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 $d\nu = f d\mu$ with $\operatorname{supp} \mu = \mathbf{K}$, are all positive semidefinite (if $\mathbf{K}$ is compact μ is an arbitrary finite Borel measure with $\operatorname{supp} \mu = \mathbf{K}$). 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. Wen 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 (in fact, a generalized eigenvalue problem). In the compact case, 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. [12].

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 [22] 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 [20, 18] 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 [14] and in Marshall and Netzer [16]. 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 $\mathbf{K}$, except if $\mathbf{K}$ is compact and basic semi-algebraic (in which case outer approximations exist). Another exception is the convex cone of quadratic forms nonnegative on $\mathbf{K} = \mathbb{R}^n_+$ for which inner and outer approximations are available; see e.g. Anstreicher and Burer [1], and Dür [7].

1991 *Mathematics Subject Classification.* 90C25 28C15.
Key words and phrases. closed sets; nonnegative functions; nonnegative polynomials; semidefinite approximations; moments.

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 involves a *single* (but known) measure μ with $\operatorname{supp}\mu = \mathbf{K}$, 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[1] with $\operatorname{supp}\mu = \mathbf{K}$. The measure μ is arbitrary if $\mathbf{K}$ is compact. If $\mathbf{K}$ is not compact then one may choose for μ the finite Borel measure:
- $d\mu = \exp(-\sum_i |x_i|)d\varphi$ if f is a polynomial, and
- $d\mu = (1+f^2)^{-1}\exp(-\sum_i |x_i|)d\varphi$, if f is not a polynomial,

where φ is any finite Borel measure with support exactly $\mathbf{K}$. But many other choices are possible.

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 [22] (which provides a certificate of nonnegativity on a semi-algebraic set $\mathbf{K}$) or Schmüdgen and Putinar's Positivstellensätze [20, 18] (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 $\mathcal{C}_d^0 \supset \cdots \mathcal{C}_d^k \supset \cdots \supset \mathcal{C}_d$ 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 $\mathbf{K}$ (neither necessarily basic semi-algebraic nor compact), provided that all moments of an appropriate measure μ (with support exactly $\mathbf{K}$) 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 computing the *global* minimum f^* of a polynomial f on a closed set $\mathbf{K}$, a notoriously difficult problem. In nonlinear programming, a sequence of upper bounds on f^* is usually obtained from a sequence of feasible points $(\mathbf{x}_k) \subset \mathbf{K}$, e.g., via some (local) minimization algorithm. But it is important to emphasize that for non convex problems, providing a sequence of upper bounds $(f(\mathbf{x}_k))$, $k \in \mathbb{N}$, that converges to f^* is in general impossible, unless one computes points on a grid whose mesh size tends to zero.

We consider the case where $\mathbf{K} \subseteq \mathbb{R}^n$ is a closed set for which one may compute all moments of a measure μ with $\operatorname{supp}\mu = \mathbf{K}$. Typical examples of such sets are e.g. $\mathbf{K} = \mathbb{R}^n$ or $\mathbf{K} = \mathbb{R}^n_+$ in the non compact case and a box, a ball, an ellipsoid, a simplex, or the vertices of an hypercube (or hyper rectangle) in the compact case.

[1] A finite Borel measure μ on $\mathbb{R}^n$ is a nonnegative set function defined on the Borel σ-algebra of $\mathbb{R}^n$ (i.e., the σ-algebra generated by the open sets), such that $\mu(\emptyset) = 0$, $\mu(\mathbb{R}^n) < \infty$, and $\mu(\bigcup_{i=1}^\infty E_i) = \sum_{i=1}^\infty \mu(E_i)$ for any collection of disjoint measurable sets E_i. Its support (denoted $\operatorname{supp}\mu$) is the smallest closed set $\mathbf{K}$ such that $\mu(\mathbb{R}^n \setminus \mathbf{K}) = 0$; see e.g. Royden [19].

We then provide a hierarchy of semidefinite programs (with only one variable!) whose optimal values form a monotone nonincreasing sequence of *upper* bounds which converges to the global minimum f^*. In fact, each semidefinite program is a very specific one as it reduces to solving a *generalized eigenvalue* problem for a pair of real symmetric matrices.. (Therefore, for efficiency one may use specialized software packages instead of a SDP solver.) However, the convergence to f^* is in general only asymptotic and not finite (except when $\mathbf{K}$ is a discrete set in which case finite convergence takes place). This is in contrast with the hierarchy of semidefinite relaxations defined in Lasserre [12, 13] which provide a nondecreasing sequence of *lower* bounds that also converges to f^*, and very often in finitely many steps. Hence, for compact basic semi-algebraic sets 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 [12] is guaranteed only for compact basic semi-algebraic sets, whereas for the new hierarchy of upper bounds, the only requirement on $\mathbf{K}$ is to know all moments of a measure μ with $\operatorname{supp}\mu = \mathbf{K}$. On the other hand, in general computing such moments is possible only for relatively simple (but not necessarily compact nor semi-algebraic) sets.

At last but not least, the nonincreasing sequence of upper bounds converges to f^* even if f^* is not attained, which when $\mathbf{K} = \mathbb{R}^n$, could provide an alternative and/or a complement to the hierarchy of convex relaxations provided in Schweighofer [21] (based on gradient tentacles) and in Hà and Pham [8] (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 provides some information on the location of global minimizers.

2. Notation, definitions and preliminary results

A Borel measure on $\mathbb{R}^n$ is understood as a *positive* Borel measure, i.e., a nonnegative set function μ on the Borel σ-algebra $\mathcal{B}$ (i.e., the σ-algebra generated by the open sets of $\mathbb{R}^n$) such that $\mu(\emptyset) = 0$, and with the countably additive property

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i),$$

for any collection of disjoint measurable sets $(E_i) \subset \mathcal{B}$; see e.g. Royden [19, pp. 253–254].

Let $\mathbb{R}[\mathbf{x}]$ be the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$, and $\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$ its subset of polynomials that are sums of squares (s.o.s.). 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 $(\mathbf{x}^\alpha)$ of monomials. Also, denote by $\Sigma[\mathbf{x}]_d \subset \Sigma[\mathbf{x}]$ the convex cone of 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. Let $\mathcal{S}_n$ denotes the vector space of $p \times p$ real symmetric matrices. For a matrix $\mathbf{A} \in \mathcal{S}_p$ the notation $\mathbf{A} \succeq 0$ (resp. $\mathbf{A} \succ 0$) stands for $\mathbf{A}$ is positive semidefinite (resp. definite).

Moment matrix. With $\mathbf{y} = (y_\alpha)$ being a sequence indexed in the canonical basis $(\mathbf{x}^\alpha)$ of $\mathbb{R}[\mathbf{x}]$, let $L_\mathbf{y} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ 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 $(\mathbf{x}^\alpha)$, and defined by:

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

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

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

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

and so $\mathbf{M}_d(\mathbf{y}) \succeq 0$. 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[2].

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 [3]). *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 $|y_\alpha| \leq c\, a^{|\alpha|}$ for every $\alpha \in \mathbb{N}^n$.*

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

$$\sum_{t=1}^\infty L_\mathbf{y}(x_i^{2t})^{-1/2t} \,=\, +\infty, \qquad \forall \, i = 1, \ldots, n. \tag{2.2}$$

In addition, in both cases (a) and (b) 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 [17]. For more details see e.g. Berg [3] and/or Maserick and Berg [11].

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 $(\mathbf{x}^\alpha)$, and defined by:

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

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 $d = 0, 1, \ldots$ because

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

[2] To see this note that (a) two measures μ_1, μ_2 on a compact set $\mathbf{K} \subset \mathbb{R}^n$ are identical if and only if $\int_\mathbf{K} f d\mu_1 = \int_\mathbf{K} f d\mu_2$ for all continuous functions f on $\mathbf{K}$, and (b) by Stone-Weierstrass, the polynomials are dense in the space of continuous functions for the sup-norm.

3. Nonnegativity on closed sets

Recall that if $\mathbf{X}$ is a separable metric space with Borel σ-field $\mathcal{B}$, the support $\operatorname{supp}\mu$ of a Borel measure μ on $\mathbf{X}$ is the (unique) smallest closed set $B \in \mathcal{B}$ such that $\mu(X \setminus B) = 0$. Given a Borel measure μ on $\mathbb{R}^n$ and a measurable function $f : \mathbb{R}^n \to \mathbb{R}$, the mapping $B \mapsto \nu(B) := \int_B f d\mu$, $B \in \mathcal{B}$, defines a set function on $\mathcal{B}$. If f is nonnegative then ν is a Borel measure (which is finite if f is μ-integrable); see e.g. Royden [19, p. 276 and p. 408]. If f is not nonnegative then setting $B_1 := \{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ and $B_2 := \{\mathbf{x} : f(\mathbf{x}) < 0\}$, the set function ν can be written as the difference

$$\nu(B) = \nu_1(B) - \nu_2(B), \qquad B \in \mathcal{B}, \tag{3.1}$$

of the two positive Borel measures ν_1, ν_2 defined by

$$\nu_1(B) = \int_{B_1 \cap B} f d\mu, \quad \nu_2(B) = -\int_{B_2 \cap B} f d\mu, \quad \forall B \in \mathcal{B}. \tag{3.2}$$

Then ν is a *signed* Borel measure provided that either $\nu_1(B_1)$ or $\nu_2(B_2)$ is finite; see e.g. Royden [19, p. 271]. 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 $\operatorname{supp}\mu = \mathbf{K}$. A continuous function $f : \mathbf{X} \to \mathbb{R}$ is nonnegative on $\mathbf{K}$ if and only if the set function $B \mapsto \nu(B) = \int_{\mathbf{K} \cap B} f d\mu$, $B \in \mathcal{B}$, is a positive measure.*

Proof. The *only if part* is straightforward. For the *if part*, if ν is a positive measure then $f(\mathbf{x}) \geq 0$ 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 $\mathbf{K} \setminus \mathbf{G}$. Indeed, otherwise suppose that there exists a Borel set B_0 with $\mu(B_0) > 0$ and $f < 0$ on B_0; then one would get the contradiction that ν is not positive because $\nu(B_0) = \int_{B_0} f d\mu < 0$. In fact, f is called the Radon-Nikodym derivative of ν with respect to μ; see Royden [19, Theorem 23, p. 276].

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 $\operatorname{supp}\mu\,(= \mathbf{K})$ 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 $\mathbf{x}_k \to \mathbf{x}$ as $k \to \infty$. But since f is continuous and $f(\mathbf{x}_k) \geq 0$ for every $k \in \mathbb{N}$, we obtain the desired result $f(\mathbf{x}) \geq 0$. □

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 $\operatorname{supp}\mu = \mathbf{K}$, and with vector of moment $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$. Let f be a continuous function on $\mathbb{R}^n$. Then:*

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

$$\int_{\mathbf{K}} g^2 f \, d\mu \geq 0, \qquad \forall\, g \in \mathbb{R}[\mathbf{x}], \tag{3.3}$$

or, equivalently, if and only if

$$\mathbf{M}_d(\mathbf{z}) \succeq 0, \qquad d = 0, 1, \ldots \tag{3.4}$$

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 (3.4) reads* $\mathbf{M}_d(f\,\mathbf{y}) \succeq 0$, $d = 0, 1, \ldots$, *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 the convex hull* $\mathrm{co}\,(\mathbf{K})$ *of* $\mathbf{K}$ *if and only if (3.3) 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 $\mathrm{co}\,(\mathbf{K})$ then $f \geq 0$ on $\mathbf{K}$ and so the "only if" part of (b) also follows.

If part. The set function $\nu(B) = \int_B f d\mu$, $B \in \mathcal{B}$, can be written as the difference $\nu = \nu_2 - \nu_2$ of the two positive finite Borel measures ν_1, ν_2 described in (3.1)-(3.2), where $B_1 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \geq 0\}$ and $B_2 := \{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) < 0\}$. As $\mathbf{K}$ is compact and f is continuous, both ν_1, ν_2 are finite, and so ν is a finite signed Borel measure; see Royden [19, p. 271]. In view of Lemma 3.1 it suffices to prove that in fact ν is a finite and positive Borel measure. So let $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, be the sequence defined by:

$$z_\alpha = \int_{\mathbf{K}} \mathbf{x}^\alpha d\nu(\mathbf{x}) := \int_{\mathbf{K}} \mathbf{x}^\alpha f(\mathbf{x}) d\mu(\mathbf{x}), \qquad \forall \alpha \in \mathbb{N}^n. \tag{3.5}$$

Every z_α, $\alpha \in \mathbb{N}^n$, is finite because $\mathbf{K}$ is compact and f is continuous. So the condition

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

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 (3.3) implies $\mathbf{M}_d(\mathbf{z}) \succeq 0$ for every $d \in \mathbb{N}$. Moreover, as $\mathbf{K} \subseteq [-1, 1]^n$,

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

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 $\mathrm{supp}\,\nu \subseteq \mathbf{K} \subseteq [-1, 1]^n$,

$$\int_{[-1,1]^n} \mathbf{x}^\alpha\, d\nu(\mathbf{x}) = \int_{[-1,1]^n} \mathbf{x}^\alpha\, d\psi(\mathbf{x}), \qquad \forall \alpha \in \mathbb{N}^n. \tag{3.6}$$

But then using (3.1) and (3.6) yields

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

which in turn implies $\nu_1 = \nu_2 + \psi$ because measures on a compact set are determinate. Next, this implies $\psi = \nu_1 - \nu_2\ (= \nu)$ and so 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

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

and so in (3.4), $\mathbf{M}_d(\mathbf{z})$ 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 (3.4) reads $\mathbf{M}_d(f\,\mathbf{y}) \succeq 0$ for all $d = 0, 1, \ldots$

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

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.

Remark 3.3. An informal alternative proof of Theorem 3.2 which does not use Lemma 3.1 is as follows. If f is not nonnegative on $\mathbf{K}$ there exists $\mathbf{a} \in \mathbf{K}$ such that $f(\mathbf{a}) < 0$, and so as $\mathbf{K}$ is compact, there is a continuous function, e.g, $\mathbf{x} \mapsto h(\mathbf{x}) := \exp(-c\|\mathbf{x} - \mathbf{a}\|^2)$ close to 1 in some open neighborhood $\mathbf{B}(\mathbf{a}, \delta)$ of $\mathbf{a}$, and very small in the rest of $\mathbf{K}$. By the Stone-Weierstrass's theorem, one may choose h to be a polynomial. Next, the complement $\mathbf{B}(\mathbf{a}, \delta)^c\,(= \mathbb{R}^n \setminus \mathbf{B}(\mathbf{a}, \delta))$ of $\mathbf{B}(\mathbf{a}, \delta)$ is closed, and so $\mathbf{K} \cap \mathbf{B}(\mathbf{a}, \delta)^c$ is a closed set contained in $\mathbf{K}$ (hence smaller than $\mathbf{K}$). Therefore $\mu(\mathbf{B}(\mathbf{a}, \delta)) > 0$ because otherwise $\mu(\mathbf{K} \cap \mathbf{B}(\mathbf{a}, \delta)^c) = \mu(\mathbf{K})$ which would imply that $\mathbf{K} \cap \mathbf{B}(\mathbf{a}, \delta)^c$ is a support of μ smaller than $\mathbf{K}$, in contradiction with $\mathrm{supp}\,\mu = \mathbf{K}$. Hence, we would get the contradiction

$$\int h^2\, f\, d\mu \;\approx\; h(\mathbf{a})^2 f(\mathbf{a})\, \mu(\mathbf{B}(\mathbf{a}, \delta)) \;<\; 0.$$

However, in the non compact case described in the next section, this argument is not valid.

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. To handle arbitrary non compact sets $\mathbf{K}$ and arbitrary continuous functions f, we need a reference measure μ with $\mathrm{supp}\,\mu = \mathbf{K}$ and with nice properties so that integrals such as $\int g^2 f d\mu$, $g \in \mathbb{R}[\mathbf{x}]$, are well-behaved.

So, let φ be an arbitrary finite Borel measure on $\mathbb{R}^n$ whose support is exactly $\mathbf{K}$, and let μ be the finite Borel measure defined by:

$$\mu(B) \;:=\; \int_B \exp\left(-\sum_{i=1}^n |x_i|\right)\, d\varphi(\mathbf{x}), \qquad \forall\, B \in \mathcal{B}(\mathbb{R}^n). \tag{3.7}$$

Observe that $\mathrm{supp}\,\mu = \mathbf{K}$ and μ satisfies Carleman's condition (2.2). Indeed, let $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, be the sequence of moments of μ. Then for every $i = 1, \ldots, n$, and every $k = 0, 1, \ldots$, using $x_i^{2k} \leq (2k)! \exp|x_i|$,

$$L_{\mathbf{z}}(x_i^{2k}) = \int_{\mathbf{K}} x_i^{2k}\; d\mu(\mathbf{x}) \;\leq\; (2k)! \int_{\mathbf{K}} \mathrm{e}^{|x_i|}\, d\mu(\mathbf{x}) \;\leq\; (2k)!\, \varphi(\mathbf{K}) \;=:\, (2k)!\, M. \tag{3.8}$$

Therefore for every $i = 1, \ldots, n$, using $(2k)! < (2k)^{2k}$ for every k, yields

$$\sum_{k=1}^\infty L_{\mathbf{z}}(x_i^{2k})^{-1/2k} \;\geq\; \sum_{k=1}^\infty M^{-1/2k}\, ((2k)!)^{-1/2k} \geq \sum_{k=1}^\infty \frac{M^{-1/2k}}{2k} \;=\; +\infty,$$

i.e., (2.2) holds. Notice also that all the moments of μ (defined in (3.7)) are finite, and so every polynomial is μ-integrable.

Theorem 3.4. *Let $\mathbf{K} \subseteq \mathbb{R}^n$ be closed and let φ be an arbitrary finite Borel measure whose support is exactly $\mathbf{K}$. Let f be a continuous function on $\mathbb{R}^n$. If $f \in \mathbb{R}[\mathbf{x}]$ (i.e., f is a polynomial) let μ be as in (3.7) whereas if f is not a polynomial let μ be defined by*

$$\mu(B) := \int_B \frac{\exp\left(-\sum_{i=1}^n |x_i|\right)}{1+f(\mathbf{x})^2}\, d\varphi(\mathbf{x}), \qquad \forall\, B \in \mathcal{B}(\mathbb{R}^n). \tag{3.9}$$

Then (a) and (b) of Theorem 3.2 hold.

For a detailed proof see §6.

It is important to emphasize that in Theorem 3.2 and 3.4, the set $\mathbf{K}$ is an *arbitrary* closed set of $\mathbb{R}^n$, and to the best of our knowledge, the characterization of nonnegativity of f in terms of positive definiteness of the moment matrices $\mathbf{M}_d(\mathbf{z})$ is new. But of course, this characterization becomes even more interesting when one knows how the compute the moment sequence $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, which is possible in a few special cases only.

Important particular cases of nice such sets $\mathbf{K}$ include boxes, hyper rectangles, ellipsoids, and simplices in the compact case, and the positive orthant, or the whole space $\mathbb{R}^n$ in the non compact case. For instance, for the whole space $\mathbf{K} = \mathbb{R}^n$ one may choose for μ in (3.7) the multivariate Gaussian (or normal) probability measure

$$\mu(B) := (2\pi)^{-n/2} \int_B \exp(-\frac{1}{2}\|\mathbf{x}\|^2)\, d\mathbf{x}, \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

which the n-times product of the one-dimensional normal distribution

$$\mu_i(B) := \frac{1}{\sqrt{2\pi}} \int_B \exp(-x_i^2/2)\, dx_i, \qquad B \in \mathcal{B}(\mathbb{R}),$$

whose moments are all easily available in closed form. In Theorem 3.4 this corresponds to the choice

$$\varphi(B) = (2\pi)^{-n/2} \int_B \frac{\exp(-\|\mathbf{x}\|^2/2)}{\exp(-\sum_{i=1}^n |x_i|)}\, d\mathbf{x}, \qquad B \in \mathcal{B}(\mathbb{R}^n). \tag{3.10}$$

When $\mathbf{K}$ is the positive orthant $\mathbb{R}^n_+$ one may choose for μ the exponential probability measure

$$\mu(B) := \int_B \exp(-\sum_{i=1}^n x_i)\, d\mathbf{x}, \qquad B \in \mathcal{B}(\mathbb{R}^n_+), \tag{3.11}$$

which the n-times product of the one-dimensional exponential distribution

$$\mu_i(B) := \int_B \exp(-x_i)\, dx_i, \qquad B \in \mathcal{B}(\mathbb{R}_+),$$

whose moments are also easily available in closed form. In Theorem 3.4 this corresponds to the choice

$$\varphi(B) = 2^n \int_B \exp(-\sum_{i=1}^n x_i)\, d\mathbf{x}, \qquad B \in \mathcal{B}(\mathbb{R}^n_+).$$

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 spectrahedron[3] in a higher dimensional space), so far there is no such simple representation for the former. In addition, when d is fixed, Blekherman [4] 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 $\mathbf{f} \in \mathbb{R}^{s(d)}$ 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 $k = 1, \ldots,$

$$\gamma_p := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^p \, \mathrm{e}^{-x^2/2} \, dx \;=\; \begin{cases} 0 & \text{if } p = 2k+1, \\ \prod_{j=1}^k (2j-1) & \text{if } p = 2k, \end{cases}$$

as $\gamma_{2k} = (2k-1)\gamma_{2(k-1)}$ for every $k \geq 1$.

Corollary 3.5. *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 $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$,*

$$y_\alpha \;=\; \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\mu \;=\; \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{\alpha_i} \mathrm{e}^{-x^2/2} dx \right), \qquad \forall\, \alpha \in \mathbb{R}^n. \tag{3.12}$$

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 $k = 0, 1, \ldots$.

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 3.5 and so of the form (3.7) with φ as in (3.10). Then μ satisfies Carleman's condition (2.2). Hence, by Theorem 3.4 with $\mathbf{K} = \mathbb{R}^n$, f is nonnegative on $\mathbf{K}$ if and only if (3.4) holds, which is equivalent to stating that $\mathbf{M}_k(f\,\mathbf{y}) \succeq 0$, $k = 0, 1, \ldots,$ which in turn is equivalent to stating that $\mathbf{f} \in \mathcal{C}_d^k$, $k = 0, 1, \ldots$ ☐

So the nested sequence of convex cones $\mathcal{C}_d^0 \supset \mathcal{C}_d^k \cdots \supset \mathcal{C}_d$ defines arbitrary close outer approximations of $\mathcal{C}_d$. In fact $\cap_{k=0}^\infty \mathcal{C}_d^k$ is closed and $\mathcal{C}_d = \cap_{k=0}^\infty \mathcal{C}_d^k$. 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).

[3]A 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).

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 $n=2$,

$$\mathcal{C}_d^1 \,=\, \left\{\mathbf{f}\in\mathbb{R}^{s(d)}\,:\,\left[\begin{array}{ccc}\displaystyle\int f d\mu & \displaystyle\int x_1 f d\mu & \displaystyle\int x_2 f d\mu\\ \displaystyle\int x_1 f d\mu & \displaystyle\int x_1^2 f d\mu & \displaystyle\int x_1x_2 f d\mu\\ \displaystyle\int x_2 f d\mu & \displaystyle\int x_1x_2 f d\mu & \displaystyle\int x_2^2 f d\mu\end{array}\right]\succeq 0\right\},$$

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

$$\begin{array}{rl}\Big\{\mathbf{f}\in\mathbb{R}^{s(2d)} \;:\; & \displaystyle\int f d\mu\,\geq\,0\\ & \displaystyle\left(\int x_i^2 f d\mu\right)\left(\int f d\mu\right)\,\geq\,\left(\int x_i f d\mu\right)^2, \quad i=1,2\\ & \displaystyle\left(\int x_1^2 f d\mu\right)\left(\int x_2^2 f d\mu\right)\,\geq\,\left(\int x_1x_2 f d\mu\right)^2\end{array}$$

$$\left(\int f d\mu\right)\left[\left(\int x_1^2 f d\mu\right)\left(\int x_2^2 f d\mu\right)-\left(\int x_1x_2 f d\mu\right)^2\right]-$$

$$\left(\int x_1 f d\mu\right)^2\left(\int x_2^2 f d\mu\right)-\left(\int x_2 f d\mu\right)^2\left(\int x_1^2 f d\mu\right)+$$

$$2\left(\int x_1 f d\mu\right)\left(\int x_2 f d\mu\right)\left(\int x_1x_2 f d\mu\right)\,\geq\,0\Big\},$$

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 $\mathbf{K}\subset\mathbb{R}^n$.

Corollary 3.6. *Let $\mathbf{K}\subset\mathbb{R}^n$ be a closed set and let μ be defined in (3.7) where φ is a an arbitrary finite Borel measure whose support is exactly $\mathbf{K}$.*

For every $k\in\mathbb{N}$, let $\mathcal{C}_d^k(\mathbf{K}):=\{\mathbf{f}\in\mathbb{R}^{s(d)}\,:\,\mathbf{M}_k(f\,\mathbf{y})\succeq 0\}$, 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 $k=0,1,\ldots$.

The proof which mimicks that of Corollary 3.5 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 the measure μ in (3.11) for which all moments are explicitly available. For compact sets $\mathbf{K}$ let us mention balls, boxes, ellipsoids, and simplices. But again, any compact set where one knows how to compute all moments of some measure with support exactly $\mathbf{K}$, is fine.

To the best of our knowledge this is the first characterization of an outer approximation of the cone $\mathcal{C}_d(\mathbf{K})$ in a relatively general context. Indeed, for the basic semi-algebraic set

$$\mathbf{K}\,=\,\{\mathbf{x}\in\mathbb{R}^n\,:\,g_j(\mathbf{x})\,\geq\,0,\quad j=1,\ldots,m\,\},\tag{3.13}$$

Stengle's Nichtnegativstellensatz [22] states that $f \in \mathbb{R}[\mathbf{x}]$ is nonnegative on $\mathbf{K}$ if and only if

$$p\,f \;=\; f^{2s} + q, \tag{3.14}$$

for some integer s and polynomials $p, q \in P(g)$, where $P(g)$ is the *preordering*[4] associated with the g_j's. In addition, there exist bounds on the integer s and the degree of the s.o.s. weights in the definition of $p, q \in P(g)$, so that in principle, when f is *known*, checking whether $f \geq 0$ on $\mathbf{K}$ reduces to solving a single SDP to compute h, p in the nonnegativity certificate (3.14). However, the size of this SDP is potentially huge and makes it unpractical. Moreover, the representation of $\mathcal{C}_d(\mathbf{K})$ in (3.14) is not convex in the vector of coefficients of f because it involves f^{2s} as well as the product pf.

Remark 3.7. If in Corollary 3.6 one replaces the finite-dimensional convex cone $\mathcal{C}_d(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_d$ with the infinite-dimensional convex cone $\mathcal{C}(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]$ of *all* polynomials nonnegative on $\mathbf{K}$, and $\mathcal{C}_d^k(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_d$ with $\mathcal{C}^k(\mathbf{K}) = \{\mathbf{f} \in \mathbb{R}^{s(2k)} : \mathbf{M}_k(f\,\mathbf{y}) \succeq 0\}$, then the nested sequence of (increasing but finite-dimensional) convex cones $\mathcal{C}^k(\mathbf{K})$, $k \in \mathbb{N}$, provides finite-dimensional approximations of $\mathcal{C}(\mathbf{K})$.

4. Application to polynomial optimization

Consider the polynomial optimization problem

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

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

If $\mathbf{K}$ is compact let μ be a finite Borel measure with $\operatorname{supp}\mu = \mathbf{K}$ and if $\mathbf{K}$ is not compact, let φ be an arbitrary finite Borel measure with $\operatorname{supp}\varphi = \mathbf{K}$ and let μ be as in (3.7). In both cases, the sequence of moments $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, is well-defined, and we assume that y_α is available or can be computed, for every $\alpha \in \mathbb{N}^n$.

Consider the sequence of semidefinite programs:

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

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 $d \in \mathbb{N}^n$.*

(ii) *The sequence (λ_d), $d \in \mathbb{N}$, is monotone nonincreasing and $\lambda_d \downarrow f^*$ as $d \to \infty$.*

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

[4]The preordering $P(g)$ associated with the g_j's is the set of polynomials of the form $\sum_{\alpha \in \{0,1\}^m} \sigma_\alpha\, g_1^{\alpha_1} \cdots g_m^{\alpha_m}$ where $\sigma_\alpha \in \Sigma[\mathbf{x}]$ for each α.

$\lambda^* > f^*$; 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 or Theorem 3.4, $f - \lambda^* \geq 0$ on $\mathbf{K}$, and so $\lambda^* \leq f^*$, in contradiction with $\lambda^* > f^*$. Therefore $\lambda^* = f^*$. □

For each $d \in \mathbb{N}$, 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.

4.1. **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

$$\lambda_d = \sup_{\lambda} \{\lambda \,:\, \lambda \mathbf{M}_d(\mathbf{y}) \preceq \mathbf{M}_d(f\,\mathbf{y})\}, \tag{4.3}$$

which in fact is a generalized eigenvalue problem for the pair of matrices $\mathbf{M}_d(\mathbf{y})$ and $\mathbf{M}_d(f\,\mathbf{y})$. 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,

$$\lambda_d^* = \inf_{\sigma} \left\{ \int_{\mathbf{K}} f\,\sigma\,d\mu \,:\, \int_{\mathbf{K}} \sigma\,d\mu = 1;\; \sigma \in \Sigma[\mathbf{x}]_d \right\}. \tag{4.4}$$

So the dual problem (4.4) 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 $\int f \sigma d\mu$, and a simple interpretation of (4.4) is as follows:

With $M(\mathbf{K})$ being the space of Borel probability measures on $\mathbf{K}$, we know that $f^* = \displaystyle\inf_{\varphi \in M(\mathbf{K})} \int_{\mathbf{K}} f d\varphi$. 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 (4.4) reads $\displaystyle\inf_{\varphi \in M_d(\mu)} \int_{\mathbf{K}} f d\varphi$, 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 (4.4) 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 (4.4) and (4.4) has an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_d$ which satisfies $\displaystyle\int_{\mathbf{K}} (f(\mathbf{x}) - \lambda_d)\,\sigma^*(\mathbf{x}) d\mu(\mathbf{x}) = 0$.*

(b) *If $\mathbf{K}$ is convex and f is convex, let $\mathbf{x}_d^* := \int \mathbf{x}\,\sigma^*(\mathbf{x}) d\mu(\mathbf{x})$. Then $\mathbf{x}_d^* \in \mathbf{K}$ and $f^* \leq f(\mathbf{x}_d^*) \leq \lambda_d$, so that $f(\mathbf{x}_d^*) \to f^*$ as $d \to \infty$. 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 $\mathbf{K}$,

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

But this means that Slater's condition[5] holds for (4.2) which in turn implies that there is no duality gap and (4.4) has an optimal solution $\sigma^* \in \Sigma[\mathbf{x}]_d$; see e.g. [23]. 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 [15]),

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

In addition, if $\mathbf{K}$ is convex then $\mathbf{x}_d^* \in \mathbf{K}$ and so, $\lambda_d \geq f(\mathbf{x}_d^*) \geq f^*$. 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 $\lambda_d \to f^*$, 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 $\mathbf{x}_{d_\ell}^* \to \mathbf{x}^*$ as $\ell \to \infty$. Continuity of f combined with the convergence $f(\mathbf{x}_d^*) \to f^*$ yields $f(\mathbf{x}_{d_\ell}^*) \to f(\mathbf{x}^*) = f^*$ as $\ell \to \infty$. As the convergent subsequence $(\mathbf{x}_{d_\ell}^*)$ was arbitrary, the proof is complete. □

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 minimizer $\mathbf{x}^* \in \mathbf{K}$.

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 (4.4). Let $\epsilon > 0$ be fixed, and suppose that d is large enough so that $f^* \leq \lambda_d \leq f^* + \epsilon$. Then, by Theorem 4.4(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^*.

The interpretation (4.4) of the dual shows that in general the monotone convergence is only asymptotic and cannot be finite. Indeed if $\mathbf{K}$ has a nonempty interior then the probability measure $d\nu = \sigma d\mu$ cannot be a Dirac measure at any global minimizer $\mathbf{x}^* \in \mathbf{K}$. An exception is the discrete case, i.e., when $\mathbf{K}$ is a finite number of points, like in e.g. 0/1 programs. Indeed we get:

Corollary 4.3. *Let $\mathbf{K} \subset \mathbb{R}^n$ be a discrete set $(\mathbf{x}(k)) \subset \mathbb{R}^n$, $k \in J$, and let μ be the probability measure uniformly distributed in $\mathbf{K}$, i.e.,*

$$\mu \,=\, \frac{1}{s}\sum_{k=1}^{s} \delta_{\mathbf{x}(k)},$$

where $s = |J|$ and $\delta_{\mathbf{x}}$ denote the Dirac measure at the point $\mathbf{x}$. Then the optimal value λ_d of (4.2) satisfies $\lambda_d = f^*$ for some integer d.

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

Proof. Let $\mathbf{x}^* = \mathbf{x}(j^*)$ (for some $j^* \in J$) be the global minimizer of $\min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$. For each $k = 1, \ldots, s$ there exists a polynomial $q_k \in \mathbb{R}[\mathbf{x}]$ such that $q_k(\mathbf{x}(j)) = \delta_{k=j}$ for every $j = 1, \ldots, s$ (where $\delta_{k=j}$ denotes the Kronecker symbol). The polynomials (q_k) are called *interpolation* polynomials. So let $\sigma^* := s q_{j^*}^2 \in \Sigma[\mathbf{x}]$, so that

$$\int_{\mathbf{K}} f(\mathbf{x})\sigma^*(\mathbf{x})d\mu(\mathbf{x}) \,=\, f(\mathbf{x}(j^*)) \,=\, f^* \quad \text{and} \quad \int_{\mathbf{K}} \sigma^* d\mu \,=\, 1.$$

Hence as soon as $d \geq \deg q_{j^*}$, $\sigma^* \in \Sigma[\mathbf{x}]_d$ is a feasible solution of (4.4), and so from $f^* = \int f\sigma^* d\mu \geq \lambda_d^* \geq \lambda_d \geq f^*$ we deduce that $\lambda_d^* = \lambda_d = f^*$, the desired result. □

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 $\mathbf{y} = (y_\alpha)$ is trivial to obtain in closed form.
- The discrete set $\{-1, 1\}^n$ with μ being uniformly distributed and normalized. Again the sequence $\mathbf{y} = (y_\alpha)$ 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 (and any nonlinear 0/1 program).
- The unit Euclidean ball $\mathbf{B} := \{\mathbf{x} : \|\mathbf{x}\|^2 \leq 1\}$ with μ uniformly distributed, and similarly the unit sphere $\mathbf{S} := \{\mathbf{x} : \|\mathbf{x}\|^2 = 1\}$, with μ being the rotation invariant probability measure on $\mathbf{S}$. In both cases the moments $\mathbf{y} = (y_\alpha)$ are obtained easily.
- 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 [2].
- 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 one then 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 [21] and Hà and Vui [8] which also work when f^* is not attained, and provide a convergent sequence of *lower* bounds.
- 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 $\nu_i(B) = \int_{\mathbb{R}^+ \cap B} \mathrm{e}^{-x} dx$, $B \in \mathcal{B}$. In particular if $\mathbf{x} \mapsto f(\mathbf{x}) := \mathbf{x}^T \mathbf{A} \mathbf{x}$ where $\mathbf{A} \in \mathcal{S}_n$, then one obtains a hierarchy of numerical tests to check whether $\mathbf{A}$ is a *copositive* matrix. Indeed, if λ_d is an optimal solution of (4.3) then $\mathbf{A}$ is copositive if and only if $\lambda_d \geq 0$ for all $d \in \mathbb{N}$. Notice that we also obtain a hierarchy of *outer approximations* $(\mathrm{COP}_d) \subset \mathcal{S}_n$ of the cone COP of $n \times n$ copositive matrices. Indeed, for every $\mathbf{A} \in \mathcal{S}_n$, let $f_\mathbf{A}$ be the quadratic form $\mathbf{x} \mapsto f_\mathbf{A}(\mathbf{x}) := \mathbf{x}^T \mathbf{A} \mathbf{x}$. Then, for every d, the set
$$\mathrm{COP}_d \,:=\, \{\mathbf{A} \in \mathcal{S}_n \,:\, \mathbf{M}_d(f_\mathbf{A}\,\mathbf{y}) \succeq 0\,\}$$

is a convex cone defined only in terms of the coefficients of the matrix $\mathbf{A}$. It is even a spectrahedron since $\mathbf{M}_d(f_{\mathbf{A}}\,\mathbf{y})$ is a linear matrix inequality in the coefficients of $\mathbf{A}$. And in view of Theorem 3.2(a), $\mathrm{COP} = \bigcap_{d\in\mathbb{N}} \mathrm{COP}_d$.

4.2. **Examples.** In this section we provide three simple examples to illustrate the above methodology.

Example 1. Consider the global minimization on $\mathbf{K} = \mathbb{R}^2_+$ of the Motzkin-like polynomial $\mathbf{x} \mapsto f(\mathbf{x}) = x_1^2 x_2^2\,(x_1^2 + x_2^2 - 1)$ whose global minimum is $f^* = -1/27 \approx -0.037$, attained at $(x_1^*, x_2^*) = (\pm\sqrt{1/3}, \pm\sqrt{1/3})$. Choose for μ the probability measure $\mu(B) := \int_B \mathrm{e}^{-x_1-x_2} d\mathbf{x}$, $B \in \mathcal{B}(\mathbb{R}^2_+)$, for which the sequence of moments $\mathbf{y} = (y_{ij})$, $i, j \in \mathbb{N}$, is easy to obtain. Namely $y_{ij} = i!j!$ for every $i, j \in \mathbb{N}$. Then the semidefinite relaxations (4.2) yield $\lambda_0 = 92$, $\lambda_1 = 1.5097$, and $\lambda_{14} = -0.0113$, showing a significant and rapid decrease in first iterations with a long tail close to f^*, illustrated in Figure 1. Then after $d = 14$, one encounters some numerical problems and we cannot trust the results anymore.

If we now minimize the same polynomial f on the box $[0, 1]^2$, one choose for μ the probability uniformly distributed on $[0, 1]^2$, whose moments $\mathbf{y} = (y_{ij})$, $i, j \in \mathbb{N}$, are also easily obtained by $y_{ij} = (i + 1)^{-1}(j + 1)^{-1}$. Then one obtains $\lambda_0 = 0.222$, $\lambda_1 = -0.055$, and $\lambda_{10} = -0.0311$, showing again a rapid decrease in first iterations with a long tail close to f^*, illustrated in Figure 2.

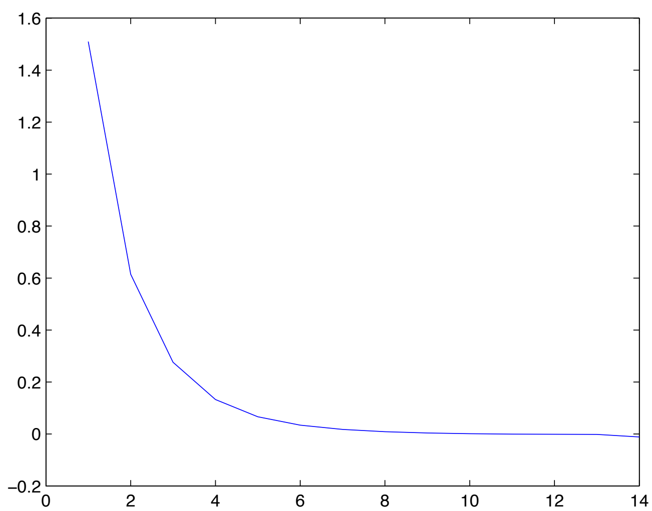

FIGURE 1. Minimizing the Motzkin-like polynomial in $\mathbb{R}^2_+$

Example 2. Still on $\mathbf{K} = \mathbb{R}^2_+$, consider the global minimization of the polynomial $\mathbf{x} \mapsto x_1^2 + (1 - x_1x_2)^2$ whose global minimum $f^* = 0$ is not attained. Again, choose for μ the probability measure $\mu(B) := \int_B \mathrm{e}^{-x_1-x_2} d\mathbf{x}$, $B \in \mathcal{B}(\mathbb{R}^2_+)$. Then the semidefinite relaxations (4.2) yield $\lambda_0 = 5$, $\lambda_1 = 1.9187$ and $\lambda_{15} = 0.4795$, showing again a significant and rapid decrease in first iterations with a long tail close to f^*,

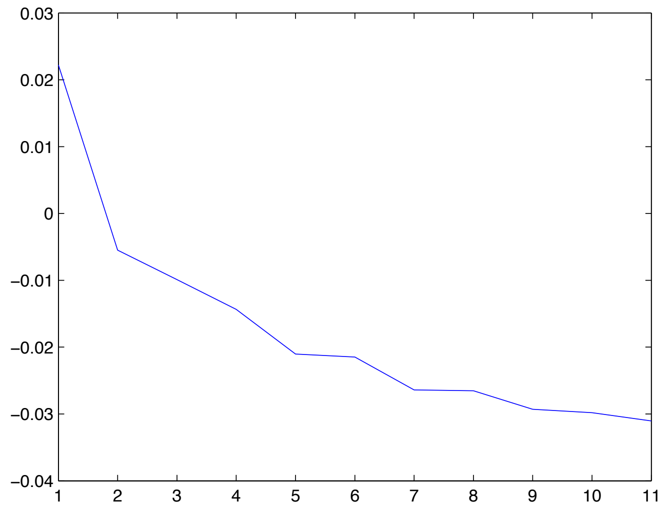

FIGURE 2. Minimizing the Motzkin-like polynomial in $[0,1]^2$

illustrated in Figure 3; numerical problems occur after $d = 15$. However, this kind of problems where the global minimum f^* is not attained, is notoriously difficult. Even the semidefinite relaxations defined in [8] (which provide lower bounds on f^*) and especially devised for such problems, encounter numerical difficulties; see [8, Example 4.8].

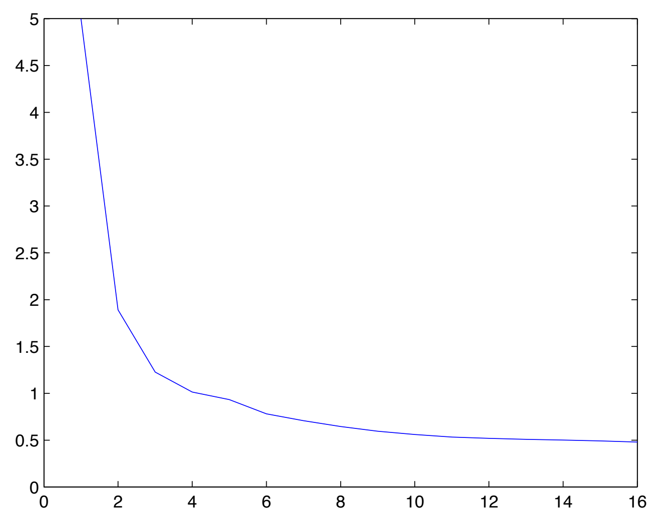

FIGURE 3. Minimizing $x_1^2 + (1 - x_1x_2)^2$ on $\mathbb{R}^2_+$

Example 3. The following example illustrates the duality results of Section §4.1. The univariate polynomial $x \mapsto f(x) := 0.375 - 5x + 21x^2 - 32x^3 + 16x^4$ displayed in Fig 4 has two global minima at $x_1^* = 0.1939$ and $x_2^* = 0.8062$, with $f^* = -0.0156$. In Fig 5 is plotted the sequence of upper bounds $\lambda_d \to f^*$ as $\hat{d} \to \infty$, with again a rapid decrease in first iterations. One has plotted in Fig 6 the s.o.s. polynomial $x \mapsto \sigma(x)$, optimal solution of (4.4) with $d = 10$, associated with the probability density $\sigma(x)dx$ as explained in §4.1. As expected, two peaks appear at the points $\tilde{x}_i \approx x_i^*$, $i = 1, 2$.

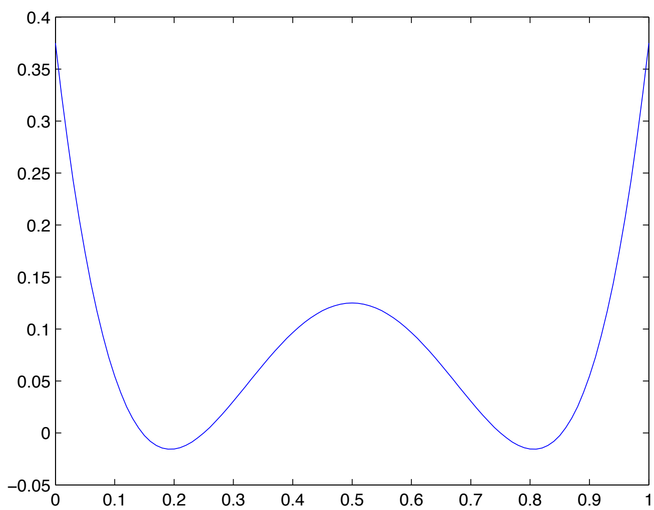

FIGURE 4. $f(x) = 0.375 - 5x + 21x^2 - 32x^3 + 16x^4$ on $[0, 1]$

Example 4. We finally consider a discrete optimization problem, namely the celebrated MAXCUT problem

$$f^* = \min_{\mathbf{x}} \{\mathbf{x}^T \mathbf{Q} \mathbf{x} \,:\, \mathbf{x} \in \{-1, 1\}^n\},$$

where $\mathbf{Q} = (Q_{ij}) \in \mathbb{R}^{n \times n}$ is a real symmetric matrix whose all diagonal elements vanish. The measure μ is uniformly distributed on $\{-1, 1\}^n$ so that its moments are readily available. We first consider the equal weights case, i.e., $Q_{ij} = 1/2$ for all (i, j) with $i \neq j$ in which case $f^* = -\lfloor n/2 \rfloor$. With $n = 11$ the successive values for λ_d, $d \leq 4$, are displayed in Table 1 and λ_4 is relatively close to f^*. Next we have generated five random instances of MAXCUT with $n = 11$ but $Q_{ij} = 0$ with probability 1/2, and if $Q_{ij} \neq 0$ it is randomly generated using the Matlab "rand" function. The successive values of λ_d, $d \leq 4$, are displayed in Table 2, and again, λ_4 is quite close to f^*[6].

The above 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 and

[6]The optimal value f^* has been computed using the GloptiPoly software [10] dedicated to solving the Generalized Problem of Moments

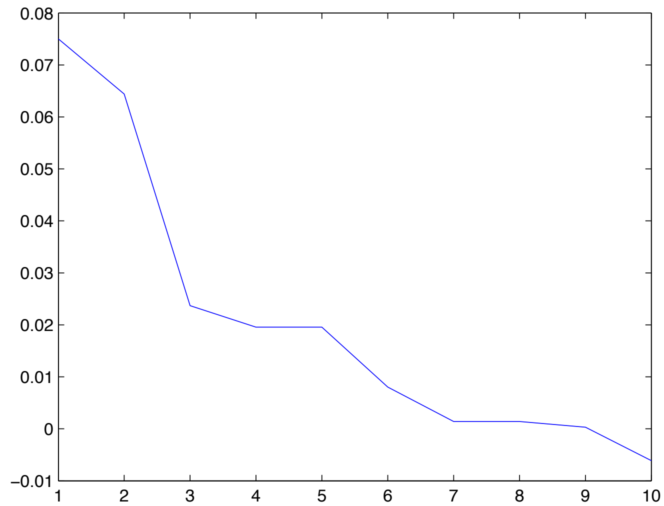

FIGURE 5. Minimizing $0.375 - 5x + 21x^2 - 32x^3 + 16x^4$ on $[0,1]$

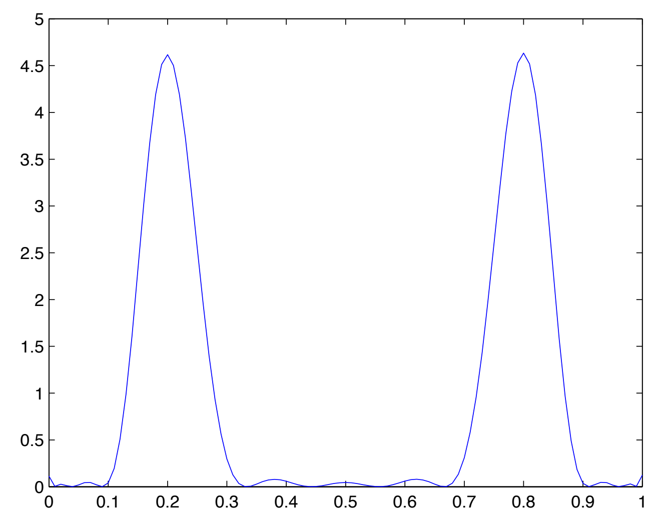

FIGURE 6. The probability density $\sigma(x)dx$ on $[0,1]$

then a slow convergence close to f^*. 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! In fact solving the semidefinite program (4.3) is computing the smallest *generalized eigenvalue* associated with the pair of real symmetric matrices $(\mathbf{M}_d(f\,\mathbf{y}), \mathbf{M}_d(\mathbf{y}))$, for which specialized codes are available (instead of using

d	$d=0$	$d=1$	$d=2$	$d=3$	$d=4$	f^*
λ_d	0	-1	-2.662	-3.22	-4	-5

TABLE 1. MAXCUT: $n = 11$; $Q(i, j) = 1$ for all $i \neq j$.

d	λ_0	λ_1	λ_2	λ_3	λ_4	f^*
Ex1	0	-1.928	-3.748	-5.22	-6.37	-7.946
Ex2	0	-1.56	-3.103	-4.314	-5.282	-6.863
Ex3	0	-1.910	-3.694	-5.078	-6.161	-8.032
Ex4	0	-2.164	-4.1664	-5.7971	-7.06	-9.198
Ex5	0	-1.825	-3.560	-4.945	-5.924	-7.467

TABLE 2. MAXCUT: $n = 11$; Q random.

a solver for semidefinite programs). However, one has to remember that the choice is limited to measures μ with $\operatorname{supp}\mu = \mathbf{K}$ and whose moments are available or easy to compute. Hence, the present methodology is so far limited to simple sets $\mathbf{K}$ as described before. Finally, analyzing how the convergence to f^* depends on μ is beyond the scope of the present paper and is a topic of further research.

4.3. **Discussssion.** In nonlinear programming, sequences of upper bounds on the global minimum f^* are usually obtained from feasible points $\mathbf{x} \in \mathbf{K}$, e.g., via some (local) minimization algorithm. But for non convex problems, providing a sequence of upper bounds that converges to the global minimum f^* is in general impossible unless one computes points on a grid whose mesh size tends to zero. In the above methodology one provides a monotone nonincreasing sequence of upper bounds converging to f^* for polynomial optimization problems on sets $\mathbf{K}$, non necessarily compact but such that one may compute all moments of some finite Borel measure μ with $\operatorname{supp}\mu = \mathbf{K}$. In fact, if there are only finitely many (say up to order $2d$) moments available then one obtains a finite sequence of upper bounds.

In contrast to the hierarchy of semidefinite relaxations in e.g. [12, 13] which provide lower bounds converging to f^* when $\mathbf{K}$ is a compact basic semi-algebraic set, the convergence of the upper bounds to f^* is only asymptotic and never finite, except when $\mathbf{K}$ is a discrete set. However, and even if we expect the convergence to be rather slow when close to f^*, to our knowledge it is the first approach of this kind, and in a few iterations one may obtain upper bounds which (even if crude) complements the lower bounds obtained in [12] (in the compact case).

Also note that to solve (4.3) several improvements are possible. For instance, we have already mentioned that it could be solved via specialized packages for generalized eigenvalue problems. Next, if instead of using the canonical basis of monomial $(\mathbf{x}^\alpha)$, one now expresses the moment matrix $\mathbf{M}_d(\mathbf{y})$ with rows and columns indexed in the basis of polynomials $(p_\alpha) \subset \mathbb{R}[\mathbf{x}]$ (up to degree d) *orthogonal*[7] with respect to μ, then $\mathbf{M}_d(\mathbf{y})$ becomes the identity matrix. And so problem (4.3) reduces to a standard eigenvalue problem, namely that of computing the smallest eigenvalue of the (real and symmetric) localizing matrix $\mathbf{M}_d(f\,\mathbf{y})$ (expressed in the basis of

[7] A family of univariate polynomials $(p_k) \subset \mathbb{R}[x]$ is orthogonal with respect to a finite measure μ on $\mathbb{R}$ if $\int p_i p_k d\mu = \delta_{i=k}$. For extensions to the multivariate case see e.g. [6, 9].

orthogonal polynomials)! And it turns out that computing the orthogonal polynomials is easy once the moment matrix $\mathbf{M}_d(\mathbf{y})$ is available, since they can be obtained via computing certain determinants, as explained in e.g. [6, 9].

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. [5] 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 polynomial 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 ome appropriate (linear or convex) objective function of its vector of coefficents $\mathbf{f}$.

5. Conclusion

In this paper we have presented a new characterization of nonnegativity on a closed set $\mathbf{K}$ which is based on the knowledge of a single finite Borel measure μ with $\operatorname{supp}\mu = \mathbf{K}$. It permits to obtain a hierarchy of spectrahedra which provides a nested sequence of outer approximations of the convex cone of polynomials of degree at most d, 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. [12, 13] when $\mathbf{K}$ is compact and basic semi-algebraic. 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

Proof of Theorem 3.4.

Proof. The *only if* part is exactly the same as in the proof of Theorem 3.2. For the *if* part, let $\mathbf{z} = (z_\alpha)$ be the sequence defined in (3.5). The sequence $\mathbf{z}$ is well defined because $\mathbf{x} \mapsto \mathbf{x}^\alpha f(\mathbf{x})$ is μ-integrable for all $\alpha \in \mathbb{N}^n$. Indeed, if f is a polynomial (so that μ is defined in (3.7)) we have seen that all moments of μ are finite and since $\mathbf{x}^\alpha f(\mathbf{x})$ is a polynomial the result follows. If f is not a polynomial (so that μ is defined in (3.9)) then

$$\int_{\mathbf{K}} |\mathbf{x}^\alpha f(\mathbf{x})| \, d\mu(\mathbf{x}) \leq \int_{\mathbf{K}} |\mathbf{x}^\alpha| \exp(-\sum_i |x_i|) \, d\varphi(\mathbf{x}) \leq \varphi(\mathbb{R}^n) \prod_{i=1}^n \alpha_i!,$$

where we have used that $|x_i^{\alpha_i}| \leq \alpha_i! \exp |x_i|$, and $|f|/(1+f^2) \leq 1$ for all $\mathbf{x}$. As in the proof of Theorem 3.2, the set function $B \mapsto \nu(B) := \int_B f d\mu$, $B \in \mathcal{B}$, is a signed Borel measure because again ν can be written as the difference $\nu_1 - \nu_2$ of the two positive Borel measures ν_1, ν_2 in (3.1)-(3.2). With same majorizations as above, both ν_1 and ν_2 are finite Borel measures and so ν is a finite signed Borel measure.

The same arguments as in the proof of Theorem 3.2 show that $\mathbf{M}_d(\mathbf{z}) \succeq 0$ for every $d \in \mathbb{N}$. Next, the sequence $\mathbf{z}$ satisfies the generalized Carleman's condition (2.2).

Indeed, first consider the case where f is a polynomial (and so μ is as in (3.7)). Let $1 \leq i \leq n$ be fixed arbitrary, and let $2s \geq \deg f$. Observe that whenever $|\alpha| \leq k$, $|\mathbf{x}|^\alpha \leq |\mathbf{x}_j|^k$ on the subset $W_j := \{\mathbf{x} \in \mathbb{R}^n \setminus [-1,1]^n \,:\, |x_j| = \max_i |x_i|\}$. And so, $|f(\mathbf{x})| \leq \|f\|_1 \, |\mathbf{x}_j|^{2s}$ for all $\mathbf{x} \in W_j$ (and where $\|f\|_1 := \sum_\alpha |f_\alpha|$). Hence,

$$\begin{aligned}
L_{\mathbf{z}}(x_i^{2k}) &= \int_{\mathbf{K}} f(\mathbf{x})\, x_i^{2k} d\mu(\mathbf{x}) \\
&\leq \int_{\mathbf{K}\cap[-1,1]^n} |f(\mathbf{x})|\, x_i^{2k} d\mu(\mathbf{x}) + \|f\|_1 \sum_{j=1}^n \int_{\mathbf{K}\cap W_J} x_j^{2(k+s)} d\mu(\mathbf{x}) \\
&\leq \|f\|_1 \mu(\mathbf{K}) + Mn\|f\|_1\, (2(k+s))! \;\leq\; 2Mn\|f\|_1\, (2(k+s))!, \qquad (6.1)
\end{aligned}$$

where M is as in (3.8) and assuming with no loss of generality that $\mu(\mathbf{K}) \leq Mn(2(k+s))!$ (otherwise rescale φ). And so we have

$$\begin{aligned}
L_{\mathbf{z}}(x_i^{2k})^{-1/2k} &\geq (2Mn\|f\|_1)^{-1/2k} \left((2(k+s))!)^{-1/2(k+s)}\right)^{(k+s)/k} \\
&\geq \frac{1}{2}\left((2(k+s))!)^{-1/2(k+s)}\right)^{(k+s)/k} \\
&\geq \frac{1}{2}\left(\frac{1}{2(k+s)}\right)^{(k+s)/k},
\end{aligned}$$

where $k \geq k_0$ is sufficiently large so that $(2Mn\|f\|_1)^{-1/2k} \geq 1/2$. Therefore,

$$\sum_{k=1}^\infty L_{\mathbf{z}}(x_i^{2k})^{-1/2k} \geq \frac{1}{2}\sum_{k=k_0}^\infty \left(\frac{1}{2(k+s)}\right)^{(k+s)/k} = +\infty,$$

where the last equality follows from $\sum_{k=1}^\infty (2k)^{-1} = +\infty$. Indeed, $(\frac{1}{2(k+s)})^{(k+s)/k} = (\frac{1}{2(k+s)})(\frac{1}{2(k+s)})^{s/k}$ and $(\frac{1}{2(k+s)})^{s/k} \geq 1/2$ whenever k is sufficiently large, say $k \geq k_1$. Hence the sequence $\mathbf{z}$ satisfies Carleman's condition (2.2).

If f is not a polynomial then μ is as in (3.9) and so

$$\begin{aligned}
L_{\mathbf{z}}(x_i^{2k}) &= \int x_i^{2k} \frac{\exp\left(-\sum_{i=1}^n |x_i|\right)}{1+f^2} f(\mathbf{x})\, d\varphi(\mathbf{x}) \\
&\leq \int x_i^{2k} \frac{\exp\left(-\sum_{i=1}^n |x_i|\right)}{1+f^2} |f(\mathbf{x})|\, d\varphi(\mathbf{x}) \\
&\leq \int x_i^{2k} \exp\left(-\sum_{i=1}^n |x_i|\right) d\varphi(\mathbf{x}) \;\leq\; (2k)!M, \qquad (6.2)
\end{aligned}$$

where we have used that $|f|/(1+f^2) \leq 1$ for all $\mathbf{x}$, and $x_i^{2k} \leq (2k)! \exp|x_i|$. And so again, the sequence $\mathbf{z}$ satisfies Carleman's condition (2.2).

Next, as $\mathbf{M}_d(\mathbf{z}) \succeq 0$ for every $d \in \mathbb{N}$, 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 (3.5) of ν and $\mathbf{z}$, we have

$$\int_{\mathbb{R}^n} \mathbf{x}^\alpha\, d\psi(\mathbf{x}) \;=\; z_\alpha \;=\; \int_{\mathbf{K}} \mathbf{x}^\alpha d\nu(\mathbf{x}), \qquad \forall\, \alpha \in \mathbb{N}^n. \qquad (6.3)$$

But then using $\nu=\nu_1-\nu_2$ in (3.1)-(3.2), (6.3) reads

$$\int_{\mathbb{R}^n}\mathbf{x}^\alpha\,d(\psi+\nu_2)(\mathbf{x})\,=\,\int_{\mathbf{K}}\mathbf{x}^\alpha d\nu_1(\mathbf{x}),\qquad\forall\alpha\in\mathbb{N}^n.\tag{6.4}$$

Let $\mathbf{v}=(v_\alpha)$, $\alpha\in\mathbb{N}^n$, be the sequence of moments associated with ν_1. Of course, $\mathbf{M}_d(\mathbf{v})\succeq 0$ for all $d\in\mathbb{N}$. Next,

$$L_{\mathbf{v}}(x_i^{2k})\,=\,\int_{B_1}x_i^{2k}\,f(\mathbf{x})\,d\mu(\mathbf{x})\,\leq\,\int_{\mathbf{K}}x_i^{2k}\,|f(\mathbf{x})|\,d\mu(\mathbf{x}),$$

and so, depending on whether f is a polynomial or not, we obtain $L_{\mathbf{v}}(x_i^{2k})\leq 2Mn\|f\|_1\,(2(k+s))!$ as in (6.1) or $L_{\mathbf{v}}(x_i^{2k})\leq(2k)!M$ as in (6.2). In both cases the sequence $\mathbf{v}$ satisfies Carleman's condition (2.2) and since $\mathbf{M}_d(\mathbf{v})\succeq 0$ for all $d\in\mathbb{N}$, by Proposition 2.1, ν_1 is moment determinate. But then (6.4) yields $\nu_1=\psi+\nu_2$, or equivalently, $\psi=\nu_1-\nu_2\,(=\nu)$, that is, ν is a positive measure. The rest of the proof is exactly the same as for proof of Theorem 3.2. □

References

[1] K.M. Anstreicher, S. Burer. Computable representations for convex hulls of low-dimensional quadratic forms, Math. Program. Ser. B **124** (2010), pp. 33?43.

[2] V. Baldoni, N. Berline, J. De Loera, M. Köppe, and M. Vergne. How to integrate a polynomial on a simplex, Math. Comp. **80** (2011), pp. 297–325.

[3] C. Berg. The multidimensional moment problem and semi-groups, Proc. Symp. Appl. Math. **37** (1980), pp. 110–124.

[4] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares, Israel J. Math. **153** (2006), pp. 355–380.

[5] A. Cuyt, G. Golub, P. Milanfar and B. Verdonk. Mutlidimensional integral inversion with application in shape reconstruction, SIAM J. Sci. Comput. **27** (2005), pp. 1058–1070.

[6] C.F. Dunkl and Y. Xu. *Orthogonal Polynomials of Several Variables*, Cambridge Univ. Press., Cambridge, 2001.

[7] M. Dür. Copositive Programming: A survey, `Optimization-online`, 2009.

[8] Hà Huy Vui and T. S. Pham. Global optimization of polynomials using the truncated tangency variety and sums of squares, SIAM J. Optim. **19** (2008), pp. 941–951.

[9] J.W. Helton, J.B. Lasserre, and M. Putinar. Measures with zeros in the inverse of their moment matrix, Ann. Probab. **36** (2008), pp. 1453–1471.

[10] D. Henrion, J.B. Lasserre, Y. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming, Optim. Methods and Softwares **24** (2009), pp. 761–779.

[11] P.H. Maserick and C. Berg. Exponentially bounded positive definite functions. Ill, J. Math. **28** (1984), pp. 162– 179.

[12] J.B. Lasserre. Global optimization with polynomials and the problem of moments, SIAM J. Optim. **11** (2001), pp. 796–817.

[13] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.

[14] J.B. Lasserre and M. Putinar. Positivity and optimization for semi-algebraic functions, SIAM J. Optim. **20** (2010), pp. 3364–3383.

[15] E.J. McShane. Jensen's inequality, Bull. Amer. Math. Soc. **43** (1937), pp. 521–527.

[16] M. Marshall and T. Netzer. Positivstellensätze for real function algebras (2010). `arXiv:1004.4521v1`.

[17] A.E. Nussbaum. Quasi-analytic vectors. Ark. Mat. **6** (1966), pp. 179–191.

[18] M. Putinar. Positive polynomials on compact semi-algebraic sets, Indiana Univ. Math. J. **42** (1993), 969–984.

[19] H.L. Royden. *Real Analysis*, 3rd. ed., Macmillan Publishing Company, New York, 1988.

[20] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets, Math. Ann. **289** (1991), 203–206.

[21] M. Schweighofer. Global optimization of polynomials using gradient tentacles and sums of squares, SIAM J. Optim. **17** (2006), pp. 920–942.

[22] G. Stengel. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry, Math. Ann. **207**, pp. 87–97.

[23] 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