CONVERGENCE RATES OF RLT AND LASSERRE-TYPE HIERARCHIES FOR THE GENERALIZED MOMENT PROBLEM OVER THE SIMPLEX AND THE SPHERE

Felix Kirschner, Etienne de Klerk

CONVERGENCE RATES OF RLT AND LASSERRE-TYPE HIERARCHIES FOR THE GENERALIZED MOMENT PROBLEM OVER THE SIMPLEX AND THE SPHERE

Felix Kirschner * **Etienne de Klerk** †

March 5, 2021

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 [*Math. Program.*, Vol. 112, 65-92, 2008] 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, Laurent and Parrilo [*J. Theoretical Computer Science*, Vol. 361, 210-225, 2006].

***K*eywords** Generalized moment problem with polynomials · linear and semidefinite programming hierarchies

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:

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

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 integrable over K with respect to all $\mu \in \mathcal{M}(K)_+$. 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. [14]. A wide range of optimization problems can be modeled as an instance of the GMP. The list includes problems from optimization,

*Tilburg University, f.c.kirschner@tilburguniversity.edu

†Tilburg University, e.deklerk@tilburguniversity.edu

This work is supported by the European Union's Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement N. 813211 (POEMA).

probability, financial economics and optimal control to name only a few, see e.g. [9]. For *polynomial data*, i.e., $f_i \in \mathbb{R}[\mathbf{x}]$ for all $i = 0, 1, \ldots, m$ and the set K being a basic closed semialgebraic set, Lasserre [8] introduced a monotone nondecreasing hierarchy of semidefinite programming (SDP) relaxations of (1). For a survey on SDP-based approximation hierarchies and their error analysis, see [3].

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 [8] for the GPM on the sphere, and for a related, RLT-type linear programming hierarchy for the GPM 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, Laurent and Parrilo [4].

Outline of the paper. First we introduce some notation in section 1.1. In section 1.2 we review the duality theory of the GMP. A brief overview of possible applications of our setting is given in section 1.3. For K the simplex we introduce a linear relaxation hierarchy in this setting in section 2 and prove a convergence rate of $O(1/r)$. Section 3 contains the new convergence analysis of the Lasserre [9] SDP hierarchies of the GPM on the sphere. In Section 4 we take a mathematical view of how the optimal measure is obtained in the limit as the level of the hierarchies approaches infinity. In section 5 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, Laurent and Parrilo [4] based on earlier results obtained by Bomze, De Klerk [2].

1.1 Notation

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ denote the set of nonnegative integers, $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ and $\mathbb{N}^n_t$ 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 $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ 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

$$[\mathbf{x}]_t = (1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_{n-1}x_n, x_n^2, \ldots, x_1^t, \ldots, x_n^t)^T.$$

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 $\Sigma[\mathbf{x}]_t$.

1.2 Duality of the generalized problem of moments

We shall briefly discuss the duality theory associated with the GMP. For this, 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 again $f_0, f_1, \ldots, f_m$ be continuous functions on K and $b_1, \ldots, b_m \in \mathbb{R}$. The dual of (1) is given by

$$\begin{aligned} \text{val}' = \sup_{(y,t) \in \mathbb{R}^m \times \mathbb{R}_+} & \sum_{i=1}^m y_i b_i - t \\ \text{s.t. } & f_0(\mathbf{x}) - \sum_{i=1}^m y_i f_i(\mathbf{x}) + t \geq 0 \quad \forall\, \mathbf{x} \in K. \end{aligned} \tag{3}$$

Note that the dual problem (3) is always strictly feasible, due to the constraint $\int_K \mathrm{d}\mu \leq 1$ 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. [9, Theorem 1.3]) Assume problem* (1) *is feasible. Then it has an optimal solution (the* inf *is attained), and* $\mathrm{val} = \mathrm{val}'$.

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

Theorem 2. *(see, e.g. [14, Proposition 2.8]) Suppose problem* (1) *is feasible. If*

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

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

As discussed in Lasserre [8], 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 on which is finite atomic.

Theorem 3. *(see, e.g. [3, 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 th 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* $\mathbf{x}^{(\ell)} (\ell \in [m])$.

1.3 Applications

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 [7, 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:

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

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).

We emphasize that minimizing a quadratic polynomial over the simplex Δ_{n-1} is already NP-hard, since it contains the problem of computing the size ($\alpha(G)$) of a maximum stable set of a graph G. Indeed, for a graph G with adjacency matrix adjacency matrix A, Motzkin and Strauss [15] showed that

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

where I is the identity matrix, which is a quadratic polynomial optimization problem over the simplex.

Similarly, deciding convexity of a homogeneous polynomial f of degree 4 or higher is known to be NP-hard [1]. It can be modeled as polynomial optimization problem over the sphere. A homogeneous polynomial f is convex if and only if

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

which in turn be cast as a GMP over the sphere.

Polynomial cubature. Another application that goes beyond polynomial optimization is concerned with polynomial cubature rules, see e.g. [5], [17]. Finding polynomial cubature rules is NP-hard in general, see [13]. 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

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

A possibility to mitigate the error in this scheme is to choose the weights and points such that the approximation is exact for polynomials up to some fixed degree. 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

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

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 solution μ^* to (7) will be of the form $\mu^* = \sum_{\ell=1}^N \omega_\ell \delta_{x^{(\ell)}}$, where $N \leq |\mathbb{N}^n_d| = \binom{n+d}{d}$ by Theorem 3. This result is known as Tchakaloff's theorem [16]. There is some freedom in the choice of the objective function, however, note that it should be linearly independent of $\{\mathbf{x}^\alpha\}$ for $\alpha \in \mathbb{N}^n_d$. Hence, our approach discussed in this paper may be applied to the problem of finding cubature rules for measures on the simplex or sphere.

2 A linear relaxation hierarchy over the simplex

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 $r \geq d$. Consider the following linear relaxation of (1):

$$\begin{aligned} \underline{f}^{(r)}_{\text{LP}} = \min \quad & L^{(r)}(f_0) \\ \text{s.t.} \quad & L^{(r)}(f_i) = b_i \quad \forall \, i \in [m] \\ & L^{(r)}(1) \leq 1 \\ & L^{(r)}(\mathbf{x}^\alpha) \geq 0 \quad \forall \, |\alpha| \leq r \\ & L^{(r)}(\mathbf{x}^\alpha) = L^{(r)}\left(\mathbf{x}^\alpha \sum_{i=1}^n x_i\right) \quad \forall \, |\alpha| \leq r-1. \end{aligned} \tag{8}$$

Every feasible solution μ' to (1) provides an upper bound for (8) by setting $L^{(r)}(\mathbf{x}^\alpha) = \langle \mathbf{x}^\alpha, \mu' \rangle$. Hence, $\underline{f}^{(r)}_{\text{LP}} \leq \text{val}$. To see it is a linear program (LP) note that each $L^{(r)}(\mathbf{x}^\alpha)$ can be replaced by a scalar variable y_α and the resulting program is an LP. The second last constraint is reflecting the necessary condition for a positive measure μ over the simplex:

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

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

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

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

$$L^{(r)}(p) = L^{(r)}(q) \Leftrightarrow p = q \mod \mathcal{I}.$$

We 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 $L^{(r)}$ 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*

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

Proof. The last equality constraint in the relaxation forces

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

Therefore, noting that $\mathbf{x}^{e_j} = x_j$ we have

$$\begin{aligned}
L^{(r)}(\mathbf{x}^\beta \mathbf{x}^{e_j}) = L^{(r)}\left(\mathbf{x}^\beta \mathbf{x}^{e_j} \sum_{i=1}^n x_i\right) \quad \forall \ |\beta| \leq r-2 \\
\Rightarrow \sum_{j=1}^n L^{(r)}(\mathbf{x}^\beta \mathbf{x}^{e_j}) = \sum_{j=1}^n L^{(r)}\left(\mathbf{x}^\beta \mathbf{x}^{e_j} \sum_{i=1}^n x_i\right) \quad \forall \ |\beta| \leq r-2 \\
\Leftrightarrow L^{(r)}\left(\mathbf{x}^\beta \sum_{j=1}^n \mathbf{x}^{e_j}\right) = L^{(r)}\left(\mathbf{x}^\beta \sum_{j=1}^n \mathbf{x}^{e_j} \sum_{i=1}^n x_i\right) \quad \forall \ |\beta| \leq r-2 \\
= L^{(r)}\left(\mathbf{x}^\beta \left(\sum_{i=1}^n x_i\right)^2\right) \quad \forall \ |\beta| \leq r-2.
\end{aligned}$$

Hence,

$$L^{(r)}(\mathbf{x}^\beta) = L^{(r)}\left(\mathbf{x}^\beta \sum_{i=1}^n x_i\right) = L^{(r)}\left(\mathbf{x}^\beta \left(\sum_{i=1}^n x_i\right)^2\right) \quad \forall \quad |\beta| \leq r-2.$$

Reiterating this procedure leads us to the desired outcome. □

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 minimization problem

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

is equivalent to

$$\inf_{\mu \in \mathcal{M}(K)_+}\left\{\int_K f_0(\mathbf{x}) - \sum_{i=1}^m y_i f_i(\mathbf{x}) \mathrm{d} + t \mathrm{d}\mu(\mathbf{x}) : \int_K \mathrm{d}\mu = 1\right\}. \tag{9}$$

By Theorem 1 there is no duality gap and there exists a primal optimal solution μ^* to the GMP (1). Set $\nu = \mu^*/\mu^*(K)$. Hence, ν is a probability measure and therefore a feasible solution to (9). We deduce

$$\begin{aligned}
0 &\leq \min_{\mathbf{x}\in K} f_0(\mathbf{x}) - \sum_{i=1}^m y_i f_i(\mathbf{x}) + t \\
&\leq \int_K f_0(\mathbf{x}) - \sum_{i=1}^m \bar{y}_i f_i(\mathbf{x}) + t \mathrm{d}\nu(\mathbf{x}) \\
&= \frac{1}{\mu^*(K)} \left(\int_K f_0(\mathbf{x}) \mathrm{d}\mu^*(\mathbf{x}) - \sum_{i=1}^m y_i \int_K f_i(\mathbf{x}) \mathrm{d}\mu^*(\mathbf{x}) + t \int_K \mathrm{d}\mu^*(\mathbf{x}) \right) \\
&\leq \frac{1}{\mu^*(K)} \left(\text{val} - y^T b + t \right) = 0,
\end{aligned}$$

where the first inequality follows from the definition of the dual (3) of the GMP and the last equality from strong duality. □

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 $\mathbf{x} \in \Delta_{n-1}$.

2.1 Convergence analysis

The following theorem is a refinement of a result by Powers and Reznick [11], obtained by de Klerk, Laurent and Parrilo [4, Theorem 1.1]. It is a quantitative version of Pólya's Positivstellensatz (see, e.g. [12] 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*

$$B(f) = \max_{|\alpha|=d} \frac{\alpha_1! \ldots \alpha_n!}{d!} f_\alpha. \tag{10}$$

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

$$k > \frac{d(d-1)}{2} \frac{B(f)}{\varepsilon} - d. \tag{11}$$

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*

$$0 \leq \text{val} - \underline{f}_{\text{LP}}^{(r)} \leq \frac{\left(\sum_{i=0}^{m+1} B(y_i f_i) + t\right) d(d-1)}{2(r-1) - d(d-1)}, \tag{12}$$

for $B(\cdot)$ as in (10) and $r > d(d-1)/2 + 1$.

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,

$$
\begin{aligned}
0 \leq \text{val} - \underline{f}_{\text{LP}}^{(r)} &= \text{val} - L^{(r)}\left(\sum_{i=1}^{m} \bar{y}_i f_i - t + f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right) \\
&= \text{val} - \sum_{i=1}^{m} \bar{y}_i L^{(r)}(f_i) + tL^{(r)}(1) - L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right) \\
&\leq \text{val} - \sum_{i=1}^{m} \bar{y}_i b_i + t - L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right) \\
&= -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right) \\
&= -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t + \varepsilon\right) + \varepsilon L^{(r)}(1) \\
&\leq -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t + \varepsilon\right) + \varepsilon,
\end{aligned}
$$

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. Hence, by Theorem 4 for k as in (11) we have

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

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

$$
\begin{aligned}
f &= \sum_{i=0}^{m+1} y_i f_i + \varepsilon \left(\sum_{i=1}^{n} x_i\right)^d \\
&= \sum_{i=0}^{m+1} y_i f_i + \varepsilon \left(\sum_{|\alpha|=d} \binom{d}{\alpha_1 \dots \alpha_n} x^\alpha\right) \\
&= \sum_{|\alpha|=d} \left(\sum_{i=0}^{m+1} y_i f_{i,\alpha} + \varepsilon \binom{d}{\alpha_1 \dots \alpha_n}\right) x^\alpha.
\end{aligned}
$$

Then,

$$
\begin{aligned}
B(f) &= \max_\alpha \left[\left(\sum_{i=0}^{m+1} y_i f_{i,\alpha} + \frac{d!}{\alpha_1! \dots \alpha_n!}\varepsilon\right) \frac{\alpha_1! \dots \alpha_n!}{d!}\right] \\
&= \left(\max_\alpha \left(\sum_{i=0}^{m+1} y_i f_{i,\alpha}\right) \frac{\alpha_1! \dots \alpha_n!}{d!}\right) + \varepsilon \\
&\leq \sum_{i=0}^{m+1} \left(\max_\alpha y_i f_{i,\alpha} \frac{\alpha_1! \dots \alpha_n!}{d!}\right) + \varepsilon \\
&= \sum_{i=0}^{m+1} B(y_i f_i) + \varepsilon.
\end{aligned}
$$

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 3 that

$$\begin{aligned} -L^{(r)}\left(f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i + \varepsilon\right) + \varepsilon &= \varepsilon - L^{(r)}(f) \\ &= \varepsilon - L^{(r)}\left(f\left(\sum_{i=1}^{n} x_i\right)^k\right) \\ &= \varepsilon - L^{(r)}\left(\sum_{\beta \in \mathbb{N}^n_{k+d}} c_\beta x^\beta\right) \leq \varepsilon, \end{aligned}$$

where the last inequality follows from the fact that $L^{(r)}(\mathbf{x}^\alpha) \geq 0$ for all $|\alpha| \leq r$. One may bound r as follows

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

concluding the proof. □

3 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 [9] of semidefinite relaxations of the GMP (1) over the sphere is given by

$$\begin{aligned} \underline{f}_{\text{SDP}}^{(2r)} = \min\ & L^{(2r)}(f_0) \\ \text{s.t.}\quad & L^{(2r)}(f_i) = b_i \quad \forall i \in [m] \\ & L^{(2r)}(1) \leq 1 \\ & L^{(2r)}\left([\mathbf{x}]_r [\mathbf{x}]_r^T\right) \succeq 0 \\ & L^{(2r)}(\mathbf{x}^\alpha) = L^{(2r)}\left(\mathbf{x}^\alpha \|\mathbf{x}\|_2^2\right) \quad \forall \quad |\alpha| \leq 2r-2, \end{aligned} \tag{13}$$

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

The following lemma enables us to use a quantitative Positivstellensatz by Fang an Fawzi [6] for positive polynomials on the sphere, to obtain a rate of convergence of the Lasserre hierarchy.

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

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

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

The quantitative Positivstellensatz by Fang and Fawzi [6] is as follows.

Theorem 6. *[6, 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 [6] and Lemma 3 to derive a rate of convergence for Lasserre hierarchy [9] 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. As in the proof of Theorem 5, Theorem 1 gives us strong duality. Let $r \geq C_d n$ and let $L^{(2r)}$ be an optimal solution to (13). Then by the same reasoning as in Theorem 5,

$$0 \leq \text{val} - \underline{f}_{\text{SDP}}^{(2r)} \leq -L^{(2r)}\left(f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i\right).$$

Set $f := f_0 - \sum_{i=1}^{m+1} y_i f_i$ and $f_{\max} = \max_{\mathbf{x}\in\mathcal{S}^{n-1}} f(\mathbf{x})$. Then $\tilde{f} = f/f_{\max}$ satisfies $0 \leq \tilde{f} \leq 1$ by Lemma 2. We find for any $\delta \geq 0$

$$\begin{aligned} -L^{(2r)}\left(f_0 - \sum_{i=1}^{m+1} \bar{y}_i f_i\right) &= -f_{\max} L^{(2r)}\left(\tilde{f}\right) \\ &\leq -f_{\max} L^{(2r)}\left(\tilde{f} + \delta\right) + \delta f_{\max}. \end{aligned}$$

Choosing $\delta = \frac{C_d' d^2}{r^2}$ and applying Theorem 6 we see that $\tilde{f} + \delta = \sigma + (1 - \|\mathbf{x}\|_2^2)h$ for $\sigma \in \Sigma[\mathbf{x}]_r$ and $h \in \mathbb{R}[\mathbf{x}]_{2r-2}$. Thus, since $L^{(2r)}(\mathbf{x}^\alpha) = L^{(2r)}(\mathbf{x}^\alpha \|\mathbf{x}\|_2^2)$ we have

$$\begin{aligned} -f_{\max} L^{(2r)}\left(\tilde{f} + \frac{C_d' d^2}{r^2}\right) + \frac{C_d' d^2}{r^2} f_{\max} &= -f_{\max} L^{(2r)}\left(\sigma + (1 - \|\mathbf{x}\|_2^2)h\right) + \frac{C_d' d^2}{r^2} f_{\max} \\ &= -f_{\max} L^{(2r)}(\sigma) + \frac{C_d' d^2}{r^2} f_{\max} \\ &\leq C_d' \frac{d^2}{r^2} f_{\max}, \end{aligned}$$

where the last inequality follows from Lemma 3. Noting that

$$f_{\max} = \max_{\mathbf{x}\in\mathcal{S}^{n-1}} \left(f_0(\mathbf{x}) - \sum_{i=1}^{m+1} \bar{y}_i f_i(\mathbf{x}) \right) \leq \sum_{i=0}^{m+1} \max_{\mathbf{x}\in\mathcal{S}^{n-1}} y_i f_i(\mathbf{x}) = \sum_{i=0}^{m+1} f_{\max}^{i,y_i}$$

we arrive at the result.

□

4 Limiting behavior of the hierarchies of linear operators

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 $\mathcal{I} = \{\mathbf{x} \mapsto p(\mathbf{x})\,(1 - \sum_{i=1}^{n} x_i) : p \in \mathbb{R}[\mathbf{x}]\}$ and let $L : \mathbb{R}[\mathbf{x}]/\mathcal{I} \to \mathbb{R}$ be a linear operator such that

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

and let

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

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

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

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. [10, 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 $L \in \mathcal{L}$ is continuous.*

Proof. By Theorem 8 it suffices to show that every $L \in \mathcal{L}$ satisfies

$$|L(f)| \leq M\|f\| = M \sup_{\mathbf{x}\in\Delta_{n-1}} |f|$$

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 L^* be the optimizer of

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

and note that $L^*(f) = f_{\min}$ as an immediate consequence of Theorem 5. Hence, for all $L \in \mathcal{L}$ we have

$$L(f) \geq L^*(f) = f_{\min} \geq -\|f\|.$$

Similarly, let L' be the optimizer of

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

By the same reasoning we have $L'(f) = f_{\max}$ and it follows that $L(f) \leq \|f\|$ for all $L \in \mathcal{L}$. Hence one can set $M = 1$ and we see

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

□

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. [10, 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 is a bounded linear operator and $\|T\| = \|T_0\|$.*

Now let

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

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

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

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

Proof. It is sufficient to show $\bar{L}(f) \geq 0$ 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 $\bar{L}(f) \geq 0$ for all $f \in \mathcal{C}(\Delta_{n-1})_+$ implies that $\bar{L} \in (\mathcal{C}(\Delta_{n-1})_+)^*$, 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

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

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

$$\begin{aligned} \bar{L}(f) &= \bar{L}(\lim_{r\to\infty} \mathcal{B}_f^r) \\ &\overset{\bar{L} \text{ cont.}}{=} \lim_{r\to\infty} \bar{L}(\mathcal{B}_f^r) \\ &= \lim_{r\to\infty} \sum_{\substack{\alpha \in \mathbb{N}_r^n \\ |\alpha| = r}} \underbrace{f\left(\frac{\alpha_1}{r}, \ldots, \frac{\alpha_n}{r}\right)}_{\geq 0} \underbrace{\binom{r}{\alpha}}_{\geq 0} \underbrace{\bar{L}(\mathbf{x}^\alpha)}_{\geq 0} \geq 0. \end{aligned}$$

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

Remark 1. *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 $\bar{L}(f) := \lim_{r\to\infty} L(\mathcal{B}_f^r)$ for $f \in \mathcal{C}(\Delta_{n-1})$ and $L \in \mathcal{L}$.*

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

Theorem 10. *(see, e.g. [9, Theorem 3.8]) Let $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}^\infty$ be a given infinite real sequence, $L : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ be the linear operator defined by*

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

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

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

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

Now, let L be a linear operator such that

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

and let $\mathcal{L}' = \{L : \mathbb{R}[\mathbf{x}] \to \mathbb{R} : L \text{ satisfies } 1. \text{ - } 3.\}$. 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 $L \in \mathcal{L}'$ 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 $L(\pm(1 - \|\mathbf{x}\|_2^2)p) = 0$ for all $p \in \mathbb{R}[\mathbf{x}]$ covering all cases except $J = \emptyset$. For $J = \emptyset$ the condition reduces to $L(p^2) \geq 0$ which holds for all $p \in \mathbb{R}[\mathbf{x}]$ because of Lemma 3. Hence, every $L \in \mathcal{L}'$ has a representing measure whose support is contained in $\mathcal{S}^{n-1}$.

5 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, Laurent and Parrilo [4].

De Klerk, Laurent and Parrilo [4] introduced the following hierarchy for minimizing a homogeneous polynomial $p \in \mathbb{R}[\mathbf{x}]$ of degree d over the simplex.

$$p^{(r)} = \max \lambda \text{ s.t. the polynomial } \left(\sum_{i=1}^n x_i\right)^r \left(p(\mathbf{x}) - \lambda \left(\sum_{i=1}^n x_i\right)^d\right) \text{ has only nonneg. coefficients.} \tag{14}$$

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 section 2 of this paper is a generalization of the hierarchy (14), 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 $\underline{f}_{\text{LP}}^{(r+d)}$ 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,

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

Proof. " $\leq$ " : Let $\lambda^* = p^{(r)}$ be optimal for (14). Then $\left(\sum_{i=1}^n x_i\right)^r \left(p(\mathbf{x}) - \lambda \left(\sum_{i=1}^n x_i\right)^d\right)$ has only negative coefficients and we find

$$\begin{aligned}
0 &\leq L^{(r+d)}\left(\left(\sum_{i=1}^n x_i\right)^r \left(p(\mathbf{x}) - \lambda^* \left(\sum_{i=1}^n x_i\right)^d\right)\right) \\
&= L^{(r+d)}\left(\left(\sum_{i=1}^n x_i\right)^r p(\mathbf{x})\right) - \lambda^* L^{(r+d)}\left(\left(\sum_{i=1}^n x_i\right)^{r+d}\right) \\
&= \underline{f}_{\text{LP}}^{(r+d)} - \lambda^*
\end{aligned}$$

for $L^{(r+d)}$ being the optimal solution to the LP relaxation.
" $\geq$ " : For the multinomial coefficient

$$\binom{k}{\alpha} = \binom{k}{\alpha_1, \ldots, \alpha_n} = \frac{k!}{\alpha_1! \ldots \alpha_n!}$$

we define $\binom{k}{\alpha} = 0$ if $\alpha_i < 0$ for some $i \in [n]$.

Consider the expansion

$$\begin{aligned}
\left(\sum_{i=1}^n x_i\right)^r \left(p(\mathbf{x}) - \lambda \left(\sum_{i=1}^n x_i\right)^d\right) &= \sum_{|\beta|=r} \binom{r}{\beta} \mathbf{x}^\beta \sum_{|\alpha|=d} p_\alpha \mathbf{x}^\alpha - \lambda \sum_{|\beta|=r+d} \binom{r+d}{\beta} \mathbf{x}^\beta \\
&= \sum_{|\beta|=r+d} \left(\sum_{|\alpha|=d} \binom{r}{\beta-\alpha} p_\alpha - \lambda \binom{r+d}{\beta}\right) \mathbf{x}^\beta.
\end{aligned}$$

Thus the LP formulation of (14) reads

$$\begin{aligned}
p^{(r)} = \max\ & \lambda \\
\text{s.t. } & \binom{r+d}{\beta} \lambda \leq \sum_{|\alpha|=d} \binom{r}{\beta-\alpha} p_\alpha \quad \forall |\beta| = r+d
\end{aligned}$$

with its dual

$$\begin{aligned} p^{(r)} = \min & \sum_{|\beta|=r+d} \sum_{|\alpha|=d} y_\beta \binom{r}{\beta-\alpha} p_\alpha \\ \text{s.t. } & y_\beta \geq 0 \quad \forall |\beta| = r+d \\ & \sum_{|\beta|=d+r} \binom{r+d}{\beta} y_\beta = 1. \end{aligned}$$

Let y be an optimal solution for the dual and define

$$L^{(r+d)}(\mathbf{x}^\beta) = y_\beta \quad \forall |\beta| = r+d.$$

Then for $|\alpha| = r+d-1$ we let

$$L^{(r+d)}(\mathbf{x}^\alpha) = \sum_{i=1}^n y_{\alpha+e_i}$$

and proceed in this manner for all $|\gamma| \leq r+d-2$. The last constraint of the dual then implies

$$1 = \sum_{|\beta|=d+r} \binom{r+d}{\beta} y_\beta = \sum_{|\beta|=d+r} \binom{r+d}{\beta} L^{(r+d)}(\mathbf{x}^\beta) = L^{(r+d)}\left(\left(\sum_{i=1}^n x_i \right)^{r+d} \right).$$

By construction we have

1. $L^{(r+d)}(\mathbf{x}^\alpha) \geq 0$ for all $|\alpha| \leq r+d$
2. $L^{(r+d)}(\mathbf{x}^\alpha) = L^{(r+d)}\left(\mathbf{x}^\alpha \sum_{i=1}^n x_i\right)$ for all $|\alpha| \leq r+d-1$
3. $1 = L^{(r+d)}\left(\left(\sum_{i=1}^n x_i\right)^{r+d} \right) \overset{2.}{=} L^{(r+d)}(1)$.

Hence, the constructed solution for the LP relaxation is feasible. Further,

$$\begin{aligned} p^{(r)} &= \sum_{|\beta|=r+d} \sum_{|\alpha|=d} y_\beta \binom{r}{\beta-\alpha} p_\alpha \\ &= \sum_{|\beta|=r+d} \sum_{|\alpha|=d} L^{(r+d)}(\mathbf{x}^\beta) \binom{r}{\beta-\alpha} p_\alpha \\ &= L^{(r+d)}\left(\sum_{|\beta|=r+d} \sum_{|\alpha|=d} \binom{r}{\beta-\alpha} p_\alpha \mathbf{x}^\beta \right) \\ &= L^{(r+d)}\left(\left(\sum_{i=1}^n x_i \right)^r p \right) \\ &= L^{(r+d)}(p) \geq \underline{f}_{\text{LP}}^{(r+d)}. \end{aligned}$$

□

In the case of polynomial optimization our estimate (12) becomes

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

and applying the inequality

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

shown in [4, Theorem 2.2], we find

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

This is essentially the same result as was obtained in [4, Theorem 1.3].

References

[1] A. Ahmadi, A. Olshevsky, P. Parrilo, and J. Tsitsiklis. Np-hardness of deciding convexity of quartic polynomials and related problems. *Mathematical programming*, 137(1):453–476, 2013.

[2] I. Bomze and E. Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24, 12 2001.

[3] E. de Klerk and M. Laurent. *A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis*, pages 17–56. Association for Women in Mathematics Series. Springer, 2019.

[4] E. de Klerk, M. Laurent, and P. Parrilo. A ptas for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, 361(2-3):210–225, 2006.

[5] C. Dunkl and Y. Xu. *Orthogonal Polynomials of Several Variables*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 2014.

[6] K. Fang and H. Fawzi. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming*, Jul 2020.

[7] D. Jibetean and E. Klerk. Global optimization of rational functions: A semidefinite programming approach. *Mathematical Programming*, 106:93–109, 05 2006.

[8] J. B. Lasserre. A semidefinite programming approach to the generalized problem of moments. *Mathematical Programming*, 112(1):65–92, 2008.

[9] J. B. Lasserre. *Moments, Positive Polynomials and Their Applications*. Imperial College Press, 10 2009.

[10] R. Megginson. *An Introduction to Banach Space Theory*. rausfinden, 1998.

[11] V. Powers and B. Reznick. A new bound for pólya's theorem with applications to polynomials positive on polyhedra. *Journal of Pure and Applied Algebra*, 164(1-2):221–229, October 2001. Copyright: Copyright 2005 Elsevier B.V., All rights reserved.

[12] B. Reznick. Some concrete aspects of hilbert's 17th problem. In *In Contemporary Mathematics*, pages 251–272. American Mathematical Society, 1996.

[13] E. Ryu and S. Boyd. Extensions of gauss quadrature via linear programming. *Foundations of Computational Mathematics*, 15(4):953–971, August 2015.

[14] Alexander Shapiro. On duality theory of conic linear problems. *Semi-Infinite Programming*, pages 135–165, 12 2000.

[15] Motzkin T. and E. Straus. Maxima for graphs and a new proof of a theorem of turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

[16] V. Tchakaloff. Formules de cubature ḿecanique cóefficients nonnégatifs. *Bulletin des Sciences Mathématiques*, 81:123–134, 1957.

[17] D. Williams, L. Shunn, and A. Jameson. Symmetric quadrature rules for simplexes based on sphere closed packed lattice arrangements. *Journal of Computational and Applied Mathematics*, 266:18–38, August 2014. Copyright: Copyright 2014 Elsevier B.V., All rights reserved.