary 3.5. *Let $\mathbf{K} \subset \mathbb{R}^n$ be a closed set such that the probability measure μ on \mathbb{R}^n defined in (3.5) (and such that (3.6) holds), satisfies $\text{supp } \mu = \mathbf{K}$.*

For every $k \in \mathbb{N}$, let $\mathcal{C}_d^k(\mathbf{K}) := \{f \in \mathbb{R}^{s(2d)} : \mathbf{M}_k(f \mathbf{y}) \succeq 0\}$, where $\mathbf{M}_k(f \mathbf{y})$ is the localizing matrix in (2.3) associated with \mathbf{y} and $f \in \mathbb{R}[\mathbf{x}]_d$. Each $\mathcal{C}_d^k(\mathbf{K})$ is a closed convex cone and a spectrahedron.

Then $\mathcal{C}_d^0(\mathbf{K}) \supset \mathcal{C}_d^1(\mathbf{K}) \cdots \supset \mathcal{C}_d^k(\mathbf{K}) \cdots \supset \mathcal{C}_d(\mathbf{K})$ and $f \in \mathcal{C}_d(\mathbf{K})$ if and only if its vector of coefficients $\mathbf{f} \in \mathbb{R}^{s(d)}$ satisfies $\mathbf{f} \in \mathcal{C}_d^k(\mathbf{K})$, for every $k = 0, 1, \dots$

The proof which mimicks that of Corollary 3.4 is omitted. Of course, for practical computation, one is restricted to sets \mathbf{K} where one may compute *effectively* the moments of the measure μ . An example of such a set \mathbf{K} is the positive orthant, in which case one may choose as measure μ , the product of the exponential measure on the half-line \mathbb{R}_+ for which all moments are explicitly available.

4. APPLICATION TO POLYNOMIAL OPTIMIZATION

Consider the polynomial optimization problem

$$(4.1) \quad \mathbf{P} : \quad f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \},$$

where $\mathbf{K} \subseteq \mathbb{R}^n$ is closed and $f \in \mathbb{R}[\mathbf{x}]$.

Let μ be a probability measure μ with $\text{supp } \mu = \mathbf{K}$, whose sequence of moments $\mathbf{y} = (y_\alpha)$ is known, and which satisfy (3.5) if \mathbf{K} is not compact. Consider the sequence of semidefinite programs:

$$(4.2) \quad \lambda_d = \sup_{\lambda \in \mathbb{R}} \{ \lambda : \mathbf{M}_d(f_\lambda \mathbf{y}) \succeq 0 \}$$

where $f_\lambda \in \mathbb{R}[\mathbf{x}]$ is the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$. Notice that (4.2) has only one variable!

Theorem 4.1. *Consider the hierarchy of semidefinite programs (4.2) indexed by $d \in \mathbb{N}$. Then:*

- (i) *(4.2) has an optimal solution $\lambda_d \geq f^*$ for every $d \in \mathbb{N}$.*
- (ii) *The sequence (λ_d) , $d \in \mathbb{N}$, is monotone nonincreasing and $\lambda_d \downarrow f^*$ as $d \rightarrow \infty$.*

Proof. (i) Since $f - f^* \geq 0$ on \mathbf{K} , by Theorem 3.2, $\lambda := f^*$ is a feasible solution of (4.2) for every d . Hence $\lambda_d \geq f^*$ for every $d \in \mathbb{N}$. Next, let $d \in \mathbb{N}$ be fixed, and let λ be an arbitrary feasible solution of (4.2). From the condition $\mathbf{M}_d(f_\lambda \mathbf{y}) \succeq 0$, the diagonal entry $\mathbf{M}_d(f_\lambda \mathbf{y})(1, 1)$ must be nonnegative, i.e., $\lambda y_0 \leq \sum_{\alpha} f_{\alpha} y_{\alpha}$, and so, as we maximize and $y_0 > 0$, (4.2) must have an optimal solution λ_d .

(ii) Obviously $\lambda_d \leq \lambda_m$ whenever $d \geq m$, because $\mathbf{M}_d(f_\lambda \mathbf{y}) \succeq 0$ implies $\mathbf{M}_m(f_\lambda \mathbf{y}) \succeq 0$. Therefore, the sequence (λ_d) , $d \in \mathbb{N}$, is monotone nonincreasing and being bounded below by f^* , converges to $\lambda^* \geq f^*$. Next, suppose that $\lambda^* > f^*$; fix $k \in \mathbb{N}$, arbitrary. The convergence $\lambda_d \downarrow \lambda^*$ implies $\mathbf{M}_k(f_{\lambda^*} \mathbf{y}) \succeq 0$.

As k was arbitrary, we obtain that $\mathbf{M}_d(f_{\lambda^*} \mathbf{y}) \succeq 0$ for every $d \in \mathbb{N}$. But then by Theorem 3.2, $f - \lambda^* \geq 0$ on \mathbf{K} , and so $\lambda^* \leq f^*$, in contradiction with $\lambda^* > f^*$. Therefore $\lambda^* = f^*$. \square

The semidefinite program (4.2) provides an upper bound on the optimal value f^* only. We next show that the dual contains some information on global minimizers, at least when d is sufficiently large.

Duality. Let \mathcal{S}_d be the space of real symmetric $s(d) \times s(d)$ matrices. One may write the semidefinite program (4.2) as

$$\lambda_d = \sup_{\lambda} \{ \lambda : \lambda \mathbf{M}_d(\mathbf{y}) \preceq \mathbf{M}_d(f \mathbf{y}) \},$$

and so its dual is the semidefinite program

$$\inf_{\mathbf{X} \in \mathcal{S}_d} \{ \langle \mathbf{X}, \mathbf{M}_d(f \mathbf{y}) \rangle : \langle \mathbf{X}, \mathbf{M}_d(\mathbf{y}) \rangle = 1; \mathbf{X} \succeq 0 \},$$

or, equivalently,

$$(4.3) \quad \inf_{\sigma} \left\{ \int_{\mathbf{K}} f \sigma d\mu : \int_{\mathbf{K}} \sigma d\mu = 1; \sigma \in \Sigma[\mathbf{x}]_d \right\}.$$

So the dual problem (4.3) is to find a sum of squares polynomial σ of degree at most $2d$ (normalized to satisfy $\int \sigma d\mu = 1$) that minimizes the integral $\int \sigma f d\mu$, and a simple interpretation of (4.3) is as follows:

With $M(\mathbf{K})$ being the space of Borel probability measures on \mathbf{K} , we know that $f^* = \inf_{\varphi \in M(\mathbf{K})} \int_{\mathbf{K}} f d\varphi$. Next, let $M_d(\mu) \subset M(\mathbf{K})$ be the space of probability measures on \mathbf{K} which have a density $\sigma \in \Sigma[\mathbf{x}]_d$ with respect to μ . Then (4.3) reads $\inf_{\varphi \in M_d(\mu)} \int_{\mathbf{K}} f d\varphi$, which clearly shows why one obtains an upper bound on f^* . Indeed, instead of searching in $M(\mathbf{K})$ one searches in its subset $M_d(\mu)$. What is not obvious at all is whether the obtained upper bound obtained in (4.3) converges to f^* when the degree of $\sigma \in \Sigma[\mathbf{x}]_d$ is allowed to increase!

Theorem 4.2. *Suppose that $f^* > -\infty$ and \mathbf{K} has nonempty interior. Then :*

(a) *There is no duality gap between (4.2) and (4.3) and (4.3) has an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_d$ which satisfies $\int_{\mathbf{K}} (f(\mathbf{x}) - \lambda_d) \sigma^*(\mathbf{x}) d\mu(\mathbf{x}) = 0$.*

(b) *If \mathbf{K} is convex and f is convex, let $\mathbf{x}_d^* := \int_{\mathbf{K}} \mathbf{x} \sigma^*(\mathbf{x}) d\mu(\mathbf{x})$. Then $\mathbf{x}_d^* \in \mathbf{K}$ and $f^* \leq f(\mathbf{x}_d^*) \leq \lambda_d$, so that $f(\mathbf{x}_d^*) \rightarrow f^*$ as $d \rightarrow \infty$. Moreover, if the set $\{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \leq f_0\}$ is compact for some $f_0 > f^*$, then any accumulation point $\mathbf{x}^* \in \mathbf{K}$ of the sequence (\mathbf{x}_d^*) , $d \in \mathbb{N}$, is a minimizer of problem (4.1), that is, $f(\mathbf{x}^*) = f^*$.*

Proof. (a) Any scalar $\lambda < f^*$ is a feasible solution of (4.2) and in addition, $\mathbf{M}_d((f - \lambda) \mathbf{y}) \succ 0$ because since \mathbf{K} has nonempty interior and $f - \lambda > 0$ on \mathbf{K} ,

$$\langle \mathbf{g}, \mathbf{M}_d((f - \lambda) \mathbf{y}) \mathbf{g} \rangle = \int_{\mathbf{K}} (f(\mathbf{x}) - \lambda) g(\mathbf{x})^2 \mu(d\mathbf{x}) > 0, \quad \forall g \in \mathbb{R}[\mathbf{x}]_d.$$

But this means that Slater's condition² holds for (4.2) which in turn implies that there is no duality gap and (4.3) has an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_d$; see e.g. [18].

²For an optimization problem $\inf_{\mathbf{x}} \{f_0(\mathbf{x}) : f_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}$, Slater's condition states that there exists \mathbf{x}_0 such that $f_j(\mathbf{x}_0) > 0$ for every $j = 1, \dots, m$.

And so,

$$\int_{\mathbf{K}} (f(\mathbf{x}) - \lambda_d) \sigma^*(\mathbf{x}) d\mu(\mathbf{x}) = \int_{\mathbf{K}} f \sigma^* d\mu - \lambda_d = 0.$$

(b) Let ν be the Borel probability measure on \mathbf{K} defined by $\nu(B) = \int_B \sigma^* d\mu$, $B \in \mathcal{B}$. As f is convex, by Jensen's inequality (see e.g. McShane [10]),

$$\int_{\mathbf{K}} f \sigma^* d\mu = \int_{\mathbf{K}} f d\nu \geq f \left(\int_{\mathbf{K}} \mathbf{x} d\nu \right) = f(\mathbf{x}_d^*).$$

In addition, if \mathbf{K} is convex then $\mathbf{x}_d^* \in \mathbf{K}$ and so, $f(\mathbf{x}_d^*) \geq f^*$. Finally if for some $f_0 > f^*$, the set $\mathcal{H} := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \leq f_0\}$ is compact, and since $\lambda_d \rightarrow f^*$, then $f(\mathbf{x}_d^*) \leq f_0$ for d sufficiently large, i.e., $\mathbf{x}_d^* \in \mathcal{H}$ for sufficiently large d . By compactness there is a subsequence (d_ℓ) , $\ell \in \mathbb{N}$, and a point $\mathbf{x}^* \in \mathbf{K}$ such that $\mathbf{x}_{d_\ell}^* \rightarrow \mathbf{x}^*$ as $\ell \rightarrow \infty$. Continuity of f combined with the convergence $f(\mathbf{x}_d^*) \rightarrow f^*$ yields $f(\mathbf{x}_{d_\ell}^*) \rightarrow f(\mathbf{x}^*) = f^*$ as $\ell \rightarrow \infty$. As the convergent subsequence $(\mathbf{x}_{d_\ell}^*)$ was arbitrary, the proof is complete. \square

So in case where f is a convex polynomial and \mathbf{K} is a convex set, Theorem 4.2 provides a means of approximating not only the optimal value f^* , but also a global minimizers $\mathbf{x}^* \in \mathbf{K}$.

In the more subtle nonconvex case, one still can obtain some information on global minimizers from an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_d$ of (4.3). Let $\epsilon > 0$ be fixed, and suppose that d is large enough so that $f^* \leq \lambda_d \leq f^* + \epsilon$. Then, by Theorem 4.3(a),

$$\int_{\mathbf{K}} (f(\mathbf{x}) - f^*) \sigma^*(\mathbf{x}) d\mu(\mathbf{x}) = \lambda_d - f^* < \epsilon.$$

As $f - f^* \geq 0$ on \mathbf{K} , necessarily the measure $d\nu = \sigma^* d\mu$ gives very small weight to regions of \mathbf{K} where $f(\mathbf{x})$ is significantly larger than f^* . For instance, if $\epsilon = 10^{-2}$ and $\Delta := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \geq f^* + 1\}$, then $\nu(\Delta) \leq 10^{-2}$, i.e., the set Δ contributes to less than 1% of the total mass of ν . So if μ is uniformly distributed on \mathbf{K} (which is a reasonable choice if one has to compute all moments of μ) then a simple inspection of the values of $\sigma^*(\mathbf{x})$ provides some rough indication on where (in \mathbf{K}) $f(\mathbf{x})$ is close to f^* .

There are several interesting cases where the above described methodology can apply, i.e., cases where \mathbf{y} can be obtained either explicitly in closed form or numerically. In particular, when \mathbf{K} is either:

- A box $\mathbf{B} := \prod_{i=1}^n [a_i, b_i] \subset \mathbb{R}^n$, with μ being the normalized Lebesgue measure on \mathbf{B} . The sequence \mathbf{y} is trivial to obtain in closed form.
- The discrete set $\{-1, 1\}^n$ with μ being uniformly distributed and normalized. Again the sequence \mathbf{y} is trivial to obtain in closed form. Notice that in this case we obtain a new hierarchy of semidefinite relaxations (with only one variable) for the celebrated MAXCUT problem.
- The unit sphere $\mathbf{S} := \{\mathbf{x} : \|\mathbf{x}\|^2 \leq 1\}$ with μ being the rotation invariant probability measure on \mathbf{S} , again \mathbf{y} is easy to obtain. And similarly the sphere $\mathbf{S} := \{\mathbf{x} : \|\mathbf{x}\|^2 = 1\}$.
- A simplex $\Delta \subset \mathbb{R}^n$, in which case if one takes μ as the Lebesgue measure then all moments of μ can be computed numerically. In particular, with d fixed, this computation can be done in time polynomial time. See e.g. the recent work of [1].

- The positive orthant \mathbb{R}_+^n , in which case μ may be chosen to be the product measure $\otimes_{i=1}^n \nu_i$ with each ν_i being the exponential distribution $\nu_i(B) = \int_{\mathbb{R}_+ \cap B} e^{-x} dx$, $B \in \mathcal{B}$.
- The whole space \mathbb{R}^n in which case μ may be chosen to be the product measure $\otimes_{i=1}^n \nu_i$ with each ν_i being the normal distribution. Observe that in this case one obtains a new hierarchy of semidefinite approximations (upper bounds) for unconstrained global optimization. The corresponding monotone sequence of upper bounds converges to f^* no matter if the problem has a global minimizer or not. This may be an alternative and/or a complement to the recent convex relaxations provided in Schweighofer [16] and Hà and Vui [5] which also work when f^* is not attained, and provide a convergent sequence of *lower* bounds.

Example 1. Consider the global minimization on $\mathbf{K} = \mathbb{R}_+^2$ of the Motzkin-like polynomial $\mathbf{x} \mapsto x_1^2 x_2^2 (x_1^2 + x_2^2 - 1)$ whose global minimum is $f^* = -1/27 \approx -0.037$. Choose for μ the probability measure $\mu(B) := \int_B e^{-x_1 - x_2} d\mathbf{x}$, $B \in \mathcal{B}(\mathbb{R}_+^2)$, for which the sequence of moments $\mathbf{z} = (z_{ij})$, $i, j \in \mathbb{N}$, is easy to obtain. Namely $z_{ij} = i!j!$ for every $i, j \in \mathbb{N}$. Then the semidefinite relaxations (4.2) yield $\lambda_0 = 92$, $\lambda_1 = 15.60$ and $\lambda_2 = 4.301$ and $\lambda_3 = 1.5097$, showing a significant and rapid decrease.

Example 2. Still on $\mathbf{K} = \mathbb{R}_+^2$, consider the global minimization of the polynomial $\mathbf{x} \mapsto x_1^2 + (1 - x_1 x_2)^2$ whose global minimum $f^* = 0$ is not attained. Again, choose for μ the probability measure $\mu(B) := \int_B e^{-x_1 - x_2} d\mathbf{x}$, $B \in \mathcal{B}(\mathbb{R}_+^2)$. Then the semidefinite relaxations (4.2) yield $\lambda_0 = 5$, $\lambda_1 = 1.9187$ and $\lambda_2 = 0.227$, showing again a significant and rapid decrease.

The above two examples seem to indicate that even though one chooses a measure μ uniformly distributed on \mathbf{K} , one obtains a rapid decrease in the first iterations. However, if on the one hand the convergence to f^* is likely to be slow, on the other hand, one has to solve semidefinite programs (4.2) with only one variable! (But one has to remember that the choice is limited to measures μ with $\text{supp } \mu = \mathbf{K}$ and whose moments are easy to compute.) Analyzing how the convergence to f^* depends on μ is beyond the scope of the present paper and is a topic of further research.

Inverse problem from moments. Finally, observe that the above methodology perfectly fits *inverse problems* from moments, where precisely some Borel measure μ is known only from its moments (via some measurement device), and one wishes to recover (or approximately recover) its support \mathbf{K} from the known moments; see e.g. the work of Cuyt et al. [4] and the many references therein. Hence if $f \in \mathbb{R}[\mathbf{x}]$ is fixed then by definition $f - f^* \geq 0$ (on \mathbf{K}) provides a strong *valid* (polynomial) *inequality* for the unknown set \mathbf{K} . So computing an optimal solution λ_d of (4.2) for d sufficiently large, will provide an almost-valid inequality $f - \lambda_d \geq 0$ for \mathbf{K} . One may even let $f \in \mathbb{R}[\mathbf{x}]_d$ be unknown and search for the “best” valid inequality $f(\mathbf{x}) - f^* \geq 0$ where f varies in some family (e.g. linear or quadratic polynomials) and minimizes some appropriate (linear or convex) objective function of its coefficients \mathbf{f} .

5. CONCLUSION

In this paper we have presented a new characterization of nonnegativity on a closed set \mathbf{K} which is based on a single finite Borel measure μ with $\text{supp } \mu = \mathbf{K}$. It

permits to obtain a hierarchy of spectrahedra which provides a nested sequence of outer approximations of the convex cone of polynomials nonnegative on \mathbf{K} . When used in polynomial optimization for certain “simple sets” \mathbf{K} , one obtains a hierarchy of semidefinite approximations (with only one variable) which provides a nonincreasing sequence of upper bounds converging to the global optimum, hence a complement to the sequence of upper bounds provided by the hierarchy of semidefinite relaxations defined in e.g. [7, 8]. A topic of further investigation is to analyze the efficiency of such an approach on a sample of optimization problems on simple closed sets like the whole space \mathbb{R}^n , the positive orthant \mathbb{R}_+^n , a box, a simplex, or an ellipsoid, as well as for some inverse problems from moments.

6. APPENDIX

Lemma 6.1. *Let $\mathbf{K} \subseteq \mathbb{R}^n$ and let μ be the measure defined in (3.5), and where the measures ν_i , $i = 1, \dots, n$, satisfy (3.6).*

Let $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, be the sequence defined by: $z_\alpha := \int_{\mathbf{K}} \mathbf{x}^\alpha f(\mathbf{x}) d\mu(\mathbf{x})$ for every $\alpha \in \mathbb{N}^n$, and assume that $L_{\mathbf{z}}(\mathbf{x}_i^{2t}) \geq 0$ for every $t \in \mathbb{N}$ and every $i = 1, \dots, n$. Then the generalized Carleman’s condition (2.2) holds.

Proof. We prove that (2.2) holds for $i = 1$ only because the same arguments hold for $i = 2, \dots, n$. Write $f(\mathbf{x}) = \sum_{k=0}^s \tilde{f}_k(\tilde{\mathbf{x}}) x_1^k$, for some polynomials $\tilde{f}_k \in \mathbb{R}[\tilde{\mathbf{x}}]$ ($= \mathbb{R}[x_2, \dots, x_n]$), $k = 0, \dots, s$. Then,

$$\begin{aligned} L_{\mathbf{z}}(x_1^{2t}) &= \int_{\mathbf{K}} x_1^{2t} f(\mathbf{x}) d\mu(\mathbf{x}) = \int_{\mathbf{K}} \left(\sum_{k=0}^s \tilde{f}_k(\tilde{\mathbf{x}}) x_1^{2t+k} \right) d\mu(\mathbf{x}) \\ &\leq \int_{\mathbf{K} \cap [-1, 1]^n} \left(\sum_{k=0}^s |\tilde{f}_k(\tilde{\mathbf{x}})| \right) d\mu(\mathbf{x}) \\ &\quad + \int_{\mathbf{K} \setminus [-1, 1]^n} \left(\sum_{k=0}^s |\tilde{f}_k(\tilde{\mathbf{x}})| |x_1^k| \right) x_1^{2t} d\mu(\mathbf{x}) \\ &\leq A + \left(\int_{\mathbb{R}} x^{2(t+s)} d\nu_1(x) \right) B, \end{aligned}$$

where

$$A = \int_{\mathbf{K} \cap [-1, 1]^n} \left(\sum_{k=0}^s |\tilde{f}_k(\tilde{\mathbf{x}})| \right) d\mu(\mathbf{x}); \quad B = \int_{\mathbb{R}^n} \left(\sum_{k=0}^s |\tilde{f}_k(\tilde{\mathbf{x}})| \right) d\nu_2(x_2) \cdots d\nu_n(x_n).$$

Hence

$$L_{\mathbf{z}}(x_1^{2t}) \leq \max(A, B) \left(1 + L_{\mathbf{z}^1}(x^{2(t+s)}) \right),$$

where $\mathbf{z}^1 = (z_k^1)$, $k \in \mathbb{N}$, is the moment sequence of the measure ν_1 which satisfies (3.6). Therefore,

$$L_{\mathbf{z}}(x_1^{2t})^{-1/2t} \geq \max(A, B)^{-1/2t} \left(1 + L_{\mathbf{z}^1}(x^{2(t+s)}) \right)^{-1/2t}.$$

If $L_{\mathbf{z}^1}(x_1^{2(t+s)}) \leq 1$ for infinitely many $t \in I$, then

$$\sum_{t=1}^{\infty} L_{\mathbf{z}}(x_1^{2t})^{-1/2t} > \sum_{t \in I} (2 \max(A, B))^{-1/2t} = +\infty,$$

otherwise, there is some $t_0 \in \mathbb{N}$ such that,

$$\sum_{t=1}^{\infty} L_{\mathbf{z}}(x_1^{2t})^{-1/2t} > \sum_{t \geq t_0} (2 \max(A, B))^{-1/2t} (L_{\mathbf{z}^1}(x^{2(t+s)})^{-1/2t} = +\infty,$$

where we have used (3.6). And so we obtain the desired result

$$\sum_{t=1}^{\infty} L_{\mathbf{z}}(x_1^{2t})^{-1/2t} = +\infty.$$

□

REFERENCES

- [1] V. Baldoni, N. Berline, J. De Loera, M. Köppe, and M. Vergne. How to integrate a polynomial on a simplex, *Math. Comp.*, to appear.
- [2] C. Berg. The multidimensional moment problem and semi-groups. *Proc. Symp. Appl. Math.* **37** (1980), pp. 110–124.
- [3] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares, *Israel J. Math.* **153** (2006), pp. 355–380.
- [4] A. Cuyt, G. Golub, P. Milanfar and B. Verdonk. Multidimensional integral inversion with application in shape reconstruction, *SIAM J. Sci. Comput.* **27** (2005), pp. 1058–1070.
- [5] Hà Huy Vui and Pham Tien So'n. Global Optimization of Polynomials Using the Truncated Tangency Variety and Sums of Squares, *SIAM J. Optim.* **19** (2008), pp. 941–951.
- [6] P.H. Maserick and C. Berg. Exponentially bounded positive definite functions. *Ill. J. Math.* **28** (1984), pp. 162–179.
- [7] J.B. Lasserre. Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11** (2001), pp. 796–817.
- [8] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.
- [9] J.B. Lasserre and M. Putinar. Positivity and optimization for semi-algebraic functions (2009). [arXiv:0910.5250](https://arxiv.org/abs/0910.5250)
- [10] E.J. McShane. Jensen's inequality. *Bull. Amer. Math. Soc.* **43** (1937), pp. 521–527.
- [11] M. Marshall and T. Netzer. Positivstellensätze for real function algebras (2010). [arXiv:1004.4521v1](https://arxiv.org/abs/1004.4521v1).
- [12] A.E. Nussbaum. Quasi-analytic vectors. *Ark. Mat.* **6** (1966), pp. 179–191.
- [13] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Indiana Univ. Math. J.* **42** (1993), 969–984.
- [14] H.L. Royden. *Real Analysis*, 3rd. ed., Macmillan Publishing Company, New York, 1988.
- [15] K. Schmüdgen. The K -moment problem for compact semi-algebraic sets, *Math. Ann.* **289** (1991), 203–206.
- [16] M. Schweighofer. Global optimization of polynomials using gradient tentacles and sums of squares, *SIAM J. Optim.* **17** (2006), pp. 920–942.
- [17] G. Stengel. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry, *Math. Ann.* **207**, pp. 87–97.
- [18] L. Vandenberghe and S. Boyd. Semidefinite programming, *SIAM Rev.* **38** (1996), pp. 49–95.

LAAS-CNRS AND INSTITUTE OF MATHEMATICS, UNIVERSITY OF TOULOUSE, LAAS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX 4, FRANCE
E-mail address: lasserre@laas.fr