Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube

Etienne de Klerk * Monique Laurent †

April 17, 2018

Abstract

We study the convergence rate of a hierarchy of upper bounds for polynomial optimization problems, proposed by Lasserre [*SIAM J. Optim.* 21(3) (2011), pp. 864 − 885], and a related hierarchy by De Klerk, Hess and Laurent [*SIAM J. Optim.* 27(1), (2017) pp. 347 − 367]. For polynomial optimization over the hypercube we show a refined convergence analysis for the first hierarchy. We also show lower bounds on the convergence rate for both hierarchies on a class of examples. These lower bounds match the upper bounds and thus establish the true rate of convergence on these examples. Interestingly, these convergence rates are determined by the distribution of extremal zeroes of certain families of orthogonal polynomials.

Keywords: Polynomial optimization; Semidefinite optimization; Lasserre hierarchy; extremal roots of orthogonal polynomials; Jacobi polynomials
AMS classification: 90C22; 90C26; 90C30

1 Introduction

We consider the problem of minimizing a polynomial $f : \mathbb{R}^n \to \mathbb{R}$ over a compact set $\mathbf{K} \subseteq \mathbb{R}^n$. That is, we consider the problem of computing the parameter:

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

We recall the following reformulation for $f_{\min,\mathbf{K}}$, established by Lasserre [12]:

$$f_{\min,\mathbf{K}} = \inf_{\sigma\in\Sigma[x]} \int_{\mathbf{K}} \sigma(x)f(x)d\mu(x) \quad \text{s.t.} \ \textstyle\int_{\mathbf{K}} \sigma(x)d\mu(x) = 1,$$

where $\Sigma[x]$ denotes the set of sums of squares of polynomials, and μ is a signed Borel measure supported on $\mathbf{K}$. Given an integer $d \in \mathbb{N}$, by bounding the degree of the polynomial $\sigma \in \Sigma[x]$ by $2d$, Lasserre [12] defined the parameter:

*Tilburg University and Delft University of Technology, E.deKlerk@uvt.nl
†Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University, monique@cwi.nl

$$\underline{f}_{\mathbf{K}}^{(d)} := \inf_{\sigma\in\Sigma[x]_d} \int_{\mathbf{K}} \sigma(x)f(x)d\mu(x) \quad \text{s.t. } \int_{\mathbf{K}} \sigma(x)d\mu(x) = 1, \tag{1}$$

where $\Sigma[x]_d$ consists of the polynomials in $\Sigma[x]$ with degree at most $2d$.

The inequality $f_{\min,\mathbf{K}} \le \underline{f}_{\mathbf{K}}^{(d)}$ holds for all $d \in \mathbb{N}$ and, in view of the identity (1), it follows that the sequence $\underline{f}_{\mathbf{K}}^{(d)}$ converges to $f_{\min,\mathbf{K}}$ as $d \to \infty$. De Klerk and Laurent [2] established the following rate of convergence for the sequence $\underline{f}_{\mathbf{K}}^{(d)}$, when μ is the Lebesgue measure and $\mathbf{K}$ is a convex body.

Theorem 1.1. *[2] Let $f \in \mathbb{R}[x]$, $\mathbf{K}$ a convex body, and μ the Lebesgue measure on $\mathbf{K}$. There exist constants $C_{f,\mathbf{K}}$ (depending only on f and $\mathbf{K}$) and $d_{\mathbf{K}} \in \mathbb{N}$ (depending only on $\mathbf{K}$) such that*

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} \le \frac{C_{f,\mathbf{K}}}{d} \quad \textit{for all } d \ge d_{\mathbf{K}}. \tag{2}$$

That is, the following asymptotic convergence rate holds: $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = O\left(\frac{1}{d}\right)$.

This result was an improvement on an earlier result by De Klerk, Laurent and Sun [5, Theorem 3], who showed a convergence rate in $O(1/\sqrt{d})$ (for $\mathbf{K}$ convex body or, more generallly, compact under a mild assumption).

As explained in [12] the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ can be computed using semidefinite programming, assuming one knows the (generalised) moments of the measure μ on $\mathbf{K}$ with respect to some polynomial basis. Set

$$m_\alpha(\mathbf{K}) := \int_{\mathbf{K}} b_\alpha(x)d\mu(x),\ m_{\alpha,\beta}(\mathbf{K}) := \int_{\mathbf{K}} b_\alpha(x)b_\beta(x)d\mu(x) \quad \text{for } \alpha,\beta \in \mathbb{N}^n,$$

where the polynomials $\{b_\alpha\}$ form a basis for the space $\mathbb{R}[x_1,\ldots,x_n]_{2d}$ of polynomials of degree at most $2d$, indexed by $N(n,2d) = \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \le 2d\}$. For example, the standard monomial basis in $\mathbb{R}[x_1,\ldots,x_n]_{2d}$ is $b_\alpha(x) = x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$ for $\alpha \in N(n,2d)$, and then $m_{\alpha,\beta}(\mathbf{K}) = m_{\alpha+\beta}(\mathbf{K})$. If $f(x) = \sum_{\beta\in N(n,d_0)} f_\beta b_\beta(x)$ has degree d_0, and writing $\sigma \in \Sigma[x]_d$ as $\sigma(x) = \sum_{\alpha\in N(n,2d)} \sigma_\alpha b_\alpha(x)$, then the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ in (1) can be computed as follows:

$$\begin{aligned} \underline{f}_{\mathbf{K}}^{(d)} \quad = \quad & \min \sum_{\beta\in N(n,d_0)} f_\beta \sum_{\alpha\in N(n,2d)} \sigma_\alpha m_{\alpha,\beta}(\mathbf{K}) \\ & \text{s.t.} \quad \sum_{\alpha\in N(n,2d)} \sigma_\alpha m_\alpha(\mathbf{K}) = 1, \\ & \qquad \sum_{\alpha\in N(n,2d)} \sigma_\alpha b_\alpha(x) \in \Sigma[x]_d. \end{aligned} \tag{3}$$

Since the sum-of-squares condition on σ may be written as a linear matrix inequality, this is a semidefinite program. In fact, since the program (3) has only one linear equality constraint, using semidefinite programming duality it can be rewritten as a generalised eigenvalue problem. In particular, $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to the the smallest generalised eigenvalue of the system:

$$Ax = \lambda Bx \qquad (x \ne 0),$$

where the symmetric matrices A and B are of order $\binom{n+d}{d}$ with rows and columns indexed by $N(n,d)$, and

$$A_{\alpha,\beta} = \sum_{\delta \in N(n,d_0)} f_\delta \int_{\mathbf{K}} b_\alpha(x) b_\beta(x) b_\delta(x) d\mu(x), \quad B_{\alpha,\beta} = \int_{\mathbf{K}} b_\alpha(x) b_\beta(x) d\mu(x) \quad \text{for } \alpha, \beta \in N(n,d). \tag{4}$$

For more details, see [12, 5]. In particular, if the basis $\{b_\alpha\}$ is orthonormal with respect to the measure μ, then B is the identity matrix, and $\underline{f}_{\mathbf{K}}^{(d)}$ is the smallest eigenvalue of the above matrix A. For further reference we summarize this result, which will play a central role in our approach.

Lemma 1.2. *Assume $\{b_\alpha : \alpha \in N(n,2d)\}$ is a basis of the space $\mathbb{R}[x_1, \ldots, x_n]_{2d}$, which is orthonormal w.r.t. the measure μ on K, i.e., $\int_K b_\alpha(x) b_\beta(x) d\mu(x) = \delta_{\alpha,\beta}$. Then the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to the smallest eigenvalue of the matrix A in (4).*

Under the conditions of the lemma, note in addition that, if the vector $u = (u_\alpha)_{\alpha \in N(n,d)}$ is an eigenvector of the matrix A in (4) for its smallest eigenvalue, then the (square) polynomial $\sigma(x) = (\sum_{\alpha \in N(n,d)} u_\alpha x^\alpha)^2$ is an optimal density function for the parameter $\underline{f}_{\mathbf{K}}^{(d)}$.

Related hierarchy by De Klerk, Hess and Laurent

For the hypercube $\mathbf{K} = [-1,1]^n$, De Klerk, Hess and Laurent [3] considered a variant on the Lasserre hierarchy (1), where the density function σ is allowed to take the more general form

$$\sigma(x) = \sum_{I \subseteq \{1,\ldots,n\}} \sigma_I(x) \prod_{i \in I} (1 - x_i^2) \tag{5}$$

and the polynomials σ_I are sum-of-squares polynomials with degree at most $2d - 2|I|$ (to ensure that the degree of σ is at most $2d$), and $I = \emptyset$ is included in the summation. Moreover the measure μ is fixed to be

$$d\mu(x) = \left(\prod_{i=1}^{n} \sqrt{1 - x_i^2} \right)^{-1} dx_1 \cdots dx_n. \tag{6}$$

As we will recall below, this measure is associated with the Chebyshev orthogonal polynomials. We let $f^{(d)}$ denote the parameter[1] obtained by using in (1) these choices (5) of density functions $\sigma(x)$ and (6) of measure μ. By construction, we have

$$f_{\min,\mathbf{K}} \leq f^{(d)} \leq \underline{f}_{\mathbf{K}}^{(d)}.$$

De Klerk, Hess and Laurent [3] proved a stronger convergence rate for the bounds $f^{(d)}$.

Theorem 1.3. *[3] Let $f \in \mathbb{R}[x]$ be a polynomial and $\mathbf{K} = [-1,1]^n$. We have*

$$f^{(d)} - f_{\min,\mathbf{K}} = O\left(\frac{1}{d^2}\right).$$

[1] We drop the dependence on $\mathbf{K}$ which is implictly selected to be the box $[-1,1]^n$.

Contribution of this paper

In this paper we investigate the rate of convergence of the hierarchies $\underline{f}_{\mathbf{K}}^{(d)}$ and $f^{(d)}$ to $f_{\min,\mathbf{K}}$ for the case of the box $\mathbf{K} = [-1,1]^n$. The above discussion raises naturally the following questions:

- Is the sublinear convergence rate $f^{(d)} - f_{\min,\mathbf{K}} = O\left(\frac{1}{d^2}\right)$ tight, or can this result be improved?
- Does this convergence rate extend to the Lasserre bounds, where we restrict to sums-of-squares density functions?

We give a positive answer to both questions. Regarding the first question we show that the convergence rate is $\Omega(1/d^2)$ when f is a linear polynomial, which implies that the convergence analysis in Theorem 1.3 for the bounds $f^{(d)}$ is tight. This relies on the eigenvalue reformulation of the bounds (from Lemma 1.2) and an additional link to the extremal zeros of the associated Chebyshev polynomials. We also show that the same lower bound holds for the convergence rate of the Lasserre bounds $\underline{f}_{\mathbf{K}}^{(d)}$ when considering measures on the hypercube corresponding to general Jacobi polynomials.

Regarding the second question we show that also the Lasserre bounds have a $O(1/d^2)$ convergence rate when using the Chebyshev type measure from (6). The starting point is again the reformulation from Lemma 1.2 in terms of eigenvalues, combined with some further analytical arguments.

The paper is organised as follows. In Section 2 we group preliminary results about orthogonal polynomials and their extremal roots. Then, in Section 3.1 we analyse the convergence rate of the Lasserre bounds $\underline{f}_{\mathbf{K}}^{(d)}$ when f is a linear polynomial and, in Section 3.2, we analyse the bounds $f^{(d)}$. In both cases we show a $\Omega(1/d^2)$ lower bound. In Section 4 we show a $O(1/d^2)$ upper bound for the convergence rate of the Lasserre bounds $\underline{f}_{\mathbf{K}}^{(d)}$, and this analysis is tight in view of the previously shown lower bounds.

Notation

We recap here some notation that is used throughout. For an integer $d \in \mathbb{N}$, $\mathbb{R}[x]_d$ denotes the set of n-variate polynomials in the variables $x = (x_1, \ldots, x_n)$ with degree at most d and $\Sigma[x]_d$ denotes the set of polynomials with degree at most $2d$ that can be written as a sum of squares of polynomials.

We use the classical Landau notation. For two functions $f, g : \mathbb{N} \to \mathbb{R}_+$, the notation $f(n) = O(g(n))$ (resp., $f(n) = \Omega(g(n))$, $f(n) = o(g(n))$) means $\limsup_{n\to\infty} f(n)/g(n) < \infty$ (resp., $\liminf_{n\to\infty} f(n)/g(n) > \infty$, $\lim_{n\to\infty} f(n)/g(n) = 0$), and $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. We also use this notation when f, g are functions of a continuous variable x and we want to indicate the behavior of $f(x)$ and $g(x)$ in the neighbourhood of a given scalar x_0 when $x \to x_0$. So, $f(x) = O(g(x))$ as $x \to x_0$ means $\limsup_{x\to x_0} f(x)/g(x) < \infty$, etc.

2 Preliminaries on orthogonal polynomials

In what follows we review some known facts on classical orthogonal polynomials that we need for our treatment. Unless we give detailed references, the relevant results may be found in the classical text by Szegö [16] (see also [8]).

We consider families of univariate polynomials $\{p_k(x)\}$ $(k = 0, 1, \ldots, d)$, that satisfy a three-term recursive relation of the form:

$$xp_k(x) = a_k p_{k+1}(x) + b_k p_k(x) + c_k p_{k-1}(x) \qquad (k = 1, \ldots, d-1), \tag{7}$$

where p_0 is a constant, $p_1(x) = (x - b_0)p_0/a_0$, and a_k, b_k and c_k are real values that satisfy $a_{k-1}c_k > 0$ for $k = 1, \ldots, d-1$. If we set $c_0 = 0$ then relation (7) also holds for $k = 0$).
Defining the $k \times k$ tri-diagonal matrix

$$A_k := \begin{pmatrix} b_0 & a_0 & 0 & \cdots & 0 \\ c_1 & b_1 & a_1 & & 0 \\ 0 & \ddots & \ddots & \ddots & \\ \vdots & & c_{k-2} & b_{k-2} & a_{k-2} \\ 0 & 0 & \cdots & c_{k-1} & b_{k-1} \end{pmatrix}, \tag{8}$$

one has the classical relation:

$$\left(\prod_{j=0}^{k-1} a_j \right) p_k(x) = \det(xI_k - A_k)p_0 \quad \text{for } k = 1, \ldots, d, \tag{9}$$

which can be easily verified using induction on $k \geq 1$ and the relation (7) (see, e.g., [11]). Therefore, the roots of the polynomial p_k are precisely the eigenvalues of the matrix A_k in (8).

Recall that the polynomials p_k ($k = 0, 1, \ldots, d$) are *orthogonal with respect to a weight function* $w : [-1, 1] \to \mathbb{R}$, that is continuous and positive on $(-1, 1)$, if

$$\langle p_i, p_j \rangle := \int_{-1}^{1} p_i(x)p_j(x)w(x)dx = 0 \quad \text{for all } \ i \neq j.$$

We denote by $\hat{p}_k := p_k/\sqrt{\langle p_k, p_k \rangle}$ the corresponding normalized polynomial, so that $\langle \hat{p}_k, \hat{p}_k \rangle = 1$.

As is well known, if the polynomials p_k are degree k polynomials that are pairwise orthogonal with respect to such a weight function then they satisfy a three-terms recurrence relation of the form (7) (see, e.g., [8, §1.3]). Of course, the corresponding orthonormal polynomials $\hat{p}_k$ also satisfy such a three-terms recurrence relation (for different scaled parameters a_k, b_k, c_k).

By taking the inner product of both sides in (7) with p_{k-1} and p_{k+1} one gets the relations $c_k\langle p_{k-1}, p_{k-1} \rangle = \langle p_k, xp_{k-1} \rangle$ and $a_k\langle p_{k+1}, p_{k+1} \rangle = \langle p_{k+1}, xp_k \rangle$, which imply $c_k\langle p_{k-1}, p_{k-1} \rangle = a_{k-1}\langle p_k, p_k \rangle$ and thus $a_{k-1}c_k > 0$. Moreover, when considering the recurrence relations associated with the orthonormal polynomials $\hat{p}_k$, we have $a_{k-1} = c_k$ for any $k \geq 1$, i.e., the matrix A_k in (8) is symmetric. We will use later the following fact.

Lemma 2.1. *Let $\{\hat{p}_k\}$ be orthonormal polynomials for the measure $d\mu(x) = w(x)dx$ on $[-1, 1]$, where $w(x)$ is continuous and positive on $(-1, 1)$, and assume they satisfy the three-terms recurrence relation (7). Then, the matrix*

$$\left(\langle x\hat{p}_i, \hat{p}_j \rangle = \int_{-1}^{1} x\hat{p}_i(x)\hat{p}_j(x)w(x)dx \right)_{i,j=0}^{k-1} \tag{10}$$

is equal to the matrix A_k in (8). In particular, its smallest eigenvalue is the smallest root of the polynomial p_k.

Proof. Using the recurrence relation (7) we obtain

$$\begin{aligned}\langle x\hat{p}_i, \hat{p}_j\rangle &= \langle a_i\hat{p}_{i+1} + b_i\hat{p}_i + c_i\hat{p}_{i-1}, \hat{p}_j\rangle \\ &= \begin{cases} a_i & \text{if } j = i+1 \\ b_i & \text{if } j = i \\ c_i & \text{if } j = i-1 \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Hence the matrix in (10) is equal to A_k and the last claim follows from (9). □

It is also known that the roots of p_k are all real, simple and lie in $(-1, 1)$, and that they interlace the roots of p_{k+1} (see, e.g., [8, §1.2]). In what follows we will use the smallest (and largest) roots to give closed-form expressions for the bounds $\underline{f}_{\mathbf{K}}^{(d)}$ and $f^{(d)}$ in some examples.

We now recall several classical univariate orthogonal polynomials on the interval $[-1, 1]$ and some information on their smallest roots.

Chebyshev polynomials

We will use the univariate Chebyshev polynomials (of the first kind), defined by:

$$T_k(x) = \cos(k \arccos(x)), \quad \text{for} \quad x \in [-1, 1], \; k = 0, 1, \ldots. \tag{11}$$

They satisfy the following three-terms recurrence relationships:

$$T_0(x) = 1, \; T_1(x) = x, \; T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x) \quad \text{for} \quad k \geq 1. \tag{12}$$

The Chebyshev polynomials are orthogonal with respect to the weight function $w(x) = \frac{1}{\sqrt{1-x^2}}$ and the roots of T_k are given by

$$\cos\left(\frac{2i-1}{2k}\pi\right) \qquad \text{for } i = 1, \ldots, k. \tag{13}$$

Jacobi polynomials

The Jacobi polynomials, denoted by $\{P_k^{\alpha,\beta}\}$ $(k = 0, 1, \ldots)$, are orthogonal with respect to the weight function

$$w_{\alpha,\beta}(x) := (1-x)^\alpha (1+x)^\beta, \quad x \in (-1, 1) \tag{14}$$

where $\alpha > -1$ and $\beta > -1$ are given parameters. The normalized Jacobi polynomials are denoted by $\hat{P}_k^{\alpha,\beta}$, so that $\int_{-1}^{1} (\hat{P}_k^{\alpha,\beta}(x))^2 w_{\alpha,\beta}(x)dx = 1$.

Thus the Chebyshev polynomials may be seen as the special case corresponding to $\alpha = \beta = -\frac{1}{2}$.

Likewise, the Legendre polynomials are the orthogonal polynomials w.r.t. the constant weight function $(w(x) = 1)$, so they correspond to the special case $\alpha = \beta = 0$.

There is no closed-form expression for the roots of Jacobi polynomials in general. But some bounds are known for the smallest root of $P_k^{\alpha,\beta}$, denoted by $\xi_k^{\alpha,\beta}$, that we recall in the next theorem.

Theorem 2.2. *The smallest root, denoted $\xi_k^{\alpha,\beta}$, of the Jacobi polynomial $P_k^{\alpha,\beta}$ satisfies the following inequalities:*

(i) ([7]) $\xi_k^{\alpha,\beta} \le -1 + \frac{2(\beta+1)(\beta+3)}{2(k-1)(k+\alpha+\beta+2)+(\beta+3)(\alpha+\beta+2)}$.

(ii) ([6]) $\xi_k^{\alpha,\beta} \ge \frac{F-4(k-1)\sqrt{\Delta}}{E}$, *where*

$$\begin{aligned} F &= (\beta-\alpha)\left((\alpha+\beta+6)k+2(\alpha+\beta)\right), \\ E &= (2k+\alpha+\beta)\left(k(2k+\alpha+\beta)+2(\alpha+\beta+2)\right) \\ \Delta &= k^2(k+\alpha+\beta+1)^2+(\alpha+1)(\beta+1)(k^2+(\alpha+\beta+4)k+2(\alpha+\beta)). \end{aligned}$$

The smallest roots $\xi_k^{\alpha,\beta}$ of the Jacobi polynomials $P_k^{\alpha,\beta}$ converge to -1 as $k\to\infty$. Using the above bounds we see that the rate of convergence is $O(1/k^2)$.

Corollary 2.3. *The smallest roots of the Jacobi polynomials* $P_k^{\alpha,\beta}$ *satisfy*

$$\xi_k^{\alpha,\beta} = -1+\Theta\left(\frac{1}{k^2}\right) \quad \textit{as} \quad k\to\infty.$$

Proof. The upper bound in Theorem 2.2(i) gives directly $\xi_k^{\alpha,\beta} = -1+O\left(\frac{1}{k^2}\right)$. We now use the lower bound in Theorem 2.2(ii) to show $\xi_k^{\alpha,\beta} = -1+\Omega\left(\frac{1}{k^2}\right)$. For this we give asymptotic estimates for the quantities E, F, Δ. First, using the expansion $\sqrt{1+x} = 1+\frac{x}{2}-\frac{x^2}{8}+o(x^2)$ as $x\to 0$ we obtain

$$\sqrt{\Delta} = k^2\left(1+\frac{\alpha+\beta+1}{k}+\frac{(\alpha+1)(\beta+1)}{2k^2}+o\left(\frac{1}{k^2}\right)\right).$$

Second, using the expansion $\frac{1}{1+x} = 1-x+x^2+o(x^2)$ as $x\to 0$ we obtain

$$\frac{1}{E} = \frac{1}{4k^3}\left(1-\frac{\alpha+\beta}{k}-\frac{4(\alpha+\beta+2)}{k^2}+o\left(\frac{1}{k^2}\right)\right).$$

Combining these two relations gives

$$\begin{aligned} \frac{4(k-1)\sqrt{\Delta}}{E} &= \left(1-\tfrac{1}{k}\right)\left(1+\tfrac{\alpha+\beta+1}{k}+\tfrac{(\alpha+1)(\beta+1)}{2k^2}+o\left(\tfrac{1}{k^2}\right)\right)\left(1-\tfrac{\alpha+\beta}{k}-\tfrac{4(\alpha+\beta+2)}{k^2}+o\left(\tfrac{1}{k^2}\right)\right) \\ &= 1+\tfrac{C}{2k^2}+o\left(\tfrac{1}{k^2}\right), \end{aligned}$$

where we set $C = (\alpha+1)(\beta+1)-8(\alpha+\beta+2)-2(\alpha+\beta)(\alpha+\beta+1)-2$. Finally, using

$$\frac{F}{E} = \frac{(\beta-\alpha)(\beta+\alpha+6)}{4k^2}+o\left(\frac{1}{k^2}\right),$$

we obtain

$$\frac{F-4(k-1)\sqrt{\Delta}}{E} = -1+\frac{1}{k^2}\left(\frac{(\beta-\alpha)(\beta+\alpha+6)}{4}-\frac{C}{2}\right)+o\left(\frac{1}{k^2}\right),$$

where the coefficient of $1/k^2$ can be verified to be strictly positive, which thus implies the estimate $\xi_k^{\alpha,\beta} = -1+\Omega(1/k^2)$. □

It is also known that $P_k^{\alpha,\beta}(x) = (-1)^k P_k^{\beta,\alpha}(-x)$. Therefore the largest root of $P_k^{\alpha,\beta}(x)$ is equal to $-\xi_k^{\beta,\alpha} = 1-\Theta(1/k^2)$.

3 Tight lower bounds for a class of examples

In this section we consider the following simple examples

$$\min\left\{\sum_{i=1}^{n} c_i x_i : \quad x \in [-1,1]^n\right\}, \tag{15}$$

asking to minimize the linear polynomial $f(x) = \sum_{i=1}^{n} c_i x_i$ over the box $\mathbf{K} = [-1,1]^n$. Here $c_i \in \mathbb{R}$ are given scalars for $i \in [n]$. Hence, $f_{\min,\mathbf{K}} = -\sum_{i=1}^{n} |c_i|$. For these examples we can obtain explicit closed-form expressions for the Lasserre bounds $\underline{f}_{\mathbf{K}}^{(d)}$ when using product measures with weight functions $w_{\alpha,\beta}$ on $[-1,1]$, and also for the strengthened bounds $f^{(d)}$ considered by De Klerk, Hess and Laurent, which use product measures with weight functions $w_{-1/2,-1/2}$. These closed-form expressions are in terms of extremal roots of Jacobi polynomials.

3.1 Tight lower bound for the Lasserre hierarchy

Here we consider the bounds $\underline{f}_{\mathbf{K}}^{(d)}$ for the example (15), when the measure μ on $\mathbf{K} = [-1,1]^n$ is a product of univariate measures given by weight functions.

First we consider the univariate case $n = 1$. When the measure μ on $\mathbf{K} = [-1,1]$ is given by a continuous positive weight function w on $(-1,1)$, one can obtain a closed form expression for $\underline{f}_{\mathbf{K}}^{(d)}$ in terms of the smallest root of the corresponding orthogonal polynomials.

Theorem 3.1. *Consider the measure $d\mu(x) = w(x)dx$ on $\mathbf{K} = [-1,1]$, where w is a positive, continuous weight function on $(-1,1)$, and let p_k be univariate degree k polynomials that are orthogonal with respect to this measure. For the univariate polynomial $f(x) = x$ (resp., $f(x) = -x$), the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to the smallest root (resp., the opposite of the largest root) of the polynomial p_{d+1}.*

Proof. Let $\hat{p}_0, \ldots, \hat{p}_{d+1}$ denote the corresponding orthonormal polynomials, with $\hat{p}_i = p_i/\sqrt{\langle p_i, p_i\rangle}$. Consider first $f(x) = x$. Using Lemma 1.2, we see that $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to the smallest eigenvalue of the matrix A in (10) (for $k = d+1$), which coincides with the matrix A_{d+1} in (8), so that its smallest eigenvalue is equal to the smallest root of p_{d+1}.
Assume now $f(x) = -x$. Then $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to $\lambda_{\min}(-A) = -\lambda_{\max}(A)$, which in turn is equal to the opposite of the largest root of p_{d+1}. □

Recall that $\xi_{d+1}^{\alpha,\beta}$ denotes the smallest root of the Jacobi polynomial $P_{d+1}^{\alpha,\beta}$ and that the largest root of $P_{d+1}^{\alpha,\beta}$ is equal to $-\xi_{d+1}^{\beta,\alpha}$.

Corollary 3.2. *Consider the measure $d\mu(x) = w_{\alpha,\beta}(x)dx$ on $\mathbf{K} = [-1,1]$ with the weight function $w_{\alpha,\beta}(x) = (1-x)^\alpha(1+x)^\beta$ and $\alpha, \beta > -1$. For the univariate polynomial $f(x) = x$ (resp., $f(x) = -x$), the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to $\xi_{d+1}^{\alpha,\beta}$ (resp., to $\xi_{d+1}^{\beta,\alpha}$) and thus we have*

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = \Theta(1/d^2).$$

In particular, $\underline{f}_{\mathbf{K}}^{(d)} = -\cos\left(\frac{\pi}{2d+2}\right)$ when $\alpha = \beta = -1/2$.

Proof. This follows directly using Theorem 3.1, Corollary 2.3, the fact that the largest root of $P_{d+1}^{\alpha,\beta}$ is equal to $-\xi_{d+1}^{\beta,\alpha}$, and the closed form expression (13) for the roots of the Chebyshev polynomials of the first kind. □

We now use the above result to show $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = \Omega(1/d^2)$ for the example (15) in the multivariate case $n \geq 2$.

Corollary 3.3. *Consider the measure $d\mu(x) = \prod_{i=1}^n w_{\alpha_i,\beta_i}(x_i)dx_i$ on the hypercube $\mathbf{K} = [-1,1]^n$, with the weight functions $w_{\alpha_i,\beta_i}(x_i) = (1-x_i)^{\alpha_i}(1+x_i)^{\beta_i}$ and $\alpha_i, \beta_i > -1$ for $i \in [n]$. For the polynomial $f(x) = \sum_{l=1}^n c_l x_l$, we have*

$$\underline{f}_{\mathbf{K}}^{(d)} \geq \sum_{l:c_l>0} c_l \xi_{d+1}^{\alpha_l,\beta_l} + \sum_{l:c_l<0} |c_l| \xi_{d+1}^{\beta_l,\alpha_l},$$

and thus $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = \Omega(1/d^2)$.

Proof. Assume $\underline{f}_{\mathbf{K}}^{(d)} = \int_{\mathbf{K}} (\sum_{l=1}^n c_l x_l)\sigma(x)d\mu(x)$, where $\sigma \in \mathbb{R}[x_1,\ldots,x_n]_{2d}$ is a sum of squares of polynomials and $\int_{\mathbf{K}} \sigma(x)d\mu(x) = 1$. For each $l \in [n]$ consider the univariate polynomial

$$\sigma_l(x_l) := \int_{[-1,1]^{n-1}} \sigma(x_1,\ldots,x_n) \prod_{i\in[n]\setminus\{l\}} w_{\alpha_i,\beta_i}(x_i)dx_i,$$

where we integrate over all variables x_i with $i \in [n]\setminus\{l\}$. Then we have $\int_{-1}^1 \sigma_l(x_l)w_{\alpha_l,\beta_l}(x_l)dx_l = 1$. Moreover, σ_l has degree at most $2d$ and, as it is a univariate polynomial which is nonnegative on $\mathbb{R}$, it is a sum of squares of polynomials. Hence, using Corollary 3.2, we can conclude that

$$\int_{-1}^1 x_l\sigma_l(x_l)w_{\alpha_l,\beta_l}(x_l)dx_l \geq \xi_{d+1}^{\alpha_l,\beta_l}, \quad \int_{-1}^1 (-x_l)\sigma_l(x_l)w_{\alpha_l,\beta_l}(x_l)dx_l \geq \xi_{d+1}^{\beta_l,\alpha_l}.$$

Combining with the definition of $\underline{f}_{\mathbf{K}}^{(d)}$ we obtain

$$\underline{f}_{\mathbf{K}}^{(d)} = \sum_{l=1}^n c_l \int_{-1}^1 x_l\sigma_l(x_l)w_{\alpha_l,\beta_l}(x_l)dx_l \geq \sum_{l:c_l>0} c_l\xi_{d+1}^{\alpha_l,\beta_l} + \sum_{l:c_l<0} |c_l|\xi_{d+1}^{\beta_l,\alpha_l}$$

and thus $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} \geq \sum_{l:c_l>0} c_l(\xi_{d+1}^{\alpha_l,\beta_l}+1) + \sum_{l:c_l<0} |c_l|(\xi_{d+1}^{\beta_l,\alpha_l}+1) = \Omega(1/d^2)$. ◻

3.2 Tight lower bound for the De Klerk, Hess and Laurent hierarchy

In this section we consider the hierarchy of bounds $f^{(d)}$ studied by De Klerk, Hess and Laurent [3], which are potentially stronger than the bounds $\underline{f}_{\mathbf{K}}^{(d)}$ since they involve the wider class of density functions in (5). Their convergence rate is known to be $O(1/d^2)$ ([3], recall Theorem 1.3).

For the example (15) we can also give an explicit expression for the bounds $f^{(d)}$ and we will show that their convergence rate to $f_{\min,\mathbf{K}}$ is also in the order $\Omega(1/d^2)$, which shows that the analysis in [3] is tight.

We first treat the univariate case, in order to introduce the main ideas, and then we extend to the multivariate case.

Theorem 3.4. *For the univariate polynomial $f(x) = \pm x$, we have*

$$f^{(d)} = \min\{\xi_{d+1}^{-1/2,-1/2}, \xi_d^{1/2,1/2}\},$$

the smallest value among the smallest roots of the Jacobi polynomials $P_{d+1}^{-1/2,-1/2}$ and $P_d^{1/2,1/2}$. In particular, we have $f^{(d)} - f_{\min,\mathbf{K}} = \Theta(1/d^2)$.

Proof. Consider first $f(x) = x$. We first recall how to compute $f^{(d)}$ as an eigenvalue problem. By definition, it is the minimum value of $\int_{-1}^{1} x(\sigma_0(x) + \sigma_1(x)(1-x^2))w_{-1/2,-1/2}(x)dx$, where $\sigma_0 \in \Sigma[x]_{2d}$, $\sigma_1 \in \Sigma[x]_{2d-2}$ and $\int_{-1}^{1}(\sigma_0(x) + \sigma_1(x)(1-x^2))w_{-1/2,-1/2}(x)dx = 1$. We express the polynomial σ_0 in the normalized Jacobi (Chebychev) basis $\{\hat{P}_k^{-1/2,-1/2}\}$ as

$$\sigma_0 = \sum_{i,j=0}^{d} M_{ij}^{(0)} \hat{P}_i^{-1/2,-1/2} \hat{P}_j^{-1/2,-1/2}$$

for some matrix $M^{(0)}$ of order $d+1$, constrained to be positive semidefinite. Based on the observation that $(1-x^2)w_{-1/2,-1/2}(x) = w_{1/2,1/2}(x)$, we express the polynomial σ_1 in the normalized Jacobi basis $\{\hat{P}_k^{1/2,1/2}\}$ as

$$\sigma_1 = \sum_{i,j=0}^{d-1} M_{ij}^{(1)} \hat{P}_i^{1/2,1/2} \hat{P}_j^{1/2,1/2}$$

for some matrix $M^{(1)}$ of order d, also constrained to be positive semidefinite. Then, we obtain

$$f^{(d)} = \min\{\langle A_d^{-1/2,-1/2}, M^{(0)}\rangle + \langle A_{d-1}^{1/2,1/2}, M^{(1)}\rangle : \mathrm{Tr}(M^{(0)}) + \mathrm{Tr}(M^{(1)}) = 1,\ M^{(0)} \succeq 0, M^{(1)} \succeq 0\},$$

where $A_d^{1/2,1/2}$ and $A_{d-1}^{-1/2,-1/2}$ are instances of (10) defined as follows:

$$A_d^{\alpha,\beta} := \left(\int_{-1}^{1} x\hat{P}_h^{\alpha,\beta}(x)\hat{P}_k^{\alpha,\beta}(x)w_{\alpha,\beta}(x)dx\right)_{h,k=0}^{d}$$

for any $\alpha, \beta > -1$ and $d \in \mathbb{N}$. Since strong duality holds we obtain

$$f^{(d)} = \max\{t : A_d^{-1/2,-1/2} - tI \succeq 0,\ A_{d-1}^{1/2,1/2} - tI \succeq 0\} = \min\{\lambda_{\min}(A_d^{-1/2,-1/2}), \lambda_{\min}(A_{d-1}^{1/2,1/2})\}.$$

By Lemma 2.1, we have $\lambda_{\min}(A_d^{-1/2,-1/2}) = \xi_{d+1}^{-1/2,-1/2}$ and $\lambda_{\min}(A_{d-1}^{1/2,1/2}) = \xi_d^{1/2,1/2}$ and thus $f^{(d)} = \min\{\xi_{d+1}^{-1/2,-1/2}, \xi_d^{1/2,1/2}\}$. The same result holds when $f(x) = -x$. Finally, by Corollary 2.3, these two smallest roots are both equal to $-1 + \Theta(1/d^2)$, which concludes the proof. □

We now extend this result to the multivariate case of example (15):

Corollary 3.5. *For the linear polynomial $f(x) = \sum_{l=1}^{n} c_l x_l$, we have*

$$f^{(d)} \geq \left(\sum_{l=1}^{n} |c_l|\right) \min\{\xi_{d+1}^{-1/2,-1/2}, \xi_d^{1/2,1/2}\}$$

and thus $f^{(d)} - f_{\min,\mathbf{K}} = \Omega(1/d^2)$.

Proof. The proof is analogous to that of Corollary 3.3, with some more technical details. Assume $f^{(d)} = \int_{\mathbf{K}} (\sum_{l=1}^{n} x_l)\sigma(x)d\mu(x)$, where $\sigma(x) = \sum_{I\subseteq[n]} \sigma_I(x) \prod_{i\in I}(1-x_i^2)$, $\sigma_I(x)$ is a sum of squares of degree at most $2d - 2|I|$ and $\int_{\mathbf{K}} \sigma(x)d\mu(x) = 1$.

Fix $l \in [n]$. Then we can write

$$\sigma(x) = \sum_{I\subseteq[n]\setminus\{l\}} \sigma_I(x) \prod_{i\in I}(1-x_i^2) \ + \ (1-x_l^2) \sum_{I\subseteq[n]:l\in I} \sigma_I(x) \prod_{i\in I\setminus\{l\}} (1-x_i^2).$$

Next, define the univariate polynomials in the variable x_l:

$$\sigma_{l,0}(x_l) := \sum_{I\subseteq[n]\setminus\{l\}} \int_{[-1,1]^{n-1}} \sigma_I(x) \prod_{i\in I}(1-x_i^2) \prod_{i\in[n]\setminus\{l\}} w_{-1/2,-1/2}(x_i)dx_i,$$

$$\sigma_{l,1}(x_l) := \sum_{I\subseteq[n]:l\in I} \int_{[-1,1]^{n-1}} \sigma_I(x) \prod_{i\in I\setminus\{l\}}(1-x_i^2) \prod_{i\in[n]\setminus\{l\}} w_{-1/2,-1/2}(x_i)dx_i,$$

$$\sigma_l(x_l) := \int_{[-1,1]^{n-1}} \sigma(x) \prod_{i\in[n]\setminus\{l\}} w_{-1/2,-1/2}(x_i)dx_i = \sigma_{l,0}(x_l) + (1-x_l^2)\sigma_{l,1}(x_l).$$

By construction, we have

$$\int_{\mathbf{K}} x_l\sigma(x)d\mu(x) = \int_{-1}^{1} x_l\sigma_l(x_l)w_{-1/2,-1/2}(x_l)dx_l, \ \int_{-1}^{1} \sigma_l(x_l)w_{-1/2,-1/2}(x_l)dx_l = \int_{\mathbf{K}} \sigma(x)d\mu(x) = 1.$$

Moreover, the polynomial $\sigma_{l,0}$ is a sum of squares (since it is univariate and nonnegative on $\mathbb{R}$) and its degree is at most $2d$, and the polynomial $\sigma_{l,1}$ is a sum of squares of degree at most $2d-2$. Hence, using Theorem 3.4, we can conclude that

$$\int_{-1}^{1} (\pm x_l)\sigma_l(x_l)w_{-1/2,-1/2}(x_l)dx_l \geq \min\{\xi_{d+1}^{-1/2,-1/2}, \xi_d^{1/2,1/2}\}.$$

This implies that

$$f^{(d)} = \int_{\mathbf{K}} (\sum_{l=1}^{n} c_l x_l)\sigma(x)d\mu(x) = \sum_{l=1}^{n} c_l \int_{-1}^{1} x_l\sigma_l(x_l)w_{-1/2,-1/2}(x_l)dx_l$$

is at least $(\sum_l |c_l|)\min\{\xi_{d+1}^{-1/2,-1/2}, \xi_d^{1/2,1/2}\}$ and the proof is complete. □

4 Tight upper bounds for the Lasserre hierarchy

In this section we analyze the rate of convergence of the Lasserre bounds $\underline{f}_{\mathbf{K}}^{(d)}$ when using the measure $d\mu(x) = \prod_{i=1}^n w_{-1/2,-1/2}(x_i)dx_i$ on the box $\mathbf{K} = [-1,1]^n$ (corresponding to the Chebyshev orthogonal polynomials). For this measure, it is known that the stronger bounds $f^{(d)}$ - that use a much richer class of density functions - enjoy a $O(1/d^2)$ rate of convergence ([3], see Theorem 1.3). We show that the convergence rate remains $O(1/d^2)$ for the weaker bounds $\underline{f}_{\mathbf{K}}^{(d)}$, which thus also implies Thoerem 1.3.

Theorem 4.1. *Consider the measure $d\mu(x) = \prod_{i=1}^n w_{-1/2,-1/2}(x_i)dx_i$ on the hypercube $\mathbf{K} = [-1,1]^n$, with the weight function $w_{-1/2,-1/2}(x_i) = (1-x_i^2)^{-1/2}$ for $i \in [n]$. For any polynomial f we have*

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = O(1/d^2).$$

It turns out that we can reduce the general result to the univariate quadratic case. In what follows we consider first the special case when f is univariate and quadratic (see Lemma 4.2) and then we indicate how to derive the result for an arbitrary multivariate polynomial f. A key tool we use for this reduction is the existence of a quadratic upper estimator for f having the same minimum as f over $\mathbf{K}$. In the quadratic univariate case we exploit again the formulation of $\underline{f}_{\mathbf{K}}^{(d)}$ in terms of the smallest eigenvalue of the associated matrix A_d in (16) (recall Lemma 1.2). This matrix A_d is now 5-diagonal, but a key feature is that it contains a large Toeplitz submatrix, whose eigenvalues can be estimated by embedding it into a circulant matrix for which closed form expressions exist for the eigenvalues. This nice structure, which allows a simple analysis, follows from the choice of the Chebyshev type measure. We expect that a similar convergence rate should hold when selecting any measure of Jacobi type, but the analysis seems more complicated.

4.1 The quadratic univariate case

Here we consider the case when $\mathbf{K} = [-1, 1]$ and f is a univariate quadratic polynomial of the form $f(x) = x^2 + \alpha x$, for some scalar $\alpha \in \mathbb{R}$.

We can first easily deal with the case when $\alpha \notin (-2, 2)$. Indeed then we have

$$f(x) \leq g(x) := \alpha x + 1 \quad \text{ for all } x \in [-1, 1],$$

and both f and g have the same minimum value on $[-1, 1]$. Namely, $f_{\min,\mathbf{K}} = g_{\min,\mathbf{K}}$ is equal to $1 - \alpha$ if $\alpha \geq 2$, and to $1 + \alpha$ if $\alpha \leq -2$. Therefore we have

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} \leq \underline{g}_{\mathbf{K}}^{(d)} - g_{\min,\mathbf{K}} = O(1/d^2),$$

where we use Corollary 3.3 for the last estimate.

We may now assume that $f(x) = x^2 + \alpha x$, where $\alpha \in [-2, 2]$. Then, $f_{\min,\mathbf{K}} = -\alpha^2/4$, which is attained at $x = -\alpha/2$. After scaling the measure μ by $2/\pi$, the Chebyshev polynomials T_i satisfy

$$\int_{-1}^{1} T_i(x) T_j(x) \frac{2}{\pi\sqrt{1-x^2}} dx = 0 \text{ if } i \neq j,\ 2 \text{ if } i = j = 0,\ 1 \text{ if } i = j \geq 1.$$

So with respect to this scaled measure the normalized Chebyshev polynomials are $\hat{T}_0 = 1/\sqrt{2}$ and $\hat{T}_i = T_i$ for $i \geq 1$, and they satisfy the 3-terms relation:

$$x\hat{T}_1 = \frac{1}{2}\hat{T}_2 + \frac{1}{\sqrt{2}}\hat{T}_0 \quad \text{ and } \quad x\hat{T}_k = \frac{1}{2}\hat{T}_{k+1} + \frac{1}{2}\hat{T}_{k-1} \ \text{ for } k \geq 2.$$

In view of Lemma 1.2 we know that the parameter $\underline{f}_{\mathbf{K}}^{(d)}$ is equal to the smallest eigenvalue of the following matrix

$$A_d = \left(\int_{-1}^{1} (x^2 + \alpha x) \hat{T}_i(x) \hat{T}_j(x) \frac{2}{\pi\sqrt{1-x^2}} dx \right)_{i,j=0}^{d}.$$

Using the above 3-terms relations one can verify that the matrix A_d has the following form:

$$
A_d = \begin{pmatrix}
\frac{1}{2} & \frac{\alpha}{\sqrt{2}} & \frac{1}{2\sqrt{2}} & & & & & & \\
\frac{\alpha}{\sqrt{2}} & \frac{3}{4} & \frac{\alpha}{2} & \frac{1}{4} & & & & & \\
\frac{1}{2\sqrt{2}} & \frac{\alpha}{2} & a & b & c & & & & \\
 & \frac{1}{4} & b & a & b & c & & & \\
 & & c & b & \ddots & \ddots & \ddots & & \\
 & & & c & \ddots & \ddots & \ddots & \ddots & \\
 & & & & \ddots & \ddots & \ddots & \ddots & c \\
 & & & & & \ddots & \ddots & \ddots & b \\
 & & & & & & c & b & a
\end{pmatrix}, \tag{16}
$$

where we set $a = 1/2$, $b = \alpha/2$ and $c = 1/4$.

Observe that if we remove the first two rows and columns of A then we obtain a principal submatrix, denoted B, which is a symmetric 5-diagonal Toeplitz matrix. Now we may embed B into a symmetric circulant matrix of size $d+1$, denoted C_d, by suitably defining the first two rows and columns. Namely,

$$
C_d = \begin{pmatrix}
a & b & c & & & & & c & b \\
b & a & b & c & & & & & c \\
c & b & a & b & c & & & & \\
 & c & b & a & b & c & & & \\
 & & c & b & \ddots & \ddots & \ddots & & \\
 & & & c & \ddots & \ddots & \ddots & \ddots & \\
 & & & & \ddots & \ddots & \ddots & \ddots & c \\
c & & & & & \ddots & \ddots & \ddots & b \\
b & c & & & & & c & b & a
\end{pmatrix}.
$$

Recall that the eigenvalues of a circulant matrix are known in closed form, see, e.g., [9]. In particular, the eigenvalues of C_d are given by

$$
a + 2b\cos(2\pi j/(d+1)) + 2c\cos(2\pi 2j/(d+1)), \quad j = 0, \ldots, d, \quad (d \geq 5). \tag{17}
$$

By the Cauchy interlacing theorem for eigenvalues (see, e.g., Corollary 2.2 in [10]), we have

$$
\underline{f}_{\mathbf{K}}^{(d)} = \lambda_{\min}(A_d) \leq \lambda_{\min}(B) \leq \lambda_3(C_d),
$$

where $\lambda_3(C_d)$ is the third smallest eigenvalue of C_d. As noted above the eigenvalues of C_d are known in closed form as in (17) and this is the key fact which enables us to conclude the analysis.

Lemma 4.2. *For any $\alpha \in [-2, 2]$, the third smallest eigenvalue of the matrix C_d satisfies*

$$
\lambda_3(C_d) = -\frac{\alpha^2}{4} + O\left(\frac{1}{d^2}\right).
$$

Therefore, if $f(x) = x^2 + \alpha x$ with $\alpha \in [-2, 2]$ then $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = O(1/d^2)$.

Proof. Setting $\vartheta_j = \frac{2\pi j}{d+1}$ for $j \in \mathbb{N}$, then by (17) the eigenvalues of the matrix C_d are the scalars

$$\frac{1}{2} + \alpha\cos(\vartheta_j) + \frac{1}{2}\cos(2\vartheta_j) = \cos^2(\vartheta_j) + \alpha\cos(\vartheta_j) \quad \text{for } 0 \le j \le d.$$

Consider the function $f(\vartheta) = \cos^2(\vartheta) + \alpha\cos(\vartheta)$ for $\vartheta \in [0, 2\pi]$. Then f satisfies: $f(\vartheta) = f(2\pi - \vartheta)$, and its minimum value is equal to $-\alpha^2/4$, which is attained at $\vartheta = \arccos(-\alpha/2) \in [0, \pi]$ and $2\pi - \vartheta$. Let j be the integer such that $\vartheta_j \le \vartheta < \vartheta_{j+1}$. Then the smallest eigenvalue of C_d is $\lambda_{\min}(C_d) = \min\{f(\vartheta_j), f(\vartheta_{j+1})\}$ and its third smallest eigenvalue is given by $\lambda_3(C_d) = \min\{f(\vartheta_{j-1}), f(\vartheta_{j+1})\}$ if $\lambda_{\min}(C_d) = f(\vartheta_j)$, and $\lambda_3(C_d) = \min\{f(\vartheta_j), f(\vartheta_{j+2})\}$ if $\lambda_{\min}(C_d) = f(\vartheta_{j+1})$. Therefore, $\lambda_3(C_d) = f(\vartheta_k)$ for some $k \in \{j-1, j, j+1, j+2\}$.

Using Taylor theorem (and the fact that $f'(\vartheta) = 0$) we can conclude that

$$\lambda_3(C_d) + \frac{\alpha^2}{4} = f(\vartheta_k) - f(\vartheta) = \frac{1}{2} f''(\xi)(\vartheta - \vartheta_k)^2,$$

for some scalar $\xi \in (\vartheta, \vartheta_k)$ (or (ϑ_k, ϑ)). Finally, $f''(\xi) = -2\cos(\xi) - \alpha\cos(\xi)$ and thus we have $|f''(\xi)| \le 2 + |\alpha|$. Also $|\vartheta - \vartheta_k| \le |\vartheta_{j+2} - \vartheta_{j-1}| = \frac{6\pi}{d+1}$. The claimed result now follows directly. □

4.2 The general case

As a direct application we can also deal with the case when f is multivariate quadratic and separable.

Corollary 4.3. *Consider the box* $\mathbf{K} = [-1, 1]^n$ *and a multivariate polynomial of the form* $f(x) = \sum_{i=1}^n x_i^2 + \alpha_i x_i$ *for some scalars* $\alpha_i \in \mathbb{R}$. *Then we have* $\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = O(1/d^2)$.

Proof. The polynomial f is separable: $f(x) = \sum_{i=1}^n f_i(x_i)$, after setting $f_i(x_i) = x_i^2 + \alpha_i x_i$. Hence its minimum over the box $\mathbf{K}$ is $f_{\min,\mathbf{K}} = \sum_{i=1}^n (f_i)_{\min,[-1,1]}$. Suppose $\sigma_i \in \Sigma[x_i]_d$ is an optimal density function for the bound $\underline{f_i}_{[-1,1]}^{(d)}$ and consider the polynomial $\sigma(x) = \prod_{i=1}^n \sigma_i(x_i) \in \Sigma[x]_{nd}$, which is a density function over $\mathbf{K}$. Then we have

$$\underline{f}_{\mathbf{K}}^{(nd)} - f_{\min,\mathbf{K}} \le \int_{\mathbf{K}} f(x)\sigma(x)d\mu(x) = \sum_{i=1}^n \left(\int_{-1}^1 f_i(x_i)d\mu(x_i) - (f_i)_{\min,[-1,1]} \right) = O(1/d^2),$$

where we use Lemma 4.2 for the last estimate. This implies the claimed convergence rate for the bounds $\underline{f}_{\mathbf{K}}^{(d)}$. □

Assume now f is an arbitrary polynomial and let $a \in \mathbf{K} = [-1, 1]^n$ be a minimizer of f over $\mathbf{K}$. Consider the following quadratic polynomial

$$g(x) = f(a) + \nabla f(a)^T (x - a) + C_f \|x - a\|_2^2,$$

where we set $C_f = \max_{x \in \mathbf{K}} \|\nabla^2 f(x)\|_2$. By Taylor's theorem we know that $f(x) \le g(x)$ for all $x \in \mathbf{K}$ and that the minimum value of $g(x)$ over $\mathbf{K}$ is $g_{\min,\mathbf{K}} = f(a) = f_{\min,\mathbf{K}}$. This implies

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} \le \underline{g}_{\mathbf{K}}^{(d)} - g_{\min,\mathbf{K}} = O(1/d^2),$$

where we use Corollary 4.3 for the last estimate. This concludes the proof of Theorem 4.1.

5 Concluding remarks

Some other hierarchical upper bounds for polynomial optimization over the hypercube have been investigated in the literature. In particular, bounds are proposed in [4], that rely on selecting density functions arising from beta distributions:

$$f_d^H := \min_{(\alpha,\beta)\in N(2n,d)} \frac{\displaystyle\int_{\mathbf{K}} f(x)\, x^\alpha (1-x)^\beta \, dx}{\displaystyle\int_{\mathbf{K}} x^\alpha (1-x)^\beta \, dx},$$

where, $\mathbf{K} = [-1,1]^n$, and $(1-x)^\beta = \prod_{i=1}^n (1-x_i)^{\beta_i}$ for $\beta \in \mathbb{N}^n$. These bounds can be computed via elementary operations only and their rate of convergence is $f_d^H - f_{\min,\mathbf{K}} = O(1/\sqrt{d})$ (or $O(1/d)$ for quadratic polynomials with rational data).

Other hierarchies involve selecting discrete measures. They rely on polynomial evaluations at rational grid points [1] or at polynomial meshes like Chebyshev grids [14]. The grids in [14] are given by the Chebyshev-Lobatto points:

$$C_d := \left\{ \cos\left(\frac{j\pi}{d}\right) \right\} \quad j = 0, \ldots, d.$$

In particular the authors of [14] show that $\min_{x \in C_d^n} f(x) - f_{\min,\mathbf{K}} = O\left(\frac{1}{d^2}\right)$, where

$$C_d^n = C_d \times \cdots \times C_d \subset [-1,1]^n.$$

Note that $|C_d^n| = (d+1)^n$, which is of course exponential in n even for fixed d.

The same $O\left(\frac{1}{d^2}\right)$ rate of convergence was shown in [1] for the regular grid (using $d+1$ evenly spaced points). We also refer to the recent work [15] where polynomial meshes are investigated for polynomial optimization over general convex bodies.

Thus the Lasserre bound $\underline{f}_{\mathbf{K}}^{(d)}$ has the same $O\left(\frac{1}{d^2}\right)$ asymptotic rate of convergence as the grid searches, but with the advantage that the computation may be done in polynomial time for fixed d.

Of course, the problem studied in this paper falls in the general framework of bound-constrained global optimization problems, and many other algorithms are available for such problems; a recent survey is given in the thesis [13]. The point is that the

methods we studied in this paper allow analysis of the convergence rate to the global minimum.

We conclude with some unresolved questions:

- Does the $O\left(\frac{1}{d^2}\right)$ rate of convergence still hold for the Lasserre bounds if $\mathbf{K}$ is a general convex body? (The best known result is the $O(1/d)$ rate from [2].)

- What is the precise influence of the choice of reference measure μ in (1) on the convergence rate?

- Is is possible to show a 'saturation' result for the Lasserre bounds of the type:

$$\underline{f}_{\mathbf{K}}^{(d)} - f_{\min,\mathbf{K}} = o\left(\frac{1}{d^2}\right) \Longleftrightarrow f \text{ is a constant polynomial?}$$

  In other words, is $O(1/d^2)$ the fastest possible convergence rate for nonconstant polynomials?

Acknowledgements. The authors would like to thank Jean-Bernard Lasserre for useful discussions.

References

[1] E. de Klerk, M. Laurent. Error bounds for some semidefinite programming approaches to polynomial optimization on the hypercube. *SIAM Journal on Optimization* 20(6), (2010) 3104–3120.

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

[3] 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(1), (2017) 347-367.

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

[5] 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(1), (2017) 363-392.

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

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

[8] W. Gautsch. *Orthogonal Polynomials - Computation and Approximation.* Oxford University Press, 2004.

[9] R.M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory* 2(3) (2006), 155–239.

[10] W.H. Haemers. Interlacing eigenvalues and graphs. *Linear Algebra and its Applications* 227-228 (1995), 593–616.

[11] M.E.H. Ismail, X. Li. Bounds on the extreme zeros of orthogonal polynomials. *Proc. Amer. Math. Soc.* 115 (1992), 131–140.

[12] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization* 21(3) (2011), 864–885.

[13] Pál, L.: *Global Optimization Algorithms for Bound Constrained Problems.* PhD thesis, University of Szeged (2010). Available at `http://www2.sci.u-szeged.hu/fokozatok/PDF/Pal_Laszlo/Diszertacio_PalLaszlo.pdf`

[14] F. Piazzon, M. Vianello. A note on total degree polynomial optimization by Chebyshev grids. *Optimization Letters* 12 (2018), 63–71.

[15] F. Piazzon and M. Vianello. Markov inequalities, Dubiner distance, norming meshes and polynomial optimization on convex bodies. Preprint, 2018. Available at `http://www.math.unipd.it/~marcov/pdf/convbodies.pdf`

[16] G. Szegö. *Orthogonal Polynomials, fourth ed.*, vol. XXIII, American Mathematical Society Colloquium Publications, Providence, RI, 1975.