A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis

Etienne de Klerk · Monique Laurent

Abstract The generalized problem of moments is a conic linear optimization problem over the convex cone of positive Borel measures with given support. It has a large variety of applications, including global optimization of polynomials and rational functions, options pricing in finance, constructing quadrature schemes for numerical integration, and distributionally robust optimization. A usual solution approach, due to J.B. Lasserre, is to approximate the convex cone of positive Borel measures by finite dimensional outer and inner conic approximations. We will review some results on these approximations, with a special focus on the convergence rate of the hierarchies of upper and lower bounds for the general problem of moments that are obtained from these inner and outer approximations.

1 Introduction

The classical problem of moments asks when a measure is determined by a set of specified moments and variants of this problem were studied (in the univariate case) by leading 19th and early 20th century mathematicians, like Hamburger, Stieltjes, Chebyshev, Hausdorff, and Markov. We refer to [1] for an early reference and to the recent monograph [52] for a comprehensive treatment of the moment problem.

The generalized problem of moments asks to optimize a linear function over the set of finite, positive Borel measures that satisfy certain moment-type conditions. More precisely, we consider continuous functions f_0 and f_i ($i \in [m]$) where $[m] =$

Etienne de Klerk
Tilburg University and Delft University of Technology, e-mail: E.deKlerk@UvT.nl

Monique Laurent
CWI Amsterdam, and Tilburg University, e-mail: M.Laurent@CWI.nl

$\{1, \ldots, m\}$, that are defined on a compact set $K \subset \mathbb{R}^n$. The generalized problem of moments (GPM) may now be defined as follows.[1]

Generalized problem of moments (GPM)

$$val := \inf_{\mu \in \mathcal{M}(K)_+} \left\{ \int_K f_0(x) d\mu(x) \; : \; \int_K f_i(x) d\mu(x) = b_i \quad \forall i \in [m] \right\}, \quad (1)$$

where

- $\mathcal{M}(K)_+$ denotes the convex cone of positive, finite, Borel measures (i.e., Radon measures) supported on the set K;[2]
- The scalars $b_i \in \mathbb{R}$ ($i \in [m]$) are given.

In this survey we will mostly consider the case where all f_i's are polynomials, and will always assume $K \subseteq \mathbb{R}^n$ to be compact. Moreover, for some of the results, we will also assume that K is a basic semi-algebraic set and we will sometimes further restrict to simple sets like a hypercube, simplex or sphere.

The generalized problem of moments has a rich history; see, e.g., [1, 31, 52] and references therein and [37] for a recent overview of many of its applications. In the recent years modern optimization approaches have been investigated in depth, in particular, by Lasserre (see [33], the monograph [34] and further references therein). Among others, there is a well-understood duality theory, and hierarchies of inner and outer approximations for the cone $\mathcal{M}(K)_+$ have been introduced that lead to converging upper and lower bounds for the problem (1). In this survey we will present these hierarchies and show how the corresponding bounds can be computed using semidefinite programming. Since several overviews are already available on general properties of these hierarchies (e.g., in [34, 35, 38, 39]), our main focus here will be on recent results that describe their rate of convergence. We will review in particular in more detail recent results on the upper bounds arising from the inner approximations, and highlight some recent links made with orthogonal polynomials and cubature rules for integration.

---

[1] We only deal with the GPM in a restricted setting; more general versions of the problem are studied in, e.g., [55].

[2] Formally, we consider the usual Borel σ-algebra, say $\mathcal{B}$, on $\mathbb{R}^n$, i.e., the smallest (or coarsest) σ-algebra that contains the open sets in $\mathbb{R}^n$. A positive, finite Borel measure μ is a nonnegative-valued set function on $\mathcal{B}$, that is countably additive for disjoint sets in $\mathcal{B}$. The support of μ is the set, denoted Supp(μ), and defined as the smallest closed set S such that $\mu(\mathbb{R}^n \setminus S) = 0$.

1.1 The dual problem of the GPM

The GPM is an infinite-dimensional conic linear program, and therefore it has an associated dual problem. Formally we introduce a duality (or pairing) between the following two vector spaces:

1. the space $\mathcal{M}(K)$ of all signed, finite, Borel measures supported on K,
2. the space $C(K)$ of continuous functions on K, endowed with the supremum norm $\|\cdot\|_\infty$.

The duality (pairing) in question is provided by the nondegenerate bilinear form $\langle \cdot, \cdot \rangle : C(K) \times \mathcal{M}(K) \to \mathbb{R}$, defined by

$$\langle f, \mu \rangle = \int_K f(x) d\mu(x) \quad (f \in C(K), \ \mu \in \mathcal{M}(K)).$$

Thus the dual cone of $\mathcal{M}(K)_+$ w.r.t. this duality is the cone of continuous functions that are nonnegative on K, and will be denoted by $C(K)_+ = (\mathcal{M}(K)_+)^*$.

In our setting of compact $K \subset \mathbb{R}^n$, $\mathcal{M}(K)$ is also the dual space of $C(K)$, i.e., $\mathcal{M}(K)$ may be associated with the space of linear functionals defined on $C(K)$. In particular, due to the Riesz-Markov-Kakutani representation theorem (e.g. [56, §1.10]), every linear functional on $C(K)$ may be expressed as

$$f \mapsto \langle f, \mu \rangle \quad \text{for a suitable } \mu \in \mathcal{M}(K).$$

As a result, we have the weak∗ topology on $\mathcal{M}(K)$ where the open sets are finite intersections of elementary sets of the form

$$\{\mu \in \mathcal{M}(K) \mid \alpha < \langle f, \mu \rangle < \beta\},$$

for given $\alpha, \beta \in \mathbb{R}$, and $f \in C(K)$, and the unions of such finite intersections.

A sequence $\{\mu_k\} \subset \mathcal{M}(K)$ converges in the weak∗ topology, say $\mu_k \rightharpoonup \mu$, if, and only if,

$$\lim_{k \to \infty} \langle f, \mu_k \rangle = \langle f, \mu \rangle \ \forall f \in C(K). \tag{2}$$

As a consequence of (2), the cone $\mathcal{M}(K)_+$ is closed and the set of probability measures in $\mathcal{M}(K)$ is closed.

By Alaoglu's theorem, e.g. [2, Theorem III(2.9)], the following set (i.e., the unit ball in $\mathcal{M}(K)$) is compact in the weak∗ topology of $\mathcal{M}(K)$:

$$\{\mu \in \mathcal{M}(K) \mid |\langle f, \mu \rangle| \le 1 \ \forall f \in C(K) \text{ with } \|f\|_\infty \le 1\}. \tag{3}$$

Hence the set of probability measures in $\mathcal{M}(K)$ is compact, since it is a closed subset of the compact set in (3), and thus it provides a compact base in the weak∗ topology for the cone $\mathcal{M}(K)_+$. This implies again that $\mathcal{M}(K)_+$ is closed in this topology (using Lemma 7.3 in [2, Part IV]) and we will also use this fact to analyze duality in the next section.

Dual linear optimization problem of (1)

Using this duality setting, the dual conic linear program of (1) reads

$$
val^* := \sup_{y \in \mathbb{R}^m} \left\{ \sum_{i \in [m]} b_i y_i \; : \; f_0 - \sum_{i \in I} y_i f_i \in C(K)_+ \right\},
$$

$$
= \sup_{y \in \mathbb{R}^m} \left\{ \sum_{i \in [m]} b_i y_i \; : \; f_0(x) - \sum_{i \in I} y_i f_i(x) \geq 0 \; \forall x \in K \right\}. \tag{4}
$$

By the duality theory of conic linear optimization, one has the following duality relations; see, e.g., [2, Section IV.7.2] or [34, Appendix C].

Theorem 1 *Consider the GPM (1) and its dual (4). Assume (1) has a feasible solution. One has* $val \geq val^*$ *(weak duality), with equality* $val = val^*$ *(strong duality) if the cone* $\{(\langle f_0, \mu \rangle, \langle f_1, \mu \rangle, \ldots \langle f_m, \mu \rangle) \; : \; \mu \in \mathcal{M}(K)_+\}$ *is a closed subset of* $\mathbb{R}^{m+1}$. *If, in addition,* $val > -\infty$ *then* (1) *has an optimal solution.*

We mention another sufficient condition for strong duality, that is a consequence of Theorem 1 in our setting.

Corollary 1 *Assume (1) has a feasible solution, and there exist* $z_0, z_1, \ldots, z_m \in \mathbb{R}$ *for which the function* $\sum_{i=0}^m z_i f_i$ *is strictly positive on* K *(i.e.,* $\sum_{i=0}^m z_i f_i(x) > 0$ *for all* $x \in K$). *Then,* $val = val^*$ *holds and* (1) *has an optimal solution.*

Hence, if in problem (1) we optimize over the probability measures (i.e., with $f_1 \equiv 1$, $b_1 = 1$) then the assumptions in Corollary 1 are satisfied.

We indicate how Corollary 1 can be derived from Theorem 1. Consider the linear map $L : \mathcal{M}(K) \to \mathbb{R}^{m+1}$ defined by $L(\mu) = (\langle f_0, \mu \rangle, \ldots, \langle f_m, \mu \rangle)$, which is continuous w.r.t. the weak* topology on $\mathcal{M}(K)$. First we claim Ker $L \cap \mathcal{M}(K)_+ = \{0\}$. Indeed, assume $L(\mu) = 0$ for some $\mu \in \mathcal{M}(K)_+$. Setting $f = \sum_{i=0}^m z_i f_i$, $L(\mu) = 0$ implies $\langle f, \mu \rangle = 0$ and thus $\mu = 0$ since f is strictly positive on K. Since the cone $\mathcal{M}(K)_*$ has a compact convex base in the weak∗ topology and the linear map L is continuous, we can conclude that the image $L(\mathcal{M}(K)_+)$ is closed (using Lemma 7.3 in [2, Part IV]). Now we can conclude using Theorem 1.

1.2 Atomic solution of the GPM

If the GPM has an optimal solution, then it has a finite atomic optimal solution, supported on at most m points (i.e., the weighted sum of at most m Dirac delta measures). This is a classical result in the theory of moments; see, e.g., [49] (univariate case), [30] (which shows an atomic measure with $m + 1$ atoms using induction on m) and a modern exposition in [55] (which shows an atomic measure with m atoms).

The result may also be obtained as a consequence of the following, dimension-free version of the Carathéodory theorem.

Theorem 2 (see, e.g., Theorem 9.2 in Chapter III of [2])

Let S be a convex subset of a vector space such that, for every line L, the intersection $S \cap L$ is a closed bounded interval. Then every extreme point of the intersection of S with m hyperplanes can be expressed as a convex combination of at most $m + 1$ extreme points of S.

Atomic solution of the (GPM)

Theorem 3 *If the GPM* (1) *has an optimal solution then it has one which is finite atomic with at most m atoms, i.e., of the form $\mu^* = \sum_{\ell=1}^{m} w_\ell \delta_{x^{(\ell)}}$ where $w_\ell \geq 0$, $x^{(\ell)} \in K$, and $\delta_{x^{(\ell)}}$ denotes the Dirac measure supported at $x^{(\ell)}$ ($\ell \in [m]$).*

This result can be derived from Theorem 2 in the following way. By assumption, the GPM has an optimal solution μ^*. Moreover, since it has one at an extreme point we may assume that μ^* is an extreme point of the feasibility region $\mathcal{M}(K)_+ \cap \cap_{i=1}^{m} H_i$ of the program (1), where H_i is the hyperplane $\langle f_i, \mu \rangle = b_i$. Then the following set $S = \{\mu \in \mathcal{M}(K)_+ : \mu(K) = \mu^*(K)\}$ meets the condition of Theorem 2, since the set of probability measures in $\mathcal{M}(K)_+$ is compact in the weak∗ topology, and any line in a topological vector space is closed (e.g. [2, p. 111]). Moreover, the extreme points of S are precisely the scaled Dirac measures supported by points in K (see, e.g., Section III.8 in [2]). In addition, μ^* is an extreme point of the set $S \cap \cap_{i=1}^{m} H_i$ and thus, by Theorem 2, μ^* is a conic combination of $m + 1$ Dirac measures supported at points $x^{(\ell)} \in K$ for $\ell \in [m + 1]$. Finally, as in [55], consider the LP

$$\min \sum_{\ell=1}^{m+1} w_\ell f_0(x^{(\ell)}) \text{ s.t. } w_\ell \geq 0 \ (\ell \in [m + 1]), \sum_{\ell=1}^{m+1} w_\ell f_i(x^{(\ell)}) = b_i \ (i \in [m])$$

whose optimum value is *val*. Then an optimal solution attained at an extreme point provides an optimal solution of the GPM (1) which is atomic with at most m atoms.

1.3 GPM in terms of moments

From now on we will assume the functions $f_0, f_1, \ldots, f_m$ in the definition of the GPM (1) are all polynomials and the set K is compact. Then the GPM may be reformulated in terms of the moments of the variable measure μ. To be precise, given a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ the moment of order α of a measure $\mu \in \mathcal{M}(K)_+$ is defined as

$$m_\alpha^\mu(K) := \int_K x^\alpha d\mu(x).$$

Here we set $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. We may write the polynomials $f_0, f_1, \ldots, f_m$ in terms of the standard monomial basis as:

$$f_i(x) = \sum_{\alpha \in \mathbb{N}_d^n} f_{i,\alpha} x^\alpha \quad \forall i = 0, \ldots, m,$$

where the $f_{i,\alpha} \in \mathbb{R}$ are the coefficients in the monomial basis, and we assume the maximum total degree of the polynomials $f_0, f_1, \ldots, f_m$ to be at most d.

Throughout we let $\mathbb{N}_d^d = \{\alpha \in \mathbb{N}^n : |\alpha| \le d\}$ denote the set of multi-indices, with $|\alpha| = \sum_{i=1}^n \alpha_i$, and $\mathbb{R}[x]_d$ denotes the set of multivariate polynomials with degree at most d.

GPM in terms of moments

We may now rewrite the GPM (1) in terms of moments:

$$\inf_{\mu \in \mathcal{M}(K)_+} \left\{ \sum_{\alpha \in \mathbb{N}_d^n} f_{0,\alpha} m_\alpha^\mu(K) \; : \; \sum_{\alpha \in \mathbb{N}_d^n} f_{i,\alpha} m_\alpha^\mu(K) = b_i \; \forall i \in [m] \right\}.$$

Here d is the maximum degree of the polynomials $f_0, f_1, \ldots, f_m$.

Thus we may consider the set of all possible truncated moments sequences:

$$\left\{ \left(m_\alpha^\mu(K) \right)_{\alpha \in \mathbb{N}_d^n} \; : \; \mu \in \mathcal{M}(K)_+ \right\},$$

and describe the inner and outer approximations for $\mathcal{M}(K)_+$ in terms of this set.

1.4 Inner and outer approximations

We will consider two types of approximations of the cone $\mathcal{M}(K)_+$, namely inner and outer conic approximations.

Inner approximations

The underlying idea, due to Lasserre [36], is to consider a subset of measures μ in $\mathcal{M}(K)_+$ of the form

$$d\mu = h \cdot d\mu_0,$$

where h is a polynomial sum-of-squares density function, and $\mu_0 \in \mathcal{M}(K)_+$ is a fixed reference measure with $\mathrm{Supp}(\mu_0) = K$.

To obtain a finite dimensional subset of measures, we will limit the total degree of h to some value $2r$ where $r \in \mathbb{N}$ is fixed. The cone of sum-of-squares polynomials of total degree at most $2r$ will be denoted by Σ_r, hence

$$\Sigma_r = \left\{ \sum_{i=1}^{k} p_i^2 : k \in \mathbb{N}, p_i \in \mathbb{R}[x]_r, i \in [k] \right\}.$$

In this way one obtains the cones

$$\mathcal{M}_{\mu_0}^r := \{ \mu \in \mathcal{M}(K)_+ \ : \ d\mu = h \cdot d\mu_0, \ h \in \Sigma_r \} \quad (r = 1, 2, \ldots) \tag{5}$$

which provide a hierarchy of inner approximations for the set $\mathcal{M}(K)_+$:

$$\mathcal{M}_{\mu_0}^r \subseteq \mathcal{M}_{\mu_0}^{r+1} \subseteq \mathcal{M}(K)_+.$$

Outer approximations

The dual GPM (4) involves the nonnegativity constraint

$$f_0(x) - \sum_{i=1}^{m} y_i f_i(x) \geq 0 \ \forall x \in K,$$

which one may relax to a sufficient condition that guarantees the nonnegativity of the polynomial $f_0 - \sum_{i=1}^{m} y_i f_i$ on K. Lasserre [32] suggested to use the following sufficient condition in the case when K is a basic closed semi-algebraic set, i.e., when we have a description of K as the intersection of the level sets of polynomials g_j ($j \in [k]$):

$$K = \left\{ x \in \mathbb{R}^n \ : \ g_j(x) \geq 0 \quad \forall j \in [k] \right\}.$$

Namely, consider the condition

$$f_0 - \sum_{i=1}^{m} y_i f_i = \sigma_0 + \sum_{j=1}^{k} \sigma_j g_j,$$

where each σ_j is a sum-of-squares polynomial and the degree of each term $\sigma_j g_j$ ($0 \leq j \leq k$) is at most $2r$, so that the degree of the right-hand-side polynomial is at most $2r$. Here we set $g_0 \equiv 1$ for notational convenience. Thus we replace the cone $C(K)_+$ by a cone of the type:

$$Q^r(g_1, \ldots, g_k) := \left\{ f \ : \ f = \sigma_0 + \sum_{j=1}^{k} \sigma_j g_j, \ \sigma_j \in \Sigma_{r_j}, \ j = 0, 1, \ldots, k \right\}, \tag{6}$$

where we set $r_j := r - \lceil \deg(g_j)/2 \rceil$ for all $j \in \{0, \ldots, m\}$.

The cone $Q^r(g_1, \ldots, g_k)$ is known as the *truncated quadratic module* generated by the polynomials $g_1, \ldots, g_k$. By definition, its dual cone consists of the signed measures μ supported on K such that $\int_K f d\mu \geq 0$ for all $f \in Q^r(g_1, \ldots, g_k)$:

$$(Q^r(g_1, \ldots, g_k))^* = \left\{ \mu \in \mathcal{M}(K) : \int_K f(x) d\mu(x) \geq 0 \quad \forall f \in Q^r(g_1, \ldots, g_k) \right\}. \tag{7}$$

This provides a hierarchy of outer approximations for the cone $\mathcal{M}(K)_+$:

$$\mathcal{M}(K)_+ \subseteq (Q^{r+1}(g_1, \ldots, g_k))^* \subseteq (Q^r(g_1, \ldots, g_k))^*.$$

We will also briefly consider the tighter outer approximations for the cone $\mathcal{M}(K)_+$ obtained by replacing the truncated quadratic module $Q^r(g_1, \ldots, g_k)$ by the larger cone $Q^r \left(\prod_{j \in J} g_j : J \subseteq [k] \right)$, thus the truncated quadratic module generated by all pairwise products of the g_j's (also known as the pre-ordering generated by the g_j's). Then we have

$$\mathcal{M}(K)_+ \subseteq \left(Q^r \left(\prod_{j \in J} g_j : J \subseteq [k] \right) \right)^* \subseteq (Q^r(g_1, \ldots, g_k))^*.$$

2 Examples of GPM

The GPM (1) has many applications. Below we will list some examples that are directly relevant to this survey; additional examples in control theory, options pricing in finance, and others, can be found in [33, 34, 37].

Global minimization of polynomials on compact sets

Consider the global optimization problem:

$$val = \min_{x \in K} p(x) \tag{8}$$

where p is a polynomial and K a compact set. This corresponds to the GPM (1) with $m = 1$, $f_0 = p$, $f_1 = 1$ and $b_1 = 1$, i.e.:

$$val = \min_{\mu \in \mathcal{M}(K)_+} \left\{ \int_K p(x) d\mu(x) \; : \; \int_K d\mu(x) = 1 \right\}.$$

In the following sections we will focus on deriving error bounds for this problem when using the inner and outer approximations of $\mathcal{M}(K)_+$.

Global minimization of rational functions on compact sets

We may generalize the previous example to rational objective functions. In particular, we now consider the global optimization problem:

$$val = \min_{x \in K} \frac{p(x)}{q(x)}, \tag{9}$$

where p, q are polynomials such that $q(x) > 0 \; \forall \; x \in K$, and $K \subseteq \mathbb{R}^n$ is compact.

This problem has applications in many areas, including signal recovery [6] and finding minimal energy configurations of point charges in a field with polynomial potential [54].

It is simple to see that we may reformulate this problem as the GPM with $m = 1$ and $f_0 = p$, $f_1 = q$, and $b_1 = 1$, i.e.:

$$val = \min_{\mu \in \mathcal{M}(K)_+} \left\{ \int_K p(x)d\mu(x) \; : \; \int_K q(x)d\mu(x) = 1 \right\}.$$

Indeed, one may readily verify that if x^* is a global minimizer of the rational function $p(x)/q(x)$ over K then an optimal solution of the GPM is given by $\mu^* = \frac{1}{q(x^*)}\delta_{x^*}$.

Polynomial cubature rules

Positive cubature (also known as multivariate quadrature) rules for numerical integration of a function f with respect to a measure μ_0 over a set K take the form

$$\int_K f(x)d\mu_0(x) \approx \sum_{\ell=1}^{N} w_\ell f(x^{(\ell)}),$$

where the points $x^{(\ell)} \in K$ and the weights $w_\ell \geq 0$ ($\ell \in [N]$) are fixed. The points (also known as the nodes of the cubature rule) and weights are typically chosen so that the approximation is exact for polynomials up to a certain degree, say d.

The problem of finding the points $x^{(\ell)} \in K$ and weights w_ℓ ($\ell \in [N]$) giving a cubature rule exact at degree d may then be written as the following GPM:

$$val := \inf_{\mu \in \mathcal{M}(K)_+} \left\{ \int_K 1 d\mu(x) \; : \; \int_K x^\alpha d\mu(x) = \int_K x^\alpha d\mu_0(x) \; \forall \alpha \in \mathbb{N}_d^n \right\}.$$

The key observation is that, by Theorem 3, this problem has an atomic solution supported on at most $N = |\mathbb{N}_d^n| = \binom{n+d}{d}$ points in K, say $\mu^* = \sum_{\ell=1}^{N} w_\ell \delta_{x^{(\ell)}}$, and this yields the cubature weights and points. This result is known as Tchakaloff's theorem [58]; see also [3, 57]. (In fact, our running assumption that K is compact may be relaxed somewhat in Tchakaloff's theorem — see, e.g. [48].)

Here we have chosen the constant polynomial 1 as objective function so that the optimal value is $val = \mu_0(K)$. Other choices of objective functions are possible as

discussed, e.g., in [50]. The GPM formulation of the cubature problem was used for the numerical calculation of cubature schemes for various sets K in [50].

3 Semidefinite programming reformulations of the approximations

The inner and outer approximations of the cone $\mathcal{M}(K)_+$ discussed in Section 1.4 lead to upper and lower bounds for the GPM (1), which may be reformulated as finite-dimensional, convex optimization problems, namely semidefinite programming (SDP) problems. These are conic linear programs over the cone of positive semidefinite matrices, formally defined as follows.

Semidefinite programming (SDP) problem

Assume we are given symmetric matrices $A_0, \ldots, A_m$ (all of the same size) and scalars $b_i \in \mathbb{R}$ ($i \in [m]$). The semidefinite programming problem in standard primal form is then defined as

$$p^* := \inf_{X \geq 0} \{\langle A_0, X\rangle \ : \ \langle A_i, X\rangle = b_i \ \forall i \in [m]\},$$

where $\langle \cdot, \cdot \rangle$ now denotes the trace inner product, i.e., the Euclidean inner product in the space of symmetric matrices, and $X \geq 0$ means that X is a symmetric positive semidefinite matrix (corresponding to the Löwner partial ordering of the symmetric matrices).

The dual semidefinite program reads

$$d^* := \sup_{y \in \mathbb{R}^m} \left\{ \sum_{i=1}^m b_i y_i : A_0 - \sum_{i=1}^m y_i A_i \geq 0 \right\}.$$

Weak duality holds: $p^* \geq d^*$. Moreover, strong duality: $p^* = d^*$ holds, e.g., if the primal problem is bounded and admits a positive definite feasible solution X (or if the dual is bounded and has a feasible solution y for which $A_0 - \sum_i y_i A_i$ is positive definite) (see, e.g., [2, 4]).

Next we recall how one can test whether a polynomial can be written as a sum of squares of polynomials using semidefinite programming. This well known fact plays a key role for reformulating the inner and outer approximations of $\mathcal{M}(K)_+$ using semidefinite programs.

Checking sums of squares with SDP

Given an integer $r \in \mathbb{N}$ let $[x]_r = \{x^\alpha : \alpha \in \mathbb{N}_r^n\}$ consist of all monomials with degree at most r, thus the monomial basis of $\mathbb{R}[x]_r$.

Proposition 1 *For a given n-variate polynomial h, one has $h \in \Sigma_r$, if and only if the following polynomial identity holds:*

$$h(x) = [x]_r^\top M[x]_r \left(= \sum_{\alpha,\beta \in \mathbb{N}_r^n} M_{\alpha,\beta} x^{\alpha+\beta} \right),$$

for some positive semidefinite matrix: $M = \left(M_{\alpha,\beta}\right)_{\alpha,\beta \in \mathbb{N}_r^n} \succeq 0$. The above identity can be equivalently written as

$$h_\gamma = \sum_{\alpha,\beta \in \mathbb{N}_r^n \,:\, \alpha+\beta=\gamma} M_{\alpha,\beta} \quad \forall \gamma \in \mathbb{N}_{2r}^n. \tag{10}$$

SDP upper bounds for GPM via the inner approximations

Recall that the inner approximations of the cone $\mathcal{M}(K)_+$ restrict the measures on K to the subsets $\mathcal{M}_{\mu_0}^r$ in (5), i.e. to those measures μ of the form $d\mu = h \cdot d\mu_0$, where μ_0 is a fixed reference measure with $\text{Supp}(\mu_0) = K$ and $h \in \Sigma_r$ is a sum-of-squares polynomial density.

Replacing the cone $\mathcal{M}(K)_+$ in the GPM (1) by its subcone $\mathcal{M}_{\mu_0}^r$ we obtain the parameter

$$val_{inner}^{(r)} := \inf_{\mu \in \mathcal{M}_{\mu_0}^r} \left\{ \int_K f_0(x) d\mu(x) : \int_K f_i(x) d\mu(x) = b_i \,\forall i \in [m] \right\}, \tag{11}$$

which provides a hierarchy of upper bounds for GPM:

$$val \leq val_{inner}^{(r+1)} \leq val_{inner}^{(r)}.$$

According to the above discussion these parameters can be reformulated as semidefinite programs involving the moments of the reference measure μ_0. Indeed, we may write the variable density function as $h(x) = [x]_r^T M[x]_r$ with $M \succeq 0$ and arrive at the following semidefinite program (in standard primal form).

SDP formulation for the inner approximations based upper bounds

$$val_{inner}^{(r)} = \inf_M \left\{ \langle A_0, M \rangle : \langle A_i, M \rangle = b_i \,\forall i \in [m], \ M = (M_{\alpha,\beta})_{\alpha,\beta \in \mathbb{N}_r^n} \succeq 0 \right\}, \tag{12}$$

where we set

$$A_i = \int_K f_i(x)[x]_r[x]_r^T d\mu_0(x) = \left(\int_K f_i(x) x^{\alpha+\beta} d\mu_0(x) \right)_{\alpha,\beta \in \mathbb{N}_r^n} \quad (0 \le i \le m).$$

Moreover, writing each polynomial f_i in the monomial basis as $f_i = \sum_\gamma f_{i,\gamma} x^\gamma$ one sees that the entries of the matrix A_i depend linearly on the moments of the reference measure μ_0, since $\int_K f_i(x) x^{\alpha+\beta} d\mu_0(x) = \sum_\gamma f_{i,\gamma} m_{\alpha+\beta+\gamma}^{\mu_0}(K)$.

To be able to compute the above SDP one needs the moments of the reference measure μ_0 to be known on the set K. This is a restrictive assumption, since even computing volumes of polytopes is an NP-hard problem. One is therefore restricted to specific choices of μ_0 and K where the moments are known in closed form (or can be derived). In Table 1 we therefore give an overview of some known moments for the Euclidean ball and sphere, the hypercube, and the standard simplex. (See [25] for an easy derivation of the moments on the ball and the sphere.) There we use the Gamma function:

$$\Gamma(k) = (k-1)!, \quad \Gamma\left(k + \frac{1}{2}\right) = \left(k - \frac{1}{2}\right)\left(k - 1 - \frac{1}{2}\right) \cdots \frac{1}{2}\sqrt{\pi} \quad \text{for } k \in \mathbb{N}.$$

K	$m_\alpha^{\mu_0}(K)$
$[0,1]^n$	$\prod_{i=1}^n \frac{1}{\alpha_i+1}$
Δ_n	$\frac{\prod_{i=1}^n \alpha_i!}{(\sum_{i=1}^n \alpha_i+n)!}$
S_n	$\begin{cases} \frac{2\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1+\ldots+\beta_n)} & \text{if } \alpha \in (2\mathbb{N})^n \quad \text{with } \beta_i = \frac{\alpha_i+1}{2} \text{ for } i \in [n] \\ 0 & \text{otherwise} \end{cases}$
B_n	$\begin{cases} \frac{1}{\alpha_1+\ldots+\alpha_n+n} \frac{2\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1+\ldots+\beta_n)} & \text{if } \alpha \in (2\mathbb{N})^n \quad \text{with } \beta_i = \frac{\alpha_i+1}{2} \text{ for } i \in [n] \\ 0 & \text{otherwise} \end{cases}$

Table 1 Examples of known moments for some choices of $K \subseteq \mathbb{R}^n$: $\Delta_n = \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$ is the standard simplex and $B_n = \{x \in \mathbb{R}^n : \|x\| \le 1\}$ is the unit Euclidean ball, in which case μ_0 is the Lebesgue measure, and $S_n = \{x \in \mathbb{R}^n : \|x\| = 1\}$ is the unit Euclidean sphere in which case μ_0 is the (Haar) surface measure on S_n.

If K is an ellipsoid, one may obtain the moments of the Lebesgue measure on K from the moments on the ball by an affine transformation of variables. Also, if K is a polytope, one may obtain the moments of the Lebesgue measure through triangulation of K, and subsequently using the formula for the simplex.

SDP lower bounds for GPM via the outer approximations

Here we assume that K is basic closed semi-algebraic, of the form

$$K = \{x \in \mathbb{R}^n : g_j(x) \geq 0 \ \forall j \in [k]\}, \quad \text{where } g_1, \ldots, g_k \in \mathbb{R}[x].$$

Recall that the dual cone of the truncated quadratic module generated by the polynomials g_j describing the set K provides an outer approximation of $\mathcal{M}(K)_+$; we repeat its definition (7) for convenience:

$$(Q^r(g_1, \ldots, g_k))^* = \left\{ \mu \in \mathcal{M}(K) \ : \ \int_K f d\mu \geq 0 \quad \forall f \in Q^r(g_1, \ldots, g_k) \right\},$$

where the quadratic module $Q^r(g_1, \ldots, g_k)$ was defined in (6).

Replacing the cone $\mathcal{M}(K)_+$ in the GPM (1) by the above outer approximations we obtain the following parameters

$$val_{outer}^{(r)} := \inf_{\mu \in (Q^r(g_1, \ldots, g_k))^*} \left\{ \int_K f_0(x) d\mu(x) : \int_K f_i(x) d\mu(x) = b_i \ \forall i \in [m] \right\}, \tag{13}$$

which provide a hierarchy of lower bounds for the GPM:

$$val_{outer}^{(r)} \leq val_{outer}^{(r+1)} \leq val.$$

Here too these parameters can be reformulated as semidefinite programs. Indeed a signed measure μ lies in the cone $(Q^r(g_1, \ldots, g_k))^*$ precisely when it satisfies the condition

$$\int_K g_j(x) \sigma_j(x) d\mu(x) \geq 0 \quad \forall \ \sigma_j \in \Sigma_{r_j}, \quad \forall j \in \{0, \ldots, k\}, \tag{14}$$

where $r_j = r - \lceil \deg(g_j)/2 \rceil$. Using Proposition 1, we may represent each sum-of-squares σ_j as

$$\sigma_j(x) = [x]_{r_j}^\top M^{(j)} [x]_{r_j}$$

for some matrix $M^{(j)} \succeq 0$ (indexed by $\mathbb{N}_{r_j}^n$). Hence we have

$$\int_K g_j(x) \sigma_j(x) d\mu(x) = \int_K g_j(x) [x]_{r_j}^T M^{(j)} [x]_{r_j} d\mu(x) = \langle B_j^\mu, M^{(j)} \rangle,$$

after setting

$$B_j^\mu = \int_K g_j(x) [x]_{r_j} [x]_{r_j}^T d\mu(x) = \left(\int_K g_j(x) x^{\alpha+\beta} d\mu(x) \right)_{\alpha, \beta \in \mathbb{N}_{r_j}^n}.$$

Hence the condition (14) can be rewritten as requiring, for each $j \in \{0, 1, \ldots, k\}$,

$$\langle B_j^\mu, M^{(j)} \rangle \geq 0 \quad \text{for all postive semidefinite matrices } M^{(j)} \text{ indexed by } \mathbb{N}_{r_j}^n,$$

which in turn is equivalent to $B_j^\mu \succeq 0$ (since the cone of positive semidefinite matrices is self-dual). Summarizing, the condition (14) on the variable measure μ

can be rewritten as

$$B_j^\mu = \left(\int_K g_j(x) x^{\alpha+\beta} d\mu(x) \right)_{\alpha,\beta \in \mathbb{N}_{r_j}^n} \geq 0 \quad \forall j \in \{0, 1, \dots, k\}.$$

Finally, observe that only the moments of μ are playing a role in the above constraints. Therefore we may introduce new variables for these moments, say

$$y_\alpha = \int_K x^\alpha d\mu(x) \quad \forall \alpha \in \mathbb{N}_{2r}^n.$$

Writing the polynomials g_j in the monomial basis as $g_j(x) = \sum_\gamma g_{j,\gamma} x^\gamma$ we arrive at the following SDP reformulation for the parameter $val_{outer}^{(r)}$.

SDP formulation for the outer approximations based lower bounds

With $r_j = r - \lceil \deg(g_j)/2 \rceil$ for $j \in \{0, 1, \dots, k\}$ and d an upper bound on the degrees of f_i for $i \in \{0, 1, \dots, m\}$ we have

$$val_{outer}^{(r)} = \inf_{(y_\alpha)_{\alpha \in \mathbb{N}_{2r}^n}} \left\{ \sum_{\alpha \in \mathbb{N}_d^n} f_{0,\alpha} y_\alpha : \sum_{\alpha \in \mathbb{N}_d^n} f_{i,\alpha} y_\alpha = b_i \quad \forall i \in [m], \right. \tag{15}$$

$$\left. \left(\int_K \sum_\gamma g_{j,\gamma} y_{\alpha+\beta+\gamma} \right)_{\alpha,\beta \in \mathbb{N}_{r_j}^n} \geq 0 \quad \forall j \in \{0, 1, \dots, k\} \right\}. \tag{16}$$

If, in the definition (13) of $val_{outer}^{(r)}$, instead of the truncated quadratic module $Q^r(g_1, \dots, g_k)$ we use the larger quadratic module $Q^r(\prod_{j \in J} g_j : J \subseteq [k])$ generated by the pairwise products of the g_j's, then we obtain a stronger bound on val, which we denote by $\overline{val}_{outer}^{(r)}$. Thus

$$\overline{val}_{outer}^{(r)} = \inf_{\mu \in (Q^r(\prod_{j \in J} g_j : J \subseteq [k]))^*} \left\{ \int_K f_0(x) d\mu(x) : \int_K f_i(x) d\mu(x) = b_i \ (i \in [m]) \right\} \tag{17}$$

and clearly we have

$$val_{outer}^{(r)} \leq \overline{val}_{outer}^{(r)} \leq val.$$

The parameter $\overline{val}_{outer}^{(r)}$ can also be reformulated as a semidefinite program, analogous to the program (15)-(16), which however now involves $2^k + 1$ semidefinite constraints instead of $k + 1$ such constraints in (16) and thus its practical implementation is feasible only for small values of k. On the other hand, as we will see later in Section 5.2, the bounds $\overline{val}_{outer}^{(r)}$ admit a much sharper error analysis than the bounds $val_{outer}^{(r)}$ for the case of polynomial optimization.

4 Convergence results for the inner approximation hierarchy

In the rest of the paper we are interested in the convergence of the respective lower and upper SDP bounds on the optimal value of the GPM, as introduced in the previous section. We will first consider in this section the upper bounds for the GPM arising from the inner approximations, since much more is known about their rate of convergence than for the lower bounds arising from the outer approximations. We deal first with the special case of polynomial optimization and then indicate how some of the results extend to the general GPM.

4.1 The special case of global polynomial optimization

Here we consider a special case of the GPM, namely global optimization of polynomials on compact sets (i.e., problem (8)) and review the main known results about the error analysis of the upper bounds $val_{inner}^{(r)}$. After that in the next section we will explain how to extend this error analysis to the bounds for the general GPM problem.

Thus we now consider the problem

$$val = \min_{x \in K} p(x), \tag{18}$$

asking to find the minimum value of the polynomial $p(x) = \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha x^\alpha$ over a compact set K.

Recall the definition of the inner approximation based upper bound (11), which can be rewritten here as

$$val_{inner}^{(r)} = \min_{h \in \Sigma_r} \left\{ \int_K p(x)h(x)d\mu_0(x) \; : \; \int_K h(x)d\mu_0(x) = 1 \right\},$$

and its SDP reformulation from (12), which now reads

$$val_{inner}^{(r)} = \min \left\{ \langle A_0, M \rangle : \langle A_1, M \rangle = 1, \; M = (M_{\alpha,\beta})_{\alpha,\beta \in \mathbb{N}_r^n} \succeq 0 \right\}, \tag{19}$$

with

$$A_0 = \left(\int_K p(x)x^{\alpha+\beta}d\mu_0(x) \right)_{\alpha,\beta \in \mathbb{N}_r^n}, \quad A_1 = \left(\int_K x^{\alpha+\beta}d\mu_0(x) \right)_{\alpha,\beta \in \mathbb{N}_r^n},$$

where as before μ_0 is a fixed reference measure on K.

A first observation made in [36] is that this semidefinite program (19) can in fact be reformulated as a generalized eigenvalue problem. Indeed, its dual semidefinite program reads

$$\max\{\lambda : A_0 - \lambda A_1 \succeq 0\},$$

whose optimal value gives again the parameter $val^{(r)}_{inner}$ (since strong duality holds). Hence $val^{(r)}_{inner}$ is equal to the smallest generalized eigenvalue of the system

$$A_0 v = \lambda A_1 v, \quad v \neq 0. \tag{20}$$

Thus one may compute $val^{(r)}_{inner}$ without having to solve an SDP problem.

In fact, if instead of the monomial basis $\{x^\alpha : \alpha \in \mathbb{R}^n_{2r}\}$ we use a polynomial basis $\{b_\alpha(x) : \alpha \in \mathbb{N}^n_{2r}\}$ of $\mathbb{R}[x]_{2r}$ that is orthonormal with respect to the reference measure μ_0 (i.e., such that $\int_K b_\alpha b_\beta d\mu_0 = 1$ if $\alpha = \beta$ and 0 otherwise), then in the above semidefinite program (19) we may set $A_1 = I$ to be the identity matrix and

$$A_0 = \left(\int_K p(x) b_\alpha(x) b_\beta(x) d\mu_0(x) \right)_{\alpha,\beta \in \mathbb{N}^n_{2r}}, \tag{21}$$

whose entries now involve the 'generalized' moments $\int_K b_\alpha(x) d\mu_0(x)$ of μ_0. Then the parameter $val^{(r)}_{inner}$ can be computed as the smallest eigenvalue of the matrix A_0:

$$val^{(r)}_{inner} = \lambda_{\min}(A_0) \quad \text{where } A_0 \text{ is as in (21).} \tag{22}$$

This fact was observed in [15] and used there to establish a link with the roots of the orthonormal polynomials, permitting to analyze the quality of the bounds $val^{(r)}_{inner}$ for the case of the hypercube $K = [-1, 1]^n$, see below for details.

In Table 2 we list the known convergence rates of the parameters $val^{(r)}_{inner}$ to the optimal value val of problem (18), i.e., we review the known upper bounds for the sequence $\{val^{(r)}_{inner} - val\}, r = 1, 2, \ldots$

$K \subseteq \mathbb{R}^n$	$\left\lvert val^{(r)}_{inner} - val \right\rvert$	measure μ_0	reference
compact	$o(1)$	positive finite Borel measure	[36]
compact, satisfies interior cone condition	$O\left(\frac{1}{\sqrt{r}}\right)$	Lebesgue measure	[18]
convex body	$O\left(\frac{1}{r}\right)$	Lebesgue measure	[14]
hypercube $[-1, 1]^n$	$\Theta\left(\frac{1}{r^2}\right)$	$\prod_{i=1}^n (1 - x_i^2)^{-1/2} dx_i$	[15]
unit sphere, p homogeneous	$O\left(\frac{1}{r}\right)$	surface measure	[22]

Table 2 Known rates of convergence for the Lasserre hierarchy of upper bounds on val in (18) based on inner approximations.

We will give some details on the proofs of each of the four results listed in Table 2. After that we will mention an interesting connection with approximations based on cubature rules.

Asymptotic convergence

The first result in Table 2 states that $\lim_{r\to\infty} val^{(r)}_{inner} = val$ if K is compact and $\mu_0 \in \mathcal{M}(K)_+$. It is a direct consequence of the following result.

Theorem 4 (Lasserre [36]) *Let $K \subseteq \mathbb{R}^n$ be compact, let μ_0 be a fixed, finite, positive Borel measure with $Supp(\mu_0) = K$. and let f be a continuous function on $\mathbb{R}^n$. Then, f is nonnegative on K if and only if*

$$\int_K g^2 f d\mu_0 \geq 0 \ \forall g \in \mathbb{R}[x].$$

The asymptotic convergence of the bounds $val^{(r)}_{inner}$ to val holds more generally for the minimization of a rational function $p(x)/q(x)$ over K (assuming $q(x) > 0$ for all $x \in K$). Indeed, using the above theorem, we obtain

$$\min_{x\in K} \frac{p(x)}{q(x)} = \sup_{t\in\mathbb{R}} t \ \text{s.t.} \ p(x) \geq tq(x) \ \forall x \in K$$

$$= \sup_{t\in\mathbb{R}} t \ \text{s.t.} \ \int_K p(x)h(x)d\mu_0(x) \geq t \int_K q(x)h(x)d\mu_0(x) \ \forall h \in \Sigma$$

$$= \inf_{h\in\Sigma} \int_K p(x)h(x)d\mu_0(x) \ \text{s.t.} \ \int_K q(x)h(x)d\mu_0(x) = 1.$$

Error analysis when K is compact and satisfies an interior cone condition

The second result in Table 2 fixes the reference measure μ_0 to the Lebesgue measure, and restricts the set K to satisfy a so-called interior cone condition.

Definition 1 (Interior cone condition) A set $K \subseteq \mathbb{R}^n$ satisfies an interior cone condition if there exist an angle $\theta \in (0, \pi/2)$ and a radius $\rho > 0$ such that, for every $x \in K$, a unit vector $\xi(x)$ exists such that

$$\{x + \lambda y : y \in \mathbb{R}^n, \|y\| = 1, y^T\xi(x) \geq \cos\theta, \lambda \in [0, \rho]\} \subseteq K.$$

For example, all full-dimensional convex sets satisfy the interior cone condition for suitable parameters θ and ρ. This assumption is used in [18] to claim that the intersection of any ball with the set K contains a positive fraction of the full ball, a fact used in the error analysis.

The main ingredient of the proof is to approximate the Dirac delta supported on a global minimizer by a Gaussian density of the form

$$G(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-\|x - x^*\|^2}{2\sigma^2}\right), \tag{23}$$

where x^* is a minimizer of p on K, and $\sigma^2 = \Theta(1/r)$. Then we approximate the Gaussian density $G(x)$ by a sum-of-squares polynomial $g_r(x)$ with degree $2r$. For

this we use the fact that the Taylor approximation of the exponential function e^{-t} is a sum of squares (since it is a univariate polynomial nonnegative on $\mathbb{R}$).

Lemma 1 *For any $r \in \mathbb{N}$ the univariate polynomial $\sum_{k=0}^{2r} \frac{(-1)^k}{k!} t^k$ (in the variable $t \in \mathbb{R}$), defined as the Taylor expansion of the function $t \in \mathbb{R} \mapsto e^{-t}$ truncated at degree $2r$, is a sum of squares of polynomials.*

Based on this the polynomial

$$g_r(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \sum_{k=0}^{2r} \frac{(-1)^k}{k!} \left(\frac{-\|x - x^*\|^2}{2\sigma^2} \right)^k$$

is indeed a sum of squares with degree $2r$, which can be used (after scaling) as feasible solution within the definition of the bound $val_{inner}^{(r)}$. We refer to [18] for the details of the analysis.

Error analysis when K is a convex body

The third item in Table 2 assumes that K is now convex, compact and full-dimensional, i.e., a convex body. The key idea is to use the following concentration result for the *Boltzman density* (or *Gibbs measure*).[3]

Theorem 5 (Kalai-Vempala [29]) *If p is a linear polynomial, K is a convex set, $T > 0$ is a fixed 'temperature' parameter, and $val = \min_{x \in K} p(x)$, then we have*

$$\int_{\mathbf{K}} p(x) H(x) dx - val \leq nT,$$

where

$$H(x) = \frac{\exp(-p(x)/T)}{\int_K \exp(-p(x)/T) dx}$$

is the Boltzman probability density supported on K.

The theorem still holds if p is convex, but not necessarily linear [14]. The proof of the third item in Table 2 now proceeds as follows:

1. Construct a sum-of-squares polynomial approximation $h_r(x)$ of the Boltzman density $H(x)$ by again using the fact that the even degree truncated Taylor expansion of e^{-t} is a sum of squares (Lemma 1); namely, consider the polynomial $h_r(x) = \sum_{k=0}^{2r} \frac{(-1)^k}{k!} \left(\frac{-p(x)}{T} \right)^k$ (up to scaling).
2. Use this construction to bound the difference between $val_{inner}^{(r)}$ and the Boltzman bound when choosing $T = O(1/r)$.
3. Use the extension of the Kalai-Vempala result to get the required result for convex polynomials p.

---

[3] This result is of independent interest in the study of *simulated annealing* algorithms.

4. When p is nonconvex, the key ingredient is to reduce to the convex case by constructing a convex (quadratic) polynomial $\hat{p}$ that upper bounds p on K and has the same minimizer on K, as indicated in the next lemma.

Lemma 2 *Assume x^* is a global minimizer of p over K. Then the following polynomial*

$$\hat{p}(x) = p(x^*) + \nabla p(x^*)^T(x - x^*) + C_p\|x - x^*\|^2$$

with $C_p = \max_{x \in K} \|\nabla^2 p(x)\|_2$, is quadratic, convex, and separable. Moreover, it satisfies: $p(x) \le \hat{p}(x)$ for all $x \in K$, and x^* is a global minimizer of $\hat{p}$ over K.

Then, in view of the inequality

$$\int_K \hat{p} h d\mu_0 \ge \int_K p h d\mu_0 \quad \forall h \in \Sigma_r, \tag{24}$$

it follows that the error analysis in the non-convex case follows directly from the error analysis in the convex case. The details of the proof are given in [14].

Error analysis for the hypercube $K = [-1, 1]^n$

The fourth result in Table 2 deals with the hypercube $K = [-1, 1]^n$. A first key idea of the proof is that it suffices to show the $O(1/r^2)$ convergence rate for a univariate quadratic polynomial. This follows from Lemma 2 above (and (24)), which implies that it suffices to analyze the case of a quadratic, separable polynomial. Hence we may further restrict to the case when $K = [-1, 1]$ and p is a quadratic univariate polynomial.

In the univariate case, the key idea is to use the eigenvalue reformulation of the bound $val_{inner}^{(r)}$ from (22). There, we use the polynomial basis $\{b_k : k \in \mathbb{N}\}$ consisting of the Chebyshev polynomials (of the first kind) which are orthonormal with respect to the Chebyshev measure $d\mu_0$ on $K = [-1, 1]$, indeed the measure used in Table 2.

Then one may use a connection to the extremal roots of these orthonormal polynomials. Namely, for the linear polynomial $p(x) = x$, the parameter $val_{inner}^{(r)}$ coincides with the smallest root of the orthonormal polynomial b_{r+1} (with degree $r + 1$); this is a well known property of orthogonal polynomials, which follows from the fact that the matrix A_0 in (21) is tri-diagonal and the 3-terms recurrence for the Chebyshev polynomials (see, e.g., [23, §1.3]). When p is a quadratic polynomial, the matrix A_0 in the eigenvalue problem (22) is now 5-diagonal and 'almost' Toepliz, properties that can be exploited to evaluate its smallest eigenvalue. See [15] for details.

Error analysis for the unit sphere

The last result in Table 2 deals with the minimization of a homogeneous polynomial p over the unit sphere $S_n = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 = 1\}$, in which case Doherty and Wehner [22] show a convergence rate in $O(1/r)$. Their construction for a suitable sum-of-squares polynomial density in Σ_r is in fact closely related to their analysis of the outer approximation based lower bounds $val_{outer}^{(r)}$. Doherty and Wehner [22] indeed show the following stronger result: $val_{inner}^{(r)} - val_{outer}^{(r)} = O(1/r)$, to which we will come back in Section 5.2 below.

Link with positive cubature rules

There is an interesting link between positive cubature formulas and the upper bound

$$val_{inner}^{(r)} = \min_{h \in \Sigma_r} \left\{ \int_K ph d\mu_0 \ : \ \int_K h d\mu_0 = 1 \right\},$$

which was recently pointed out in [40] and is summarized in the next result.

Theorem 6 (Martinez et al. [40])

Let $x^{(1)}, \ldots, x^{(N)} \in K$ and weights $w_1 > 0, \ldots, w_N > 0$ give a positive cubature rule on K for the measure μ_0, that is exact for polynomials of total degree at most $d + 2r$, where $d > 0$ and $r > 0$ are given integers. Let p be a polynomial of degree d.

Then, if h is a polynomial nonnegative on K and of degree at most $2r$, one has

$$\int_K ph d\mu_0 \geq \min_{\ell \in [N]} p(x^{(\ell)}).$$

In particular, the inner approximation bounds therefore satisfy

$$val_{inner}^{(r)} \geq \min_{\ell \in [N]} p(x^{(\ell)}).$$

The proof is an immediate consequence of the definitions, but this result has several interesting implications.

- First of all, one may derive information about the rate of convergence for the scheme $\min_{\ell \in [N]} p(x^{(\ell)})$ from the error bounds in Table 2. For example, if K is a convex body, the implication is that $\min_{\ell \in [N]} p(x^{(\ell)}) - val = O(1/r)$.
- Also, if a positive cubature rule is known for the pair (K, μ_0), and the number of points N meets the Tchakaloff bound $N = \binom{n+2r+d}{2r+d}$, then there is no point in computing the parameter $val_{inner}^{(r)}$. Indeed, as

$$val_{inner}^{(r)} \geq \min_{\ell \in [N]} p(x^{(\ell)}) \geq val,$$

the right-hand-side bound is stronger and can be computed more efficiently. Having said that, positive cubature rules that meet the Tchakaloff bound are only known in special cases, typically in low dimension and degree; see e.g. [7, 57, 9], and the references therein.

- Theorem 6 also shows why the last convergence rate in Table 2 is tight for $K = [-1, 1]^n$. Indeed if we consider the univariate example $p(x) = x$ and the Chebyshev probability measure $d\mu_0(x) = \frac{1}{\pi\sqrt{1-x^2}}dx$ on $K = [-1, 1]$, then a positive cubature scheme is given by

$$x^{(\ell)} = \cos\left(\frac{2\ell - 1}{2N}\pi\right), \ w_\ell = \frac{1}{N} \quad \forall \ell \in [N],$$

and it is exact at degree $2N - 1$. This is known as the Chebyshev-Gauss quadrature, and the points are precisely the roots of the degree N Chebyshev polynomial of the first kind. Thus, with $N = r + 1$, in this case we have

$$val_{inner}^{(r)} \geq \min_{\ell \in [N]} p(x^{(\ell)}) = \min_{\ell \in [N]} \cos\left(\frac{(2\ell - 1)\pi}{2N}\right) = \cos\left(-\pi/(2N)\right) = -1 + \Omega\left(\frac{1}{N^2}\right).$$

This explains that the $\Theta(1/r^2)$ result in Table 2 holds for $p(x) = x$. A different proof of this result is given in [15], where it is shown that for this example one actually has equality $val_{inner}^{(r)} = \cos\left(-\pi/(2N)\right)$.

- Finally, Theorem 6 shows that there is not much gain in using a larger set of densities than Σ_r in the definition of the inner approximations $\mathcal{M}_{\mu_0}^r$ since the statement of the theorem holds for any nonnegative polynomial h on K. For example, for the hypercube $K = [-1, 1]^n$, if we use the larger set of densities $h \in Q^r(\prod_{j\in J}(1 - x_j^2) : J \subseteq [k])$ and the Chebyshev measure as reference measure μ_0 on $[-1, 1]^n$, then we obtain upper bounds with convergence rate in $O(1/r^2)$ [10]. This also follows from the later results in [15] where in addition it is shown that this convergence result is tight for linear polynomials. By the above discussion tightness also follows from Theorem 6.

Upper bounds using grid point sets

Of course one may also obtain upper bounds on val, the minimum value taken by a polynomial p over a compact set K, by evaluating p at any suitably selected set of points in K. This corresponds to restricting the optimization over selected finite atomic measures in the definition of val.

A first basic idea is to select the grid point sets consisting of all rational points in K with denominator r for increasing values of $r \in \mathbb{N}$. For the standard simplex $K = \Delta_n$ and the hypercube $K = [0, 1]^n$ this leads to upper bounds that satisfy:

$$\min_{x\in K, rx\in\mathbb{N}^n} p(x) - \min_{x\in K} p(x) \leq \frac{C_d}{r}\left(\max_{x\in K} p(x) - \min_{x\in K} p(x)\right) \quad \text{for all } r \geq d, \quad (25)$$

where C_d is a constant that depends only on the degree d of p; see [17] for $K = \Delta_n$ and [13] for $K = [0, 1]^n$. A faster regime in $O(1/r^2)$ can be shown when allowing a constant that depends on the polynomial p (see [20] for Δ_n and [12] for $[0, 1]^n$). Note that the number of rational points with denominator r in the simplex Δ_n is $\binom{n+r-1}{r} = O(n^r)$ and thus the computation time for these upper bounds is polynomial in the dimension n for any fixed order r. On the other hand, there are $(r+1)^n = O(r^n)$ such grid points in the hypercube $[0, 1]^n$ and thus the computation time of the upper bounds grows exponentially with the dimension n.

For a general convex body K some constructions are proposed recently in [45] for suitable grid point sets (so-called meshed norming sets) $X_d(\epsilon) \subseteq K$ where $d \in \mathbb{N}$ and $\epsilon > 0$. Namely, whenever p has degree at most d, by minimizing p over $X_d(\epsilon)$ one obtains an upper bound on the minimum of p over K satisfying

$$\min_{x \in X_d(\epsilon)} p(x) - \min_{x \in K} p(x) \leq \epsilon \left(\max_{x \in K} p(x) - \min_{x \in K} p(x) \right),$$

where the computation involves $|X_d(\epsilon)| = O\left(\left(\frac{d}{\sqrt{\epsilon}} \right)^{2n} \right)$ point evaluations, thus exponential in the dimension n for fixed precison ϵ.

In comparison, the computation of the upper bound $val_{outer}^{(r)}$ relies on a semidefinite program involving a matrix of size $\binom{n+r}{r} = O(n^r)$, which is polynomial in the dimension n for any fixed order r.

4.2 The general problem of moments (GPM)

One may extend the results of the last section to the inner approximations for the general GPM (1). In other words, we now consider the upper bounds (11) obtained using the inner approximations of the cone $\mathcal{M}(K)_+$, which we repeat for convenience:

$$val_{inner}^{(r)} = \inf_{h \in \Sigma_r} \left\{ \int_K f_0(x) h(x) d\mu_0(x) \; : \; \int_K f_i(x) h(x) d\mu_0(x) = b_i \quad \forall i \in [m] \right\}.$$

A first observation is that this program may not have a feasible solution, even if the GPM (1) does. For example, two constraints like

$$\int_0^1 x d\mu(x) = 0, \; \int_0^1 d\mu(x) = 1$$

admit the Dirac measure $\mu = \delta_{\{0\}}$ as solution but they do not admit any solution of the form $d\mu = h dx$ with $h \in \Sigma_r$ for any $r \in \mathbb{N}$. Thus any convergence result must relax the equality constraints of the GPM (1) in some way, or involve additional assumptions.

We now indicate how one may use the convergence results of the last section to derive an error analysis for the inner approximations of the GPM when relaxing the equality constraints.

Theorem 7 (De Klerk-Postek-Kuhn [21])

Assume that $f_0, \ldots, f_m$ are polynomials, K is compact and the GPM (1) has an optimal solution. Let $b_0 := val$ denote the optimum value of (1) and for any integer $r \in \mathbb{N}$ define the parameter

$$\Delta^{(r)} := \min_{h \in \Sigma_r} \max_{i \in \{0, 1, \ldots, m\}} \left| \int_K f_i(x) h(x) d\mu_0(x) - b_i \right|.$$

Then the following assertions hold:

(1) $\lim_{r \to \infty} \Delta^{(r)} = 0.$

(2) $\Delta^{(r)} = O\left(\frac{1}{r^{1/4}}\right)$ *if K satisfies an interior cone assumption and μ_0 is the Lebesgue measure;*

(3) $\Delta^{(r)} = O\left(\frac{1}{r^{1/2}}\right)$ *if K is a convex body and μ_0 is the Lebesgue measure;*

(4) $\Delta^{(r)} = O\left(\frac{1}{r}\right)$ *if $K = [-1, 1]^n$ and $d\mu_0(x) = \prod_i (1 - x_i^2)^{-1/2} dx_i.$*

We will derive this from the convergence results for global polynomial optimization in Table 2. By assumption, problem (1) has an optimal solution and by Theorem 3 we may assume it has an atomic optimal solution $\mu^* = \sum_\ell \lambda_\ell \delta_{x_\ell^*}$ with $\lambda_\ell > 0$ and $x_\ell^* \in K$. We now sketch the proof.

1. For each atom x_ℓ^* of the optimal measure μ^* consider the polynomial

$$p_\ell(x) = \sum_{i=0}^m \left(f_i(x) - f_i(x_\ell^*)\right)^2,$$

   whose minimum value over K is equal to 0 (attained at x_ℓ^*).
2. We apply the error analysis of the previous section to the problem of minimizing the polynomial p_ℓ over K. In particular, the asymptotic convergence of the upper bounds implies that for any given $\epsilon > 0$

$$\exists r \in \mathbb{N} \; \exists h_\ell \in \Sigma_r \; \text{ s.t. } \int_K p_\ell(x) h_\ell(x) d\mu_0(x) \leq \epsilon^2, \int_K h_\ell(x) d\mu_0(x) = 1$$

   and, therefore,

$$\int_K (f_i(x) - f_i(x_\ell^*))^2 h_\ell(x) d\mu_0(x) \leq \epsilon^2 \quad \forall i \in \{0, \ldots, m\}. \tag{26}$$

3. Using the Jensen inequality, one obtains

$$\left| \int_K f_i(x) h_\ell(x) d\mu_0(x) - f_i(x_\ell^*) \right| = \left| \int_K (f_i(x) - f_i(x_\ell^*)) h_\ell(x) d\mu_0(x) \right| \leq \epsilon$$

for each $i \in \{0, \ldots, m\}$.

4. We now consider the sum-of-squares density $h := \sum_\ell \lambda_\ell h_\ell \in \Sigma_r$. Then we have $b_i = \int_K f_i(x) d\mu^*(x) = \sum_\ell \lambda_\ell f_i(x_\ell^*)$ for each $i \in \{0, \ldots, m\}$. Moreover, the above argument shows that for any $i \in \{0, \ldots, m\}$

$$\left| \int_K f_i(x) h(x) d\mu_0(x) - b_i \right| = \left| \sum_\ell \lambda_\ell \left(\int_K f_i(x) h_\ell(x) d\mu_0(x) - f_i(x_\ell^*) \right) \right| \le \epsilon \mu^*(K)$$

with $\mu^*(K) = \sum_\ell \lambda_\ell$. This shows that $\Delta^{(r)} \le \epsilon \mu^*(K)$ and thus the desired asymptotic result (1).

5. The additional three claims (2)-(4) follow in the same way using the results in Table 2. For instance, in case (1) when K satisfies an interior cone condition and μ_0 is the Lebesgue measure, we replace the estimate (26) by

$$\left| \int_K (f_i(x) - f_i(x_\ell^*))^2 h_\ell(x) d\mu_0(x) \right| = O\left(\frac{1}{\sqrt{r}} \right),$$

which leads to $\Delta^{(r)} = O\left(\frac{1}{r^{1/4}} \right)$ (since we 'loose a square root' when applying Jensen inequality).

We may also use the relation with positive cubature rules discussed in the previous section (Theorem 6) to obtain the following cubature-based approximations for the GPM (1).

Corollary 2 *Assume the GPM (1) admits an optimal solution and let d denote the maximum degree of the polynomials $f_0, \ldots, f_m$. For any integer $r \in \mathbb{N}$ assume we have a cubature rule for (K, μ_0) that is exact for degree $d + 2r$, consisting of the points $x^{(\ell)} \in K$ and weights $w_\ell > 0$ for $\ell \in [N]$, and define the parameter*

$$\Delta_{cub}^{(r)} := \min_{\nu} \max_{i \in \{0, 1, \ldots, m\}} \left| \int_K f_i(x) d\nu - b_i \right|,$$

where in the outer minimization we minimize over all atomic measures ν whose atoms all belong to the set $\{x^{(\ell)} : \ell \in [N]\}$. Then the following assertions hold:

(1) $\lim_{r \to \infty} \Delta_{cub}^{(r)} = 0$;

(2) $\Delta_{cub}^{(r)} = O\left(\frac{1}{r^{1/4}} \right)$ *if K satisfies an interior cone assumption and μ_0 is the Lebesgue measure;*

(3) $\Delta_{cub}^{(r)} = O\left(\frac{1}{\sqrt{r}} \right)$ *if K is a convex body and μ_0 is the Lebesgue measure;*

1. $\Delta_{cub}^{(r)} = O\left(\frac{1}{r} \right)$ *if $K = [-1, 1]^n$ and $d\mu_0(x) = \prod_i (1 - x_i^2)^{-1/2} dx_i$.*

This result follows from Theorem 7. Indeed, for any polynomial $h \in \Sigma_r$, the polynomials $f_i h$ have degree at most $d + 2r$ so that using the cubature rule we obtain

$$\int_K f_i(x) h(x) d\mu_0(x) = \sum_{\ell=1}^N w_\ell f_i(x^{(\ell)}) h(x^{(\ell)}) = \int_K f_i(x) d\nu(x),$$

where ν is the atomic measure with atoms $x^{(\ell)}$ and weights $\alpha_\ell := w_\ell h(x^{(\ell)})$ for $\ell \in [N]$. Therefore, the parameter $\Delta_{cub}^{(r)}$ in Corollary 2 is upper bounded by the parameter $\Delta^{(r)}$ in Theorem 7. The claims (1)-(4) now follow directly from the corresponding claims in Theorem 7.

Note that, for any fixed $r \in \mathbb{N}$, in order to find the best atomic measure ν in the definition of $\Delta_{cub}^{(r)}$ we need to find the best weights α_ℓ ($\ell \in [N]$) giving the measure $\nu = \sum_{\ell=1}^N \alpha_\ell \delta_{x^{(\ell)}}$. This can be done by solving the following linear program:

$$\Delta_{cub}^{(r)} = \min_{t, \alpha_\ell \in \mathbb{R}} t \ \text{ s.t. } \alpha_\ell \geq 0 \ (\ell \in [N]), \ \left| \sum_{\ell=1}^N \alpha_\ell f_i(x^{(\ell)}) - b_i \right| \leq t \ \forall i \in \{0, 1, \ldots, m\}.$$

(This is similar to an idea used in [50].)

5 Convergence results for the outer approximations

In this last section we consider the convergence of the lower bounds for the GPM (1), that are obtained by using outer approximations for the cone of positive measures. We first mention properties dealing with asymptotic and finite convergence for the general GPM and after that we mention some known results on the error analysis in the special case of polynomial optimization.

Here we assume K is a compact semi-algebraic set, defined as before by

$$K = \{x \in \mathbb{R}^n : g_j(x) \geq 0 \quad \forall j \in [k]\},$$

where $g_1, \ldots, g_k \in \mathbb{R}[x]$. We will consider the following (*Archimedean*) condition:

$$\exists r \in \mathbb{N} \ \exists u \in Q^r(g_1, \ldots, g_k) \ \text{ s.t. the set } \{x \in \mathbb{R}^n : u(x) \geq 0\} \text{ is compact.} \quad (27)$$

This condition clearly implies that K is compact. Moreover, it does not depend on the set K but on the choice of the polynomials used to describe K. Note that it is easy to modify the presentation of K so that the condition (27) holds. Indeed, if we know the radius R of a ball containing K then, by adding to the description of K the (redundant) polynomial constraint $g_{k+1}(x) := R^2 - \sum_{i=1}^n x_i^2 \geq 0$, we can ensure that assumption (27) holds for this enriched presentation of K.

For convenience we recall the definition of the bounds $val_{outer}^{(r)}$ from (13):

$$val_{outer}^{(r)} = \inf_{\mu \in (Q^r(g_1, \ldots, g_k))^*} \left\{ \int_K f_0(x) d\mu(x) : \int_K f_i(x) d\mu(x) = b_i \ \forall i \in [m] \right\},$$

where we refer to (6) and (7) for the definitions of the truncated quadratic module $Q^r(g_1, \ldots, g_k)$ and of its dual cone $(Q^r(g_1, \ldots, g_k))^*$.

We also recall the stronger bounds $\overline{val}_{outer}^{(r)}$, introduced in (17), and obtained by replacing in the definition of $val_{outer}^{(r)}$ the cone $Q^r(g_1, \ldots, g_k)$ by the larger cone

$Q^r(\prod_{j \in J} g_j : J \subseteq [k]))$, so that we have

$$val_{outer}^{(r)} \leq \overline{val}_{outer}^{(r)} \leq val.$$

5.1 Asymptotic and finite convergence

Here we present some results on the asymptotic and finite convergence of the lower bounds on val obtained by considering outer approximations of the cone $\mathcal{M}(K)_+$.

Asymptotic convergence

The parameters $val_{outer}^{(r)}$ form a non-decreasing sequence of lower bounds for the optimal value val of problem (1), which converge to it under assumption (27). This asymptotic convergence result relies on the following representation result of Putinar [47] for positive polynomials.

Theorem 8 (Putinar)
Assume K is compact and assumption (27) holds. Any polynomial f that is strictly positive on K (i.e., $f(x) > 0$ for all $x \in K$) belongs to $Q^r(g_1, \ldots, g_k)$ for some $r \in \mathbb{N}$.

The following result can be found in [33, 34] for the general GPM and in [32] for the case of global polynomial optimization.

Asymptotic convergence for the bounds $val_{outer}^{(r)}$

Theorem 9 *Assume K is compact and assumption (27) holds. Then we have*

$$val^* \leq lim_{r \to \infty} val_{outer}^{(r)} \leq val,$$

with equality: $val^* = lim_{r \to \infty} val_{outer}^{(r)} = val$ if, in addition, there exists $z \in \mathbb{R}^{m+1}$ such that $\sum_{i=0}^m z_i f_i(x) > 0$ for all $x \in K$.

This result follows using Theorem 8. Observe that it suffices to show the inequality: $val^* \leq \sup_r val_{outer}^{(r)}$ (as the rest follows using Corollary 1). For this let $\epsilon > 0$ and let $y \in \mathbb{R}^m$ be feasible for val^*, i.e., $f_0(x) - \sum_{i=1}^m y_i f_i(x) \geq 0$ for all $x \in K$; we will show the inequality $b^T y \leq \sup_r val_{outer}^{(r)} + \epsilon \mu(K)$. Then, letting ϵ tend to 0 gives $b^T y \leq \sup_r val_{outer}^{(r)}$ and thus the desired result: $val^* \leq \sup_r val_{outer}^{(r)} = lim_{r \to \infty} val_{outer}^{(r)}$.

As the polynomial $f_0 + \epsilon - \sum_i y_i f_i$ is strictly positive on K, it belongs to $Q^r(g_1, \ldots, g_k)$ for some $r \in \mathbb{N}$ in view of Theorem 8. Then, for any measure

μ feasible for $val_{outer}^{(r)}$, we have $\int_K (f_0 + \epsilon - \sum_i y_i f_i) d\mu \geq 0$, which implies $b^T y \leq \int_K f_0 d\mu + \epsilon\mu(K)$ and thus the desired inequality:

$$b^T y \leq val_{outer}^{(r)} + \epsilon\mu(K) \leq \sup_r val_{outer}^{(r)} + \epsilon\mu(K).$$

When assuming only K compact (thus not assuming condition (27)), the following representation result of Schmüdgen [51] permits to show the asymptotic convergence of the stronger bounds $\overline{val}_{outer}^{(r)}$ to val (in the same way as Theorem 9 follows from Putinar's theorem).

Theorem 10 (Schmüdgen)

Assume K is compact. Any polynomial f that is strictly positive on K (i.e., $f(x) > 0$ for all $x \in K$) belongs to $\mathcal{Q}^r (\prod_{j\in J} g_j : J \subseteq [k])$ for some $r \in \mathbb{N}$.

Asymptotic convergence for the bounds $\overline{val}_{outer}^{(r)}$

Theorem 11 *Assume K is compact. Then we have*

$$val^* \leq lim_{r\to\infty} \overline{val}_{outer}^{(r)} \leq val,$$

with equality: $val^* = lim_{r\to\infty} \overline{val}_{outer}^{(r)} = val$ if, in addition, there exists $z \in \mathbb{R}^{m+1}$ such that $\sum_{i=0}^m z_i f_i(x) > 0$ for all $x \in K$.

Finite convergence

A remarkable property of the lower bounds $val_{outer}^{(r)}$ is that they often exhibit finite convergence. Indeed, there is an easily checkable criterion, known as the *flatness condition*, that permits to conclude that the bound is exact: $val_{outer}^{(r)} = val$, and to extract an (atomic) optimal solution to the GPM. This is condition (28) below, which permits to claim that a given truncated sequence is indeed the sequence of moments of a positive measure; it goes back to work of Curto and Fialkow ([8], see also [34, 38] for details). To expose it we use the SDP formulation (15)-(16) for the parameter $val_{outer}^{(r)}$.

Finite convergence

Theorem 12 *(see [34, Theorem 4.1]) Set $d_K := \max\{\lceil \deg(g_j/2) : j \in [k]\}$ and let $r \in \mathbb{N}$ such that $2r \geq \max\{\deg(f_i) : i \in \{0, \ldots, m\}\}$ and $r \geq d_K$. Assume*

*the program (15)-(16) defining the parameter $val_{outer}^{(r)}$ has an optimal solution
$y = (y_\alpha)_{\alpha \in \mathbb{N}_{2r}^n}$ that satisfies the following (flatness) condition:*

$$rankM_s(y) = rankM_{s-d_K}(y) \quad \text{for some integer } s \text{ s.t. } d_K \le s \le r, \qquad (28)$$

where

$$M_s(y) = (y_{\alpha+\beta})_{\alpha,\beta \in \mathbb{N}_s^n} \quad \text{and} \quad M_{s-d_K}(y) = (y_{\alpha+\beta})_{\alpha,\beta \in \mathbb{N}_{s-d_K}^n}.$$

*Then equality $val_{outer}^{(r)} = val$ holds and the GPM problem (1) has an optimal
solution $\mu \in \mathcal{M}(K)_+$ which is atomic and supported on $rankM_s(y)$ points in K.*

Under the flatness condition (28) there is an algorithmic procedure to find the
atoms and weights of the optimal atomic measure (see, e.g., [34, 38] for details).

In addition, for the special case of the polynomial optimization problem (8), Nie
[43] shows that the flatness condition is a generic property, so that finite conver-
gence of the lower bounds $val_{outer}^{(r)}$ to the minimum of a polynomial over K holds
generically.

Note that analogous results also hold for the stronger bounds $\overline{val}_{outer}^{(r)}$ on val.

5.2 Error analysis for the case of polynomial optimization

We now consider the special case of global polynomial optimization, i.e., problem
(8), which is the case of GPM with only one affine constraint, requiring that μ is a
probability measure on K:

$$val = \min_{x \in K} p(x) = \min_{\mu \in \mathcal{M}(K)_+} \int_K p(x)d\mu(x) \text{ s.t. } \int_K d\mu(x) = 1.$$

Recall the definition of the bound $val_{outer}^{(r)}$ from (13), which now reads

$$val_{outer}^{(r)} = \inf_{\mu \in (Q^r(g_1,...,g_k))^*} \left\{ \int_K p(x)d\mu(x) : \int_K d\mu(x) = 1 \right\}.$$

It can be reformulated via an SDP as in (15)-(16), whose dual SDP reads

$$\sup_{\lambda \in \mathbb{R}} \{\lambda : p - \lambda \in Q^r(g_1, \ldots, g_k)\}. \qquad (29)$$

By weak duality $val_{outer}^{(r)}$ is at least the optimum value of (29). Strong duality holds
for instance if the set K has a non-empty interior (since then the primal SDP is
strictly feasible), or if there is a ball constraint present in the description of the set
K (as shown in [28]). Then, $val_{outer}^{(r)}$ is also given by the program (29), which is the
case, e.g., when K is a simplex, a hypercube, or a sphere.

As we saw above, the bounds $val_{outer}^{(r)}$ converge asymptotically to the minimum value val taken by the polynomial p over the set K when condition (27) holds. We now indicate some known results on the rate of convergence of these bounds.

For a polynomial $p = \sum_\alpha p_\alpha x^\alpha \in \mathbb{R}[x]_d$, we set

$$L_p := \max_\alpha |p_\alpha| \frac{\alpha_1! \cdots \alpha_n!}{|\alpha|!}.$$

Error analysis for the bounds $val_{outer}^{(r)}$

Theorem 13 *[44]Assume $K \subseteq (-1, 1)^n$. There exists a constant $c > 0$ (depending only on K) such that, for any polynomial p with degree d, we have*

$$val - val_{outer}^{(r)} \leq 6d^3 n^{2d} L_p \frac{1}{\left(\log \frac{r}{c}\right)^{1/c}} \quad \text{for all integers} \quad r \geq c \exp\left((2d^2 n^d)^c\right).$$

Note that this result displays a very slow convergence rate, which does not reflect the good behaviour of the bounds often observed in practice.

On the other hand, a sharper error analysis holds for the stronger bounds $\overline{val}_{outer}^{(r)}$, obtained by using the larger set $Q^r \left(\prod_{j \in J} g_j : J \subseteq [k]\right)$ instead of $Q^r(g_1, \ldots, g_k)$.

Error analysis for the bounds $\overline{val}_{outer}^{(r)}$

Theorem 14 *[53]Assume $K \subseteq (-1, 1)^n$. There exists a constant $c > 0$ (depending only on K) such that, for any polynomial p with degree d, we have*

$$val - \overline{val}_{outer}^{(r)} \leq cd^4 n^{2d} L_p \frac{1}{r^{1/c}} \quad \text{for all integers} \quad r \geq cd^c n^{cd}.$$

We now recap some known sharper results for the case of polynomial optimization over special sets K like the simplex, the hypercube and the sphere. As motivation recall that this already captures well known hard combinatorial optimization problems such as the maximum independence number in a graph.

Given a graph $G = (V = [n], E)$ let $\alpha(G)$ denote the largest cardinality of an independent set in G, i.e., of a set $I \subseteq V$ that does not contain any edge of E. In fact the parameter $\alpha(G)$ can be reformulated via polynomial optimization over the simplex Δ_n, the hypercube $[0, 1]^n$, or the unit sphere S_n. Indeed the following results are known:

$$\frac{1}{\alpha(G)} = \min_{x \in \Delta_n} x^T (I_n + A_G)x, \quad \alpha(G) = \max_{x \in [0,1]^n} \sum_{i \in V} x_i - \sum_{\{i,j\} \in E} x_i x_j,$$

$$\frac{2\sqrt{2}}{3\sqrt{3}} \sqrt{1 - \frac{1}{\alpha(G)}} = \max_{y \in \mathbb{R}^n, z \in \mathbb{R}^m} \left\{ 2 \sum_{\{i,j\} \in \overline{E}} z_{ij} y_i y_j : (y, z) \in S_{n+m} \right\}$$

(see [41, 42]). Here I_n is the identity matrix of size n, A_G is the adjacency matrix of G (with entries $A_{ij} = A_{ji} = 1$ if $\{i, j\} \in E$ and 0 otherwise), $\overline{E}$ is the set of pairs of distinct elements $i, j \in V$ such that $\{i, j\} \notin E$ and $m = |\overline{E}|$.

Error analysis for the sphere

We first consider the case of the sphere $K = S_n = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 = 1\}$. Then an error analysis for the bounds $val_{outer}^{(r)}$ is known when p is a homogeneous polynomial.

First, one may reduce to the case when p has even degree. Indeed, as shown in [22], if p has odd degree d then we have

$$\max \left\{ p(x) : \sum_{i=1}^n x_i^2 = 1 \right\} = \frac{d^{d/2}}{(d+1)^{(d+1)/2}} \max \left\{ x_{n+1} p(x) : \sum_{i=1}^{n+1} x_i^2 = 1 \right\}.$$

Another useful observation is that, for a homogeneous polynomial q of even degree d, q belongs to the truncated quadratic module of the sphere:

$$Q^r \left(\pm \left(1 - \sum_{i=1}^n x_i^2 \right) \right) = \Sigma_r + \left(1 - \sum_{i=1}^n x_i^2 \right) \mathbb{R}[x]$$

if and only if the polynomial $q(x) \left(\sum_{i=1}^n x_i^2 \right)^r$ is a sum of squares of polynomials (see [16]). Therefore, when p is a homogeneous polynomial of even degree $d = 2a$, the parameter $val_{outer}^{(r)}$ can be reformulated as

$$val_{outer}^{(r)} = \min \left\{ t : t \in \mathbb{R}, \ t \left(\sum_{i=1}^n x_i^2 \right)^r - \left(\sum_{i=1}^n x_i^2 \right)^{r-a} p(x) \in \Sigma_r \right\}. \quad (30)$$

Based on this, the following error bounds for the parameters $val_{outer}^{(r)}$ are shown in [24, 22] (for general polynomials) and in [17] (for even polynomials).

Theorem 15 *Let p be a homogeneous polynomial of even degree d.*

(i) ([24, 22]) There exist constants $C_{n,d}$ and $r_{n,d}$ (depending on n and d) such that

$$\min_{x \in S_n} p(x) - val_{outer}^{(r)} \leq \frac{C_{n,d}}{r} \quad \text{for all integers } r \geq r_{n,d}.$$

(ii) ([17]) If p is an even polynomial (i.e., of the form $p = \sum_{\alpha \in \mathbb{N}^n_{d/2}} p_\alpha x^{2\alpha}$), then the above holds where the constant $C_{n,d}$ depends only on d and $r_{n,d} = d$.

We briefly discuss the approach in [22], which in fact provides an error analysis for the larger range $val^{(r)}_{inner} - val^{(r)}_{outer}$.

For an integer a let $\mathrm{MSym}((\mathbb{R}^n)^{\otimes a})$ denote the set of matrices acting on $(\mathbb{R}^n)^{\otimes a}$ that are maximally symmetric, which means the associated $2a$-tensor is fully symmetric (i.e., invariant under the action of the symmetric group $\mathrm{Sym}(2a)$). Any homogeneous polynomial p of degree $2a$ can be written as $p(x) = (x^{\otimes a})^T Z_p x^{\otimes a}$ for a (unique) $Z_p \in \mathrm{MSym}((\mathbb{R}^n)^{\otimes a})$. Then, defining the polynomial $p_r(x) = (\sum_i x_i^2)^{r-a} p(x)$, the program (30) can be reformulated as

$$val^{(r)}_{outer} = \min \left\{ \langle Z_{p_r}, M \rangle : M \succeq 0, \mathrm{Tr}(M) = 1, \ M \in \mathrm{MSym}((\mathbb{R}^n)^{\otimes r}) \right\}.$$

Let M be an optimal solution to this program. As $M \succeq 0$ the polynomial $(x^{\otimes r})^T M x^{\otimes r}$ is a sum of squares. One can scale it to obtain $h \in \Sigma_r$ which provides a probability density function on S_n, i.e., $\int_{S_n} h(x) d\mu_0(x) = 1$ (with μ_0 the surface measure on S_n), and thus $val^{(r)}_{inner} \leq \int_{S_n} h(x) d\mu_0$. Using the orthogonal polynomial basis with respect to μ_0 (consisting of spherical harmonic polynomials), Doherty and Wehner [22] show a de Finetti type result, which permits to upper bound the range $\int_{S_n} h(x) d\mu_0 - \langle Z_{p_r}, M \rangle$ and thus $val^{(r)}_{inner} - val^{(r)}_{outer}$.

Error analysis for the simplex and the hypercube

For the simplex $K = \Delta_n = \{x \in \mathbb{R}^n : x_i \geq 0 \ (i \in [n]), 1 - \sum_{i=1}^n x_i = 0\}$ and the hypercube $K = [0,1]^n = \{x \in \mathbb{R}^n : x_i \geq 0, 1 - x_i \geq 0 \ (i \in [n])\}$, a refined error analysis is known only for the stronger bounds $\overline{val}^{(r)}_{outer}$, where we use the larger quadratic module generated by all pairwise products of the constraints defining K.

Error analysis for the simplex

Theorem 16 *[17] Assume $K = \Delta_n$ and p is a homogeneous polynomial with degree d. Then we have*

$$\min_{x \in \Delta_n} p(x) - \overline{val}^{(r)}_{outer} \leq \frac{C_d}{r} \left(\max_{x \in \Delta_n} p(x) - \min_{x \in \Delta_n} p(x) \right) \quad \text{for all } r \geq d,$$

where $C_d > 0$ is an absolute constant depending only on d.

Error analysis for the hypercube

Theorem 17 *[13] Assume $K = [0,1]^n$. For any polynomial p with degree d we have*

$$\min_{x \in [0,1]^n} p(x) - \overline{val}_{outer}^{(r)} \leq n^d \binom{d+1}{3} L_p \frac{1}{r} \quad \textit{for all } r \geq d.$$

The above results show that in Theorem 14 one may choose the unknown constant to be $c = 1$ (roughly) if K is a hypercube or simplex. In both cases the proof relies on showing this error analysis for a weaker bound, which is obtained by using only nonnegative scalar multipliers (instead of sum-of-squares multipliers) in the definition of the quadratic module. See [17, 13] for details.

6 Concluding remarks

We conclude with a few remarks on available software and future research directions.

Software

The bounds based on the outer approximations (13) described here have been implemented in the software *Gloptipoly3* [27]. The software can in fact deal with a more general version of the GPM (1) than presented here. Namely it can deal with the problem

$$val = \inf_{\mu_i \in \mathcal{M}(K_i)_+ \ \forall i \in \{0\} \cup [m]} \left\{ \int_{K_0} f_0(x) d\mu_0(x) \ : \ \int_{K_i} f_i(x) d\mu_i(x) = b_i \quad \forall i \in [m] \right\},$$

where we have a variable measure $\mu_i \in \mathcal{M}(K_i)_+$ for each index $i \in \{0, \ldots, m\}$, with $K_i \subseteq \mathbb{R}^n$ being basic closed semi-algebraic sets defined by (possibly different) sets of polynomial inequalities.

Due to the sizes of the resulting semidefinite programs that are solved, applicability is typically limited to $n \leq 20$ variables and low order, say $r \leq 4$. This is due to the fact the matrix variables in the semidefinite programs are roughly of order $\binom{n+r}{r}$. Solving larger instances requires exploiting additional structure (like sparsity) leading to more economical semidefinite programs. We refer, e.g., to [34] and references therein for further details.

Error bounds for the inner approximation hierarchy

The known error bounds for the inner approximation, presented earlier in Table 2, are for specific choices of the set K and reference measure $\mu_0 \in \mathcal{M}(K)_+$. More work is

required to understand the role of the reference measure in the convergence analysis, and to extend the regime in $O(1/r^2)$ to more classes of sets K. In particular, an obvious choice is whether one can sharpen the analysis of the the convergence rate for the Euclidean unit sphere. As explained, such results would also have implications for grid search on cubature points on the sphere. Cubature on the sphere is a vast research topic (see, e.g., [9, Chapter 6]), even in the special case of spherical t-designs [9, §6.5], where all cubature weights are equal and positive. Moreover, the complexity of polynomial optimization on spheres is not fully understood; indeed the problem is NP-hard, but allows polynomial-time approximation schemes in special cases (see [17, 11]). Sharpening the analysis of the inner approximations for polynomial optimization over spheres may help to gain a more complete understanding.

Error bounds for the outer approximation hierarchy

The bounds based on the outer approximation presented here are more practically suited for computation, in particular since they (sometimes) enjoy finite convergence and permit to extract the global minimizers; moreover, as mentioned above, the dedicated software *Gloptipoly3* is available for this purpose. On the other hand, the known results on the rate of convergence are somewhat disappointing (as discussed in Section 5.2), and in general much weaker than those known for the inner approximation. There is certainly room for a breakthrough here; new ideas are needed to obtain convergence rates that match the performance observed in practice.

Acknowledgement

References

1. Akhiezer, N.I. *The classical moment problem.* Hafner, New York (1965)
2. Barvinok, A. *A course in Convexity.* Graduate Study in Mathematics, Volume 54, AMS, Providence, Rhode Island (2002)
3. Bayer, C., and Teichmann, J. The proof of Tchakaloff's theorem. *Proceedings of the American Mathematical Society*, 134:3035–3040 (2006)
4. Ben Tal, A., and Nemirovski, A. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* MPS/SIAM Series on Optimization, 2, SIAM (2001)
5. Bhattiprolu, V., Ghosh, M., Guruswami, V., Lee, E., Tulsiani, M. Weak decoupling, polynomial folds and approximate optimization over the sphere. arXiv:1611.05998v2 (2017)
6. Castella, M. Rational optimization for nonlinear reconstruction with approximate ℓ_0 penalization. Preprint version available at arXiv:1808.00724 (2018)
7. Cools, R. An encyclopaedia of cubature formulas, *J. Complexity*, 19, 445–453 (2003).
8. Curto, R.E., and Fialkow, L.A. Solution of the truncated complex moment problem for flat data, *Memoirs of the American Mathematical Society* 119 (568) (1996)

9. Dai, F., and Xu, Y. *Approximation Theory and Harmonic Analysis on Spheres and Balls*, Springer, New York (2013)

10. De Klerk, E., Hess, R., and Laurent, M. Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization. *SIAM Journal on Optimization* 27(1), 347–367 (2017)

11. De Klerk, E. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research*, 16(2), 111–125 (2008)

12. De Klerk, E., Lasserre, J.B., Laurent, M., and Sun, S. Bound-constrained polynomial optimization using only elementary calculations. *Mathematics of Operations Research,* 42(3), 834–853 (2017)

13. De Klerk, E., Laurent, M. Error bounds for some semidefinite programming approaches to polynomial minimization on the hypercube. *SIAM Journal on Optimization* 20(6), 3104–3120 (2010)

14. De Klerk, E., Laurent, M. Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing. *Mathematics of Operations Research*, to appear. Preprint version available at http://arxiv.org/abs/1703.00744 (2017)

15. De Klerk, E., Laurent, M. Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube. *Mathematics of Operations Research*, to appear. Preprint version available at http://arxiv.org/abs/1804.05524 (2018)

16. De Klerk, E., Laurent, M., and Parrilo, P. On the equivalence of algebraic approaches to the miniization of forms on the simplex. In D. Henrion and A. Garulli (eds), *Positive Polynomials in Control*, 121–133, Springer (2005)

17. De Klerk, E., Laurent, M., and Parrilo, P. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science,* 361(2-3), 210–225 (2006)

18. De Klerk, E., Laurent, M., Sun, Z. Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization. *Mathematical Programming Series A* 162(1), 363–392 (2017)

19. De Klerk, E., Laurent, M., Sun, Z. An error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution. *SIAM Journal on Optimization,* 25(3), 1498–1514 (2015)

20. De Klerk, E., Laurent, M., Sun, Z., and Vera, J. On the convergence rate of grid search for polynomial optimization over the simplex. *Optimization Letters,* 11(3), 597–608 (2017)

21. De Klerk, E., Postek, K., and Kuhn, D. Distributionally robust optimization with polynomial densities: theory, models and algorithms. Preprint version available at arXiv:1805.03588 (2018)

22. Doherty, A.C., Wehner, S. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. arXiv:1210.5048v2 (2013)

23. Dunkl, C.F., and Xu., Y. *Orthogonal Polynomials of Several Variables,* Cambridge University Press (2001)

24. Faybusovich, L. Global optimization of homogeneous polynomials on the simplex and on the sphere. In C. Floudas and P. Pardalos (eds), *Frontiers in Global Optimization*, Kluwer (2003)

25. Folland, G.B. How to integrate a polynomial over a sphere? *The American Mathematical Monthly*, 108(5), 446–448 (2001)

26. Handelman, D. Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific J. Math.*, 132(1), 35–62 (1988)

27. Henrion, D., Lasserre, J.B., and Loefberg, J. GloptiPoly 3: moments, optimization and semidefinite programming. *Optimization Methods and Software*, 24(4-5), 761–779 (2009). Software download: www.laas.fr/~henrion/software/gloptipoly3

28. Josz, C., Henrion, D. Strong duality in Lasserre's hierarchy for polynomial optimization. *Optim. Letters*, 10, 3–10 (2016)

29. Kalai. A. T., and Vempala, S. Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2), 253–266 (2006)

30. Kemperman, J.H.B. The general moment problem, a geometric approach. *The Annals of Mathematics Statistics*, 39, 93–122 (1968)

31. Landau, H. *Moments in Mathematics*, *Proc. Sympos. Appl. Math.*, 37 (1987)

32. Lasserre, J.B. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11, 796–817 (2001)
33. Lasserre, J.B. A semidefinite programming approach to the generalized problem of moments. *Mathematical Programming Series B* 112, 65–92 (2008)
34. Lasserre, J.B. *Moments, Positive Polynomials and Their Applications.* Imperial College Press (2009)
35. Lasserre, J.B. *Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press (2015)
36. Lasserre, J.B. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization* 21(3), 864–885 (2011)
37. Lasserre, J.B. The moment-SOS hierarchy. *Proc. Int. Cong. of Math. ? 2018*, Rio de Janeiro, 3, 3761–3784 (2018)
38. Laurent, M. Sums of squares, moment matrices and optimization over polynomials. In *Emerging Applications of Algebraic Geometry,* Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, 157-270 (2009)
39. Laurent, M. Optimization over polynomials: Selected topics. In Chapter 16 (Control Theory and Optimization) of *Proc. Int. Cong. of Math. 2014*. Jang, S. Y., Kim, Y. R., Lee, D-W. & Yie, I. (eds.). Seoul: Kyung Moon SA Co. Ltd., p. 843–869 (2014)
40. Martinez, A., Piazzon, F., Sommariva, A., and Vianello, M. Quadrature-based polynomial optimization. Manuscript (2018). http://www.math.unipd.it/ marcov/pdf/quadropt.pdf
41. Motzkin, T.S., Sraus, E.G. Maxima for graphs and a new proof of a theorem of Túran. *Canadian J. Math.*, 17, 533–540 (1965)
42. Nesterov, Yu. Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper 2003/71, CORE-UCL, Louvain-La-Neuve (2003)
43. Nie, J. Optimality conditions and finite convergence of Lasserre's hierarchy. *Mathematical Programming, Ser. A,*146(1-2), 97–121 (2014)
44. Nie, J., and Schweighofer, M. On the complexity of Putinar's positivstellensatz *Journal of Complexity* 23, 135–150 (2007)
45. Piazzon, F., Vianello, M. Markov inequalities, Dubiner distance, norming meshes and polynomial optimization on convex bodies. Preprint at Optimization Online (2018)
46. Powers, V., Reznick, B. A new bound for Pólya's theorem with applications to polynomials positive on polyhedra. *J. Pure Appl. Algebra*, 164, 221–229 (2001)
47. Putinar, M. Positive polynomials on compact semi-algebraic sets. *Ind. Univ. Math. J.* 42, 969–984 (1993)
48. Putinar, M. A note on Tchakaloff's theorem. Proceedings of the American Mathematical Society, 125(8), 2409–2414 (1997)
49. Rogosinski, W.W. Moments of non-negative mass, *Proceedings of the Royal Society A* 245, 1–27 (1958)
50. Ryu, E.K. and Boyd, S.P. Extensions of Gauss Quadrature Via Linear Programming. *Foundations of Computational Mathematics* 15(4), 953–971 (2015)
51. Schmüdgen, K. The *K*-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289, 203–206 (1991)
52. Schmüdgen, K. *The moment problem*. Springer (2017)
53. Schweighofer, M. On the complexity of Schmüdgen's Positivstellensatz *Journal of Complexity* 20(4), 529-543 (2004)
54. Schwartz, R.E. The 5 electron case of Thomson's problem. *Exp. Math* 22(2), 157–186 (2013)
55. Shapiro, A. *On duality theory of conic linear problems*, Semi-Infinite Programming: Recent Advances (M,Á. Goberna and M.A. López, eds.), Springer, 135–165 (2001)
56. Tao, T. An Epsilon of Room, I: Real Analysis: pages from year three of a mathematical blog. AMS, Graduate Studies in Mathematics Volume: 117 (2010)
57. Trefethen, L.N. Cubature, Approximation, and Isotropy in the Hypercube. *SIAM Review*, 59(3), 469–491 (2017)
58. Tchakaloff, V. Formules de cubature ḿecanique c̀oefficients non négatifs, *Bull. Sci. Math.*, 81, 123–134 (1957)