Optimization Letters (2023) 17:515–530
https://doi.org/10.1007/s11590-022-01922-5

ORIGINAL PAPER

An effective version of Schmüdgen's Positivstellensatz for the hypercube

Monique Laurent[1,2] · Lucas Slot[1,2]

Received: 21 September 2021 / Accepted: 15 August 2022 / Published online: 15 September 2022

Abstract
Let $S \subseteq \mathbb{R}^n$ be a compact semialgebraic set and let f be a polynomial nonnegative on S. Schmüdgen's Positivstellensatz then states that for any $\eta > 0$, the nonnegativity of $f + \eta$ on S can be certified by expressing $f + \eta$ as a conic combination of products of the polynomials that occur in the inequalities defining S, where the coefficients are (globally nonnegative) sum-of-squares polynomials. It does not, however, provide explicit bounds on the degree of the polynomials required for such an expression. We show that in the special case where $S = [-1, 1]^n$ is the hypercube, a Schmüdgen-type certificate of nonnegativity exists involving only polynomials of degree $O(1/\sqrt{\eta})$. This improves quadratically upon the previously best known estimate in $O(1/\eta)$. Our proof relies on an application of the polynomial kernel method, making use in particular of the Jackson kernel on the interval $[-1, 1]$.

Keywords Schmüdgen's Positivstellensatz · Sum-of-squares polynomials · Lasserre hierarchy · Polynomial kernel method · Jackson kernel · Semidefinite programming

Mathematics Subject Classification 90C22 · 90C23 · 90C26

1 Introduction

Consider the problem of computing the global minimum:

$$f_{\min} := \min_{\mathbf{x} \in \mathrm{B}^n} f(\mathbf{x}) \tag{1}$$

✉ Lucas Slot
lucas.slot@cwi.nl

Monique Laurent
monique.laurent@cwi.nl

1 Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

2 Tilburg University, Tilburg, The Netherlands

of a polynomial f of degree $d \in \mathbb{N}$ over the hypercube $\mathrm{B}^n := [-1,1]^n \subseteq \mathbb{R}^n$. The program (1) can be reformulated as finding the largest $\lambda \in \mathbb{R}$ for which the function $f - \lambda$ is nonnegative on B^n. That is, writing $\mathcal{P}(\mathrm{B}^n) \subseteq \mathbb{R}[x]$ for the cone of all polynomials that are nonnegative on B^n, we have:

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in \mathcal{P}(\mathrm{B}^n)\}. \tag{2}$$

By replacing $\mathcal{P}(\mathrm{B}^n)$ in (2) by a smaller subset of $\mathbb{R}[\mathbf{x}]$ one may obtain lower bounds on $f_{\min}$. One way of obtaining such subsets is based on the following description of B^n as a semialgebraic set:

$$\mathrm{B}^n = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) := (1 - x_i^2) \geq 0 \quad \forall i \in [n]\}. \tag{3}$$

In light of this description, we see that the *preordering* $Q(\mathrm{B}^n)_r$, truncated at degree r, defined by[1]:

$$Q(\mathrm{B}^n)_r := \left\{ \sum_{J \subseteq [n]} \sigma_J g_J : \sigma_J \in \Sigma[\mathbf{x}],\ \deg(\sigma_J g_J) \leq r \right\} \quad \left(g_J := \prod_{j \in J} g_j \right), \tag{4}$$

satisfies $Q(\mathrm{B}^n)_r \subseteq \mathcal{P}(\mathrm{B}^n)$ for all $r \in \mathbb{N}$. Here, $\Sigma[\mathbf{x}]$ is the set of sum-of-squares polynomials (i.e., of the form $p = p_1^2 + p_2^2 + \ldots + p_m^2$ for certain $p_i \in \mathbb{R}[\mathbf{x}]$). When no degree bounds are imposed (i.e., $r = \infty$) we obtain the full preordering $Q(\mathrm{B}^n)$ generated by the polynomials $g_i(\mathbf{x}) = 1 - x_i^2$ ($i \in [n]$), which coincides with the quadratic module generated by the products $\prod_{i \in I} g_i(\mathbf{x})$ ($I \subseteq [n]$). We thus obtain the following hierarchy of lower bounds for $f_{\min}$, due to Lasserre [1]:

$$f_{(r)} := \max\{\lambda \in \mathbb{R} : f - \lambda \in Q(\mathrm{B}^n)_r\}. \tag{5}$$

If the program (5) is feasible, its maximum is attained. By definition, we have $f_{\min} \geq f_{(r+1)} \geq f_{(r)}$ for all $r \in \mathbb{N}$. Furthermore, we have $\lim_{r\to\infty} f_{(r)} = f_{\min}$, which follows directly from the following special case of *Schmüdgen's Positvstellensatz*.

Theorem 1 (Special case of Schmüdgen's Positivstellensatz [2]) *Let* $f \in \mathcal{P}(\mathrm{B}^n)$ *be a polynomial. Then for any* $\eta > 0$ *there exists an* $r \in \mathbb{N}$ *such that* $f + \eta \in Q(\mathrm{B}^n)_r$.

1.1 Main result

We show a bound on the convergence rate of the lower bounds $f_{(r)}$ to the global minimum $f_{\min}$ of f over B^n in $O(1/r^2)$. Alternatively, our result can be interpreted as a bound on the degree r in Schmüdgen's Positivstellensatz of the order $O(1/\sqrt{\eta})$ of a positivity certificate for $f + \eta$ when $f \in \mathcal{P}(\mathrm{B}^n)$.

[1] Sometimes the index r is used in the literature to denote the truncation where all summands have degree at most $2r$. For our treatment here it is more convenient to let r denote the truncation where all summands have degree at most r, the main reason being our use later of Theorem 8.

Theorem 2 *Let f be a polynomial of degree $d \in \mathbb{N}$. Then there exists a constant $C(n,d) > 0$, depending only on n and d, such that*:

$$f_{\min} - f_{((r+1)n)} \leq \frac{C(n,d)}{r^2} \cdot (f_{\max} - f_{\min}) \quad \text{for all } r \geq \pi d\sqrt{2n}. \tag{6}$$

Furthermore, the constant $C(n, d)$ may be chosen such that it either depends polynomially on n (for fixed d) or it depends polynomially on d (for fixed n), see relation (20) for details.

Corollary 3 *Let $f \in \mathcal{P}(\mathrm{B}^n)$ with degree d. Then, for any $\eta > 0$, we have*:

$$f + \eta \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n} \quad \text{for all } r \geq \max\left\{\pi d\sqrt{2n}, \frac{1}{\sqrt{\eta}}\sqrt{C(n,d)(f_{\max} - f_{\min})}\right\},$$

where $C(n, d)$ is the constant from Theorem 2. Hence we have $f + \eta \in \mathcal{Q}(\mathrm{B}^n)_r$ for $r = O(1/\sqrt{\eta})$.

Proof Let $\eta > 0$ and set $C_f := C(n,d) \cdot (f_{\max} - f_{\min})$. Pick an integer $r \geq \max\{\pi d\sqrt{2n}, \sqrt{C_f/\eta}\}$. Then we have:

$$f + \eta = \underbrace{f - f_{((r+1)n)}}_{\in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n}} + \underbrace{\left(f_{((r+1)n)} - f_{\min} + \frac{C_f}{r^2}\right)}_{\geq 0 \text{ by Theorem 2}} + \underbrace{f_{\min}}_{\geq 0} + \underbrace{\left(\eta - \frac{C_f}{r^2}\right)}_{\geq 0},$$

which shows $f + \eta \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n}$. □

1.2 Outline of the proof

Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial of degree d. To simplify our arguments and notation, we will work with the scaled function:

$$F := \frac{f - f_{\min}}{f_{\max} - f_{\min}},$$

for which $F_{\min} = 0$ and $F_{\max} = 1$. Since the inequality (6) is invariant under a positive scaling of f and adding a constant, it indeed suffices to show the result for the function F.

The idea of the proof is as follows. Let $\epsilon > 0$ and consider the polynomial $\tilde{F} := F + \epsilon$. Let $r \geq d$. Suppose that we are able to construct a (nonsingular) linear operator $\mathbf{K}_r : \mathbb{R}[\mathbf{x}]_r \to \mathbb{R}[\mathbf{x}]_r$ which has the following two properties:

$$\mathbf{K}_r p \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n} \quad \text{for all } p \in \mathcal{P}(\mathrm{B}^n)_r, \tag{P1}$$

$$\|\mathbf{K}_r^{-1}\tilde{F} - \tilde{F}\|_\infty := \max_{\mathbf{x}\in \mathrm{B}^n} |\mathbf{K}_r^{-1}\tilde{F}(\mathbf{x}) - \tilde{F}(\mathbf{x})| \leq \epsilon. \tag{P2}$$

Then, by (P2), we have $\mathbf{K}_r^{-1}\tilde{F} \in \mathcal{P}(\mathrm{B}^n)_r$. Indeed, as F is nonnegative on B^n, $\tilde{F}(\mathbf{x}) = F(\mathbf{x}) + \epsilon$ is greater than or equal to ϵ for all $\mathbf{x} \in \mathrm{B}^n$, and so (P2) tells us that after application of the operator $\mathbf{K}_r^{-1}$, the resulting polynomial $\mathbf{K}_r^{-1}\tilde{F}$ is nonnegative on B^n. Using (P1), we may then conclude that $\tilde{F} = \mathbf{K}_r(\mathbf{K}_r^{-1}\tilde{F}) \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n}$. It follows that $-\epsilon \le F_{((r+1)n)}$, i.e., $F_{\min} - F_{((r+1)n)} \le \epsilon$, and thus $f_{\min} - f_{((r+1)n)} \le \epsilon \cdot (f_{\max} - f_{\min})$. We collect this in the next lemma for future reference.

Lemma 4 *Assume that for some $r \ge d$ and $\epsilon > 0$ there exists a nonsingular operator $\mathbf{K}_r : \mathbb{R}[\mathbf{x}]_r \to \mathbb{R}[\mathbf{x}]_r$ which satisfies the properties (P1) and (P2). Then we have*

$$f_{\min} - f_{((r+1)n)} \le \epsilon \cdot (f_{\max} - f_{\min}).$$

In what follows, we will construct such an operator $\mathbf{K}_r$ for each $r \ge \pi d\sqrt{2n}$ and the parameter $\epsilon := C(n,d)/r^2$, where the constant $C(n, d)$ will be specified later. Our main Theorem 2 then follows after applying Lemma 4.

We make use of the *polynomial kernel method* for our construction: after choosing a suitable kernel $K_r : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, we define the linear operator $\mathbf{K}_r : \mathbb{R}[\mathbf{x}]_r \to \mathbb{R}[\mathbf{x}]_r$ via the integral transform:

$$\mathbf{K}_r p(\mathbf{x}) := \int_{\mathrm{B}^n} K_r(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mu(\mathbf{y}) \quad (p \in \mathbb{R}[\mathbf{x}]_r).$$

Here, μ is the *Chebyshev measure* on B^n as defined in (7) below. A good choice for the kernel K_r is a multivariate version (see Sect. 3.1) of the well-known *Jackson kernel K_r^{ja}* of degree r (see Sect. 2.3). For this choice of kernel, the operator $\mathbf{K}_r$ naturally satisfies (P1) (see Sect. 3.2). Furthermore, it diagonalizes with respect to the basis of $\mathbb{R}[\mathbf{x}]$ given by the (multivariate) *Chebyshev polynomials* (see Sect. 2.2). Property (P2) can then be verified by analyzing the eigenvalues of $\mathbf{K}_r$, which are closely related to the expansion of K_r^{ja} in the basis of (univariate) Chebyshev polynomials (see Sect. 3.3). We end this section by illustrating our method of proof with a small example.

Example 5 Consider the polynomial $f(x) = 1 - x^2 - x^3 + x^4$, which is nonnegative on $[-1, 1]$. For $r \in \mathbb{N}$, let $\mathbf{K}_r$ be the operator associated to the univariate Jackson kernel (11) of degree r, which satisfies (P1) (see Sect. 3.2). For $\eta = 0.1$, we observe that applying $\mathbf{K}_7^{-1}$ to $f + \eta$ yields a nonnegative function on $[-1, 1]$, whereas applying $\mathbf{K}_5^{-1}$ does not (see Fig. 1). Applying the arguments of Sect. 1.2, we may thus conclude that $f + \eta \in \mathcal{Q}(\mathrm{B}^n)_8$, but not that $f + \eta \in \mathcal{Q}(\mathrm{B}^n)_6$.

1.3 Related work

The polynomial kernel method, which forms the basis of our analysis, is widely used in functional approximation, see, e.g., [3]. In the present context, the method has already been employed for the analysis of the sum-of-squares hierarchy for optimization over the hypersphere S^{n-1} in [4] (where a rate in $O(1/r^2)$ was shown as well)

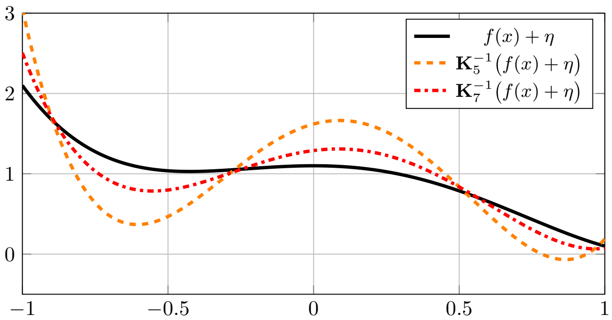

Fig. 1 The polynomial $f(x) + \eta$ of Example 5 and its transformations under the inverse operators $\mathbf{K}_5^{-1}$ and $\mathbf{K}_7^{-1}$ associated to the Jackson kernels of degree 5 and 7

and for optimization over the *binary* cube $\{-1, 1\}^n$ in [5]. There, the authors use kernels that are invariant under the symmetry of S^{n-1} and $\{-1, 1\}^n$, respectively.

In [6], the polynomial kernel method, and the Jackson kernel in particular, were used to analyze the quality of a related Lasserre-type hierarchy of *upper* bounds on $f_{\min}$ over $\mathrm{B}^n = [-1, 1]^n$, where one searches for a density in the truncated preordering $Q(\mathrm{B}^n)_r$ minimizing the expected value of f over B^n (showing again a convergence rate in $O(1/r^2)$).

For a general compact semialgebraic set S, a polynomial f nonnegative on S and $\eta > 0$, existence of Schmüdgen-type certificates of positivity for $f + \eta$ with degree bounds in $O(1/\eta^c)$ was shown in [7], where $c > 0$ is a constant depending on S. This result uses different tools, including in particular a representation result for polynomial optimization over the simplex by Pólya [8] and the effective degree bounds by Powers and Reznick [9].

For the case of the hypercube[2] a degree bound in $O(1/\eta)$ for Schmüdgen-type certificates is obtained in [10], thus showing that one can take $c \leq 1$ in the above mentioned result of [7]. This result holds in fact for a weaker hierarchy of bounds obtained by restricting in (5) to decompositions of the polynomial $f - \lambda$ involving factors σ_J that are nonnegative scalars (instead of sums of squares), also known as Handelman-type decompositions (thus replacing the preordering $Q(\mathrm{B}^n)_r$ by its subset H_r of polynomials having a Handelman-type decomposition). The analysis in [10] relies on employing the *Bernstein operator* $\mathbf{B}_r$, which has the property of mapping a polynomial nonnegative over the hypercube to a polynomial in the set $H_{rn} \subseteq Q(\mathrm{B}^n)_{rn}$.

In this paper, we can show a further improvement by using a different type of kernel operator; namely we show that we can take the constant $c \leq 1/2$ in the special case $S = [-1, 1]^n$.

[2] The hypercube $[0, 1]^n$ is considered in [10] but the results extend to the hypercube $[-1, 1]^n$ by an affine change of variables.

2 Preliminaries

2.1 Notation

Throughout, $\mathrm{B}^n := [-1,1]^n \subseteq \mathbb{R}^n$ is the n-dimensional hypercube. We write $\mathbb{R}[x]$ for the univariate polynomial ring, while reserving the bold-face notation $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, x_2, \ldots, x_n]$ to denote the ring of polynomials in n variables. Similarly, $\Sigma[x] \subseteq \mathbb{R}[x]$ and $\Sigma[\mathbf{x}] \subseteq \mathbb{R}[\mathbf{x}]$ denote the sets of univariate and n-variate sum-of-squares polynomials, respectively, consisting of all polynomials of the form $p = p_1^2 + p_2^2 + \cdots + p_m^2$ for certain polynomials $p_1, \ldots, p_m$ and $m \in \mathbb{N}$. For a polynomial $p \in \mathbb{R}[\mathbf{x}]$, we write $p_{\min}, p_{\max}$ for its minimum and maximum over B^n, respectively, and $\|p\|_\infty := \sup_{\mathbf{x} \in \mathrm{B}^n} |p(\mathbf{x})|$ for its sup-norm on B^n.

2.2 Chebyshev polynomials

Let μ be the normalized *Chebyshev measure* on $\mathrm{B}^n = [-1,1]^n$, defined by:

$$d\mu(\mathbf{x}) = \frac{dx_1}{\pi\sqrt{1-x_1^2}} \cdots \frac{dx_n}{\pi\sqrt{1-x_n^2}}. \tag{7}$$

Note that μ is a probability measure on B^n, meaning that $\int_{\mathrm{B}^n} d\mu = 1$. We write $\langle \cdot, \cdot \rangle_\mu$ for the corresponding inner product on $\mathbb{R}[\mathbf{x}]$, given by:

$$\langle f, g \rangle_\mu := \int_{\mathrm{B}^n} f(\mathbf{x}) g(\mathbf{x}) d\mu(\mathbf{x}).$$

For $k \in \mathbb{N}$, let T_k be the univariate *Chebyshev polynomial* (see, e.g., [11]) of degree k, defined by:

$$T_k(\cos\theta) := \cos(k\theta) \quad (\theta \in \mathbb{R}).$$

Note that $|T_k(x)| \leq 1$ for all $x \in [-1,1]$ and that $T_0 = 1$. The Chebyshev polynomials satisfy the orthogonality relations:

$$\langle T_a, T_b \rangle_\mu = \int_{-1}^{1} T_a(x) T_b(x) d\mu(x) = \begin{cases} 0 & a \neq b, \\ 1 & a = b = 0, \\ \frac{1}{2} & a = b \neq 0. \end{cases} \tag{8}$$

A univariate polynomial p may therefore be expanded as:

$$p = p_0 + \sum_{k=1}^{\deg(p)} 2 p_k T_k, \quad \text{where } p_k := \langle T_k, p \rangle_\mu.$$

For $\kappa \in \mathbb{N}^n$, we consider the *multivariate* Chebyshev polynomial T_κ, defined by setting:

$$T_\kappa(\mathbf{x}) := \prod_{i=1}^{n} T_{\kappa_i}(x_i).$$

The multivariate Chebyshev polynomials form a basis for $\mathbb{R}[\mathbf{x}]$ and satisfy the orthogonality relations:

$$\langle T_\alpha, T_\beta \rangle_\mu = \int_{\mathrm{B}^n} T_\alpha(\mathbf{x}) T_\beta(\mathbf{x}) d\mu(\mathbf{x}) = \begin{cases} 0 & \alpha \neq \beta, \\ 1 & \alpha = \beta = 0, \\ 2^{-w(\alpha)} & \alpha = \beta \neq 0. \end{cases} \tag{9}$$

Here, $w(\alpha) := |\{i \in [n] : \alpha_i \neq 0\}|$ denotes the Hamming weight of $\alpha \in \mathbb{N}^n$.

We use the notation $\mathbb{N}^n_d \subseteq \mathbb{N}^n$ to denote the set of n-tuples $\alpha \in \mathbb{N}^n$ with $|\alpha| = \sum_{i=1}^n \alpha_i \leq d$. As in the univariate case, we may expand any n-variate polynomial p as:

$$p = \sum_{\kappa \in \mathbb{N}^n_{\deg(p)}} 2^{w(\kappa)} p_\kappa T_\kappa, \quad \text{where } p_\kappa := \langle T_\kappa, p \rangle_\mu. \tag{10}$$

2.3 The Jackson kernel

For $r \in \mathbb{N}$ and for coefficients $\lambda^r_k \in \mathbb{R}$ to be specified below in (12), consider the kernel $K^{\mathrm{ja}}_r : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by:

$$K^{\mathrm{ja}}_r(x, y) := 1 + 2 \sum_{k=1}^{r} \lambda^r_k T_k(x) T_k(y). \tag{11}$$

We associate a linear operator $\mathbf{K}^{\mathrm{ja}}_r : \mathbb{R}[x]_r \to \mathbb{R}[x]_r$ to this kernel by setting:

$$\mathbf{K}^{\mathrm{ja}}_r p(x) := \int_{-1}^{1} K^{\mathrm{ja}}_r(x, y) p(y) d\mu(y) \quad (p \in \mathbb{R}[x]_r).$$

Using the orthogonality relations (8), and writing $\lambda^r_0 := 1$, we see that:

$$\mathbf{K}^{\mathrm{ja}}_r T_k(x) := \int_{-1}^{1} K^{\mathrm{ja}}_r(x, y) T_k(y) d\mu(y) = \lambda^r_k T_k(x) \quad (0 \leq k \leq r).$$

In other words, $\mathbf{K}^{\mathrm{ja}}_r$ is a diagonal operator with respect to the Chebyshev basis of $\mathbb{R}[x]_r$, and its eigenvalues are given by $\lambda^r_0 = 1, \lambda^r_1, \ldots, \lambda^r_r$. In what follows, we set:

$$\lambda^r_k = \frac{1}{r+2}\Big((r+2-k)\cos(k\theta_r) + \frac{\sin(k\theta_r)}{\sin(\theta_r)}\cos(\theta_r)\Big) \quad (1 \leq k \leq r), \tag{12}$$

with $\theta_r = \frac{\pi}{r+2}$. We then obtain the so-called *Jackson kernel* (see, e.g., [3]). The following properties of the Jackson kernel are crucial to our analysis.

Proposition 6 *For every $d, r \in \mathbb{N}$ with $d \leq r$, we have:*

(i) $K_r^{\mathrm{ja}}(x, y) \geq 0$ *for all* $x, y \in [-1, 1]$,
(ii) $1 \geq \lambda_k^r > 0$ *for all* $0 \leq k \leq r$, *and*
(iii) $|1 - \lambda_k^r| = 1 - \lambda_k^r \leq \frac{\pi^2 d^2}{(r+2)^2}$ *for all* $0 \leq k \leq d$.

Proof Nonnegativity of the Jackson kernel is a well-known fact, and is verified, e.g., in [6]. We check that the other properties (ii)-(iii) hold as well.

Second property (ii): Note that when $k \leq (r+2)/2$, both terms of (12) are positive, and so certainly $\lambda_k^r > 0$. So assume $(r+2)/2 < k \leq r$. Set $h = r + 2 - k$, so that $k\theta_r = \pi - h\theta_r$, $2 \leq h < (r+2)/2$, and

$$(r+2)\lambda_k^r = -h\cos(h\theta_r) + \frac{\sin(h\theta_r)}{\sin(\theta_r)}\cos(\theta_r). \tag{13}$$

It remains to show that the RHS of (13) is positive for all $2 \leq h < (r+2)/2$. Note that $1 > \cos(\theta_r) > 0$, $\sin(\theta_r) > 0$ and that $\sin(h\theta_r) \geq 0$ for all $2 \leq h < (r+2)/2$. We proceed by induction on h. For $h = 2$, we compute:

$$\begin{aligned} -h\cos(h\theta_r) + \frac{\sin(h\theta_r)}{\sin(\theta_r)}\cos(\theta_r) &= -2(2\cos^2(\theta_r) - 1) + 2\cos^2(\theta_r) \\ &= -2\cos^2(\theta_r) + 2 > 0, \end{aligned} \tag{14}$$

which settles the base of induction. For $h \geq 2$, we compute:

$$\begin{aligned} &-(h+1)\cos((h+1)\theta_r) + \sin((h+1)\theta_r)\frac{\cos(\theta_r)}{\sin(\theta_r)} \\ &= -(h+1)\big(\cos(h\theta_r)\cos(\theta_r) - \sin(h\theta_r)\sin(\theta_r)\big) \\ &\quad + \big(\sin(h\theta_r)\cos(\theta_r) + \cos(h\theta_r)\sin(\theta_r)\big)\frac{\cos(\theta_r)}{\sin(\theta_r)} \\ &= -h\cos(h\theta_r)\cos(\theta_r) + (h+1)\sin(h\theta_r)\sin(\theta_r) + \frac{\sin(h\theta_r)}{\sin(\theta_r)}\cos^2(\theta_r) \\ &= \underbrace{\cos(\theta_r)}_{>0}\underbrace{\Big(-h\cos(h\theta_r) + \frac{\sin(h\theta_r)}{\sin(\theta_r)}\cos(\theta_r)\Big)}_{\geq 0 \text{ by the induction assumption}} + (h+1)\underbrace{\sin(h\theta_r)\sin(\theta_r)}_{\geq 0} \\ &\geq 0. \end{aligned}$$

We conclude that $\lambda_k^r > 0$ for all $k \in [r]$. To see that $\lambda_k^r \leq 1$, note that for all $k \in \mathbb{N}$, $T_k(x) \leq 1$ for $-1 \leq x \leq 1$ and $T_k(1) = 1$. We can thus compute:

$$\lambda_k^r = \lambda_k^r T_k(1) = \int_{-1}^{1} K_r^{\mathrm{ja}}(1, y) T_k(y) d\mu(y) \leq \int_{-1}^{1} K_r^{\mathrm{ja}}(1, y) d\mu(y) = \lambda_0^r = 1,$$

making use of the nonnegativity of $K_r^{\mathrm{ja}}(x, y)$ on $[-1, 1]^2$ for the inequality.

Third property (iii): Using the expression of λ_r^k in (12) we have

$$1-\lambda_r^k = 1-\frac{r+2-k}{r+2}\cos(k\theta_r)-\frac{1}{r+2}\frac{\sin(k\theta_r)\cos(\theta_r)}{\sin(\theta_r)}.$$

We now bound each trigonometric term using the fact that:

$$\cos(x)\geq 1-\frac{1}{2}x^2,\quad x-\frac{1}{6}x^3\leq \sin(x)\leq x\quad (x\in\mathbb{R}). \tag{15}$$

When $k=1$ we immediately get:

$$1-\lambda_r^1 = 1-\cos(\theta_r)\leq \frac{1}{2}\theta_r^2=\frac{\pi^2}{2(r+2)^2}\leq\frac{d^2\pi^2}{(r+2)^2}.$$

Assume now $2\leq k\leq d$. Using (15) combined with $\cos(\theta_r), \sin(\theta_r), \sin(k\theta_r)>0$ we obtain:

$$\frac{\sin(k\theta_r)\cos(\theta_r)}{\sin(\theta_r)}\geq \left(k\theta_r-\frac{1}{6}k^3\theta_r^3\right)\left(1-\frac{1}{2}\theta_r^2\right)\frac{1}{\theta_r}\geq k-\frac{k}{2}\theta_r^2\left(1+\frac{k^2}{3}\right)$$

and thus:

$$\begin{aligned}1-\lambda_r^k &\leq 1-\frac{r+2-k}{r+2}\left(1-\frac{k^2\theta_r^2}{2}\right)-\frac{1}{r+2}\left(k-\frac{k}{2}\theta_r^2\left(1+\frac{k^2}{3}\right)\right)\\ &= \underbrace{\frac{r+2-k}{r+2}}_{\leq 1}\frac{k^2\theta_r^2}{2}+\underbrace{\frac{k}{2(r+2)}}_{\leq 1/2}\theta_r^2\underbrace{\left(1+\frac{k^2}{3}\right)}_{\leq\frac{2}{3}k^2 \text{ if } k\geq 2}\\ &\leq k^2\theta_r^2\leq\frac{d^2\pi^2}{(r+2)^2}.\end{aligned}$$

This concludes the proof if $k\geq 2$. □

3 Proof of the main theorem

3.1 Construction of the linear operator $\mathbf{K}_r$

As noted before, in order to prove Theorem 2 it suffices to construct a linear operator $\mathbf{K}_r : \mathbb{R}[\mathbf{x}]_r\to\mathbb{R}[\mathbf{x}]_r$ that is nonsingular and satisfies (P1) and (P2). For this purpose we define the multivariate Jackson kernel $K_r : \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ by setting:

$$K_r(\mathbf{x},\mathbf{y}) := \prod_{i=1}^n K_r^{\mathrm{ja}}(x_i,y_i), \tag{16}$$

where K_r^{ja} is the (univariate) Jackson kernel from (11). Now let $\mathbf{K}_r$ be the corresponding kernel operator defined by:

$$\mathbf{K}_r p(\mathbf{x}) = \int_{\mathbf{y}\in \mathrm{B}^n} K_r(\mathbf{x},\mathbf{y})p(\mathbf{y})d\mu(\mathbf{y}) \quad (p \in \mathbb{R}[\mathbf{x}]_r).$$

The operator $\mathbf{K}_r$ is diagonal w.r.t. the (multivariate) Chebyshev basis, and its eigenvalues can be expressed in terms of the coefficients λ_k^r of the Jackson kernel, as the following lemma shows.

Lemma 7 *The operator* $\mathbf{K}_r$ *is diagonal w.r.t. the Chebyshev basis for* $\mathbb{R}[\mathbf{x}]_r$, *and its eigenvalues are given by*:

$$\lambda_\kappa^r := \prod_{i=1}^n \lambda_{\kappa_i}^r \quad (\kappa \in \mathbb{N}_r^n).$$

Proof For $\kappa \in \mathbb{N}_r^n$, we see that:

$$\begin{aligned}\mathbf{K}_r T_\kappa(\mathbf{x}) &= \int_{\mathbf{y}\in \mathrm{B}^n} K_r(\mathbf{x},\mathbf{y})T_\kappa(\mathbf{y})d\mu(\mathbf{y})\\ &= \prod_{i=1}^n \left(\int_{y_i\in[-1,1]} K_r^{\mathrm{ja}}(x_i,y_i)T_{\kappa_i}(y_i)d\mu(y_i)\right) = \prod_{i=1}^n \lambda_{\kappa_i}^r T_{\kappa_i}(x_i) = \lambda_\kappa^r T_\kappa(\mathbf{x}),\end{aligned}$$

as required. □

It follows immediately from Proposition 6(ii) that $\mathbf{K}_r$ has only nonzero eigenvalues and thus is non-singular. We show that $\mathbf{K}_r$ further satisfies (P1) and (P2).

3.2 Verification of property (P1)

Consider the following strengthening of Schmüdgen's Positivstellensatz in the univariate case.

Theorem 8 (Fekete, Markov-Lukácz (see [12])) *Let* p *be a univariate polynomial of degree* r, *and assume that* $p \geq 0$ *on the interval* $[-1,1]$. *Then* p *admits a representation of the form*:

$$p(x) = \sigma_0(x) + \sigma_1(x)(1-x^2), \tag{17}$$

where $\sigma_0, \sigma_1 \in \Sigma[x]$ *and* σ_0 *and* $\sigma_1 \cdot (1-x^2)$ *are of degree at most* $r+1$. *In other words*, *in view of* (4), *we have* $p \in Q([-1,1])_{r+1}$.

By Proposition 6(i), for any $y \in [-1,1]$, the polynomial $x \mapsto K_r^{\mathrm{ja}}(x,y)$ is nonnegative on $[-1,1]$ and thus, by Theorem 8, it belongs to $Q([-1,1])_{r+1}$. This implies directly that the multivariate polynomial $\mathbf{x} \mapsto K_r(\mathbf{x},\mathbf{y}) = \prod_{i=1}^n K_r^{\mathrm{ja}}(x_i,y_i)$ belongs to $Q(\mathrm{B}^n)_{(r+1)n}$ for all $\mathbf{y} \in [-1,1]^n$.

Lemma 9 *The operator* $\mathbf{K}_r$ *satisfies property* (*P*1), *that is, we have* $\mathbf{K}_r p \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n}$ *for all* $p \in \mathcal{P}(\mathrm{B}^n)_r$.

Proof One way to see this is as follows. Let $\{\mathbf{y}_i : i \in [N]\} \subseteq \mathrm{B}^n$ and $w_i > 0$ ($i \in [N]$) form a quadrature rule for integration of degree $2r$ polynomials over B^n; that is, $\int_{\mathrm{B}^n} p(\mathbf{x}) d\mu(\mathbf{x}) = \sum_{i=1}^N w_i p(\mathbf{y}_i)$ for any $p \in \mathbb{R}[\mathbf{x}]_{2r}$. Then, for any $p \in \mathcal{P}(\mathrm{B}^n)_r$, we have $\mathbf{K}_r p(\mathbf{x}) = \sum_{i=1}^N K_r(\mathbf{x}, \mathbf{y}_i) p(\mathbf{y}_i) w_i$ with $p(\mathbf{y}_i) w_i \geq 0$ for all i, which shows that $\mathbf{K}_r p \in \mathcal{Q}(\mathrm{B}^n)_{(r+1)n}$. □

3.3 Verification of property (P2)

We may decompose the polynomial $\tilde{F} = F + \epsilon$ into the multivariate Chebyshev basis (10):

$$\tilde{F} = \epsilon + \sum_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)} F_\kappa T_\kappa, \quad \text{where } F_\kappa = \langle F, T_\kappa \rangle_\mu.$$

By Lemma 7, we then have:

$$\begin{aligned} \|\mathbf{K}_r^{-1} \tilde{F} - \tilde{F}\|_\infty &= \| \sum_{\kappa \in \mathbb{N}_d^n} (1/\lambda_\kappa^r) 2^{w(\kappa)} F_\kappa T_\kappa - 2^{w(\kappa)} F_\kappa T_\kappa \|_\infty \\ &\leq \sum_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)} |F_\kappa| |1 - 1/\lambda_\kappa^r|, \end{aligned} \tag{18}$$

making use of the fact that $\lambda_0 = 1$ and $|T_\kappa(x)| \leq 1$ for all $x \in \mathrm{B}^n$. It remains to analyze the expression at the right-hand side of (18). First, we bound the size of $|F_\kappa|$ for $\kappa \in \mathbb{N}^n$.

Lemma 10 *We have* $|F_\kappa| = |\langle F, T_\kappa \rangle_\mu| \leq 2^{-w(\kappa)/2}$ *for all* $\kappa \in \mathbb{N}^n$.

Proof Since μ is a probability measure on B^n, we have $\|F\|_\mu \leq \|F\|_\infty \leq 1$. Using the Cauchy-Schwarz inequality and (9), we then find:

$$\langle F, T_\kappa \rangle_\mu \leq \|F_\kappa\|_\mu \|T_\kappa\|_\mu \leq \|T_\kappa\|_\mu = 2^{-w(\kappa)/2}.$$

□

To bound the parameter $|1 - 1/\lambda_\kappa^r|$, we first prove a bound on $|1 - \lambda_\kappa^r|$, which we obtain by applying Bernoulli's inequality.

Lemma 11 (Bernoulli's inequality) *For any* $x \in [0, 1]$ *and* $t \geq 1$, *we have*:

$$1 - (1 - x)^t \leq tx. \tag{19}$$

Lemma 12 *For any* $\kappa \in \mathbb{N}_d^n$ *and* $r \geq \pi d$, *we have*:

$$|1-\lambda_\kappa^r| \leq \frac{n\pi^2 d^2}{r^2}.$$

Proof By Proposition 6, we know that $0 \leq \gamma_k := (1-\lambda_k^r) \leq \pi^2 d^2/r^2 \leq 1$ for $0 \leq k \leq d$. Writing $\gamma := \max_{0 \leq k \leq d} \gamma_k$, we compute:

$$1-\lambda_\kappa^r = 1-\prod_{i=1}^n \lambda_{\kappa_i}^r = 1-\prod_{i=1}^n (1-\gamma_{\kappa_i}) \leq 1-(1-\gamma)^n \leq n\gamma \leq \frac{n\pi^2 d^2}{r^2},$$

making use of (19) for the second to last inequality. □

Lemma 13 *Assuming that* $r \geq \pi d\sqrt{2n}$, *we have*:

$$|1-1/\lambda_\kappa^r| \leq \frac{2n\pi^2 d^2}{r^2}.$$

Proof Under the assumption, and using the previous lemma, we have $|1-\lambda_\kappa^r| \leq 1/2$, which implies that $\lambda_\kappa^r \geq 1/2$. We may then bound:

$$|1-1/\lambda_\kappa^r| = |\frac{1-\lambda_\kappa^r}{\lambda_\kappa^r}| \leq 2|1-\lambda_\kappa^r| \leq \frac{2n\pi^2 d^2}{r^2}.$$

□

Putting things together and using (18), Lemma 10 and Lemma 12 we find that:

$$\begin{aligned}\|\mathbf{K}_r^{-1}\tilde{F}-\tilde{F}\|_\infty &\leq \sum_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)}|F_\kappa||1-1/\lambda_\kappa^r| \\ &\leq \sum_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)/2} \cdot \frac{2n\pi^2 d^2}{r^2} \leq |\mathbb{N}_d^n| \cdot \max_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)/2} \cdot \frac{2n\pi^2 d^2}{r^2}.\end{aligned}$$

Hence $\mathbf{K}_r$ satisfies (P2) with $\epsilon = C(n,d)/r^2$, where:

$$C(n,d) := |\mathbb{N}_d^n| \cdot \max_{\kappa \in \mathbb{N}_d^n} 2^{w(\kappa)/2} \cdot 2n\pi^2 d^2.$$

In view of Lemma 4, we have thus proven Theorem 2. Finally, we can bound the constant $C(n, d)$ in two ways. On the one hand, we have:

$$|\mathbb{N}_d^n| = \binom{n+d}{n} = \prod_{i=1}^n \frac{d+i}{i} \leq (d+1)^n \text{ and } \max_{\kappa \in \mathbb{N}_d^n} w(\kappa) \leq n,$$

resulting in a polynomial dependence of $C(n, d)$ on d for fixed n. On the other hand, we have:

$$|\mathbb{N}_d^n| = \binom{n+d}{d} \leq (n+1)^d \text{ and } \max_{\kappa \in \mathbb{N}_d^n} w(\kappa) \leq d,$$

resulting in a polynomial dependence of $C(n, d)$ on n for fixed d. Namely, we have:

$$C(n,d) \leq 2\pi^2 d^2 n 2^{n/2}(d+1)^n \quad \text{and} \quad C(n,d) \leq 2\pi^2 d^2 n 2^{d/2}(n+1)^d. \tag{20}$$

4 Concluding remarks

We have shown that the error of the degree r Lasserre-type bound (5) for the minimization of a polynomial over the hypercube $[-1,1]^n$ is of the order $O(1/r^2)$ when using a sum-of-squares decomposition in the truncated preordering. Alternatively, if f is a polynomial nonnegative on $[-1,1]^n$ and $\eta > 0$, our result may be interpreted as showing a bound in $O(1/\sqrt{\eta})$ on the degree of a Schmüdgen-type certificate of positivity for $f + \eta$. The dependence on the dimension n and the degree d of f in the constants of our result is both polynomial in n (for fixed d), and polynomial in d (for fixed n).

4.1 The constant $C(n, d)$

A question left open in this work is whether it is possible to show Theorem 2 with a constant $C(d)$ that only depends on the degree d of f, and not on the number of variables n (cf. (20)). This question is motivated by the fact that for the analysis of the analogous hierarchies for the unit sphere in [4] and for the boolean hypercube in [5] the existence of such a constant (depending only on d) was in fact shown.

4.2 Relation to recent developments

Recently, there has been growing interest in obtaining a sharper convergence analysis for various Lasserre-type hierarchies for the minimization of a polynomial f over a semialgebraic set $S = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0 \ \ (j \in [m])\}$. Our work thus contributes to this research area. We outline some recent developments.

We refer to the works [13, 14] (and further references therein) for the analysis of hierarchies of upper bounds (obtained by minimizing the expected value of f on S with respect to a sum-of-squares density).

The most commonly used hierarchies of lower bounds are defined in terms of sums-of-squares decompositions in the *quadratic module* of S, being the set of conic combinations of the form $\sigma_0 + \sum_{j=1}^m \sigma_j g_j$ with $\sigma_j \in \Sigma[\mathbf{x}]$. Such decompositions are called *Putinar-type* certificates. In comparison, the *preordering* $Q(S)$ also involves conic combinations of the *products* of the g_j. In [15] a degree bound in $O(\exp(\eta^{-c}))$ is given for the quadratic module, where $c > 0$ is a constant depending on S.

In a recent work [16], Baldi & Mourrain are able to improve this result to obtain a bound with a polynomial dependency on η. Roughly speaking, their method of proof relies on embedding the semialgebraic set S in a box $[-R, R]^n$ of large enough size $R > 0$, and then relating positivity certificates on S to those on $[-R, R]^n$. Our present result on $[-1, 1]^n$ then allows them to conclude their analysis. Their argument relies on the fact the constant $C(n, d)$ in Theorem 2 may be chosen to depend polynomially on the degree d of f. Such a dependence was not shown in the earlier work [10].

Note that it has been shown in [17] that the hierarchies of bounds based on Putinar type representations have *finite* convergence for *generic* problems. However, and perhaps somewhat surprisingly, their convergence analysis (for general problems) has remained a challenging problem.

We also wish to note that a polynomial degree bound was shown already in [18] for a slightly different hierarchy, based on Putinar-Vasilescu type representations, which give a decomposition in the quadratic module after multiplying the polynomial $f + \eta$ by a suitable power $(1 + \sum_{i=1}^{n} \|\mathbf{x}\|^2)^k$ (under some conditions).

4.3 Putinar vs. Schmüdgen on the hypercube

As mentioned, Putinar-type hierarchies (making use of the quadratic module) are more commonly applied in practice than the Schmüdgen-type hierarchy (making use of the preordering) that we consider in this paper. It is therefore natural to consider the status of convergence results for Putinar-type hierarchies on the hypercube B^n.

Magron [19] shows a degree bound in $O(\exp(c\eta^{-1}))$ for Putinar-type certificates of $f + \eta$ on B^n, improving the general result of [15] in this special case[3]. His result relies on the degree bound in $O(\eta^{-1})$ for *Schmüdgen*-type certificates on B^n shown in [10]. Importantly, it is contingent on an unresolved conjecture also posed in [10]: For each $n \in \mathbb{N}$ even, the polynomial $2^{-n}(1 - x_1)(1 - x_2) \dots (1 - x_n) + \eta$ lies in the quadratic module of B^n truncated at degree n for $\eta = \frac{1}{n(n+2)}$. This open question, which asks for an exact estimation of the constant that needs to be added to each generator of the preordering of $Q([-1, 1]^n)$ in order to ensure membership in the quadratic module, remains interesting in itself.

In principle, our new degree bounds for Schmüdgen-type certificates on B^n could (slightly) improve the result of Magron (which relies on the weaker bounds of [10]). However, such an improvement would still depend exponentially on $1/\eta$, in addition to being contingent on a conjecture. Furthermore, it seems to us that it is in any case superseded by the new result of Baldi & Mourrain [16] mentioned above, which (when specialized to the hypercube) shows degree bounds for Putinar-type certificates with *polynomial* dependency on $1/\eta$. It is an open question whether the degree bound in $O(1/\sqrt{\eta})$ we have shown here for Schmüdgen-type certificates on B^n may be extended to Putinar-type certificates.

Lastly, we wish to mention that error bounds for the Putinar-type Lasserre hierarchy on the hypercube B^n were already provided in [20]. There, however, the author considers a regime where the order r of the relaxation is fixed, while the dimension n tends to infinity. His results are therefore not directly comparable to those of the present paper or to those discussed above.

[3] The cube $[0, 1]^n$ is considered in [19], but all results carry over immediately to $[-1, 1]^n$ after an affine change of variables.

4.4 Negative results

We have so far focused our discussion on *positive* results concerning sum-of-squares representations. That is, results that give *upper* bounds on the error of Lasserre's bound (5); or equivalently on the required degree of Schmüdgen-type positivity certificates. In order to put these results in context, it would be interesting to have complementary *negative* results, thus giving *lower* bounds on the convergence rate of the Lasserre hierarchy.

The only applicable negative result known to the authors is due to Stengle [21]. He considers the interval $[-1,1] \subseteq \mathbb{R}$ with the semialgebraic description:

$$[-1,1] = \{x \in \mathbb{R} \,:\, (1-x^2)^3 \geq 0\}.$$

Note that this description is different from the (more natural) description (3) that we have used in this paper. In particular, Theorem 8 does not apply to it. Writing $Q((1-x^2)^3)_r$ for the corresponding (truncated) preordering, Stengle shows that

$$1 - x^2 + \eta \in Q((1-x^2)^3)_r$$

only when $r = \Omega(1/\sqrt{\eta})$. In other words, he shows for $f(x) = 1 - x^2$ that the Lasserre-type bound $f_{(r)}$ obtained by replacing $Q(1-x^2)_r$ in (5) by $Q((1-x^2)^3)_r$ satisfies:

$$f_{\min} - f_{(r)} = \Omega(1/r^2).$$

On the one hand, it is remarkable that Stengle's lower bound in $\Omega(1/r^2)$ matches the upper bound in $O(1/r^2)$ we show in this paper exactly. On the other hand, we emphasize that Stengle's result relies heavily on the nonstandard description of $[-1,1]$ as a semialgebraic set. We leave the question of proving negative results for the standard description (3) for future research.

Acknowledgements We thank Lorenzo Baldi and Bernard Mourrain for their helpful suggestions. We also thank Etienne de Klerk and Felix Kirschner for useful discussions. Furthermore, we thank Markus Schweighofer for bringing to our attention the paper of Stengle [21]. We are grateful to two reviewers for their careful reading and useful suggestions; in particular we thank the referees for bringing the papers [19, 20] to our attention.

Funding This work is supported by the European Union's Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No. 764759 (MINOA).

Availability of data and material Not applicable.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

1. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
2. Schmüdgen, K.: The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**(2), 203–206 (1991)
3. Weiße, A., Wellein, G., Alvermann, A., Fehske, H.: The kernel polynomial method. Rev. Mod. Phys. **78**, 275 (2006)
4. Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01537-7
5. Slot, L., Laurent, M.: Sum-of-squares hierarchies for binary polynomial optimization. Singh M., Williamson D.P. (eds) Integer Programming and Combinatorial Optimization (IPCO 2021). Lecture Notes in Computer Science
6. de Klerk, E., Hess, R., Laurent, M.: Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization. SIAM J. Opti. **27**(1), 347–367 (2017)
7. Schweighofer, M.: On the complexity of Schmüdgen's Positivstellensatz. J. Complex. **20**(4), 529–543 (2004)
8. Pólya, G.: Uber positive Darstellung von Polynomen. Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich, 73, 14–145, 1928. Reprinted. In: Collected Papers. Vol 2, pp. 309–313. MIT Press, Cambridge (1974)
9. Powers, V., Reznick, B.: A new bound for Pólya's Theorem with applications to polynomials positive on polyhedra. J. Pure Appl. Algebra **164**(1–2), 221–229 (2001)
10. de Klerk, E., Laurent, M.: Error bounds for some semidefinite programming approaches to polynomial minimization on the hypercube. SIAM J. on Optim. **20**(6), 3104–3120 (2010)
11. Szegö, G.: Orthogonal Polynomials. vol. 23 in American Mathematical Society colloquium publications
12. Powers, V., Reznick, B.: Polynomials that are positive on an interval. Trans. Amer. Math. Soc. **352**, 4677–4692 (2000)
13. de Klerk, E., Laurent, M.: Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube. Math. Oper. Res. **45**(1), 86–98 (2020)
14. Slot, L., Laurent, M.: Near-optimal analysis of univariate moment bounds for polynomial optimization. Math. Program. **188**, 443–460 (2021)
15. Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. J. Complex. **23**(1), 135–150 (2007)
16. Baldi, L., Mourrain, B.: On moment approximation and the effective Putinar's Positivstellensatz, arXiv:2111.11258, 2021
17. Nie, J.: Optimality conditions and finite convergence of Lasserre's hierarchy. Math. program. **146**(1), 97–121 (2014)
18. Mai, N.H.A., Magron V.: On the complexity of Putinar-Vasilescu's Positivstellensatz. Journal of Complexity, 10.1016/j.jco.2022.101663
19. Magron, V.: Error bounds for polynomial optimization over the hypercube using Putinar type representations. arXiv:1404.6145, 2014
20. Nie, J.: An approximation bound analysis for Lasserre's relaxation. J. Oper. Res. Soc. China **1**(3), 313–332 (2013)
21. Stengle, G.: Complexity Estimates for the Schmüdgen Positivstellensatz. J. Complex. **12**, 167–174 (1996)