HAL
open science

Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere

Etienne de Klerk, Monique Laurent

To cite this version:

Etienne de Klerk, Monique Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. 2019, 10.1007/s10107-019-01465-1 . hal-02407335

HAL Id: hal-02407335

https://hal.science/hal-02407335

Submitted on 17 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1904.08828v1 [math.OC] 18 Apr 2019

CONVERGENCE ANALYSIS OF A LASSERRE HIERARCHY OF UPPER BOUNDS FOR POLYNOMIAL MINIMIZATION ON THE SPHERE

Etienne de Klerk * **Monique Laurent** †

April 19, 2019

ABSTRACT

We study the convergence rate of a hierarchy of upper bounds for polynomial minimization problems, proposed by Lasserre [*SIAM J. Optim.* $21(3)$ (2011), pp. $864-885$], for the special case when the feasible set is the unit (hyper)sphere. The upper bound at level $r \in \mathbb{N}$ of the hierarchy is defined as the minimal expected value of the polynomial over all probability distributions on the sphere, when the probability density function is a sum-of-squares polynomial of degree at most $2r$ with respect to the surface measure.

We show that the exact rate of convergence is $\Theta(1/r^2)$, and explore the implications for the related rate of convergence for the generalized problem of moments on the sphere.

Keywords polynomial optimization on sphere · Lasserre hierarchy · semidefinite programming · generalized eigenvalue problem

AMS subject classification 90C22; 90C26; 90C30

1 Introduction

We consider the problem of minimizing an n-variate polynomial $f : \mathbb{R}^n \to \mathbb{R}$ over a compact set $K \subseteq \mathbb{R}^n$, i.e., the problem of computing the parameter:

$$f_{\min,K} := \min_{x \in K} f(x). \tag{1}$$

In this paper we will focus on the case when K is the unit sphere: $K = \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$, in which case we will omit the subscript K and simply write $f_{\min} = \min_{x \in \mathbb{S}^{n-1}} f(x)$.

Problem (1) is in general a computationally hard problem, already for simple sets K like the hypercube, the standard simplex, and the unit ball or sphere. For instance, the problem of finding the maximum cardinality $\alpha(G)$ of a stable set in a graph $G = ([n], E)$ can be expressed as optimizing a quadratic polynomial over the standard simplex [19], or

*Tilburg University and Delft University of Technology, E.deKlerk@uvt.nl

†Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University, monique@cwi.nl

a degree 3 polynomial over the unit sphere [20]:

$$\begin{aligned}\frac{1}{\alpha(G)} &= \min_{x\in\mathbb{R}^n}\left\{x^T(I+A_G)x : x\geq 0, \sum_{i=1}^n x_i = 1\right\}\\ &= \min_{y\in\mathbb{S}^{n-1}}\left(\sum_{i\neq j:\{i,j\}\in E} y_i^2y_j^2 + \sum_{i\in[n]} y_i^4\right),\\ \frac{\sqrt{2}}{3\sqrt{3}}\sqrt{1-\frac{1}{\alpha(G)}} &= \max_{(y,z)\in\mathbb{S}^{n+m-1}}\sum_{ij\in\overline{E}} y_iy_jz_{ij},\end{aligned}$$

where A_G is the adjacency matrix of G, $\overline{E}$ is the set of non-edges of G and $m = |\overline{E}|$. Other applications of polynomial optimization over the unit sphere include deciding whether homogeneous polynomials are positive semidefinite. Indeed, a homogeneous polynomial f is defined as positive semidefinite precisely if

$$f_{\min} = \min_{x\in\mathbb{S}^{n-1}} f(x) \geq 0,$$

and positive definite if the inequality is strict; see e.g. [23]. As special case, one may decide if a symmetric matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ is copositive, by deciding if the associated form $f(x) = \sum_{i,j\in[n]} a_{ij}x_i^2x_j^2$ is positive semidefinite; see, e.g. [21].

Another special case is to decide the convexity of a homogeneous polynomial f, by considering the parameter

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

which is nonnegative if and only if f is convex. This decision problem is known to be NP-hard, already for degree 4 forms [1].

As shown by Lasserre [16], the parameter (1) can be reformulated via the infinite dimensional program

$$f_{\min,K} = \inf_{h\in\Sigma[x]}\int_K h(x)f(x)d\mu(x) \text{ s.t. } \int_K h(x)d\mu(x) = 1, \tag{2}$$

where $\Sigma[x]$ denotes the set of sums of squares of polynomials, and μ is a given Borel measure supported on K. Given an integer $r\in\mathbb{N}$, by bounding the degree of the polynomial $h\in\Sigma[x]$ by $2r$, Lasserre [16] defined the parameter:

$$\overline{f}_K^{(r)} := \min_{h\in\Sigma[x]_r}\int_K h(x)f(x)d\mu(x) \text{ s.t. } \int_K h(x)d\mu(x) = 1, \tag{3}$$

where $\Sigma[x]_r$ consists of the polynomials in $\Sigma[x]$ with degree at most $2r$. Here we use the 'overline' symbol to indicate that the parameters provide *upper* bounds for $f_{\min,K}$, in contrast to the parameters $\underline{f}^{(r)}$ in (9) below, which provide *lower* bounds for it.

Since sums of squares of polynomials can be formulated using semidefinite programming, the parameter (3) can be expressed via a semidefinite program. In fact, since this program has only one affine constraint, it even admits an eigenvalue reformulation [16], which will be mentioned in (12) in Section 2.2 below. Of course, in order to be able to compute the parameter (3) in practice, one needs to know explicitly (or via some computational procedure) the moments of the reference measure μ on K. These moments are known for simple sets like the simplex, the box, the sphere, the ball and some simple transforms of them (they can be found, e.g., in Table 1 in [10]).

As a direct consequence of the formulation (2), the bounds $\overline{f}_K^{(r)}$ converge asymptotically to the global minimum $f_{\min,K}$ when $r\to\infty$. How fast the bounds converge to the global minimum in terms of the degree r has been investigated in the papers [12, 7, 9], which show, respectively, a convergence rate in $O(1/\sqrt{r})$ for general compact K (satisfying a minor geometric condition), a convergence rate in $O(1/r)$ when K is a convex body, and a convergence rate in $O(1/r^2)$ when K is the box $[-1,1]^n$. In these works the reference measure μ is the Lebesgue measure, except for the box $[-1,1]^n$ where more general measures are considered (see Theorem 3 below for details).

In this paper we are interested in analyzing the worst-case convergence of the bounds (3) in the case of the unit sphere $K = \mathbb{S}^{n-1}$, when selecting as reference measure the surface (Haar) measure $d\sigma(x)$ on $\mathbb{S}^{n-1}$. We let σ_{n-1} denote the surface measure of $\mathbb{S}^{n-1}$, so that $d\sigma(x)/\sigma_{n-1}$ is a probability measure on $\mathbb{S}^{n-1}$, with

$$\sigma_{n-1} := \int_{\mathbb{S}^{n-1}} d\sigma(x) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}. \tag{4}$$

(See, e.g., [6, relation (2.2.3)].) To simplify notation we will throughout omit the subscript $K = \mathbb{S}^{n-1}$ in the parameters (1) and (3), which we simply denote as

$$f_{\min} = \min_{x \in \mathbb{S}^{n-1}} f(x), \;\; \overline{f}^{(r)} = \inf_{h \in \Sigma[x]_r} \Big\{ \int_{\mathbb{S}^{n-1}} h(x) f(x) d\sigma(x) : \int_{\mathbb{S}^{n-1}} h(x) d\sigma(x) = 1 \Big\}. \tag{5}$$

Example 1. *Consider the minimization of the Motzkin form*

$$f(x_1, x_2, x_3) = x_3^6 + x_1^4 x_2^2 + x_1^2 x_2^4 - 3 x_1^2 x_2^2 x_3^2$$

on $\mathbb{S}_2$*. This form has* 12 *minimizers on the sphere, namely* $\frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1)$ *as well as* $(\pm 1, 0, 0)$ *and* $(0, \pm 1, 0)$*, and one has* $f_{\min} = 0$*.*

In Table 1 we give the bounds $\overline{f}^{(r)}$ *for the Motzkin form for* $r \leq 9$*. In Figure 1 we show a contour plot of the Motzkin*

r	0	1	2	3	4	5	6	7	8	9
$\overline{f}^{(r)}$	0.1714	0.0952	0.0519	0.0457	0.0287	0.0283	0.0193	0.0177	0.0139	0.0122

Table 1: Upper bounds for the Motzkin form

form on the sphere (top left), as well as a contour plot of the optimal density function for $r = 3$ *(top right),* $r = 6$ *(bottom left), and* $r = 9$ *(bottom right). In the figure, the red end of the spectrum denotes higher function values. Some local maximimizers of the Motzkin form are visible that correspond to* $|x_3| = 1$ *(at the poles) and* $x_3 = 0$ *(on the equator).*

When $r = 3$ *and* $r = 6$*, the modes of the optimal density are at the global minimizers* $(\pm 1, 0, 0)$ *and* $(0, \pm 1, 0)$ *(one may see the contours of two of these modes in one hemisphere). On the other hand, when* $r = 9$*, the mass of the distribution is concentrated at the* 8 *global minimizers* $\frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1)$ *(one may see* 4 *of these in one hemisphere), and there are no modes at the global minimizers* $(\pm 1, 0, 0)$ *and* $(0, \pm 1, 0)$*.*

It is also illustrative to do the same plots using spherical coordinates:

$$\begin{aligned} x_1 &= \sin\theta \sin\phi \\ x_2 &= \sin\theta \cos\phi \\ x_3 &= \cos\theta \\ \theta &\in [0, \pi] \\ \phi &\in [0, 2\pi]. \end{aligned}$$

In Figure 2 we plot the Motzkin form in spherical coordinates (top left), as well as the optimal density function that corresponds to $r = 3$ *(top right),* $r = 6$ *(bottom left), and* $r = 9$ *(bottom right). For example, when* $r = 9$ *one can see the* 8 *modes (peaks) of the density that correspond to the* 8 *global minimizers* $\frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1)$*. (Note that the peaks at* $\phi = 0$ *and* $\phi = 2\pi$ *correspond to the same mode of the density, due to periodicity.) Likewise when* $r = 3$ *and* $r = 6$ *one may see* 4 *modes corresponding to* $(\pm 1, 0, 0)$ *and* $(0, \pm 1, 0)$*.*

The convergence rate of the bounds $\overline{f}^{(r)}$ was investigated by Doherty and Wehner [4], who showed

$$\overline{f}^{(r)} - f_{\min} = O\left(\frac{1}{r}\right) \tag{6}$$

when f is a *homogeneous* polynomial. As we will briefly recap in Section 2.1, their result follows in fact as a byproduct of their analysis of another Lasserre hierarchy of bounds for $f_{\min}$, namely the *lower* bounds (9) below.

Our main contribution in this paper is to show that the convergence rate of the bounds $\overline{f}^{(r)}$ is $O(1/r^2)$ for *any* polynomial f and, moreover, that this analysis is tight for any (nonzero) linear polynomial f. This is summarized in the following theorem.

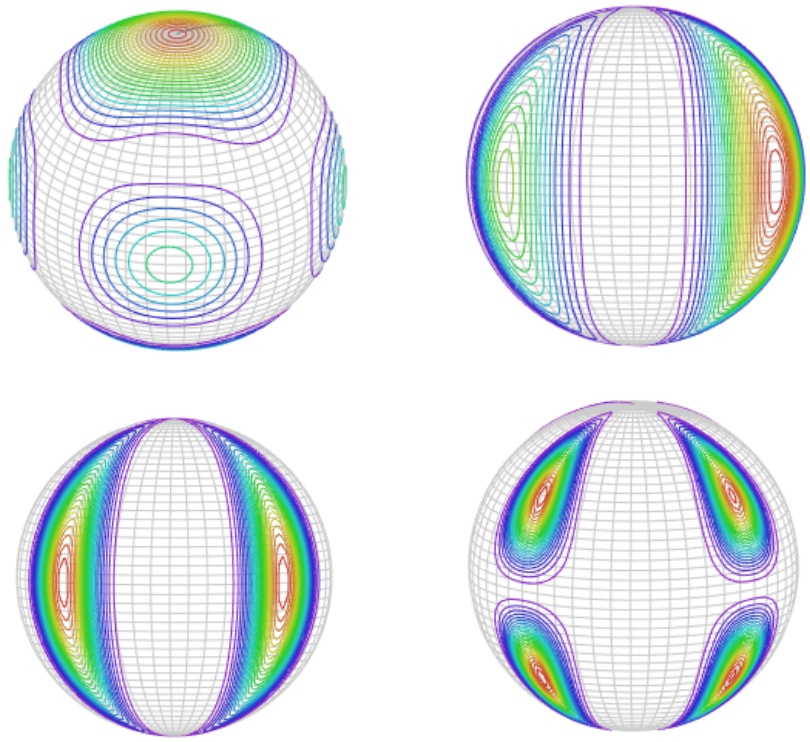

Figure 1: Contour plots of the Motzkin form on the sphere (top left) and optimal density for $r = 3$ (top right), $r = 6$ (bottom left), and $r = 9$ (bottom right).

Theorem 1. *(i) For any polynomial f we have*

$$\overline{f}^{(r)} - f_{\min} = O\left(\frac{1}{r^2}\right). \tag{7}$$

(ii) For any (nonzero) linear polynomial f we have

$$\overline{f}^{(r)} - f_{\min} = \Omega\left(\frac{1}{r^2}\right). \tag{8}$$

Let us say a few words about the proof technique. For the first part (i), our analysis relies on the following two basic steps: first, we observe that it suffices to consider the case when f is linear (which follows using Taylor's theorem), and then we show how to reduce to the case of minimizing a linear univariate polynomial over the interval $[-1, 1]$, where we can rely on the analysis completed in [9]. For the second part (ii), by exploiting a connection recently mentioned in [18] between the bounds (3) and cubature rules, we can rely on known results for cubature rules on the unit sphere to show tightness of the bounds.

Organization of the paper. In Section 2 we recall some previously known results that are most relevant to this paper. First we give in Section 2.1 a brief recap of the approach of Doherty and Wehner [4] for analysing bounds for polynomial optimization over the unit sphere. After that, we recall our earlier results about the quality of the bounds (3) in the case of the interval $K = [-1, 1]$. Section 3 contains our main results about the convergence analysis of the bounds (3) for the unit sphere: after showing in Section 3.1 that the convergence rate is in $O(1/r^2)$ we prove in Section 3.2 that the analysis is tight for nonzero linear polynomials.

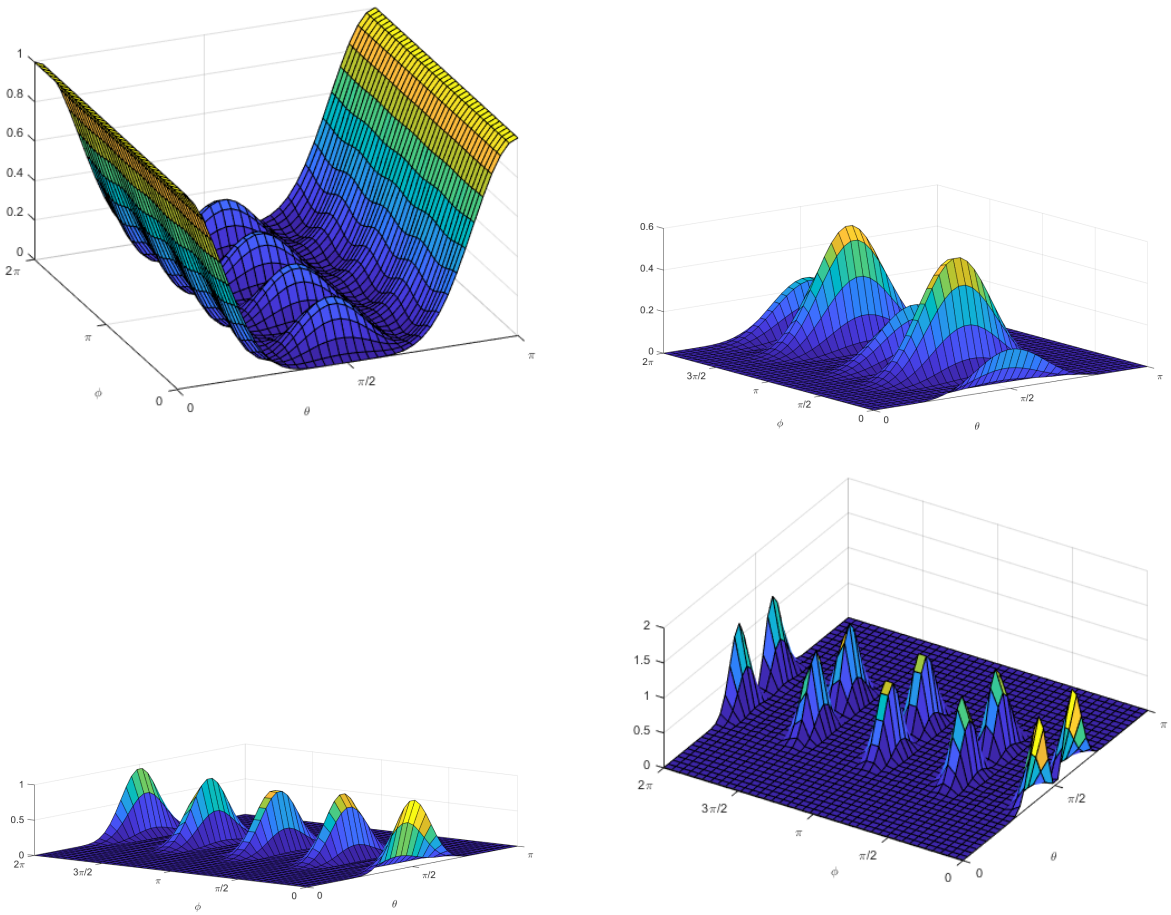

Figure 2: Plots of the Motzkin form on the sphere (top left) and optimal density for $r = 3$ (top right), $r = 6$ (bottom left), and $r = 9$ (bottom right), in spherical coordinates.

2 Preliminaries

2.1 The approach of Doherty & Wehner for the sphere

Here we briefly sketch the approach followed by Doherty and Wehner [4] for showing the convergence rate $O(1/r)$ mentioned above in (6). Their approach applies to the case when f is a *homogeneous* polynomial, which enables using the tensor analysis framework. A first observation made in [4] is that we may restrict to the case when f has even degree, because if f is homogeneous with odd degree d then we have

$$\max_{x\in\mathbb{S}^{n-1}} f(x) = \frac{d^{d/2}}{(d+1)^{(d+1)/2}} \max_{(x,x_{n+1})\in\mathbb{S}^n} x_{n+1} f(x).$$

So we now assume that f is homogeneous with even degree $d = 2a$.

The approach in [4] in fact also permits to analyze the following hierarchy of lower bounds on $f_{\min}$:

$$\underline{f}^{(r)} := \sup_{\lambda\in\mathbb{R}} \lambda \ \text{ s.t. } \ f(x) - \lambda \in \Sigma[x]_r + (1 - \|x\|^2)\mathbb{R}[x], \tag{9}$$

which are the usual sums-of-squares bounds for polynomial optimization (as introduced in [14, 22]). Here and throughout, $\|x\|$ denotes the Euclidean norm for real vectors. One can verify that (9) can be reformulated as

$$\begin{aligned} \underline{f}^{(r)} &= \sup_{\lambda\in\mathbb{R}} \lambda \ \text{ s.t. } \ (f(x) - \lambda\|x\|^{2a})\|x\|^{2r-2a} \in \Sigma[x]_r + (1 - \|x\|^2)\mathbb{R}[x] \\ &= \sup_{\lambda\in\mathbb{R}} \lambda \ \text{ s.t. } \ f(x)\|x\|^{2r-2a} - \lambda\|x\|^{2r} \in \Sigma[x] \end{aligned} \tag{10}$$

(see [11]). For any integer $r \in \mathbb{N}$ we have

$$\underline{f}^{(r)} \leq f_{\min} \leq \overline{f}^{(r)}.$$

The following error estimate is shown on the range $\overline{f}^{(r)} - \underline{f}^{(r)}$ in [4].

Theorem 2. *[4] Assume $n \geq 3$ and f is a homogeneous polynomial of degree $2a$. There exists a constant $C_{n,a}$ (depending only on n and a) such that, for any integer $r \geq a(2a^2+n-2)-n/2$, we have*

$$\overline{f}^{(r)} - \underline{f}^{(r)} \leq \frac{C_{n,a}}{r}(f_{\max} - f_{\min}),$$

where $f_{\max}$ is the maximum value of f taken over $\mathbb{S}^{n-1}$.

The starting point in the approach in [4] is reformulating the problem in terms of tensors. For this we need the following notion of 'maximally symmetric matrix'. Given a real symmetric matrix $M = (M_{\underline{i},\underline{j}})$ indexed by sequences $\underline{i} \in [n]^a$, M is called *maximally symmetric* if it is invariant under action of the permutation group $\mathrm{Sym}(2a)$ after viewing M as a $2a$-tensor acting on $\mathbb{R}^n$. This notion is the analogue of the 'moment matrix' property, when expressed in the tensor setting. To see this, for a sequence $\underline{i} = (i_1, \ldots, i_a) \in [n]^a$, define $\alpha(\underline{i}) = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ by letting α_ℓ denote the number of occurrences of ℓ within the multi-set $\{i_1, \ldots, i_a\}$ for each $\ell \in [n]$, so that $a = |\alpha| = \sum_{i=1}^n \alpha_i$. Then, the matrix M is maximally symmetric if and only if each entry $M_{\underline{i},\underline{j}}$ depends only on the n-tuple $\alpha(\underline{i}) + \alpha(\underline{j})$. Following [4] we let $\mathrm{MSym}((\mathbb{R}^n)^{\otimes a})$ denote the set of maximally symmetric matrices acting on $(\mathbb{R}^n)^{\otimes a}$.

It is not difficult to see that any degree $2a$ homogeneous polynomial f can be represented in a unique way as

$$f(x) = (x^{\otimes a})^T Z_f x^{\otimes a},$$

where the matrix Z_f is maximally symmetric.

Given an integer $r \geq a$, define the polynomial $f_r(x) = f(x)\|x\|^{2r-2a}$, thus homogeneous with degree $2r$. The parameter (10) can now be reformulated as

$$\underline{f}^{(r)} = \sup\{\langle Z_{f_r}, M\rangle : M \in \mathrm{MSym}((\mathbb{R}^n)^{\otimes r}),\ M \succeq 0,\ \mathrm{Tr}(M) = 1\}. \tag{11}$$

The approach in [4] can be sketched as follows. Let M be an optimal solution to the program (11) (which exists since the feasible region is a compact set). Then the polynomial $Q_M(x) := (x^{\otimes r})^T M x^{\otimes r}$ is a sum of squares since $M \succeq 0$. After scaling, we obtain the polynomial

$$h(x) = Q_M(x) / \int_{\mathbb{S}^{n-1}} Q_M(x) d\sigma(x) \in \Sigma[x]_r,$$

which defines a probability density function on $\mathbb{S}^{n-1}$, i.e., $\int_{\mathbb{S}^{n-1}} h(x) d\sigma(x) = 1$. In this way h provides a feasible solution for the program defining the upper bound $\overline{f}^{(r)}$. This thus implies the chain of inequalities

$$\langle Z_{f_r}, M\rangle = \underline{f}^{(r)} \leq f_{\min} \leq \overline{f}^{(r)} \leq \int_{\mathbb{S}^{n-1}} f(x)h(x)d\sigma(x).$$

The main contribution in [4] is their analysis for bounding the range between the two extreme values in the above chain and showing Theorem 2, which is done by using, in particular, Fourier analysis on the unit sphere.

Using different techniques we will show below a rate of convergence in $O(1/r^2)$ for the upper bounds $\overline{f}^{(r)}$, thus stronger than the rate $O(1/r)$ in Theorem 2 above and applying to any polynomial (not necessarily homogeneous). On the other hand, while the constant involved in Theorem 2 depends only on the degree of f and the dimension n, the constant in our result depends also on other characteristics of f (its first and second order derivatives). A key ingredient in our analysis will be to reduce to the univariate case, namely to the optimization of a linear polynomial over the interval $[-1, 1]$. Thus we next recall the relevant known results that we will need in our treatment.

2.2 Convergence analysis for the interval $[-1, 1]$

We start with recalling the following eigenvalue reformulation for the bound (3), which holds for general K compact and plays a key role in the analysis for the case $K = [-1, 1]$. For this consider the following inner product

$$(f, g) \mapsto \int_K f(x)g(x)d\mu(x)$$

on the space of polynomials on K and let $\{b_\alpha(x) : \alpha \in \mathbb{N}^n\}$ denote a basis of this polynomial space that is orthonormal with respect to the above inner product; that is, $\int_K b_\alpha(x) b_\beta(x) d\mu(x) = \delta_{\alpha,\beta}$. Then the bound (2) can be equivalently rewritten as

$$\overline{f}^{(r)} = \lambda_{\min}(A_f), \quad \text{where } A_f = \left(\int_K f(x) b_\alpha(x) b_\beta(x) d\mu(x) \right)_{\substack{\alpha,\beta \in \mathbb{N}^n \\ |\alpha|,|\beta| \le r}} \tag{12}$$

(see [16, 7]). Using this reformulation we could show in [7] that the bounds (3) have a convergence rate in $O(1/r^2)$ for the case of the interval $K = [-1, 1]$ (and as an application also for the n-dimensional box $[-1, 1]^n$).

This result holds for a large class of measures on $[-1, 1]$, namely those which admit a weight function $w(x) = (1 - x)^a (1 + x)^b$ (with $a, b > -1$) with respect to the Lebesgue measure. The corresponding orthogonal polynomials are known as the Jacobi polynomials $P_d^{a,b}(x)$ where $d \ge 0$ is their degree. The case $a = b = -1/2$ (resp., $a = b = 0$) corresponds to the Chebychev polynomials (resp., the Legendre polynomials), and when $a = b = \lambda - 1/2$, the corresponding polynomials are the Gegenbauer polynomials $C_d^\lambda(x)$ where d is their degree. See, e.g., [6, Chapter 1] for a general reference about orthogonal polynomials.

The key fact is that, in the case of the univariate polynomial $f(x) = x$, the matrix A_f in (12) has a tri-diagonal shape, which follows from the 3-term recurrence relationship satisfied by the orthogonal polynomials. In fact, A_f coincides with the so-called Jacobi matrix of the orthogonal polynomials in the theory of orthogonal polynomials and its eigenvalues are given by the roots of the degree $r + 1$ orthogonal polynomial (see, e.g. [6, Chapter 1]). This fact is key to the following result.

Theorem 3. *[7] Consider the measure $d\mu(x) = (1 - x)^a (1 + x)^b dx$ on the interval $[-1, 1]$, where $a, b > -1$. For the univariate polynomial $f(x) = x$, the parameter $\overline{f}^{(r)}$ is equal to the smallest root of the Jacobi polynomial $P_{r+1}^{a,b}$ (with degree $r + 1$). In particular, $\overline{f}^{(r)} = -\cos\left(\frac{\pi}{2r+2}\right)$ when $a = b = -1/2$. For any $a, b > -1$ we have*

$$\overline{f}^{(r)} - f_{\min} = \overline{f}^{(r)} + 1 = \Theta\Big(\frac{1}{r^2}\Big).$$

3 Convergence analysis for the unit sphere

In this section we analyze the quality of the bounds $\overline{f}^{(r)}$ when minimizing a polynomial f over the unit sphere $\mathbb{S}^{n-1}$. In Section 3.1 we show that the range $\overline{f}^{(r)} - f_{\min}$ is in $O(1/r^2)$ and in Section 3.2 we show that the analysis is tight for linear polynomials.

3.1 The bound $O(1/r^2)$

We first deal with the n-variate linear (coordinate) polynomial $f(x) = x_1$ and after that we will indicate how the general case can be reduced to this special case. The key idea is to get back to the analysis in Section 2.2, for the interval $[-1, 1]$ with an appropriate weight function. We begin with introducing some notation we need.

To simplify notation we set $d = n - 1$ (which also matches the notation customary in the theory of orthogonal polynomials where d usually is the number of variables). We let $\mathbb{B}^d = \{x \in \mathbb{R}^d : \|x\| \le 1\}$ denote the unit ball in $\mathbb{R}^d$, where $\|x\|^2 = \sum_{i=1}^d x_i^2$ for $x \in \mathbb{R}^d$. Given a scalar $\lambda > -1/2$, define the d-variate weight function

$$w_{d,\lambda}(x) = (1 - \|x\|^2)^{\lambda - 1/2} \tag{13}$$

(well-defined when $\|x\| < 1$) and set

$$C_{d,\lambda} := \int_{\mathbb{B}^d} w_{d,\lambda}(x_1, \ldots, x_d) dx_1 \cdots dx_d = \frac{\pi^{d/2} \Gamma\Big(\lambda + \frac{1}{2}\Big)}{\Gamma\Big(\lambda + \frac{d+1}{2}\Big)} \tag{14}$$

so that $C_{d,\lambda}^{-1} w_{d,\lambda}(x_1, \ldots, x_d) dx_1 \cdots dx_d$ is a probability measure over the unit ball $\mathbb{B}^d$. See, e.g., [6, Section 2.3.2] or [2, Section 11].

We will use the following simple lemma, which indicates how to integrate the d-variate weight function $w_{d,\lambda}$ along $d - 1$ variables.

Lemma 1. *Fix $x_1 \in [-1,1]$ and let $d \geq 2$. Then we have:*

$$\int_{\{(x_2,\ldots,x_d):x_2^2+\ldots+x_d^2\leq 1-x_1^2\}} w_{d,\lambda}(x_1,\ldots,x_d)dx_2\cdots dx_d = C_{d-1,\lambda}(1-x_1^2)^{\lambda+\frac{d-2}{2}},$$

which is thus equal to $C_{d-1,\lambda}w_{1,\lambda+(d-1)/2}(x_1)$.

Proof. Change variables and set $u_j = x_j/\sqrt{1-x_1^2}$ for $2 \leq j \leq d$. Then we have $w_{d,\lambda}(x) = (1-x_1^2-x_2^2+\ldots-x_d^2)^{\lambda-\frac{1}{2}} = (1-x_1^2)^{\lambda-\frac{1}{2}}(1-u_2^2-\ldots-u_d^2)^{\lambda-\frac{1}{2}}$ and $dx_2\cdots dx_d = (1-x_1^2)^{\frac{d-1}{2}}du_2\cdots du_d$. Putting things together and using relation (14) we obtain the desired result. □

We also need the following lemma, which relates integration over the unit sphere $\mathbb{S}^d \subseteq \mathbb{R}^{d+1}$ and integration over the unit ball $\mathbb{B}^d \subseteq \mathbb{R}^d$ and can be found, e.g., in [6, Lemma 3.8.1] and [2, Lemma 11.7.1].

Lemma 2. *Let g be a $(d+1)$-variate integrable function defined on S^d and $d \geq 1$. Then we have:*

$$\int_{\mathbb{S}^d} g(x)d\sigma(x) = \int_{\mathbb{B}^d}\left(g(x,\sqrt{1-\|x\|^2})+g(x,-\sqrt{1-\|x\|^2})\right)\frac{dx_1\cdots dx_d}{\sqrt{1-\|x\|^2}}.$$

By combining these two lemmas we obtain the following result.

Lemma 3. *Let $g(x_1)$ be a univariate polynomial and $d \geq 1$. Then we have:*

$$\sigma_d^{-1}\int_{\mathbb{S}^d} g(x_1)d\sigma(x_1,\ldots,x_{d+1}) = C_{1,\nu}^{-1}\int_{-1}^{1} g(x_1)w_{1,\nu}(x_1)dx_1,$$

where we set $\nu = \frac{d-1}{2}$.

Proof. Applying Lemma 2 to the function $x \in \mathbb{R}^{d+1} \mapsto g(x_1)$ we get

$$\sigma_d^{-1}\int_{\mathbb{S}^d} g(x_1)d\sigma(x_1,\ldots,x_{d+1}) = 2\sigma_d^{-1}\int_{\mathbb{B}^d} g(x_1)w_{d,0}(x)dx_1\cdots dx_d. \tag{15}$$

If $d=1$ then $\nu=0$ and the right hand side term in (15) is equal to

$$2\sigma_1^{-1}\int_{-1}^{1} g(x_1)w_{1,0}(x_1)dx_1 = C_{1,0}^{-1}\int_{-1}^{1} g(x_1)w_{1,0}(x_1)dx_1,$$

as desired, since $2\sigma_1^{-1}C_{1,0} = 1$ using $\sigma_1 = 2\pi$ and $C_{1,0} = \pi$ (by (14) and $\Gamma(1/2) = \sqrt{\pi}$). Assume now $d \geq 2$. Then the right hand side in (15) is equal to

$$2\sigma_d^{-1}\int_{-1}^{1} g(x_1)\left(\int_{x_2^2+\ldots+x_d^2\leq 1-x_1^2} w_{d,0}(x_1,\ldots,x_d)dx_2\cdots dx_d\right)dx_1$$

$$= 2\sigma_d^{-1}C_{d-1,0}\int_{-1}^{1} g(x_1)(1-x_1^2)^{(d-2)/2}dx_1 = 2\sigma_d^{-1}C_{d-1,0}\int_{-1}^{1} g(x_1)w_{1,\nu}(x_1)dx_1,$$

where we have used Lemma 1 for the first equality. Finally we verify that the constant $2\sigma_d^{-1}C_{d-1,0}C_{1,\nu}$ is equal to 1:

$$2\sigma_d^{-1}C_{d-1,0}C_{1,\nu} = 2\frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{\frac{d+1}{2}}}\frac{\pi^{\frac{d-1}{2}}\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)} = 1$$

(using relations (4) and (14)), and thus we arrive at the desired identity. □

We can now complete the convergence analysis for the minimization of x_1 on the unit sphere.

Lemma 4. *For the minimization of the polynomial $f(x) = x_1$ over $\mathbb{S}^d$ with $d \geq 1$, the order r upper bound (3) satisfies*

$$\overline{f}^{(r)} = -1 + O\left(\frac{1}{r^2}\right).$$

Proof. Let $h(x_1)$ be an optimal univariate sum-of-squares polynomial of degree $2r$ for the order r upper bound corresponding to the minimization of x_1 over $[-1,1]$, when using as reference measure on $[-1,1]$ the measure with weight function $w_{1,\nu}(x_1)C_{1,\nu}^{-1}$ and $\nu = (d-1)/2$ (thus $\nu > -1$). Applying Lemma 3 to the univariate polynomials $h(x_1)$ and $x_1 h(x_1)$, we obtain

$$\sigma_d^{-1}\int_{\mathbb{S}^d} h(x_1)d\sigma(x) = C_{1,\nu}^{-1}\int_{-1}^{1} h(x_1)w_{1,\nu}(x_1)dx_1 = 1$$

and

$$\overline{f}^{(r)} \leq \sigma_d^{-1}\int_{\mathbb{S}^d} x_1 h(x_1)d\sigma(x) = C_{1,\nu}^{-1}\int_{-1}^{1} x_1 h(x_1)w_{1,\nu}(x_1)dx_1.$$

Since the function x_1 has the same global minimum -1 over $[-1,1]$ and over the sphere $\mathbb{S}^d$, we can apply Theorem 3 to conclude that

$$\overline{f}^{(r)} + 1 \leq 1 + C_{1,\nu}^{-1}\int_{-1}^{1} x_1 h(x_1)w_{1,\nu}(x_1)dx_1 = O\Big(\frac{1}{r^2}\Big).$$

□

We now indicate how the analysis for an arbitrary polynomial f reduces to the case of the linear coordinate polynomial x_1. To see this, suppose $a \in \mathbb{S}^{n-1}$ is a global minimizer of f over $\mathbb{S}^{n-1}$. Then, using Taylor's theorem, we can upper estimate f as follows:

$$\begin{array}{rcll} f(x) & \leq & f(a) + \nabla f(a)^T(x-a) + \frac{1}{2}C_f\|x-a\|^2 & \forall x \in \mathbb{S}^{n-1} \\ & = & f(a) + \nabla f(a)^T(x-a) + C_f(1-a^Tx) =: g(x) & \forall x \in \mathbb{S}^{n-1}, \end{array}$$

setting $C_f = \max_{x\in\mathbb{S}^{n-1}} \|\nabla^2 f(x)\|_2$. Note that the upper estimate $g(x)$ is a linear polynomial, which has the same minimum value as $f(x)$ on $\mathbb{S}^{n-1}$, namely $f(a) = f_{\min} = g_{\min}$. From this it follows that $\overline{f}^{(r)} - f_{\min} \leq \overline{g}^{(r)} - g_{\min}$ and thus we may restrict to analyzing the bounds for a linear polynomial.

Next, assume f is a linear polynomial, of the form $f(x) = c^Tx$ with (up to scaling) $\|c\| = 1$. We can then apply a change of variables to bring $f(x)$ into the form x_1. Namely, let U be an orthogonal $n \times n$ matrix such that $Uc = e_1$. Then the polynomial $g(x) := f(U^Tx) = x_1$ has the desired form and it has the same minimum value -1 over $\mathbb{S}^{n-1}$ as $f(x)$. As the sphere is invariant under any orthogonal transformation it follows that $\overline{f}^{(r)} = \overline{g}^{(r)} = -1 + O(1/r^2)$ (applying Lemma 4 to $g(x) = x_1$). Summarizing, we have shown the following.

Theorem 4. *For the minimization of any polynomial $f(x)$ over $\mathbb{S}^{n-1}$ with $n \geq 2$, the order r upper bound (3) satisfies*

$$\overline{f}^{(r)} - f_{\min} = O\left(\frac{1}{r^2}\right).$$

Note the difference to Theorem 2 where the constant depends only on the degree of f and the number n of variables; here the constant in $O(1/r^2)$ does also depend on the polynomial f, namely it depends on the norm of $\nabla f(a)$ at a global minimizer a of f in $\mathbb{S}^{n-1}$ and on $C_f = \max_{x\in\mathbb{S}^{n-1}} \|\nabla^2 f(x)\|_2$.

3.2 The analysis is tight for linear polynomials

In this section we show — through an example — that the convergence rate cannot be better than $\Omega\left(1/r^2\right)$. The example is simply minimizing x_1 over the sphere $\mathbb{S}^{n-1}$. The key tool we use is a link between the bounds $\overline{f}^{(r)}$ and properties of some known cubature rules on the unit sphere. This connection, recently mentioned in [18], holds for any compact set K. It goes as follows.

Suppose the points $x^{(1)}, \ldots, x^{(N)} \in K$ and the weights $w_1, \ldots, w_N > 0$ provide a (positive) cubature rule for K for a given measure μ, which is exact up to degree $d + 2r$, that is,

$$\int_K g(x)d\mu(x) = \sum_{i=1}^{N} w_i g(x^{(i)})$$

for all polynomials g with degree at most $d + 2r$. Then, for any polynomial f with degree at most d, we have

$$\overline{f}^{(r)} \geq \min_{i=1}^{N} f(x^{(i)}). \tag{16}$$

The argument is simple: if $h \in \Sigma[x]_r$ is an optimal sum-of-squares density for the parameter $\overline{f}^{(r)}$, then we have

$$1 = \int_K h(x)d\mu(x) = \sum_{i=1}^N w_i h(x^{(i)}),$$

$$\overline{f}^{(r)} = \int_K f(x)h(x)d\mu(x) = \sum_{i=1}^N w_i f(x^{(i)})h(x^{(i)}) \geq \min_i f(x^{(i)}).$$

As a warm-up we consider the case $n = 2$, where we can use the cubature rule in Theorem 5 below for the unit circle. We use spherical coordinates $(x_1, x_2) = (\cos\theta, \sin\theta)$ to express a polynomial f in x_1, x_2 as a polynomial g in $\cos\theta, \sin\theta$.

Theorem 5. [2, Proposition 6.5.1] *For each $d \in \mathbb{N}$, the cubature formula*

$$\frac{1}{2\pi}\int_0^{2\pi} g(\theta)d\theta = \frac{1}{d}\sum_{j=0}^{d-1} g\left(\frac{2\pi j}{d}\right)$$

is exact for all $g \in$ span$\{1, \cos\theta, \sin\theta, \ldots, \cos(d\theta), \sin(d\theta)\}$, i.e. for all polynomials of degree at most d, restricted to the unit circle.

Using this cubature rule on $\mathbb{S}^1$ we can lower bound the parameters $\overline{f}^{(r)}$ for the minimization of $f(x) = x_1$ over $\mathbb{S}^1$. Namely, by setting $x_1 = \cos\theta$, we derive directly from the above theorem combined with relation (16) that

$$\overline{f}^{(r)} \geq \min_{0\leq j\leq 2r} \cos\left(\frac{2\pi j}{2r+1}\right) = \cos\left(\frac{2\pi r}{2r+1}\right) = -1 + \Omega\Big(\frac{1}{r^2}\Big).$$

This reasoning extends to any dimension $n \geq 2$, by using product-type cubature formulas on the sphere $\mathbb{S}^{n-1}$. In particular we will use the cubature rule described in [2, Theorem 6.2.3], see Theorem 7 below.

We will need the generalized spherical coordinates given by

$$\left.\begin{array}{rcl} x_1 & = & r\sin\theta_{n-1}\cdots\sin\theta_3\sin\theta_2\sin\theta_1 \\ x_2 & = & r\sin\theta_{n-1}\cdots\sin\theta_3\sin\theta_2\cos\theta_1 \\ x_3 & = & r\sin\theta_{n-1}\cdots\sin\theta_3\cos\theta_2 \\ & \vdots & \\ x_n & = & r\cos\theta_{n-1}, \end{array}\right\} \tag{17}$$

where $r \geq 0$ ($r = 1$ on $\mathbb{S}^{n-1}$), $0 \leq \theta_1 \leq 2\pi$, and $0 \leq \theta_i \leq \pi$ ($i = 2, \ldots, n-1$).

To define the nodes of the cubature rule on $\mathbb{S}^{n-1}$ we need the Gegenbauer polynomials $C_d^\lambda(x)$, where $\lambda > -1/2$. Recall that these are the orthogonal polynomials with respect to the weight function

$$w_{1,\lambda}(x) = (1-x^2)^{\lambda-1/2} \quad x \in (-1,1)$$

on $[-1, 1]$. We will not need the explicit expressions for the polynomials $C_d^\lambda(x)$, we only need the following information about their extremal roots, shown in [7] (for general Jacobi polynomials, using results of [3, 5]). It is well known that each $C_d^\lambda(x)$ has d distinct roots, lying in $(-1, 1)$.

Theorem 6. *Denote the roots of the polynomial $C_d^\lambda(x)$ by $t_{1,d}^{(\lambda)} < \ldots < t_{d,d}^{(\lambda)}$. Then, $t_{1,d}^{(\lambda)} + 1 = \Theta(1/d^2)$.*

The cubature rule we will use may now be stated.

Theorem 7. [2, Theorem 6.2.3] *Let $f : \mathbb{S}^{n-1} \to \mathbb{R}$ be a polynomial of degree at most $2d-1$, and let*

$$g(\theta_1, \ldots, \theta_{n-1}) := f(x_1, \ldots, x_n),$$

be the expression of f in the generalized spherical coordinates (17)*. Then*

$$\int_{\mathbb{S}^{n-1}} f(x)d\sigma(x) = \frac{\pi}{d}\sum_{k=0}^{2d-1}\sum_{j_2=1}^{d}\cdots\sum_{j_{n-1}=1}^{d}\prod_{i=2}^{n-1}\mu_{i,d}^{((i-1)/2)} g\left(\frac{\pi k}{d}, \theta_{j_2,d}^{(1/2)}, \ldots, \theta_{j_{n-1},d}^{((n-2)/2)}\right), \tag{18}$$

where $\cos\left(\theta_{j,d}^{(\lambda)}\right) := t_{j,d}^{(\lambda)}$ and the parameters $\mu_{i,d}^{((i-1)/2)}$ are positive scalars as in relation (6.2.3) of [2].

We can now show the tightness of the convergence rate $\Omega(1/r^2)$ for the minimization of a coordinate polynomial on $\mathbb{S}^{n-1}$.

Theorem 8. *Consider the problem of minimizing the coordinate polynomial x_n on the unit sphere $\mathbb{S}^{n-1}$ with $n \geq 2$. The convergence rate for the parameters (3) satisfies*

$$\overline{f}^{(d)} - f_{\min} = \overline{f}^{(d)} + 1 = \Omega\left(\frac{1}{d^2}\right).$$

Proof. We have $f(x_1, \ldots, x_n) = x_n$, so that $g(\theta_1, \ldots, \theta_{n-1}) = \cos\theta_{n-1}$. Using (16) we obtain that

$$\overline{f}^{(r)} \geq \min_{1\leq j\leq d} \cos\theta_{j,d}^{((n-2)/2)} = \min_{1\leq j\leq d} t_{j,d}^{((n-2)/2)} = t_{1,d}^{((n-2)/2)} = -1 + \Omega\Big(\frac{1}{d^2}\Big),$$

where we use the fact that $t_{1,d}^{(\lambda)} + 1 = \Theta(1/d^2)$ (Theorem 6).

□

4 Implications for the generalized problem of moments

In this section, we describe the implications of our results for the generalized problem of moments (GPM), defined as follows for a compact set $K \subset \mathbb{R}^n$.

$$val := \inf_{\nu\in\mathcal{M}(K)_+} \left\{ \int_K f_0(x)d\nu(x) \ : \ \int_K f_i(x)d\nu(x) = b_i \quad \forall i \in [m] \right\}, \tag{19}$$

where

- the functions f_i $(i = 0, \ldots, m)$ are continuous on K;
- $\mathcal{M}(K)_+$ denotes the convex cone of probability measures supported on the set K;
- the scalars $b_i \in \mathbb{R}$ $(i \in [m])$ are given.

As before, we are interested in the special case where $K = \mathbb{S}^{n-1}$. This special case is already of independent interest, since it contains the problem of finding cubature schemes for numerical integration on the sphere, see e.g. [10] and the references therein. Our main result in Theorem 4 has the following implication for the GPM on the sphere, as a corollary of the following result in [13] (which applies to any compact K, see also [10] for a sketch of the proof in the setting described here).

Theorem 9 (De Klerk-Postek-Kuhn [13])**.** *Assume that $f_0, \ldots, f_m$ are polynomials, K is compact, μ is a Borel measure supported on K, and the GPM* (19) *has an optimal solution. Given $r \in \mathbb{N}$, define the parameter*

$$\Delta(r) = \min_{h\in\Sigma_r} \max_{i\in\{0,1,\ldots,m\}} \Big| \int_K f_i(x)h(x)d\mu(x) - b_i \Big|,$$

setting $b_0 = val$. If, for any polynomial f, we have

$$\overline{f}_K^{(r)} - f_{\min} = O(\varepsilon(r)),$$

where $\lim_{r\to\infty} \varepsilon(r) = 0$, then the parameters $\Delta(r)$ satisfy: $\Delta(r) = O(\sqrt{\varepsilon(r)})$.

As a consequence of our main result in Theorem 4, combined with Theorem 10, we immediately obtain the following corollary.

Corollary 1. *Assume that $f_0, \ldots, f_m$ are polynomials, $K = \mathbb{S}^{n-1}$, and the GPM* (19) *has an optimal solution. Then, for any integer $r \in \mathbb{N}$, there is an $h_r \in \Sigma_r$ such that*

$$\left| \int_{\mathbb{S}^{n-1}} f_0(x)h_r(x)d\sigma(x) - val \right| = O(1/r), \ \left| \int_{\mathbb{S}^{n-1}} f_i(x)h_r(x)d\sigma(x) - b_i \right| = O(1/r) \quad \forall i \in [m].$$

Minimization of a rational function on K is a special case of the GPM where we may prove a better rate of convergence. In particular, we now consider the global optimization problem:

$$val = \min_{x\in K} \frac{p(x)}{q(x)}, \tag{20}$$

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

It is well-known that one may reformulate this problem as the GPM with $m = 1$ and $f_0 = p$, $f_1 = q$, and $b_1 = 1$, i.e.:

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

Analogously to (3), we now define the hierarchy of upper bounds on val as follows:

$$\overline{p/q}_K^{(r)} := \min_{h \in \Sigma[x]_r} \int_K p(x)h(x)d\mu(x) \;\; \text{s.t.} \int_K q(x)h(x)d\mu(x) = 1, \tag{21}$$

where μ is a Borel measure supported on K.

Theorem 10. *Consider the rational optimization problem* (20)*. If, for any polynomial f, it holds that*

$$\overline{f}_K^{(r)} - f_{\min} = O(\varepsilon(r))$$

where $\lim_{r \to \infty} \varepsilon(r) = 0$*, then one also has* $\overline{p/q}_K^{(r)} - val = O(\varepsilon(r))$*. In particular, if* $K = \mathbb{S}^{n-1}$*, then* $\overline{p/q}_K^{(r)} - val = O(1/r^2)$*.*

Proof. Consider the polynomial

$$f(x) = p(x) - val \cdot q(x).$$

Then $f(x) \geq 0$ for all $x \in K$, and $f_{\min,K} = 0$, with global minimizer given by the minimizer of problem (20).

Now, for given $r \in \mathbb{N}$, let $h \in \Sigma_r$ be such that $\overline{f}_K^{(r)} = \int_K f(x)h(x)d\mu(x)$, and $\int_K h(x)d\mu(x) = 1$, where μ is the reference measure for K. Setting

$$h^* = \frac{1}{\int_K h(x)q(x)d\mu(x)} h,$$

one has $h^* \in \Sigma_r$ and $\int_K h^*(x)q(x)d\mu(x) = 1$. Thus h^* is feasible for problem (21). Moreover, by construction,

$$\begin{aligned} \int_K p(x)h^*(x)d\mu(x) - val &= \frac{\overline{f}_K^{(r)}}{\int_K h(x)q(x)d\mu(x)} \\ &\leq \frac{\overline{f}_K^{(r)}}{\min_{x \in K} q(x)} = O(\varepsilon(r)). \end{aligned}$$

The final result for the special case $K = \mathbb{S}^{n-1}$ and $\mu = \sigma$ (surface measure) now follows from our main result in Theorem 4. □

5 Concluding remarks

In this paper we have improved on the $O(1/r)$ convergence result of Doherty and Wehner [4] for the Lasserre hierarchy of upper bounds (3) for (homogeneous) polynomial optimization on the sphere. Having said that, Doherty and Wehner also showed that the hierarchy of *lower* bounds (9) of Lasserre satisfies the same rate of convergence, due to Theorem 2. In view of the fact that we could show the improved $O(1/r^2)$ rate for the upper bounds, and the fact that the lower bounds hierarchy empirically converges much faster in practice, one would expect that the lower bounds (9) also converge at a rate no worse than $O(1/r^2)$. However, our analysis does not allow us to analyse the convergence of the lower bound hierarchy, and this remains an interesting open problem.

Another open problem is the exact rate of convergence of the bounds in Theorem 10 for the generalized problem of moments (GPM). In our analysis of the GPM on the sphere in Corollary 1, we could only obtain $O(1/r)$ convergence, which is a square root worse than the special cases for polynomial and rational function minimization. We do not know at the moment if this is a weakness of the analysis or inherent to the GPM.

Note that if we pick another reference measure $d\mu(x) = q(x)d\sigma(x)$, where q is strictly positive on the sphere, then the convergences rates with respect to both measures σ and μ have the same behaviour (up to multiplicative constant). It would be interesting to understand the convergence rate for more general reference measures.

Acknowledgement

This work has been supported by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 813211 (POEMA).

References

[1] A. A. Ahmadi, A. Olshevsky, P. A. Parrilo, and J. N. Tsitsiklis, *NP-hardness of deciding convexity of quartic polynomials and related problems*, Mathematical Programming 137 (2013), no. 1-2, 453–476.

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

[3] D.K. Dimitrov, G.P. Nikolov. Sharp bounds for the extreme zeros of classical orthogonal polynomials, Journal of Approximation Theory 162 (2010), 1793–1804.

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

[5] K. Driver, K. Jordaan. Bounds for extreme zeros of some classical orthogonal polynomials. Journal of Approximation Theory 164 (2012), 1200–1204.

[6] C.F. Dunkl and Y. Xu. Orthogonal Polynomials of Several Variables. Encyclopedia of Mathematics, Cambridge University Press (2001).

[7] E. de Klerk and M. Laurent. Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing. arXiv:1703.00744, to appear in *Mathematics of Operations Research*.

[8] E. de Klerk, R. Hess and M. Laurent. Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization. *SIAM Journal on Optimization* 27 (2017), no. 1, 347–367.

[9] E. de Klerk, M. Laurent. Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube. arXiv:1804.05524, to appear in *Mathematics of Operations Research*.

[10] E. de Klerk, M. Laurent. A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. arXiv:1811.05439 (2018).

[11] 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, Springer (2005), 121–133.

[12] E. de Klerk, M. Laurent, Z. Sun. Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization, *Mathematical Programming Ser. A* 162 (2017), no. 1, 363-392.

[13] De Klerk, E., Postek, K., and Kuhn, D. Distributionally robust optimization with polynomial densities: theory, models and algorithms. arXiv:1805.03588 (2018).

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

[15] Lasserre, J.B. Moments, Positive Polynomials and Their Applications. Imperial College Press (2009).

[16] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization* 21 (2011), no. 3, 864–885.

[17] M. Laurent. 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, (2009), 157–270.

[18] A. Martinez, F. Piazzon, A. Sommariva, and M. Vianello. Quadrature-based polynomial optimization. Manuscript (2018). http://www.math.unipd.it/ marcov/pdf/quadropt.pdf

[19] T.S. Motzkin, E.G. Straus. Maxima for graphs and a newproof of a theorem of Túran. Canadian J. Math. 17(1965), 533–540.

[20] Nesterov, Yu. Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper 2003/71, CORE-UCL, Louvain-La-Neuve (2003).

[21] P.A. Parrilo. *Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, Pasadena, California, USA (2000).

[22] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Ser. B* 96 (2003), no. 2, 293–320.

[23] Reznick, B. Some concrete aspects of Hilbert's 17th Problem. In *Real algebraic geometry and ordered structures (Baton Rouge, LA, 1996)*, Amer. Math. Soc., Providence, RI (2000), 251–272.