ORIGINAL PAPER

Convergence rates of RLT and Lasserre-type hierarchies for the generalized moment problem over the simplex and the sphere

Felix Kirschner[1] · **Etienne de Klerk**[1]

Abstract

We consider the generalized moment problem (GMP) over the simplex and the sphere. This is a rich setting and it contains NP-hard problems as special cases, like constructing optimal cubature schemes and rational optimization. Using the reformulation-linearization technique (RLT) and Lasserre-type hierarchies, relaxations of the problem are introduced and analyzed. For our analysis we assume throughout the existence of a dual optimal solution as well as strong duality. For the GMP over the simplex we prove a convergence rate of $O(1/r)$ for a linear programming, RLT-type hierarchy, where r is the level of the hierarchy, using a quantitative version of Pólya's Positivstellensatz. As an extension of a recent result by Fang and Fawzi (Math Program, 2020. https://doi.org/10.1007/s10107-020-01537-7) we prove the Lasserre hierarchy of the GMP (Lasserre in Math Program 112(1):65–92, 2008. https://doi.org/10.1007/s10107-006-0085-1) over the sphere has a convergence rate of $O(1/r^2)$. Moreover, we show the introduced linear RLT-relaxation is a generalization of a hierarchy for minimizing forms of degree d over the simplex, introduced by De Klerk et al. (J Theor Comput Sci 361(2–3):210–225, 2006).

✉ Felix Kirschner
  f.c.kirschner@tilburguniversity.edu

  Etienne de Klerk
  e.deklerk@tilburguniversity.edu

[1] Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

1 Introduction

For a compact set $K \subset \mathbb{R}^n$ let $\mathcal{M}(K)$ denote the (infinite-dimensional) vector space of signed finite Borel measures with support contained in K. Let $[m] = \{1, \ldots, m\}$ for $m \in \mathbb{N}$. The generalized moment problem (GMP) is an optimization problem of the following form:

$$\text{val} := \inf_{\mu \in \mathcal{M}(K)_+} \int_K f_0(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$$
$$\text{s.t.} \int_K f_i(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) = b_i \ \forall i \in [m]$$
$$\int_K \mathrm{d}\mu(\mathbf{x}) \leq 1, \tag{1}$$

where $m \in \mathbb{N}$, $b_i \in \mathbb{R}$ for all $i \in [m]$, $\mathcal{M}(K)_+$ is the convex cone of positive finite Borel measures supported on K, and $f_0, f_1, \ldots, f_m$ are continuous on K. We will always assume the GMP (1) has a feasible solution, which implies that it has an optimal solution as well (see Theorem 1).

The constraint $\int_K \mathrm{d}\mu(\mathbf{x}) \leq 1$ essentially means that we know an upper bound on the measure of K for the optimal solution, since, in this case, we may scale the functions f_i a priori to satisfy this condition.

The GMP is a conic linear optimization problem whose duality theory is well understood, see e.g. [18]. A wide range of optimization problems can be modeled as an instance of the GMP. The list includes problems from optimization, probability, financial economics and optimal control to name only a few, see e.g. [11].

For *polynomial data*, i.e. when all the f_i's are polynomials ($i \in \{0, 1, \ldots, m\}$), and the set K is a basic closed semialgebraic set, Lasserre [10] introduced a monotone nondecreasing hierarchy of semidefinite programming (SDP) relaxations of (1). For a survey on SDP approaches to the GMP with polynomial data and their error analysis, we refer to the survey of De Klerk and Laurent [4].

In this paper, we will consider the case where K is the standard (probability) simplex

$$\Delta_{n-1} = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : x_1 + \cdots + x_n = 1 \right\},$$

where $\mathbb{R}^n_+$ is the nonnegative orthant, or the Euclidean sphere

$$\mathcal{S}^{n-1} = \left\{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2^2 = x_1^2 + \cdots + x_n^2 = 1 \right\}.$$

Our main result is to establish a rate of convergence for the Lasserre hierarchy [10] for the GPM with polynomial data on the sphere, and for a related, RLT (reformulation-linearization technique)-type linear programming hierarchy for the GPM with polynomial data on the simplex. This RLT hierarchy is in fact a generalisation of LP hierarchies for polynomial optimization on the simplex, as introduced by Bomze and De Klerk [2], and De Klerk et al. [5], and is closely related to the original work on RLT hierarchies by Sherali and Adams [19].

Outline of the paper

First we introduce some notation in Sect. 2.1. In Sect. 2.2 we review the duality theory of the GMP. A brief overview of possible applications of our setting is given in Sect. 2.3. For K the simplex we introduce a linear relaxation hierarchy in this setting in Sect. 3 and prove a convergence rate of $O(1/r)$. Section 4 contains the new convergence analysis of the Lasserre [11] SDP hierarchies of the GPM on the sphere. In Sect. 5 we take a mathematical view of how the optimal measure is obtained in the limit as the level of the hierarchies approaches infinity. In Sect. 6 we explain how our LP hierarchy is a generalization of an approximation hierarchy for the problem of minimizing a form of degree d over the simplex introduced by De Klerk et al. [5] based on earlier results obtained by Bomze and De Klerk [2].

2 Preliminaries

2.1 Notation

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ denote the set of nonnegative integers, $\mathbb{N}_+ = \mathbb{N}\backslash\{0\}$ and $\mathbb{N}_t^n$ the set of sequences $\alpha \in \mathbb{N}^n$ for which $|\alpha| = \sum_{i=1}^n \alpha_i \leq t$ for $t \in \mathbb{N}$. For $\alpha \in \mathbb{N}^n$, $\mathbf{x}^\alpha$ denotes the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and its degree is $|\alpha|$. The ring of multivariate polynomials in n variables $\mathbf{x} = (x_1, \ldots, x_n)$ is denoted by $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$ and $\mathbb{R}[\mathbf{x}]_t$ is its subspace of polynomials of degree at most t. The (total) degree of a polynomial is the maximal degree of its appearing monomials. A monomial basis vector of order t is given by

$$[\mathbf{x}]_t = (1, x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_{n-1} x_n, x_n^2, \ldots, x_1^t, \ldots, x_n^t)^T.$$

Any polynomial $p \in \mathbb{R}[\mathbf{x}]$ can be written as $p = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha$, where only finitely many p_α are non-zero. A polynomial $p \in \mathbb{R}[\mathbf{x}]$ is a sum of squares (sos) if $p = \sum_{j=1}^k (h_j)^2$ for $h_j \in \mathbb{R}[\mathbf{x}]$ and $k \geq 1$. The set of sos polynomials is denoted by $\Sigma[\mathbf{x}]$ and the set of sos polynomials of degree at most t is denoted by $\Sigma[\mathbf{x}]_t$.

2.2 Duality of the generalized problem of moments

We shall briefly discuss the duality theory associated with the GMP (1). To this end, let $\mathcal{C}(K)$ denote the space of bounded continuous functions on K endowed with the supremum norm $\|\cdot\|_\infty$. For two vector spaces E, F of arbitrary dimension, a non-degenerate bilinear form $\langle\rangle : E \times F \to \mathbb{R}$ is called a duality of E and F. The spaces $\mathcal{M}(K)$ and $\mathcal{C}(K)$ can be put in duality by defining $\langle\rangle : \mathcal{C}(K) \times \mathcal{M}(K) \to \mathbb{R}$ as

$$\langle f, \mu \rangle = \int_K f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}). \tag{2}$$

Let $f_0, f_1, \ldots, f_m$ be continuous functions on K and $b_1, \ldots, b_m \in \mathbb{R}$. The dual of (1) is given by

$$
\text{val}' = \sup_{(y,t) \in \mathbb{R}^m \times \mathbb{R}_+} \sum_{i=1}^m y_i b_i - t
$$

$$
\text{s.t. } f_0(\mathbf{x}) - \sum_{i=1}^m y_i f_i(\mathbf{x}) + t \geq 0 \quad \forall \, \mathbf{x} \in K. \tag{3}
$$

Note that the dual problem (3) is always strictly feasible, due to the constraint $\int_K d\mu \leq 1$ in the primal GMP (1).

Weak duality holds for this pair of problems, meaning $\text{val}' \leq \text{val}$. The difference $\text{val} - \text{val}'$ is called *duality gap*. In fact, the duality gap is always zero, as the next theorem shows. Note that a zero duality gap does *not* imply the existence of a dual optimal solution.

Theorem 1 (see, e.g. [11, Theorem 1.3]) *Assume problem (1) is feasible. Then it has an optimal solution (the* inf *is attained), and* $\text{val} = \text{val}'$.

We continue by recalling a sufficient condition for a dual optimal solution to exist.

Theorem 2 (see, e.g. [18, Proposition 2.8]) *Suppose problem (1) is feasible. If*

$$
b \in \text{int}((\langle f_1, \mu \rangle, \ldots, \langle f_m, \mu \rangle) : \mu \in \mathcal{M}(K)_+) \tag{4}
$$

then the set of optimal solutions of (3) is nonempty and bounded.

As discussed in Lasserre [10], it is customary in the literature to assume that condition (4) holds, but in practice it may be a non-trivial task to check whether it does. We do stress, however, that condition (4) does hold for the applications discussed in the next subsection.

Another result worth mentioning is that if the GMP (1) has an optimal solution, it has one which is finite atomic.

Theorem 3 (see, e.g. [4, Theorem 3]) *If the GMP (1) has an optimal solution, then it has one which is finite atomic with at most m atoms, i.e., of the form* $\mu^* = \sum_{\ell=1}^m \omega_\ell \delta_{\mathbf{x}^{(\ell)}}$ *where* $\omega_\ell \geq 0$, $\mathbf{x}^{(\ell)} \in K$ *and* $\delta_{\mathbf{x}^{(\ell)}}$ *denotes the Dirac measure supported at* $\mathbf{x}^{(\ell)}(\ell \in [m])$.

2.3 Applications

We now review some examples of problems which can be formulated as a GMP with polynomial data, and discuss the special cases considered in this paper, namely when the set K is a simplex or sphere.

2.3.1 Polynomial and rational optimization

Consider the problem of minimizing a rational function over K:

$$p^* = \inf_{\mathbf{x} \in K} \frac{p(\mathbf{x})}{q(\mathbf{x})}, \tag{5}$$

where $q, p \in \mathbb{R}[\mathbf{x}]$ are relatively prime and we may assume $q(\mathbf{x}) > 0$ for all $\mathbf{x} \in K$. Indeed, if q changes signs on K, Jibetean and De Klerk [9, Corollary 1] showed that $p^* = -\infty$. We will in fact make the stronger assumption that $q(\mathbf{x}) \geq 1$ on K, i.e. that we know a positive lower bound on the minimum of q over K. The optimization problem (5) can be modeled as a GMP:

$$\text{val} = \inf_{\mu \in \mathcal{M}(K)_+} \left\{ \int_K p(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) : \int_K q(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) = 1 \right\}. \tag{6}$$

The inequality constraint $\int_K \mathrm{d}\mu(\mathbf{x}) \leq 1$ is redundant if $q(\mathbf{x}) \geq 1 \; \forall \mathbf{x} \in K$ and can be added to obtain a problem of form (1). Note that setting $q(\mathbf{x}) = 1$ for all $\mathbf{x} \in K$ problem (5) becomes a polynomial optimization problem.

We now consider the special case where K is the simplex. Motzkin and Strauss [12] showed that the maximum stable set problem can be formulated as a quadratic polynomial optimization problem over the simplex. Indeed, for a graph G with adjacency matrix adjacency matrix A,

$$\frac{1}{\alpha(G)} = \min_{\mathbf{x} \in \Delta_{n-1}} \mathbf{x}^T (A + I)\mathbf{x},$$

where I is the identity matrix, and $\alpha(G)$ is the stable set number (independence number) of G. This gives a quadratic polynomial optimization problem over the simplex, that may be written as the GMP (6) with $p(\mathbf{x}) = \mathbf{x}^T (A + I)\mathbf{x}$ and $q(\mathbf{x}) \equiv 1$.

To give an example for the special case when K is a sphere, recall that deciding convexity of a homogeneous polynomial f of degree 4 or higher is known to be NP-hard [1]. A homogeneous polynomial f is convex if and only if

$$\min_{(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^{2n-1}} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} \geq 0,$$

which can again be cast as a GMP over the sphere via (6).

2.3.2 Polynomial cubature

Another application that goes beyond polynomial optimization is concerned with polynomial cubature rules, see e.g. [7, 21]. Let $N \in \mathbb{N}$. Consider the problem of multivariate numerical integration of a function f over a set K with respect to a given (reference) measure $\mu_0 \in \mathcal{M}(K)_+$. Loosely speaking, a cubature scheme consists of a set of nodes $\mathbf{x}^{(\ell)} \in K$ and weights $\omega_\ell \geq 0$ for $\ell \in [N]$, respectively, such that

$$\int_K f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) \approx \sum_{\ell=1}^{N} \omega_\ell f\left(\mathbf{x}^{(\ell)}\right).$$

We call a rule consisting of nodes $\mathbf{x}^{(\ell)}$ and weights ω_ℓ for $\ell \in [N]$ a polynomial cubature scheme of degree d if it is exact for polynomials up to degree d. Finding polynomial cubature rules is NP-hard in general, see [3]. The problem of finding such weights and nodes can be cast as a GMP. Let $d \in \mathbb{N}$ and $\beta \in \mathbb{N}^n$ any vector such that $|\beta| > d$. Assume the reference measure μ_0 is a probability measure, otherwise set $\mu_0 \leftarrow \mu_0/\mu_0(K)$. In the GMP given by

$$\begin{aligned}
\text{val} := &\inf_{\mu \in \mathcal{M}(K)_+} \int_K \mathbf{x}^\beta \mathrm{d}\mu(\mathbf{x}) \\
&\text{s.t.} \int_K \mathbf{x}^\alpha \mathrm{d}\mu(\mathbf{x}) = \int_K \mathbf{x}^\alpha \mathrm{d}\mu_0(\mathbf{x}) \ \forall \alpha \in \mathbb{N}_d^n
\end{aligned} \tag{7}$$

the redundant constraint $\int_K \mathrm{d}\mu(\mathbf{x}) \leq 1$ can be added to turn it into a GMP of form (1). The the solution μ^* to (7) will be of the form $\mu^* = \sum_{\ell=1}^{N} \omega_\ell \delta_{x^{(\ell)}}$, where $N \leq |\mathbb{N}_d^n| = \binom{n+d}{d}$ by Theorem 3. This result is known as Tchakaloff's theorem [20]. There is some freedom in the choice of the objective function, however, note that it should be linearly independent of $\{\mathbf{x}^\alpha\}$ for $\alpha \in \mathbb{N}_d^n$.

In the special cases where K is a simplex or sphere, many cubature schemes are known, but this remains an active field of study. The interested reader is referred to the book [6] for more details.

3 A linear relaxation hierarchy over the simplex

In the remainder of the paper we will only deal with the GMP (1) with polynomial data, i.e. we assume in what follows that all f_i's are polynomials ($i \in \{0, \ldots, m\}$).

A moment sequence $(y_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}$ of a measure $\mu \in \mathcal{M}(K)$ is an infinite sequence such that

$$y_\alpha = \int_K \mathbf{x}^\alpha \mathrm{d}\mu(\mathbf{x}) \ \forall \alpha \in \mathbb{N}^n.$$

Let $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be a linear operator

$$p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha \mapsto L(p) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha y_\alpha$$

that maps monomials to their respective moments. Thus, to an optimal solution μ^* of a GMP there is an associated linear functional L^* such that $L^*(f_0) = \text{val}$ and $L^*(f_i) = b_i$ for all $i \in [m]$ as well as $L^*(1) \leq 1$. The idea of the relaxation we are about to introduce is to approximate the optimal solution by a sequence (hierarchy) of linear functionals $L^{(r)}$ that depend on $r = 1, 2, \ldots$. Let $K = \Delta_{n-1}$. For $i = 0, 1, \ldots, m$

let f_i be a real homogeneous polynomial of degree d and let $r \geq d$. Let $L^{(r)}$ be the optimal solution of the following RLT-type relaxation of (1):

$$\underline{f}_{LP}^{(r)} = \min_{\substack{L:\mathbb{R}[\mathbf{x}]\to\mathbb{R} \\ L \text{ linear}}} L(f_0)$$

$$\text{s.t. } L(f_i) = b_i \quad \forall\, i \in [m]$$

$$L(1) \leq 1$$

$$L(\mathbf{x}^\alpha) \geq 0 \quad \forall\, |\alpha| \leq r$$

$$L(\mathbf{x}^\alpha) = L\left(\mathbf{x}^\alpha \sum_{i=1}^n x_i\right) \quad \forall\, |\alpha| \leq r-1. \tag{8}$$

Every feasible solution μ' to (1) provides an upper bound for (8) by setting $L(\mathbf{x}^\alpha) = \langle \mathbf{x}^\alpha, \mu' \rangle$. Hence, $\underline{f}_{LP}^{(r)} \leq$ val. The second last constraint is reflecting the necessary condition for a positive measure μ over the simplex:

$$\langle \mathbf{x}^\alpha, \mu \rangle = \int_{\Delta_{n-1}} \mathbf{x}^\alpha d\mu \geq 0 \quad \forall \alpha \in \mathbb{N}^n.$$

The last constraint in (8) arises from the fact that

$$L(p) = L(q) \text{ if } p(\mathbf{x}) = q(\mathbf{x}) \quad \forall \mathbf{x} \in \Delta_{n-1}.$$

Equivalently, defining the ideal $\mathcal{I} = \{\mathbf{x} \mapsto p(\mathbf{x})\left(1 - \sum_{i=1}^n x_i\right) \; : \; p \in \mathbb{R}[\mathbf{x}]\}$ we require

$$L(p) = L(q) \Leftrightarrow p = q \mod \mathcal{I},$$

where $p = q \mod \mathcal{I}$ means $p(\mathbf{x}) = q(\mathbf{x}) + (1 - \sum_{i=1}^n x_i)h(\mathbf{x})$ for some $h \in \mathbb{R}[\mathbf{x}]$.

Our formulation (8) is closely related to the RLT approach by Sherali and Adams [19], that was originally introduced for 0–1 mixed integer linear programming problems and subsequently extended for more general problems (but not to the GMP, to the best of the authors' knowledge). In fact, for the special case of polynomial optimization, problem (8) is essentially a Sherali–Adams RLT approach. To see this, note that our linearization operator L corresponds to the approximation $L(\mathbf{x}^\alpha) \approx \langle \mathbf{x}^\alpha, \mu^* \rangle$, where μ^* again denotes an optimal solution to the GMP (1). For the special case of polynomial optimization, we may assume that μ^* is a Dirac delta centered at an optimal solution, say $\mathbf{x}^*$. In this case, $L(\mathbf{x}^\alpha) \approx \langle \mathbf{x}^\alpha, \mu^* \rangle = \mathbf{x}^{*\alpha}$, i.e. L corresponds to the type of linearization operator introduced by Sherali and Adams [19].

We now state two lemmas that will come in handy in our later analysis.

Lemma 1 *Let $r, k \in \mathbb{N}$ with $k \leq r$ and let L be a feasible solution to the linear relaxation (8) for some $f_0, f_1, \ldots, f_m$. Then for all $\mathbf{x}^\gamma$ with $\gamma \in \mathbb{N}^n$ and $|\gamma| \leq r-k$ we have*

$$L\left(\mathbf{x}^\gamma\right) = L\left(\mathbf{x}^\gamma \left(\sum_{i=1}^n x_i\right)^k\right).$$

Proof The proof is immediate, by using induction on k.   □

Lemma 2 *Consider the GMP given in (1) and let $(y,t) \in \mathbb{R}^m \times \mathbb{R}_+$. Then the pair (y,t) is dual optimal only if*

$$0 = \min_{\mathbf{x} \in K}\left(f_0(\mathbf{x}) - \sum_{i=1}^m y_i f_i(\mathbf{x}) + t\right).$$

Proof The proof is a direct consequence of the GPM duality theory, and is omitted here.   □

When we consider the case where $K = \Delta_{n-1}$, we may, without loss of generality, assume the f_i to be homogeneous of the same degree for all $i = 0, 1, \ldots, m$. Indeed, let $f(\mathbf{x}) = \sum_{j=0}^d f_j(\mathbf{x})$, where $\deg(f_j) = j$. Then, $g(\mathbf{x}) := \sum_{j=0}^d f_j(\mathbf{x})\left(\sum_{i=1}^n x_i\right)^{d-j}$ is homogeneous of degree d and $f(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x} \in \Delta_{n-1}$.

3.1 Convergence analysis

The following theorem is a refinement of a result by Powers and Reznick [15], obtained by de Klerk et al. [5, Theorem 1.1]. It is a quantitative version of Pólya's Positivstellensatz (see, e.g. [17] for a survey), and it will be crucial in our analysis of the simplex case.

Theorem 4 *Suppose $f \in \mathbb{R}[\mathbf{x}]$ is a homogeneous polynomial of degree d of the form $f(\mathbf{x}) = \sum_{|\alpha|=d} f_\alpha \mathbf{x}^\alpha$. Let $\varepsilon = \min_{\Delta_{n-1}} f(\mathbf{x})$ and define*

$$B(f) = \max_{|\alpha|=d} \frac{\alpha_1! \cdots \alpha_n!}{d!} f_\alpha. \tag{9}$$

Then the polynomial $(x_1 + \cdots + x_n)^k f(\mathbf{x})$ has only positive coefficients if

$$k > \frac{d(d-1)}{2}\frac{B(f)}{\varepsilon} - d. \tag{10}$$

We continue by stating and proving one of the main results of this paper.

Theorem 5 *Let* val *be the optimal value of the GMP (1) for input data $K = \Delta_{n-1}, f_0, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ homogeneous of degree d and $b_1, \ldots, b_m \in \mathbb{R}$. Assume there exists a dual optimal solution $(\bar{y}, t)$ and let $f_{m+1}(\mathbf{x}) := 1$ for every $\mathbf{x} \in \Delta_{n-1}$ and set $\bar{y}_{m+1} = -t$. Then, setting $y_0 = 1$ and $y_i = -\bar{y}_i$ for $i \in [m+1]$ we have*

$$0 \leq \mathrm{val} - \underline{f}_{\mathrm{LP}}^{(r)} \leq \frac{\left(\sum_{i=0}^{m+1} B(y_i f_i) + t\right) d(d-1)}{2(r-1) - d(d-1)}, \tag{11}$$

for $B(\cdot)$ as in (9) and $r > d(d-1)/2 + 1$.

Remark 1 The bound we give in Theorem 5 depends on the dual optimal solution (y, t). We cannot bound the dual variables in terms of the problem data a priori in general, as they may become arbitrarily large. There are, however, cases in which one can bound the variables in terms of the problem data. An example of this case can be found in Sect. 6.

Proof By Theorem 1 there is no duality gap. Let $r > d(d-1)/2 + 1$ and let $L^{(r)}$ be an optimal solution to (8). Fix some $\varepsilon > 0$. Then,

$$0 \leq \mathrm{val} - \underline{f}_{\mathrm{LP}}^{(r)} = \mathrm{val} - L^{(r)}\left(\sum_{i=1}^{m} \bar{y}_i f_i - t + f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right)$$

$$= \mathrm{val} - \sum_{i=1}^{m} \bar{y}_i L^{(r)}(f_i) + t L^{(r)}(1) - L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right)$$

$$\leq \mathrm{val} - \sum_{i=1}^{m} \bar{y}_i b_i + t - L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right)$$

$$= -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right)$$

$$= -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t + \varepsilon\right) + \varepsilon L^{(r)}(1)$$

$$\leq -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t + \varepsilon\right) + \varepsilon,$$

where both inequalities follow from the fact that $L^{(r)}(1) \leq 1$. By Lemma 2 we have $\min_{\mathbf{x} \in \Delta_{n-1}} f_0(\mathbf{x}) - \sum_{i=1}^{m+1} \bar{y}_i f_i(\mathbf{x}) + \varepsilon = \varepsilon$. We assume wlog that $f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i$ is homogeneous of degree d. Define

$$f := f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i + \varepsilon \left(\sum_{i=1}^{n} x_i\right)^d,$$

which is homogeneous as well and its minimum over the simplex is ε. The aim now is to show that $L^{(r)}(f) \geq 0$ for the appropriate choice of r and then bound r in terms of ε. By Theorem 4 for k as in (10) we have

$$f(\mathbf{x})\left(\sum_{i=1}^{n} x_i\right)^k = \sum_{\beta \in \mathbb{N}_{d+k}^n} c_\beta x^\beta$$

with $c_\beta > 0$ for all $\beta \in \mathbb{N}_{d+k}^n$. To determine the smallest integer k for which the theorem holds we will first bound $B(f)$. For this, set $y_0 = 1$ and $y_i = -\bar{y}_i$. We may rewrite f as

$$f = \sum_{i=0}^{m+1} y_i f_i + \varepsilon \left(\sum_{i=1}^{n} x_i \right)^d$$

$$= \sum_{i=0}^{m+1} y_i f_i + \varepsilon \left(\sum_{|\alpha|=d} \binom{d}{\alpha_1 \cdots \alpha_n} x^\alpha \right)$$

$$= \sum_{|\alpha|=d} \left(\sum_{i=0}^{m+1} y_i f_{i,\alpha} + \varepsilon \binom{d}{\alpha_1 \cdots \alpha_n} \right) x^\alpha.$$

Then,

$$B(f) = \max_{\alpha} \left[\left(\sum_{i=0}^{m+1} y_i f_{i,\alpha} + \frac{d!}{\alpha_1! \cdots \alpha_n!} \varepsilon \right) \frac{\alpha_1! \cdots \alpha_n!}{d!} \right]$$

$$= \left(\max_{\alpha} \left(\sum_{i=0}^{m+1} y_i f_{i,\alpha} \right) \frac{\alpha_1! \cdots \alpha_n!}{d!} \right) + \varepsilon$$

$$\leq \sum_{i=0}^{m+1} \left(\max_{\alpha} y_i f_{i,\alpha} \frac{\alpha_1! \cdot \cdots \cdot \alpha_n!}{d!} \right) + \varepsilon$$

$$= \sum_{i=0}^{m+1} B(y_i f_i) + \varepsilon.$$

With this bound on $B(f)$ we find that if r is large enough, i.e.,

$$r \geq \left\lceil \frac{d(d-1)}{2} \frac{\sum_{i=0}^{m+1} B(y_i f_i) + \varepsilon}{\varepsilon} \right\rceil \geq \left\lceil \frac{d(d-1)}{2} \frac{B(f)}{\varepsilon} \right\rceil,$$

it follows from Lemma 1 that

$$-L^{(r)} \left(f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i + \varepsilon \right) + \varepsilon = \varepsilon - L^{(r)}(f)$$

$$= \varepsilon - L^{(r)} \left(f \left(\sum_{i=1}^{n} x_i \right)^k \right)$$

$$= \varepsilon - L^{(r)} \left(\sum_{\beta \in \mathbb{N}_{k+d}^n} c_\beta x^\beta \right) \leq \varepsilon,$$

where the last inequality follows from the fact that $L^{(r)}(\mathbf{x}^\alpha) \geq 0$ for all $|\alpha| \leq r$. To find a bound on r in terms of ε we set

$$r = \left\lceil \frac{d(d-1)}{2} \frac{\sum_{i=0}^{m+1} B(y_i f_i) + \varepsilon}{\varepsilon} \right\rceil.$$

Then, one may bound r as follows

$$r - 1 \leq \frac{d(d-1)}{2} \left(\frac{\sum_{i=0}^{m+1} B(y_i f_i)}{\varepsilon} + 1 \right)$$

$$\Leftrightarrow \varepsilon \leq \frac{\sum_{i=0}^{m+1} B(y_i f_i) d(d-1)}{2(r-1) - d(d-1)},$$

concluding the proof. □

4 Lasserre hierarchy over the sphere

We now consider the GMP (1) over the sphere, i.e. we consider the case $K = \mathcal{S}^{n-1}$. Additionally, we assume the $f_0, f_1, \ldots, f_m$ in (1) are homogeneous polynomials of even degree $2d$.

The Lasserre hierarchy [11] of semidefinite relaxations of the GMP (1) over the sphere is given by

$$\underline{f}_{\text{SDP}}^{(2r)} = \min_{\substack{L:\mathbb{R}[\mathbf{x}]\to\mathbb{R} \\ L \text{ linear}}} L(f_0)$$

$$\text{s.t.} \quad L(f_i) = b_i \quad \forall i \in [m]$$

$$L(1) \leq 1$$

$$L\left([\mathbf{x}]_r [\mathbf{x}]_r^T\right) \succeq 0$$

$$L(\mathbf{x}^\alpha) = L\left(\mathbf{x}^\alpha \|\mathbf{x}\|_2^2\right) \quad \forall \ |\alpha| \leq 2r - 2, \qquad (12)$$

where the L operator is now applied entry-wise to matrix-valued functions, where needed and the optimal solution is denoted by $L^{(2r)}$.

The following lemma enables us to use a quantitative Positivstellensatz by Fang and Fawzi [8] for positive polynomials on the sphere, to obtain a rate of convergence of the Lasserre hierarchy. It is a folklore result and certainly known to be true, however we did not find a suitable reference. Hence, we give a short proof for completeness.

Lemma 3 *Let* $L : \mathbb{R}[\mathbf{x}]_{2k} \to \mathbb{R}$ *be a linear operator and suppose* $L\left([\mathbf{x}]_k [\mathbf{x}]_k^T\right) \succeq 0,$ *where the operator is applied entrywise to the matrix* $[\mathbf{x}]_k [\mathbf{x}]_k^T$. *Then,* $L(\sigma) \geq 0$ *for all* $\sigma \in \Sigma[\mathbf{x}]_k$.

Proof Let $\sigma \in \Sigma[\mathbf{x}]_k$ be a sum of squares of degree $2k$. Then there exists $A \succeq 0$ such that $\sigma = [\mathbf{x}]_k^T A [\mathbf{x}]_k$. Let $\langle \cdot, \cdot \rangle$ denote the trace inner product. We have

$$L(\sigma) = L\left([\mathbf{x}]_k^T A [\mathbf{x}]_k\right) = \sum_{i,j} A_{i,j} L\left(([\mathbf{x}]_k)_i ([\mathbf{x}]_k)_j\right) = \langle A, L\left([\mathbf{x}]_k [\mathbf{x}]_k^T\right) \rangle \geq 0,$$

since both A and $L\left([\mathbf{x}]_k [\mathbf{x}]_k^T\right)$ are psd. □

The quantitative Positivstellensatz by Fang and Fawzi [8] is as follows.

Theorem 6 [8, Theorem 3.8] *Assume f is a homogeneous polynomial of degree $2d$ such that $0 \leq f(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathcal{S}^{n-1}$ and $d \leq n$. There are constants C_d, C'_d that depend only on d such that if $r \geq C_d n$ then*

$$f + C'_d (d/r)^2 = \sigma(\mathbf{x}) + (1 - \|\mathbf{x}\|_2^2) h(\mathbf{x})$$

for $\sigma(\mathbf{x}) \in \Sigma[\mathbf{x}]_r$ and $h \in \mathbb{R}[\mathbf{x}]_{2r-2}$.

We may now use the theorem by Fang and Fawzi [8] and Lemma 3 to derive a rate of convergence for Lasserre hierarchy [11] of the GMP on the sphere as follows.

Theorem 7 *Let* val *be the optimal value of the GMP (1) for input data $K = \mathcal{S}^{n-1}$, $f_0, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ homogeneous of even degree $2d$, $b_1, \ldots, b_m \in \mathbb{R}$ and $d \leq n$. Let $(\bar{y}, t)$ be a dual optimal solution and let $f_{m+1}(\mathbf{x}) := 1$ for every $\mathbf{x} \in \mathcal{S}^{n-1}$, set $\bar{y}_{m+1} = -t$ and set $y_0 = 1$ and $y = -\bar{y}$. Further, let $f_{\max}^{i, y_i} = \max_{\mathbf{x} \in \mathcal{S}^{n-1}} y_i f_i(\mathbf{x})$. There exist constants C_d, C'_d, only dependent on d, such that if $r \geq C_d n$ we have*

$$0 \leq \text{val} - \underline{f}_{\text{SDP}}^{(2r)} \leq \frac{C'_d d^2 \sum_{i=0}^{m+1} f_{\max}^{i, y_i}}{r^2}.$$

Proof The proof is similar to that of Theorem 5, essentially the only difference being that Lemma 3 is used, and we omit the details. □

5 Limiting behavior of the hierarchies of linear operators

The purpose of this section is to show that the limit functionals of the introduced hierarchies correspond to measures, in the sense that they are the Riesz functional of the optimal solution of the corresponding GMP. In the following we will define the limit of the optimal solutions $L^{(r)}$ of the introduced hierarchies in a meaningful way and prove that the corresponding moment sequences have a representing measure.

5.1 The simplex case

Consider the case when $K = \Delta_{n-1}$. When looking at the linear operators in the relaxation hierarchies (8) one would expect that in the limit, i.e. for $r \to \infty$, the operators $L^{(r)}(\cdot)$ behave like $\langle \cdot, \mu \rangle$ for some positive measure μ. In the rest of this section we prove that this is in fact the case and we will define the limit in a meaningful

way. Consider again the ideal $\mathcal{I} = \{\mathbf{x} \mapsto p(\mathbf{x}) \left(1 - \sum_{i=1}^{n} x_i\right) : p \in \mathbb{R}[\mathbf{x}]\}$ and let $\bar{L} : \mathbb{R}[\mathbf{x}]/\mathcal{I} \to \mathbb{R}$ be a linear operator such that

1. $\bar{L}(\mathbf{x}^\alpha) \geq 0$ for all $\alpha \in \mathbb{N}^n$
2. $\bar{L}(1) \leq 1$

and let

$$\mathcal{L} = \{\bar{L} : \mathbb{R}[\mathbf{x}]/\mathcal{I} \to \mathbb{R} : \bar{L} \text{ fulfills conditions 1. and 2.}\}$$

be the class of all linear operators that satisfy the conditions above. Note that for every $\bar{L} \in \mathcal{L}$ the relation

$$\bar{L}\left(\left(1 - \sum_{i=1}^{n} x_i\right) \mathbf{x}^\alpha\right) = 0 \text{ for all } \alpha \in \mathbb{N}^n$$

trivially holds. If $\|f\| = \sup_{\mathbf{x} \in \Delta_{n-1}} |f(\mathbf{x})|$, then $(\mathbb{R}[\mathbf{x}]/\mathcal{I}, \|\cdot\|)$ is a normed vector space.

Theorem 8 (see, e.g. [13, Theorem 1.4.2]) *Suppose $F : X \to Y$ is a linear operator between two normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, then the following are equivalent*

1. *F is continuous*
2. *$\|Fx\|_Y \leq M\|x\|_X$ for some $M \in \mathbb{R}$.*

Using Theorem 8 we can prove that the operators we consider are continuous in the limit.

Lemma 4 *Every $\bar{L} \in \mathcal{L}$ is continuous.*

Proof By Theorem 8 it suffices to show that every $\bar{L} \in \mathcal{L}$ satisfies
$$|\bar{L}(f)| \leq M\|f\| = M \sup_{\mathbf{x} \in \Delta_{n-1}} |f|$$

for all $f \in \mathbb{R}[\mathbf{x}]/\mathcal{I}$. Hence, let $f \in \mathbb{R}[\mathbf{x}]/\mathcal{I}$ and let $\|f\| = \sup_{\mathbf{x} \in \Delta_{n-1}} |f(\mathbf{x})|$. Also set

$$f_{\min} = \min_{x \in \Delta_{n-1}} f(\mathbf{x}) \geq -\|f\| \text{ and } f_{\max} = \max_{x \in \Delta_{n-1}} f(\mathbf{x}) \leq \|f\|.$$

Let $\bar{L}^*$ be the optimizer of

$$\min \bar{L}(f) \text{ s.t. } \bar{L} \in \mathcal{L}.$$

Then $\bar{L}^*(f) \geq -\|f\|$. To see this suppose $f_{\min} \geq 0$, from which follows that $\bar{L}^*(f) \geq 0 \geq -\|f\|$ by Theorem 5. If $f_{\min} < 0$, consider $\bar{L}^*(f - f_{\min}) \geq 0$ and so $\bar{L}^*(f) \geq f_{\min} \geq -\|f\|$. Hence, for all $\bar{L} \in \mathcal{L}$ we have
$$\bar{L}(f) \geq \bar{L}^*(f) \geq -\|f\|.$$

Similarly, let $\bar{L}'$ be the optimizer of

$$\max \bar{L}(f) \text{ s.t. } \bar{L} \in \mathcal{L}.$$

By the same reasoning we have $\bar{L}'(f) \leq \|f\|$. Hence one can set $M = 1$ and we see

$$|L(f)| \leq \|f\|.$$

$\square$

The set $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is dense in $\mathcal{C}(\Delta_{n-1})$. This means we can employ the following theorem in the next step.

Theorem 9 (see, e.g. [13, Theorem 1.9.1]) *Suppose that M is a dense subspace of a normed space X, that Y is a Banach space, and that $T_0 : M \to Y$ is a bounded linear operator. Then there is a unique continuous function $T : X \to Y$ that agrees with T_0 on M. This function T, called a continuous linear extension of T_0, is a bounded linear operator and $\|T\| = \|T_0\|$.*

Now let

$$\mathcal{T} = \left\{ T : \mathcal{C}(\Delta_{n-1}) \to \mathbb{R} : T \text{ is the continuous linear extension of some } \bar{L} \in \mathcal{L} \right\}.$$

Proposition 1 *Let $T \in \mathcal{T}$ and $f \in \mathcal{C}(\Delta_{n-1})$. Then*

$$T(f) = \int_{\Delta_{n-1}} f(\mathbf{x}) d\mu(\mathbf{x})$$

for some positive measure μ supported on Δ_{n-1}, satisfying $\mu(\Delta_{n-1}) \leq 1$.

Proof It is sufficient to show $T(f) \geq 0$ for all $f \in \mathcal{C}(\Delta_{n-1})_+ = \{f \in \mathcal{C}(\Delta_{n-1}) : f(\mathbf{x}) \geq 0 \; \forall \mathbf{x} \in \Delta_{n-1}\}$. To see this, note that the space $\mathcal{C}(\Delta_{n-1})$ can be ordered by the convex cone $\mathcal{C}(\Delta_{n-1})_+$. Now $T(f) \geq 0$ for all $f \in \mathcal{C}(\Delta_{n-1})_+$ implies that $T \in (\mathcal{C}(\Delta_{n-1})_+)^*$, i.e. the dual cone of $\mathcal{C}(\Delta_{n-1})_+$ which is known to be the set of finite Borel measures on Δ_{n-1}. Let f be a homogeneous continuous function that is non-negative on the simplex and consider its Bernstein approximation of order r given by

$$\mathcal{B}_f^r(\mathbf{x}) = \sum_{\substack{\alpha \in \mathbb{N}_r^n \\ |\alpha| = r}} f\left(\frac{\alpha}{r}\right) \binom{r}{\alpha} \mathbf{x}^\alpha.$$

The approximation converges uniformly to f as $r \to \infty$ since f is continuous. Using Lemma 4 we see

$$T(f) = T\left(\lim_{r \to \infty} \mathcal{B}_f^r\right)$$

$$\stackrel{T \text{ cont.}}{=} \lim_{r \to \infty} T(\mathcal{B}_f^r)$$

$$= \lim_{r \to \infty} \sum_{\substack{\alpha \in \mathbb{N}_r^n \\ |\alpha| = r}} \underbrace{f\left(\frac{\alpha_1}{r}, \dots, \frac{\alpha_n}{r}\right)}_{\geq 0} \underbrace{\binom{r}{\alpha}}_{\geq 0} \underbrace{T(\mathbf{x}^\alpha)}_{\geq 0} \geq 0.$$

Hence, it follows that $T(f) = \langle f, \mu \rangle$ for some positive measure μ, such that $\mu(\Delta_{n-1}) \leq 1$. □

Remark 2 By the proof given above, it becomes clear that the continuous linear extension can in fact be defined in terms of the limit of the Bernstein approximation, i.e., define $T(f) := \lim_{r \to \infty} \bar{L}(\mathcal{B}_f^r)$ for $f \in \mathcal{C}(\Delta_{n-1})$ and $\bar{L} \in \mathcal{L}$.

5.2 The sphere case

For the sphere case, i.e. $K = \mathcal{S}^{n-1}$ consider the following theorem.

Theorem 10 (see, e.g. [11, Theorem 3.8]) *Let $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}^\infty$ be a given infinite real sequence, $\bar{L} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be the linear operator defined by*

$$p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha \mapsto \bar{L}(p) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha y_\alpha,$$

and let $K = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \dots, g_m(\mathbf{x}) \geq 0\}$. The sequence $\mathbf{y}$ has a finite Borel representing measure with support contained in K if and only if

$$\bar{L}(f^2 g_J) \geq 0 \ \forall J \subseteq \{1, \dots, m\} \ and \ f \in \mathbb{R}[\mathbf{x}],$$

where $g_J(\mathbf{x}) = \prod_{j \in J} g_j(\mathbf{x})$.

Now, let $\bar{L}$ be a linear operator such that

1. $\bar{L}(1) \leq 1$
2. $\bar{L}([\mathbf{x}]_t [\mathbf{x}]_t^T) \succeq 0 \ \forall t \in \mathbb{N}$
3. $\bar{L}(\mathbf{x}^\alpha) = \bar{L}(\mathbf{x}^\alpha \|\mathbf{x}\|_2^2) \ \forall \alpha \in \mathbb{N}^n$

and let $\mathcal{L}' = \{\bar{L} : \mathbb{R}[\mathbf{x}] \to \mathbb{R} : \bar{L} \text{ satisfies } 1. - 3.\}$. Recall that as a semialgebraic set the sphere can be written as $\mathcal{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) := 1 - \|\mathbf{x}\|_2^2 \geq 0, g_2(\mathbf{x}) := \|\mathbf{x}\|_2^2 - 1 \geq 0\}$. Then for $K = \mathcal{S}^{n-1}$ every $\bar{L} \in \mathcal{L}'$ satisfies all conditions of Theorem 10. To see this, note that the only possibilities for J are $\{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Because of condition 3 we have that $\bar{L}(\pm(1 - \|\mathbf{x}\|_2^2)p) = 0$ for all $p \in \mathbb{R}[\mathbf{x}]$ covering all cases except $J = \emptyset$. For $J = \emptyset$ the condition reduces to $\bar{L}(p^2) \geq 0$ which holds for all $p \in \mathbb{R}[\mathbf{x}]$ because of Lemma 3. Hence, every $\bar{L} \in \mathcal{L}'$ has a representing measure whose support is contained in $\mathcal{S}^{n-1}$.

6 Concluding remarks

In this last section we conclude by outlining the connection of our results to previous work. We show that—in the special case of polynomial optimization on the simplex—

our RLT hierarchy reduces to one studied earlier by Bomze and De Klerk [2], and De Klerk et al. [5].

De Klerk et al. [5] introduced the following hierarchy for minimizing a homogeneous polynomial $p \in \mathbb{R}[x]$ of degree d over the simplex.

$$p^{(r)} = \max \lambda \text{ s.t. the polynomial } \left(\sum_{i=1}^{n} x_i \right)^r \left(p(\mathbf{x}) - \lambda \left(\sum_{i=1}^{n} x_i \right)^d \right) \tag{13}$$
$$\text{has only nonneg. coefficients.}$$

It was proved that $\lim_{r \to \infty} p^{(r)} = p_{\min} = \min_{\mathbf{x} \in \Delta_{n-1}} p(\mathbf{x})$. The LP hierarchy introduced in Sect. 3 of this paper is a generalization of the hierarchy (13), in the sense made precise in the following theorem.

Theorem 11 *For some homogeneous polynomial $p \in \mathbb{R}[\mathbf{x}]$ of degree d let $\underline{f}_{\mathrm{LP}}^{(r+d)}$ be the solution to the LP relaxation of the problem*

$$\min_{\mathbf{x} \in \Delta_{n-1}} p(\mathbf{x}) = \text{val} = \inf_{\mu \in \mathcal{M}(\Delta_{n-1})_+} \left\{ \int_{\Delta_{n-1}} p(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) : \int_{\Delta_{n-1}} \mathrm{d}\mu(\mathbf{x}) = 1 \right\}$$

for some $r \in \mathbb{N}$. Then,

$$p^{(r)} = \underline{f}_{\mathrm{LP}}^{(r+d)}.$$

Proof The proof is straightforward, and omitted for the sake of brevity. □

As has been noted before, the estimate of Theorem (5) depends on the dual variables. While it is in general not possible to get rid of these variables in the estimate there are cases in which we can. In the following we present an example of such a case.

Example 1 Consider the case of polynomial optimization over the simplex. Let $f \in \mathbb{R}[\mathbf{x}]$ be of degree d and set

$$f_{\min} = \min_{\mathbf{x} \in \Delta_{n-1}} f(\mathbf{x}),$$

and analogously define $f_{\max}$. We can cast this as a GMP of type (1)

$$f_{\min} = \inf_{\mu \in \mathcal{M}(\Delta_{n-1})} \left\{ \int_{\Delta_{n-1}} f(\mathbf{x}) \mathrm{d}\mu : \int_{\Delta_{n-1}} \mathrm{d}\mu = 1, \int_{\Delta_{n-1}} \mathrm{d}\mu \le 1 \right\}.$$

A dual optimal solution is in this case given by $(y^*, t^*) = (f_{\min}, 0)$. Noting that in the estimate we set $y_0 = 1$, our estimate (11) becomes

$$f_{\min} - \underline{f}_{\mathrm{LP}}^{(r+d)} \le \frac{d(d-1)}{2(r+d-1) - d(d-1)} (B(f) - f_{\min})$$

and applying the inequality

$$B(f) - f_{\min} \leq \binom{2d-1}{d} d^d \left(f_{\max} - f_{\min} \right),$$

shown in [5, Theorem 2.2], we find

$$f_{\min} - \underline{f}_{LP}^{(r+d)} \leq \frac{d(d-1)}{2(r+d-1) - d(d-1)} \binom{2d-1}{d} d^d \left(f_{\max} - f_{\min} \right).$$

This is essentially the same result as was obtained in [5, Theorem 1.3]. The presented example highlights the fact that results for convergence rates of the GMP may not be as clean as for simpler problems like polynomial optimization, even though the tools that are used to obtain these results are the same. This, of course, is due to the fact that the GMP is much more complicated in general.

Moreover, we would like to emphasize that the conceptual tools of this paper are not limited to the cases that were treated. In fact, given a quantitative version of a Positivstellensatz, it is possible to perform a convergence analysis of the kinds we proposed in this paper as long as the nature of the relaxation hierarchy, i.e. linear or semidefinite, is coherent with the positivity certificate given by the Positivstellensatz. For example, for more general sets K there is a (much weaker) quantitative Positivstellensatz available found by Nie and Schweighofer [14]. This result can be used to bound the rate of convergence of the GMP for more general sets. We chose to discuss the simplex and the sphere as there are strong Positivstellensätze available in these cases and to expose the fact that the relaxation must be in line with the certificate. For the sphere case, one could also use the following Positivstellensatz by Reznick.

Theorem 12 (cf. Theorem 3.12 in Reznick [16]) *Assume f is a homogeneous polynomial of degree $2d$ such that $0 \leq f(x) \leq 1$ for all $x \in \mathcal{S}^{n-1}$. Then one has*

$$f(\mathbf{x}) + \frac{d(d-1)n}{r \log 2} = \sigma(\mathbf{x}) + (1 - \|\mathbf{x}\|^2)\|h(\mathbf{x})$$

for some $\sigma \in \Sigma[\mathbf{x}]_{2(r+d)}$ and $h \in \mathbb{R}[\mathbf{x}]_{2(r+d)-2}$.

By using this theorem instead of Theorem 6, one obtains a convergence result with fewer assumptions than the one presented in Theorem 7, but at the cost of a worse convergence rate. In particular, one may avoid the assumption $n \leq d$ in Theorem 6 by using the result by Reznick, leading to a convergence rate of $O(1/r)$ on the sphere (as opposed to the $O(1/r^2)$ in Theorem 7).

References

1. Ahmadi, A., Olshevsky, A., Parrilo, P., Tsitsiklis, J.: NP-hardness of deciding convexity of quartic polynomials and related problems. Math. Program. **137**(1), 453–476 (2013)
2. Bomze, I., Klerk, E.: Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. J. Glob. Optim. (2001). https://doi.org/10.1023/A:1020209017701
3. Boyd, S., Ryu, E.: Extensions of Gauss quadrature via linear programming. Found. Comput. Math. **15**(4), 953–971 (2015). https://doi.org/10.1007/s10208-014-9197-9
4. de Klerk, E., Laurent, M.: A Survey of Semidefinite Programming Approaches to the Generalized Problem of Moments and Their Error Analysis. Association for Women in Mathematics Series, pp. 17–56. Springer, Berlin (2019)
5. de Klerk, E., Laurent, M., Parrilo, P.: A PTAS for the minimization of polynomials of fixed degree over the simplex. Theor. Comput. Sci. **361**(2–3), 210–225 (2006)
6. Dai, F., Xu, Y.: Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer, Berlin (2013)
7. Dunkl, C., Xu, Y.: Orthogonal polynomials of several variables. In: Encyclopedia of Mathematics and Its Applications, 2 edn. Cambridge University Press (2014). https://doi.org/10.1017/CBO9781107786134
8. Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere and applications in quantum information theory. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01537-7
9. Jibetean, D., Klerk, E.: Global optimization of rational functions: a semidefinite programming approach. Math. Program. **106**, 93–109 (2006). https://doi.org/10.1007/s10107-005-0589-0
10. Lasserre, J.B.: A semidefinite programming approach to the generalized problem of moments. Math. Program. **112**(1), 65–92 (2008). https://doi.org/10.1007/s10107-006-0085-1
11. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press, London (2009). https://doi.org/10.1142/p665
12. Motzkin, M., Straus, E.: Maxima for graphs and a new proof of a Theorem of Turán. Can. J. Math. **17**, 533–540 (1965)
13. Megginson, R.: An Introduction to Banach Space Theory. Springer, Berlin (1998)
14. Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. J. Complex. **23**, 135–150 (2007). https://doi.org/10.1016/j.jco.2006.07.002
15. Powers, V., Reznick, B.: A new bound for Pólya's Theorem with applications to polynomials positive on polyhedra. J. Pure Appl. Algebra **164**(1–2), 221–229 (2001). https://doi.org/10.1016/S0022-4049(00)00155-9. (**Copyright: Copyright 2005 Elsevier B.V., All rights reserved**)
16. Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Math. Z. **220**, 75–97 (1995)
17. Reznick, B.: Some concrete aspects of Hilbert's 17th Problem. In: In Contemporary Mathematics, pp. 251–272. American Mathematical Society (1996)
18. Shapiro, A.: On duality theory of conic linear problems. Semi-Infinite Programming pp. 135–165 (2000). https://doi.org/10.1007/978-1-4757-3403-47
19. Sherali, H., Adams, W.: A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems, Nonconvex Optimization and Its Applications, vol. 31. Springer, Berlin (1999)
20. Tchakaloff, V.: Formules de cubature ?ecanique cóefficients nonnégatifs. Bulletin des Sciences Mathématiques **81**, 123–134 (1957)
21. Williams, D., Shunn, L., Jameson, A.: Symmetric quadrature rules for simplexes based on sphere closed packed lattice arrangements. J. Comput. Appl. Math. **266**, 18–38 (2014). https://doi.org/10.1016/j.cam.2014.01.007. (**Copyright: Copyright 2014 Elsevier B.V., All rights reserved**)