FULL LENGTH PAPER

Series B

Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets

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

Abstract

We consider the problem of computing the minimum value $f_{\min, K}$ of a polynomial f over a compact set $K \subseteq \mathbb{R}^n$, which can be reformulated as finding a probability measure ν on K minimizing $\int_K f \, d\nu$. Lasserre showed that it suffices to consider such measures of the form $\nu = q\mu$, where q is a sum-of-squares polynomial and μ is a given Borel measure supported on K. By bounding the degree of q by $2r$ one gets a converging hierarchy of upper bounds $f^{(r)}$ for $f_{\min, K}$. When K is the hypercube $[-1, 1]^n$, equipped with the Chebyshev measure, the parameters $f^{(r)}$ are known to converge to $f_{\min, K}$ at a rate in $O(1/r^2)$. We extend this error estimate to a wider class of convex bodies, while also allowing for a broader class of reference measures, including the Lebesgue measure. Our analysis applies to simplices, balls and convex bodies that locally look like a ball. In addition, we show an error estimate in $O(\log r/r)$ when K satisfies a minor geometrical condition, and in $O(\log^2 r/r^2)$ when K is a convex body, equipped with the Lebesgue measure. This improves upon the currently best known error estimates in $O(1/\sqrt{r})$ and $O(1/r)$ for these two respective cases.

Keywords Polynomial optimization · Sum-of-squares polynomial · Lasserre hierarchy · Semidefinite programming · Needle polynomial

Mathematics Subject Classification 90C22 · 90C26 · 90C30

✉ Monique Laurent
Monique.Laurent@cwi.nl

Lucas Slot
Lucas.Slot@cwi.nl

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

[2] Tilburg University, Tilburg, The Netherlands

🌀 Springer

1 Introduction

1.1 Lasserre's measure-based hierarchy

Let $K \subseteq \mathbb{R}^n$ be a compact set and let $f \in \mathbb{R}[x]$ be a polynomial. We consider the minimization problem

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

Computing $f_{\min, K}$ is a hard problem in general, and some well-known problems from combinatorial optimization are among its special cases. For example, it is shown in [13,24] that the stability number $\alpha(G)$ of a graph $G = ([n], E)$ is given by

$$\frac{1}{\alpha(G)} = \min_{x \in K} \sum_{i \in V} x_i^2 + 2 \sum_{\{i,j\} \in E} x_i x_j,$$

where we take $K = \{x \in \mathbb{R}^n : x \geq 0, \ \sum_{i=1}^n x_i = 1\}$ to be the standard simplex in $\mathbb{R}^n$.

Problem (1) may be reformulated as the problem of finding a probability measure ν on K for which the integral $\int_K f d\nu$ is minimized. Indeed, for any such ν we have $\int_K f d\nu \geq f_{\min, K} \int_K d\nu = f_{\min, K}$. On the other hand, if $a \in K$ is a global minimizer of f in K, then we have $\int_K f d\delta_a = f(a) = f_{\min, K}$, where δ_a is the Dirac measure centered at a.

Lasserre [21] showed that it suffices to consider measures of the form $\nu = q\mu$, where $q \in \Sigma$ is a sum-of-squares polynomial and μ is a (fixed) reference Borel measure supported by K. That is, we may reformulate (1) as

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

For each $r \in \mathbb{N}$ we may then obtain an upper bound $f_{K,\mu}^{(r)}$ for $f_{\min, K}$ by limiting our choice of q in (2) to polynomials of degree at most $2r$:

$$f_{K,\mu}^{(r)} := \inf_{q \in \Sigma_r} \int_K f(x) q(x) d\mu(x) \quad \text{s.t.} \int_K q(x) d\mu(x) = 1. \tag{3}$$

Here, Σ_r denotes the set of all sum-of-squares polynomials of degree at most $2r$. We shall also write $f^{(r)} = f_{K,\mu}^{(r)}$ for simplicity. As detecting sum-of-squares polynomials is possible using semidefinite programming, the program (3) can be modeled as an SDP [21]. Moreover, the special structure of this SDP allows a reformulation to an eigenvalue minimization problem [21], as will be briefly described below.

By definition, we have $f_{\min, K} \leq f^{(r+1)} \leq f^{(r)}$ for all $r \in \mathbb{N}$ and

$$\lim_{r \to \infty} f^{(r)} = f_{\min, K}.$$

In this paper we are interested in upper bounding the convergence rate of the sequence $(f^{(r)})_r$ to $f_{\min,K}$ in terms of r. That is, we wish to find bounds in terms of r for the parameter:

$$E_{K,\mu}^{(r)}(f) := f^{(r)} - f_{\min,K},$$

often also denoted $E^{(r)}(f)$ for simplicity when there is no ambiguity on K, μ.

1.2 Related work

Bounds on the parameter $E_{K,\mu}^{(r)}(f)$ have been shown in the literature for several different sets of assumptions on K, μ and f. Depending on these assumptions, two main strategies have been employed, which we now briefly discuss.

Algebraic analysis via an eigenvalue reformulation The first strategy relies on a reformulation of the optimization problem (3) as an eigenvalue minimization problem (see [12,21]). We describe it briefly, in the univariate case $n = 1$ only, for simplicity and since this is the case we need. Let $\{p_r \in \mathbb{R}[x]_r : r \in \mathbb{N}\}$ be the (unique) orthonormal basis of $\mathbb{R}[x]$ w.r.t. the inner product $\langle p_i, p_j \rangle = \int_K p_i p_j d\mu$. For each $r \in \mathbb{N}$, we then define the (generalized) *truncated moment matrix* $M_{r,f}$ of f by setting

$$M_{r,f}(i,j) := \int_K p_i p_j f d\mu \quad \text{for } 0 \leq i, j \leq r.$$

It can be shown that $f^{(r)} = \lambda_{\min}(M_{r,f})$, the smallest eigenvalue of the matrix $M_{r,f}$. Any bounds on the eigenvalues of $M_{r,f}$ thus immediately translate to bounds on $f^{(r)}$.

In [10], the authors determine the exact asymptotic behaviour of $\lambda_{\min}(M_{r,f})$ in the case that f is a quadratic polynomial, $K = [-1, 1]$ and $d\mu(x) = (1-x^2)^{-\frac{1}{2}}dx$, known as the Chebyshev measure. Based on this, they show that $E^{(r)}(f) = O(1/r^2)$ and extend this result to arbitrary multivariate polynomials f on the hypercube $[-1, 1]^n$ equipped with the product measure $d\mu(x) = \prod_i^n (1 - x_i)^{-1/2}dx_i$. In addition, they prove that $E^{(r)}(f) = \Theta(1/r^2)$ for linear polynomials, which thus shows that in some sense quadratic convergence is the best we can hope for. (This latter result is shown in [10] for all measures with Jacobi weight on $[-1, 1]$).

The orthogonal polynomials corresponding to the measure $(1 - x^2)^{-1/2}dx$ on $[-1, 1]$ are the Chebyshev polynomials of the first kind, denoted by T_r. They are well-studied objects (see, e.g., [25]). In particular, they satisfy the following *three-term recurrence* relation

$$T_0(x) = 1, T_1(x) = x, \text{ and } T_{r+1}(x) = 2xT_r(x) - T_{r-1}(x) \text{ for } r \geq 1. \quad (4)$$

This imposes a large amount of structure on the matrix $M_{r,f}$ when f is quadratic, which has been exploited in [10] to obtain information on its smallest eigenvalue.

The main disadvantage of the eigenvalue strategy is that it requires the moment matrix of f to have a closed form expression which is sufficiently structured so as to allow for an analysis of its eigenvalues. Closed form expressions for the entries of the

matrix $M_{r,f}$ are known only for special sets K, such as the interval $[-1, 1]$, the unit ball, the unit sphere, or the simplex, and only with respect to certain measures.

However, as we will see in this paper, the convergence analysis from [10] in $O(1/r^2)$ for the interval $[-1, 1]$ equipped with the Chebyshev measure, can be transported to a large class of compact sets, such as the interval $[-1, 1]$ with more general measures, the ball, the simplex, and 'ball-like' convex bodies.

Analysis via the construction of feasible solutions A second strategy to bound the convergence rate of the parameters $E^{(r)}(f)$ is to construct explicit sum-of-squares density functions $q_r \in \Sigma_r$ for which the integral $\int_K q_r f d\mu$ is close to $f_{\min, K}$. In contrast to the previous strategy, such constructions will only yield upper bounds on $E^{(r)}(f)$.

As noted earlier, the integral $\int_K f d\nu$ may be minimized by selecting the probability measure $\nu = \delta_a$, the Dirac measure at a global minimizer a of f on K. When the reference measure μ is the Lebesgue measure, it thus intuitively seems sensible to consider sum-of-squares densities q_r that approximate the Dirac delta in some way.

This approach is followed in [12]. There, the authors consider truncated Taylor expansions of the Gaussian function $e^{-t^2/2\sigma}$, which they use to define the sum-of-squares polynomials

$$\phi_r(t) = \sum_{k=0}^{2r} \frac{1}{k!} \left(\frac{-t^2}{2\sigma} \right)^k \in \Sigma_{2r} \quad \text{for } r \in \mathbb{N}.$$

Setting $q_r(x) \sim \phi_r(\|x - a\|)$ for carefully selected standard deviation $\sigma = \sigma(r)$, they show that $\int_K f(x)q_r(x)dx - f(a) = O(1/\sqrt{r})$ when K satisfies a minor geometrical assumption (Assumption 1 below), which holds, e.g., if K is a convex body or if it is star-shaped with respect to a ball.

In subsequent work [8], the authors show that if K is assumed to be a convex body, then a bound in $O(1/r)$ may be obtained by setting $q_r \sim \phi_r(f(x))$. As explained in [8], the sum-of-squares density q_r in this case can be seen as an approximation of the Boltzman density function for f, which plays an important role in simulated annealing.

The advantage of this second strategy seems to be its applicability to a broad class of sets K with respect to the natural Lebesgue measure. This generality, however, is so-far offset by significantly weaker bounds on $E^{(r)}(f)$. Another main contribution of this paper will be to show improved bounds on $E^{(r)}(f)$ for this broad class of sets K.

Analysis for the hypersphere Tight results are known for polynomial minimization on the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \sum_i x_i^2 = 1\}$, equipped with the Haar surface measure. Doherty and Wehner [15] have shown a convergence rate in $O(1/r)$, by using harmonic analysis on the sphere and connections to quantum information theory. In the very recent work [11], the authors show an improved convergence rate in $O(1/r^2)$, by using a reduction to the case of the interval $[-1, 1]$ and the above mentioned convergence rate in $O(1/r^2)$ for this case. This reduction is based on replacing f by an easy (linear) upper estimator. This idea was already exploited in [8,12] (where a quadratic upper estimator was used) and we will also exploit it in this paper.

1.3 Our contribution

The contribution of this paper is showing improved bounds on the convergence rate of the parameters $E_{K,\mu}^{(r)}(f)$ for a wide class of sets K and measures μ. It is twofold.

Firstly, we extend the known bound from [10] in $O(1/r^2)$ for the hypercube $[-1, 1]^n$ equipped with the Chebyshev measure, to a wider class of convex bodies. Our results hold for the ball B^n, the simplex Δ^n, and 'ball-like' convex bodies (see Definition 3) equipped with the Lebesgue measure. For the ball and hypercube, they further hold for a wider class of measures; namely for the measures given by

$$w_\lambda(x)dx := \left(1 - \|x\|^2\right)^\lambda dx \quad (\lambda \geq 0)$$

on the ball, and the measures

$$\widehat{w_\lambda}(x)dx := \prod_{i=1}^{n} \left(1 - x_i^2\right)^\lambda dx \quad \left(\lambda \geq -\frac{1}{2}\right)$$

on the hypercube. Note that for the hypercube, setting $\lambda = -\frac{1}{2}$ yields the Chebyshev measure, and that for both the ball and the hypercube, setting $\lambda = 0$ yields the Lebesgue measure. The rate $O(1/r^2)$ also holds for any compact K equipped with the Lebesgue measure under the assumption of existence of a global minimizer in the interior of K. These results are presented in Sect. 3.

Secondly, we improve the known bounds in $O(1/\sqrt{r})$ and $O(1/r)$ for general compact sets (under Assumption 1) and convex bodies equipped with the Lebesgue measure, established in [8,12], respectively. For general compact sets, we prove a bound in $O(\log r/r)$, and for convex bodies we show a bound in $O(\log^2 r/r^2)$. These results are exposed in Sect. 4.

For our results in Sect. 3, we will use several tools that will enable us to reduce to the case of the interval $[-1, 1]$ equipped with the Chebyshev measure. These tools are presented in Sects. 2 and 3. They include: (a) replacing K by an affine linear image of it (Sect. 2.3); (b) replacing f by an upper estimator (easier to analyze, obtained via Taylor's theorem) (Sect. 2.4); (c) transporting results between two comparable weight functions on K and between two convex sets $K, \widehat{K}$ which look locally the same in the neighbourhood of a global minimizer (Sects. 3.1, 3.2). In particular, the result of Proposition 1 will play a key role in our treatment.

To establish our results in Sect. 4 we will follow the second strategy sketched above, namely we will define suitable sum-of-squares polynomials that approximate well the Dirac delta at a global minimizer. However, instead of using truncations of the Taylor expansion of the Gaussian function or of the Boltzman distribution as was done in [8,12], we will now use the so-called needle polynomials from [19] (constructed from the Chebyshev polynomials, see Sect. 4.1). In Table 1 we provide an overview of both known and new results.

Finally, we illustrate some of the results in Sects. 3 and 4 with numerical examples in Sect. 5.

Table 1 Known and new convergence rates for the Lasserre hierarchy of upper bounds

$K \subseteq \mathbb{R}^n$ compact	$E_{K,\mu}^{(r)}(f)$	μ	Reference
General	$o(1)$	Borel	[21]
Assumption 1	$O(1/\sqrt{r})$	Lebesgue	[12]
Convex body	$O(1/r)$	Lebesgue	[8]
Hypersphere	$O(1/r)$	Haar	[15]
Hypersphere	$O(1/r^2)$	Haar	[11]
Hypercube	$O(1/r^2)$	Chebyshev	[10]
Hypercube	$O(1/r^2)$	$\widehat{w_\lambda}(x)dx(\lambda \geq -1/2)$	Theorem 3
Ball	$O(1/r^2)$	$w_\lambda(x)dx(\lambda \geq 0)$	Theorem 4
Simplex	$O(1/r^2)$	Lebesgue	Theorem 9
Ball-like convex body	$O(1/r^2)$	Lebesgue	Theorem 6
Global minimizer in the interior	$O(1/r^2)$	Lebesgue	Theorem 5
Assumption 1	$O(\log r/r)$	Lebesgue	Theorem 10
Convex body	$O(\log^2 r/r^2)$	Lebesgue	Theorem 11

2 Preliminaries

In this section, we first introduce some notation that we will use throughout the rest of the paper and recall some basic terminology and results about convex bodies. We then show that the error $E^{(r)}(f)$ is invariant under nonsingular affine transformations of $\mathbb{R}^n$. Finally, we introduce the notion of *upper estimators* for f. Roughly speaking, this tool will allow us to replace f in the analysis of $E^{(r)}(f)$ by a simpler function (usually a quadratic, separable polynomial). We will make use of this extensively in both Sects. 3 and 4.

2.1 Notation

For $x, y \in \mathbb{R}^n$, $\langle x, y \rangle$ denotes the standard inner product and $\|x\|^2 = \langle x, x \rangle$ the corresponding norm. We write $B_\rho^n(c) := \{x \in \mathbb{R}^n : \|x-c\| \leq \rho\}$ for the n-dimensional ball of radius ρ centered at c. When $\rho = 1$ and $c = 0$, we also write $B^n := B_1^n(0)$.

Throughout, $K \subseteq \mathbb{R}^n$ is always a compact set with non-empty interior, and f is an n-variate polynomial. We let $\nabla f(x)$ (resp., $\nabla^2 f(x)$) denote the gradient (resp., the Hessian) of f at $x \in \mathbb{R}^n$, and introduce the parameters

$$\beta_{f,K} := \max_{x \in K} \|\nabla f(x)\| \quad \text{and} \quad \gamma_{f,K} := \frac{1}{2} \max_{x \in K} \|\nabla^2 f(x)\|. \tag{5}$$

Whenever we write an expression of the form

$$\text{``}E^{(r)}(f) = O\left(1/r^2\right)\text{''},$$

we mean that there exists a constant $c > 0$ such that $E^{(r)}(f) \leq c/r^2$ for all $r \in \mathbb{N}$, where c depends only on K, μ, and the parameters $\beta_{f,K}$, $\gamma_{f,K}$. Some of our results are obtained by embedding K into a larger set $\widehat{K} \subseteq \mathbb{R}^n$. If this is the case, then c may depend on $\beta_{f,\widehat{K}}$, $\gamma_{f,\widehat{K}}$ as well. If there is an additional dependence of c on the global minimizer a of f on K, we will make this explicit by using the notation "O_a".

2.2 Convex bodies

Let $K \subseteq \mathbb{R}^n$ be a convex body, i.e., a compact, convex set with non-empty interior. We say $v \in \mathbb{R}^n$ is an (inward) normal of K at $a \in K$ if $\langle v, x - a \rangle \geq 0$ holds for all $x \in K$. We refer to the set of all normals of K at a as the normal cone, and write

$$N_K(a) := \{v \in \mathbb{R}^n : \langle v, x - a \rangle \geq 0 \text{ for all } x \in K\}.$$

We will make use of the following basic result.

Lemma 1 (e.g., [2, Prop. 2.1.1]) *Let K be a convex body and let $g : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with local minimizer $a \in K$. Then $\nabla g(a) \in N_K(a)$.*

Proof Suppose not. Then, by definition of $N_K(a)$, there exists an element $y \in K$ such that $\langle \nabla g(a), y - a \rangle < 0$. Expanding the definition of the gradient this means that

$$0 > \langle \nabla g(a), y - a \rangle = \lim_{t \downarrow 0} \frac{g(ty + (1-t)a) - g(a)}{t},$$

which implies $g(ty + (1-t)a) < g(a)$ for all $t > 0$ small enough. But $ty + (1-t)a \in K$ by convexity, contradicting the fact that a is a local minimizer of g on K. $\qquad \square$

The set K is *smooth* if it has a unique unit normal $v(a)$ at each boundary point $a \in \partial K$. In this case, we denote by $T_a K$ the (unique) hyperplane tangent to K at a, defined by the equation $\langle x - a, v(a) \rangle = 0$.

For $k \geq 1$, we say K is of class C^k if there exists a convex function $\Psi \in C^k(\mathbb{R}^n, \mathbb{R})$ such that $K = \{x \in \mathbb{R}^n : \Psi(x) \leq 0\}$ and $\partial K = \{x \in \mathbb{R}^n : \Psi(x) = 0\}$. If K is of class C^k for some $k \geq 1$, it is automatically smooth in the above sense.

We refer, e.g., to [1] for a general reference on convex bodies.

2.3 Linear transformations

Suppose that $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a nonsingular affine transformation, given by $\phi(x) = Ux + c$. If q is a sum-of-squares density function w.r.t. the Lebesgue measure on $\phi(K)$, then we have

$$\int_{\phi(K)} q(y) f(\phi^{-1}(y)) dy = |\det U| \cdot \int_K q(\phi(x)) f(x) dx \quad \text{and}$$

$$1 = \int_{\phi(K)} q(y) dy = |\det U| \cdot \int_K q(\phi(x)) dx.$$

As a result, the polynomial $\widehat{q} := (q \circ \phi) / \int_K q(\phi(x)) dx = (q \circ \phi) \cdot |\det U|$ is a sum of squares density function w.r.t. the Lebesgue measure on K. It has the same degree as q, and it satisfies

$$\int_K \widehat{q}(x) f(y) dx = \int_{\phi(K)} q(x) f(\phi^{-1}(y)) dx.$$

We have just shown the following.

Lemma 2 *Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a non-singular affine transformation. Write $g := f \circ \phi^{-1}$. Then we have*

$$E_K^{(r)}(f) = E_{\phi(K)}^{(r)}(g).$$

2.4 Upper estimators

Given a point $a \in K$ and two functions $f, g : K \to \mathbb{R}$, we write $f \leq_a g$ if $f(a) = g(a)$ and $f(x) \leq g(x)$ for all $x \in K$; we then say that g is an *upper estimator* for f on K, which is *exact at a*. The next lemma, whose easy proof is omitted, will be very useful.

Lemma 3 *Let $g : K \to \mathbb{R}$ be an upper estimator for f, exact at one of its global minimizers on K. Then we have $E^{(r)}(f) \leq E^{(r)}(g)$ for all $r \in \mathbb{N}$.*

Remark 1 We make the following observations for future reference.

1. Lemma 3 tells us that we may always replace f in our analysis by an upper estimator which is exact at one of its global minimizers. This is useful if we can find an upper estimator that is significantly simpler to analyze.
2. We may always assume that $f_{\min, K} = 0$, in which case $f(x) \geq 0$ for all $x \in K$ and $E^{(r)}(f) = f^{(r)}$. Indeed, if we consider the function g given by $g(x) = f(x) - f_{\min, K}$, then $g_{\min, K} = 0$, and for every density function q on K, we have

$$\int_K g(x) q(x) d\mu(x) = \int_K f(x) q(x) d\mu(x) - f_{\min, K},$$

showing that $E^{(r)}(f) = E^{(r)}(g) = g^{(r)}$ for all $r \in \mathbb{N}$.

In the remainder of this section, we derive some general upper estimators based on the following variant of Taylor's theorem for multivariate functions.

Theorem 1 (Taylor's theorem) *For $f \in C^2(\mathbb{R}^n, \mathbb{R})$ and $a \in K$ we have*

$$f(x) \leq f(a) + \langle \nabla f(a), x - a \rangle + \gamma_{K, f} \|x - a\|^2 \quad \text{for all } x \in K,$$

where $\gamma_{K, f}$ is the constant from (5).

Lemma 4 *Let $a \in K$ be a global minimizer of f on K. Then f has an upper estimator g on K which is exact at a and satisfies the following properties:*

(i) *g is a quadratic, separable polynomial.*
(ii) *$g(x) \geq f(a) + \gamma_{K,f} \|x - a\|^2$ for all $x \in K$.*
(iii) *If $a \in \mathrm{int}\, K$, then $g(x) \leq f(a) + \gamma_{K,f} \|x - a\|^2$ for all $x \in K$.*

Proof Consider the function g defined by

$$g(x) := f(a) + \langle \nabla f(a), x - a \rangle + \gamma_{K,f} \|x - a\|^2, \tag{6}$$

which is an upper estimator of f exact at a by Theorem 1. As we have $\|x - a\|^2 = \sum_i^n (x_i - a_i)^2$, g is indeed a quadratic, separable polynomial.

As a is a global minimizer of f on K, we know by Lemma 1 that $\nabla f(a) \in N_K(a)$. This means that $\langle \nabla f(a), x - a \rangle \geq 0$ for all $x \in K$, which proves the second property.

If $a \in \mathrm{int}\, K$, we must have $\nabla f(a) = 0$, and the third property follows. □

In the special case that K is a ball and f has a global minimizer a on the boundary of K, we have an upper estimator for f, exact at a, which is a linear polynomial.

Lemma 5 *Assume that $f(a) = f_{\min, B_\rho^n(c)}$ for some $a \in \partial B_\rho^n(c)$. Then there exists a linear polynomial g with $f \leq_a g$ on $B_\rho^n(c)$.*

Proof Write $K = B_\rho^n(c)$ and $\gamma = \gamma_{K,f}$ for simplicity. In view of Lemma 4, we have $f(x) \leq g(x)$ for all $x \in K$, where g is the quadratic polynomial from relation (6). Since $a \in \partial K$ is a global minimizer of f on K, we have $\nabla f(a) \in N_K(a)$ by Lemma 1, and thus $\nabla f(a) = \lambda(c - a)$ for some $\lambda \geq 0$. Therefore we have

$$\langle \nabla f(a), x - a \rangle = \langle \lambda(c - a), x - a \rangle = \lambda \rho^2 + \lambda \langle x - c, c - a \rangle.$$

On the other hand, for any $x \in K$ we have

$$\|x - a\|^2 = \|x - c\|^2 + \|c - a\|^2 + 2\langle x - c, c - a \rangle \leq 2\rho^2 + 2\langle x - c, c - a \rangle.$$

Combining these facts we get

$$f(x) \leq g(x) \leq f(a) + (\lambda + 2\gamma)\left(\rho^2 + \langle x - c, c - a \rangle\right) =: h(x).$$

So $h(x)$ is a linear upper estimator of f with $h(a) = f(a)$, as desired. □

Remark 2 As can be seen from the above proof, the assumption in Lemma 5 that $a \in \partial K = \partial B_\rho^n(c)$ is a global minimizer of f on K may be replaced by the weaker assumption that $\nabla f(a) \in N_K(a)$.

Finally, we give a very simple upper estimator, which will be used in Sect. 4.

Lemma 6 *Recall the constant $\beta_{K,f}$ from (5). Let a be a global minimizer of f on K. Then we have*

$$f(x) \leq_a f(a) + \beta_{K,f} \|x - a\| \quad \text{for all } x \in K.$$

3 Special convex bodies

In this section we extend the bound $O(1/r^2)$ from [10] on $E_{K,\mu}^{(r)}(f)$, when $K = [-1, 1]^n$ is equipped with the Chebyshev measure $d\mu(x) = \prod_{i=1}^n (1 - x_i^2)^{-\frac{1}{2}} dx_i$, to a broader class of convex bodies K and reference measures μ.

First, we show that, for the hypercube $K = [-1, 1]^n$, we still have $E_{K,\mu}^{(r)}(f) = O(1/r^2)$ for all polynomial f and all measures of the form $d\mu(x) = \prod_{i=1}^n (1 - x_i^2)^\lambda dx_i$ with $\lambda > -1/2$. Previously this was only known to be the case when f is a linear polynomial. Note that, for $\lambda = 0$, we obtain the Lebesgue measure on $[-1, 1]^n$. Next, we use this result to show that $E_{B^n,\mu}^{(r)}(f) = O(1/r^2)$ for all measures μ on the unit ball B^n of the form $d\mu(x) = (1 - ||x||^2)^\lambda dx$ with $\lambda \geq 0$. We apply this result to also obtain $E_{K,\mu}^{(r)}(f) = O(1/r^2)$ when μ is the Lebesgue measure and K is a 'ball-like' convex body, meaning it has inscribed and circumscribed tangent balls at all boundary points (see Definition 3 below). The primary new tool we use to obtain these results is Proposition 1, which tells us that the behaviour of $E_{K,\mu}^{(r)}(f)$ essentially only depends on the *local* behaviour of f and μ in a neighbourhood of a global minimizer a of f on K.

3.1 Measures and weight functions

A function $w : \text{int } K \to \mathbb{R}_{>0}$ is a *weight function* on K if it is continuous and satisfies $0 < \int_K w(x)dx < \infty$. A weight function w gives rise to a measure μ_w on K defined by $d\mu_w(x) := w(x)dx$. We note that if $K \subseteq \widehat{K}$, and $\widehat{w}$ is a weight function on $\widehat{K}$, it can naturally be interpreted as a weight function on K as well, by simply restricting its domain (assuming $\int_K \widehat{w}(x)dx > 0$). In what follows we will implicitly make use of this fact.

Definition 1 Given two weight functions $w, \widehat{w}$ on K and a point $a \in K$, we say that $\widehat{w} \preceq_a w$ on K if there exist constants $\epsilon, m_a > 0$ such that

$$m_a \widehat{w}(x) \leq w(x) \text{ for all } x \in B_\epsilon^n(a) \cap \text{int } K. \tag{7}$$

If the constant m_a can be chosen uniformly, i.e., if there exists a constant $m > 0$ such that

$$m\widehat{w}(x) \leq w(x) \text{ for all } x \in \text{int } K, \tag{8}$$

then we say that $\widehat{w} \preceq w$ on K.

Remark 3 We note the following facts for future reference:

(i) As weight functions are continuous on the interior of K by definition, we always have $\widehat{w} \preceq_a w$ if $a \in \text{int } K$.
(ii) If w is bounded from below, and $\widehat{w}$ is bounded from above on int K, then we automatically have $\widehat{w} \preceq w$.
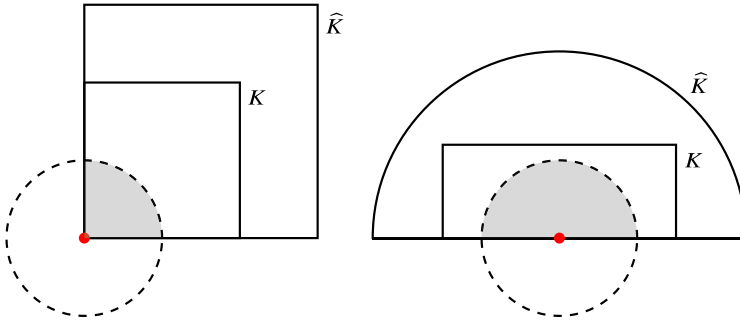
Fig. 1 Some examples of sets K, $\widehat{K}$ for which $K \subseteq_a \widehat{K}$. The dot indicates the point a, and the gray area indicates $B_\epsilon^n(a) \cap K$

3.2 Local similarity

Assuming that the global minimizer a of f on K is unique, sum-of-squares density functions q for which the integral $\int_K q(x)f(x)d\mu(x)$ is small should in some sense approximate the Dirac delta function centered at a. With this in mind, it seems reasonable to expect that the quality of the bound $f^{(r)}$ depends in essence only on the *local* properties of K and μ around a. We formalize this intuition here.

Definition 2 Suppose $K \subseteq \widehat{K} \subseteq \mathbb{R}^n$. Given $a \in K$, we say that K and $\widehat{K}$ are *locally similar* at a, which we denote by $K \subseteq_a \widehat{K}$, if there exists $\epsilon > 0$ such that

$$B_\epsilon^n(a) \cap K = B_\epsilon^n(a) \cap \widehat{K}.$$

Clearly, $K \subseteq_a \widehat{K}$ for any point $a \in \text{int } K$.

Figure 1 depicts some examples of locally similar sets.

Proposition 1 *Let $K \subseteq \widehat{K} \subseteq \mathbb{R}^n$, let $a \in K$ be a global minimizer of f on K and assume $K \subseteq_a \widehat{K}$. Let $w, \widehat{w}$ be two weight functions on K, $\widehat{K}$, respectively. Assume that $\widehat{w}(x) \geq w(x)$ for all $x \in \text{int } K$, and that $\widehat{w} \preceq_a w$. Then there exists an upper estimator g of f on $\widehat{K}$ which is exact at a and satisfies*

$$E_{K,w}^{(r)}(g) \leq \frac{2}{m_a} E_{\widehat{K},\widehat{w}}^{(r)}(g)$$

for all $r \in \mathbb{N}$ large enough. Here $m_a > 0$ is the constant defined by (7).

Recall that if g is an upper estimator for f which is exact at one of its global minimizers, we then have $E_{K,w}^{(r)}(f) \leq E_{K,w}^{(r)}(g)$ by Lemma 3. Proposition 1 then allows us to bound $E_{K,w}^{(r)}(f)$ in terms of $E_{\widehat{K},\widehat{w}}^{(r)}(g)$. For its proof, we first need the following lemma.

Lemma 7 *Let $a \in K$, and assume that $K \subseteq_a \widehat{K}$. Then any normal vector of K at a is also a normal vector of $\widehat{K}$. That is, $N_K(a) \subseteq N_{\widehat{K}}(a)$.*

Proof Let $v \in N_K(a)$. Suppose for contradiction that $v \notin N_{\widehat{K}}(a)$. Then, by definition of the normal cone, there exists $y \in \widehat{K}$ such that $\langle v, y - a \rangle < 0$. As $K \subseteq_a \widehat{K}$, there exists $\epsilon > 0$ for which $K \cap B_a^n(\epsilon) = \widehat{K} \cap B_a^n(\epsilon)$. Now choose $1 > \eta > 0$ small enough such that $y' := \eta y + (1 - \eta)a \in B_a^n(\epsilon)$. Then, by convexity, we have $y' \in \widehat{K} \cap B_a^n(\epsilon) = K \cap B_a^n(\epsilon)$. Now, we have $\langle v, y' - a \rangle = \eta \langle v, y - a \rangle < 0$. But, as $y' \in K$, this contradicts the assumption that $v \in N_K(a)$. $\qquad\square$

Proof (of Proposition 1) For simplicity, we assume here $f(a) = 0$, which is without loss of generality by Remark 1. Consider the quadratic polynomial g from (6):

$$g(x) = \langle \nabla f(a), x - a \rangle + \gamma \|x - a\|^2,$$

where $\gamma := \gamma_{\widehat{K}, f}$ is defined in (5). By Taylor's theorem (Theorem 1), we have that $g(x) \geq f(x)$ for all $x \in \widehat{K}$, and clearly $g(a) = f(a)$. That is, g is an upper estimator for f on $\widehat{K}$, exact at a (cf. Lemma 4). We proceed to show that

$$E_{K,w}^{(r)}(g) \leq \frac{2}{m_a} E_{\widehat{K}, \widehat{w}}^{(r)}(g).$$

We start by selecting a degree $2r$ sum-of-squares polynomial $\widehat{q_r}$ satisfying

$$\int_{\widehat{K}} \widehat{q_r}(x) \widehat{w}(x) dx = 1 \quad \text{and} \quad \int_{\widehat{K}} g(x) \widehat{q_r}(x) \widehat{w}(x) dx = E_{\widehat{K}, \widehat{w}}^{(r)}(g).$$

We may then rescale $\widehat{q_r}$ to obtain a density function $q_r \in \Sigma_r$ on K w.r.t. w by setting

$$q_r := \frac{\widehat{q_r}}{\int_K \widehat{q_r}(x) w(x) dx}.$$

By assumption, $w(x) \leq \widehat{w}(x)$ for all $x \in \operatorname{int} K$. Moreover, $g(x) \geq f(a) = 0$ for all $x \in \operatorname{int} K$. This implies that

$$E_{K,w}^{(r)}(g) \leq \int_K g(x) q_r(x) w(x) dx \leq \frac{\int_{\widehat{K}} g(x) \widehat{q_r}(x) \widehat{w}(x) dx}{\int_K \widehat{q_r}(x) w(x) dx} = \frac{E_{\widehat{K}, \widehat{w}}^{(r)}(g)}{\int_K \widehat{q_r}(x) w(x) dx}$$

and thus it suffices to show that $\int_K \widehat{q_r}(x) w(x) dx \geq \frac{1}{2} m_a$. The key to proving this bound is the following lemma, which tells us that optimum sum-of-squares densities should assign rather high weight to the ball $B_\epsilon^n(a)$ around a. $\qquad\square$

Lemma 8 *Let $\epsilon > 0$. Then, for any $r \in \mathbb{N}$, we have*

$$\int_{B_\epsilon^n(a) \cap \widehat{K}} \widehat{q_r}(x) \widehat{w}(x) dx \geq 1 - \frac{E_{\widehat{K}, \widehat{w}}^{(r)}(g)}{\gamma \epsilon^2}.$$

Proof By Lemma 1, we have $\nabla f(a) \in N_K(a)$ and so $\nabla f(a) \in N_{\widehat{K}}(a)$ by Lemma 7. As a result, we have $g(x) \geq \gamma ||x - a||^2$ for all $x \in \widehat{K}$ (cf. Lemma 4). In particular, this implies that $g(x) \geq \gamma ||x - a||^2 \geq \gamma \epsilon^2$ for all $x \in \widehat{K} \setminus B_\epsilon^n(a)$ and so

$$E_{\widehat{K},\widehat{w}}^{(r)}(g) \geq \int_{\widehat{K} \setminus B_\epsilon^n(a)} g(x)\widehat{q}_r(x)\widehat{w}(x)dx \geq \gamma \epsilon^2 \int_{\widehat{K} \setminus B_\epsilon^n(a)} \widehat{q}_r(x)\widehat{w}(x)dx$$

$$= \gamma \epsilon^2 \left(1 - \int_{B_\epsilon^n(a) \cap \widehat{K}} \widehat{q}_r(x)\widehat{w}(x)dx \right).$$

The statement now follows from reordering terms. $\qquad\square$

As $K \subseteq_a \widehat{K}$, there exists $\epsilon_1 > 0$ such that $B_{\epsilon_1}^n(a) \cap K = B_{\epsilon_1}^n(a) \cap \widehat{K}$. As $\widehat{w} \preceq_a w$, there exist $\epsilon_2 > 0$, $m_a > 0$ such that $m_a \widehat{w}(x) \leq w(x)$ for $x \in B_{\epsilon_2}^n(a) \cap \text{int } K$. Set $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Choose $r_0 \in \mathbb{N}$ large enough such that $E_{\widehat{K},\widehat{w}}^{(r)}(g) < \frac{\epsilon^2 \gamma}{2}$ for all $r \geq r_0$, which is possible since $E_{\widehat{K},\widehat{w}}^{(r)}(g)$ tends to 0 as $r \to \infty$. Then, Lemma 8 yields

$$\int_{B_\epsilon^n(a) \cap \widehat{K}} \widehat{q}_r(x)\widehat{w}(x)dx \geq \frac{1}{2}$$

for all $r \geq r_0$. Putting things together yields the desired lower bound:

$$\int_K \widehat{q}_r(x)w(x)dx \geq \int_{B_\epsilon^n(a) \cap K} \widehat{q}_r(x)w(x)dx \geq m_a \int_{B_\epsilon^n(a) \cap \widehat{K}} \widehat{q}_r(x)\widehat{w}(x)dx \geq \frac{1}{2}m_a.$$

for all $r \geq r_0$. $\qquad\square$

Corollary 1 *Let $K \subseteq \widehat{K} \subseteq \mathbb{R}^n$, let $a \in K$ be a global minimizer of f on K, and assume that $K \subseteq_a \widehat{K}$. Let $w, \widehat{w}$ be two weight functions on $K, \widehat{K}$, respectively. Assume that $\widehat{w}(x) \geq w(x)$ for all $x \in \text{int } K$ and that $\widehat{w} \preceq w$. Then there exists an upper estimator g of f on $\widehat{K}$, exact at a, such that*

$$E_{K,w}^{(r)}(g) \leq \frac{2}{m} E_{\widehat{K},\widehat{w}}^{(r)}(g)$$

for all $r \in \mathbb{N}$ large enough. Here $m > 0$ is the constant defined by (8).

3.3 The unit cube

Here we consider optimization over the hypercube $K = [-1, 1]^n$ and we restrict to reference measures on K having a weight function of the form

$$\widehat{w_\lambda}(x) := \prod_{i=1}^n w_\lambda(x_i) = \prod_{i=1}^n \left(1 - x_i^2 \right)^\lambda \qquad (9)$$

with $\lambda > -1$. The following result is shown in [10] on the convergence rate of the bound $E^{(r)}_{K,\widehat{w_\lambda}}(f)$ when using the measure $\widehat{w_\lambda}(x)dx$ on $K = [-1, 1]^n$.

Theorem 2 ([10]) *Let $K = [-1, 1]^n$ and consider the weight function $\widehat{w_\lambda}$ from* (9).

(i) *If $\lambda = -\frac{1}{2}$, then we have:*

$$E^{(r)}_{K,\widehat{w_\lambda}}(f) = O\left(\frac{1}{r^2}\right). \tag{10}$$

(ii) *If $n = 1$ and f has a global minimizer on the boundary of $[-1, 1]$, then* (10) *holds for all $\lambda > -1$.*

The key ingredients for claim (ii) above are: (a) when the global minimizer is a boundary point of $[-1, 1]$ then f has a linear upper estimator (recall Lemma 5), and (b) the convergence rate of (10) holds for any linear function and any $\lambda > -1$ (see [10]).

In this section we show Theorem 3 below, which extends the above result to all weight functions $\widehat{w_\lambda}(x)$ with $\lambda \geq -\frac{1}{2}$. Following the approach in [10], we proceed in two steps: first we reduce to the univariate case, and then we deal with the univariate case. Then the new situation to be dealt with is when $n = 1$ and the minimizer lies in the interior of $[-1, 1]$, which we can settle by getting back to the case $\lambda = -\frac{1}{2}$ through applying Proposition 1, the 'local similarity' tool, with $K = \widehat{K} = [-1, 1]$.

Reduction to the univariate case Let $a \in K$ be a global minimizer of f in $K = [-1, 1]^n$. Following [10] (recall Remark 1 and Lemma 4), we consider the upper estimator $f(x) \leq_a g(x) := f(a) + \langle \nabla f(a), x - a \rangle + \gamma_{f,K} ||x - a||^2$. This g is separable, i.e., we can write $g(x) = \sum_{i=1}^n g_i(x_i)$, where each g_i is quadratic univariate with a_i as global minimizer over $[-1, 1]$. Let q_r^i be an optimum solution to the problem (3) corresponding to the minimization of g_i over $[-1, 1]$ w.r.t. the weight function $w_\lambda(x_i) = (1 - x_i^2)^\lambda$. If we set $q_r(x) = \prod_{i=1}^n q_r^i(x_i)$, then q_r is a sum of squares with degree at most nr, such that $\int_K q_r(x)\widehat{w_\lambda}(x)dx = 1$. Hence we have

$$
\begin{aligned}
f^{(rn)}_{K,\widehat{w_\lambda}} - f(a) &\leq \int_K f(x)q_r(x)\widehat{w_\lambda}(x)dx - f(a) \\
&\leq \int_K g(x)q_r(x)\widehat{w_\lambda}(x)dx - g(a) \\
&= \sum_{i=1}^n \left(\int_{-1}^1 g_i(x)q_r^i(x_i)w_\lambda(x_i)dx_i - g_i(a_i) \right) \\
&= \sum_{i=1}^n \left((g_i)^{(r)}_{[-1,1],w_\lambda} - g_i(a_i) \right) = \sum_{i=1}^n E^{(r)}_{[-1,1],w_\lambda}(g_i).
\end{aligned}
$$

As a consequence, we need only to consider the case of a quadratic univariate polynomial f on $K = [-1, 1]$. We distinguish two cases, depending whether the global minimizer lies on the boundary or in the interior of K. The case when the global minimizer lies on the boundary of $[-1, 1]$ is settled by Theorem 2(ii) above, so we next assume the global minimizer lies in the interior of $[-1, 1]$.

Case of a global minimizer in the interior of $K = [-1, 1]$ To deal with this case we make use of Proposition 1 with $K = \widehat{K} = [-1, 1]$, weight function $w(x) := w_\lambda(x)$ on K, and weight function $\widehat{w}(x) := w_{-1/2}(x)$ on $\widehat{K}$. We check that the conditions of the proposition are met. As $\widehat{K} = K$, clearly we have $K \subseteq_a \widehat{K}$. Further, for any $\lambda \geq -\frac{1}{2}$, we have

$$w_\lambda(x) = \left(1 - x^2\right)^\lambda \leq \left(1 - x^2\right)^{-\frac{1}{2}} = w_{-1/2}(x)$$

for all $x \in (-1, 1) = \text{int } K$. As $a \in \text{int } K$, we also have $w_\lambda \preceq_a w_{-1/2}$ [see Remark 3(i)]. Hence we may apply Proposition 1 to find that there exists a polynomial upper estimator g of f on $[-1, 1]$, exact at a, and having

$$E_{K,w}^{(r)}(g) \leq \frac{2}{m_a} E_{\widehat{K},\widehat{w}}^{(r)}(g)$$

for all $r \in \mathbb{N}$ large enough. Now, (the univariate case of) Theorem 2(i) allows us to claim $E_{\widehat{K},\widehat{w}}^{(r)}(g) = O(1/r^2)$, so that we obtain:

$$E_{K,w_\lambda}^{(r)}(f) \leq E_{K,w_\lambda}^{(r)}(g) = O_a\left(E_{\widehat{K},\widehat{w}}^{(r)}(g)\right) = O_a\left(1/r^2\right).$$

In summary, in view of the above, we have shown the following extension of Theorem 2.

Theorem 3 *Let* $K = [-1, 1]^n$ *and* $\lambda \geq -\frac{1}{2}$. *Let a be a global minimizer of f on K. Then we have*

$$E_{K,\widehat{w_\lambda}}^{(r)}(f) = O_a\left(\frac{1}{r^2}\right).$$

The constant m_a involved in the proof of Theorem 3 depends on the global minimizer a of f on $[-1, 1]$. It is introduced by the application of Proposition 1 to cover the case where a lies in the interior of $[-1, 1]$. When $\lambda = 0$ (i.e., when $w = w_0 = 1$ corresponds to the Lebesgue measure), one can replace m_a by a uniform constant $m > 0$, as we now explain.

Consider $\widehat{K} := [-2, 2] \supseteq [-1, 1] = K$, equipped with the scaled Chebyshev weight $\widehat{w}(x) := w_{-1/2}(x/2) = (1 - x^2/4)^{-1/2}$. Of course, Theorem 2 applies to this choice of $\widehat{K}, \widehat{w}$ as well. Further, we still have $\widehat{w}(x) \geq w(x) = w_0(x) = 1$ for all $x \in [-1, 1]$. However, we now have a *uniform* upper bound $\widehat{w}(x) \leq \widehat{w}(1)$ for $\widehat{w}$ on K, which means that $\widehat{w} \preceq w$ on K [see Remark 3(ii)]. Indeed, we have

$$\widehat{w}(x)/\widehat{w}(1) \leq 1 = w_0(x) = w(x) \quad \text{for all } x \in [-1, 1].$$

We may thus apply Corollary 1 (instead of Proposition 1) to obtain the following.

Corollary 2 *If* $K = [-1, 1]^n$ *is equipped with the Lebesgue measure then*

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

3.4 The unit ball

We now consider optimization over the unit ball $K = B^n \subseteq \mathbb{R}^n$ $(n \geq 2)$; we restrict to reference measures on B^n with weight function of the form

$$w_\lambda(x) = (1 - ||x||^2)^\lambda, \tag{11}$$

where $\lambda > -1$. For further reference we recall (see e.g. [16, §6.3.2]) or [3, §11]) that

$$C_{n,\lambda} := \int_{B^n} w_\lambda(x) dx = \frac{\pi^{\frac{n}{2}} \Gamma(\lambda + 1)}{\Gamma\left(\lambda + 1 + \frac{n}{2}\right)}. \tag{12}$$

For the case $\lambda \geq 0$, we can analyze the bounds and show the following result.

Theorem 4 *Let* $K = B^n$ *be the unit ball. Let* a *be a global minimizer of* f *on* K. *Consider the weight function* w_λ *from* (11) *on* K.

(i) If $\lambda = 0$*, we have*

$$E_{K, w_\lambda}^{(r)}(f) = O\left(\frac{1}{r^2}\right).$$

(ii) If $\lambda > 0$*, we have*

$$E_{K, w_\lambda}^{(r)}(f) = O_a\left(\frac{1}{r^2}\right).$$

For the proof, we distinguish the two cases when a lies in the interior of K or on its boundary.

Case of a global minimizer in the interior of K Our strategy is to reduce this to the case of the hypercube with the help of Proposition 1. Set $\widehat{K} := [-1, 1]^n \supseteq B^n = K$. As $a \in$ int K, we have $K \subseteq_a \widehat{K}$. Consider the weight function $w(x) := w_\lambda(x) = (1 - ||x||^2)^\lambda$ on K, and $\widehat{w}(x) := 1$ on the hypercube $\widehat{K}$. Since $\lambda \geq 0$, we have $w_\lambda(x) \leq 1 \leq \widehat{w}(x)$ for all $x \in K$. Furthermore, as $a \in$ int K, we also have $\widehat{w} \preceq_a w$. Hence we may apply Proposition 1 to find a polynomial upper estimator g of f on $\widehat{K}$, exact at a, satisfying

$$E_{K, w}^{(r)}(g) \leq \frac{2}{m_a} E_{\widehat{K}, \widehat{w}}^{(r)}(g)$$

for all $r \in \mathbb{N}$ large enough. Here $m_a > 0$ is the constant from (7). Now, Theorem 3 allows us to claim $E_{\widehat{K},\widehat{w}}^{(r)}(g) = O_a(1/r^2)$. Hence we obtain:

$$E_{K,w}^{(r)}(f) \leq E_{K,w}^{(r)}(g) = O_a\left(E_{\widehat{K},\widehat{w}}^{(r)}(g)\right) = O_a\left(1/r^2\right).$$

As in the previous section, it is possible to replace the constant m_a by a uniform constant $m > 0$ in the case that $\lambda = 0$, i.e., in the case that we have the Lebesgue measure on K. Indeed, in this case we have $\widehat{w} = w \ (= w_0 = 1)$, and so in particular $\widehat{w} \preceq w$. We may thus invoke Corollary 1 (instead of Proposition 1) to obtain

$$E_{K,w}^{(r)}(g) \leq 2E_{\widehat{K},\widehat{w}}^{(r)}(g)$$

and so

$$E_{K,w}^{(r)}(f) = O\left(E_{\widehat{K},\widehat{w}}^{(r)}(g)\right) = O\left(1/r^2\right).$$

Note that in this case, we do not actually make use of the fact that $K = B^n$. Rather, we only need that a lies in the interior of K and that $K \subseteq [-1, 1]^n$. As we may freely apply affine transformations to K (by Lemma 2), the latter is no true restriction. We have thus shown the following result.

Theorem 5 *Let $K \subseteq \mathbb{R}^n$ be a compact set, with non-empty interior, equipped with the Lebesgue measure. Assume that f has a global minimizer a on K with $a \in \mathrm{int}\, K$. Then we have*

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

Case of a global minimizer on the boundary of K Our strategy is now to reduce to the univariate case of the interval $[-1, 1]$. For this, we use Lemma 5, which claims that f has a linear upper estimator g on K, exact at a. Up to applying an orthogonal transformation (and scaling) we may assume that g is of the form $g(x) = x_1$. It therefore suffices now to analyze the behaviour of the bounds for the function x_1 minimized on the ball B^n. Note that when minimizing x_1 on B^n or on the interval $[-1, 1]$ the minimum is attained at the boundary in both cases. The following technical lemma will be useful for reducing to the case of the interval $[-1, 1]$.

Lemma 9 *Let h be a univariate polynomial and let $\lambda > -1$. Then we have*

$$\int_{B^n} h(x_1)w_\lambda(x)dx = C_{n-1,\lambda} \int_{-1}^{1} h(x_1)w_{\lambda+\frac{n-1}{2}}(x_1)dx_1,$$

where $C_{n-1,\lambda}$ is given in (12).

Proof Change variables and set $u_j = \dfrac{x_j}{\sqrt{1-x_1^2}}$ for $2 \le j \le d$. Then we have

$$w_\lambda(x) = \left(1 - x_1^2 - x_2^2 + \cdots - x_n^2\right)^\lambda = \left(1 - x_1^2\right)^\lambda \left(1 - u_2^2 - \ldots - u_n^2\right)^\lambda$$

and $dx_2 \ldots dx_n = (1 - x_1^2)^{\frac{n-1}{2}} du_2 \ldots du_n$. Putting things together we obtain the desired result. □

Let $q_r(x_1)$ be an optimal sum-of-squares density with degree at most $2r$ for the problem of minimizing x_1 over the interval $[-1, 1]$, equipped with the weight function $w(x) := w_{\lambda + \frac{n-1}{2}}(x)$. Then, its scaling $C_{n-1,\lambda}^{-1} q_r(x_1)$ provides a feasible solution for the problem of minimizing $g(x) = x_1$ over the ball $K = B^n$. Indeed, using Lemma 9, we have $\int_{B^n} C_{n-1,\lambda}^{-1} q_r(x_1) w_\lambda(x) dx = \int_{-1}^{1} q_r(x_1) w(x) dx_1 = 1$, and so

$$g_{K,w_\lambda}^{(r)} \le \int_{B^n} x_1 C_{n-1,\lambda}^{-1} q_r(x_1) w_\lambda(x) dx = \int_{-1}^{1} x_1 q_r(x_1) w(x_1) dx_1.$$

The proof is now concluded by applying Theorem 2(ii).

3.5 Ball-like convex bodies

Here we show a convergence rate of $E_K^{(r)}(f)$ in $O(1/r^2)$ for a special class of smooth convex bodies K with respect to the Lebesgue measure. The basis for this result is a reduction to the case of the unit ball.

We say K has an *inscribed tangent ball* (of radius ϵ) at $x \in \partial K$ if there exists $\epsilon > 0$ and a closed ball B_{insc} of radius ϵ such that $x \in \partial B_{insc}$ and $B_{insc} \subseteq K$. Similarly, we say K has a *circumscribed tangent ball* (of radius ϵ) at $x \in \partial K$ if there exists $\epsilon > 0$ and a closed ball B_{circ} of radius ϵ such that $x \in \partial B_{circ}$ and $K \subseteq B_{circ}$.

Definition 3 We say that a (smooth) convex body K is *ball-like* if there exist (uniform) $\epsilon_{insc}, \epsilon_{circ} > 0$ such that K has inscribed and circumscribed tangent balls of radii $\epsilon_{insc}, \epsilon_{circ}$, respectively, at all points $x \in \partial K$.

Theorem 6 *Assume that K is a (smooth) ball-like convex body, equipped with the Lebesgue measure. Then we have*

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

Proof Let $a \in K$ be a global minimizer of f on K. We again distinguish two cases depending on whether a lies in the interior of K or on its boundary.

Case of a global minimizer in the interior of K This case is covered directly by Theorem 5.

Case of a global minimizer on the boundary of K By applying a suitable affine transformation, we can arrange that the following holds: $f(a) = 0$, $a = 0$, e_1 is an inward normal of K at a, and the radius of the circumscribed tangent ball B_{circ} at a is equal to 1, i.e., $B_{circ} = B_1^n(e_1)$. See Fig. 2 for an illustration. Now, as a is a global minimizer of f on K, we have $\nabla f(a) \in N_K(a)$ by Lemma 1. But $N_K(a) = N_{B_{circ}}(a)$, and so $\nabla f(a) \in N_{B_{circ}}(a)$. As noted in Remark 2, we may thus use Lemma 5 to find that $f(x) \leq_a c\langle e_1, x \rangle = cx_1$ on B_{circ} for some constant $c > 0$. In light of Remark 1(i), and after scaling, it therefore suffices to analyze the function $f(x) = x_1$.

Again, we will use a reduction to the univariate case, now on the interval $[0, 2]$. For any $r \in \mathbb{N}$, let $q_r \in \Sigma_r$ be an optimum sum-of-squares density of degree $2r$ for the minimization of x_1 on $[0, 2]$ with respect to the weight function

$$w'(x_1) := w_{\frac{n-1}{2}}(x_1 - 1) = \left[1 - (x_1 - 1)^2\right]^{\frac{n-1}{2}} = \left[2x_1 - x_1^2\right]^{\frac{n-1}{2}}.$$

That is, $q_r \in \Sigma_r$ satisfies

$$\int_0^2 x_1 q_r(x_1) w'(x_1) dx_1 = O\left(1/r^2\right) \quad \text{and} \quad \int_0^2 q_r(x_1) w'(x_1) dx_1 = 1, \quad (13)$$

where the first equality relies on Theorem 2(ii). As $x \mapsto q_r(x_1)/(\int_K q_r(x_1)dx)$ is a sum-of-squares density on K with respect to the Lebesgue measure, we have

$$E_K^{(r)}(f) \leq \frac{\int_K x_1 q_r(x_1)dx}{\int_K q_r(x_1)dx}. \quad (14)$$

We will now show that, on the one hand, the numerator $\int_K x_1 q_r(x_1)dx$ in (14) has an upper bound in $O(1/r^2)$ and that, on the other hand, the denominator $\int_K q_r(x_1)dx$ in (14) is lower bounded by an absolute constant that does not depend on r. Putting these two bounds together then yields $E_K^{(r)}(f) = O(1/r^2)$, as desired.

The upper bound We make use of the fact that $K \subseteq B_{circ}$ to compute:

$$\int_K x_1 q_r(x_1)dx \leq \int_{B_{circ}} x_1 q_r(x_1)dx$$
$$= \int_{B^n} (y_1 + 1)q_r(y_1 + 1)dy \qquad [y = x - e_1]$$
$$= C_{n-1,0} \int_{-1}^1 (y_1 + 1)q_r(y_1 + 1)w_{\frac{n-1}{2}}(y_1)dy_1 \quad [\text{by Lemma 9}]$$
$$= C_{n-1,0} \int_0^2 zq_r(z)w'(z)dz \qquad [z = y_1 + 1]$$
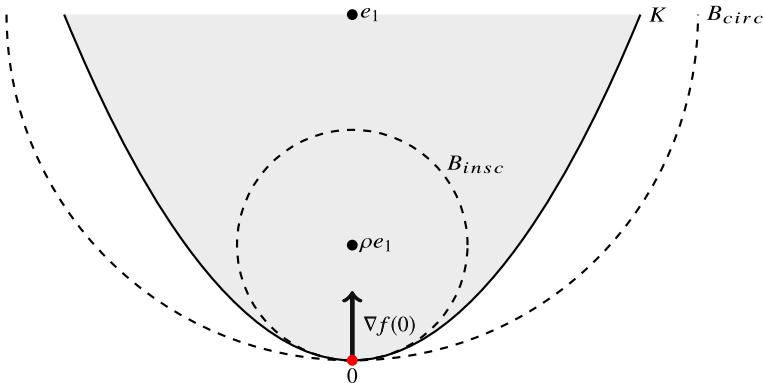$$= O\left(1/r^2\right). \qquad [\text{by (13)}]$$

Fig. 2 An overview of the situation in the second case of the proof of Theorem 6

The lower bound Here, we consider an inscribed tangent ball B_{insc} of K at $a = 0$. Say $B_{insc} = B_\rho^n(\rho e_1)$ for some $\rho > 0$. See again Fig. 2. We may then compute:

$$
\begin{aligned}
\int_K q_r(x_1)dx &\geq \int_{B_{insc}} q_r(x_1)dx \\
&= \int_{B^n} q_r\big(\rho(y_1+1)\big)\rho^n dy && \left[y = \frac{x-\rho e_1}{\rho}\right] \\
&= \rho^n C_{n-1,0}\int_{-1}^1 q_r\big(\rho(y_1+1)\big)w_{\frac{n-1}{2}}(y_1)dy_1 && \text{[by Lemma 9]} \\
&= \rho^{n-1}C_{n-1,0}\int_0^{2\rho} q_r(z)w_{\frac{n-1}{2}}(z/\rho-1)dz && [z = \rho(y_1+1)] \\
&\geq \rho^{n-1}C_{n-1,0}\int_0^\rho q_r(z)w'(z)\frac{w_{\frac{n-1}{2}}(z/\rho-1)}{w_{\frac{n-1}{2}}(z-1)}dz && \left[w'(z) = w_{\frac{n-1}{2}}(z-1)\right] \\
&\geq \left(\frac{\rho}{2-\rho}\right)^{\frac{n-1}{2}} C_{n-1,0}\int_0^\rho q_r(z)w'(z)dz,
\end{aligned}
$$

where the last inequality follows using the fact that $\frac{1-(z/\rho-1)^2}{1-(z-1)^2} \geq \frac{1}{\rho(2-\rho)}$ for $z \in [0,\rho]$. It remains to show that

$$
\int_0^\rho q_r(z)w'(z)dz \geq \frac{1}{2} \quad \text{for all } r \text{ large enough.}
$$

The argument is similar to the one used for the proof of Lemma 8. By (13), there is a constant $C > 0$ such that $\int_0^2 zq_r(z)w'(z)dz \leq \frac{C}{r^2}$ for all $r \in \mathbb{N}$. So we have

$$
\frac{C}{r^2} \geq \int_\rho^2 zq_r(z)w'(z)dz \geq \rho\int_\rho^2 q_r(z)w'(z)dz = \rho\left(1 - \int_0^\rho q_r(z)w'(z)dz\right),
$$

which implies $\int_0^\rho q_r(z)w'(z)dz \geq 1 - \frac{C}{\rho r^2} \geq \frac{1}{2}$ for r large enough.

This concludes the proof of Theorem 6. $\qquad\square$

Classification of ball-like sets With Theorem 6 in mind, it is interesting to understand under which conditions a convex body K is ball-like. Under the assumption that K has a C^2-boundary, the well-known Rolling Ball Theorem (cf., e.g., [17]) guarantees the existence of inscribed tangent balls.

Theorem 7 (Rolling Ball Theorem) *Let $K \subseteq \mathbb{R}^n$ be a convex body with C^2- boundary. Then there exists $\epsilon_{insc} > 0$ such that K has an inscribed tangent ball of radius ϵ_{insc} for each $x \in \partial K$.*

Classifying the existence of circumscribed tangent balls is somewhat more involved. Certainly, we should assume that K is *strictly* convex, which means that its boundary should not contain any line segments. This assumption, however, is not sufficient. Instead we need the following stronger notion of *2-strict convexity* introduced in [4].

Definition 4 Let $K \subseteq \mathbb{R}^n$ be a convex body with C^2-boundary and let $\Psi \in C^2(\mathbb{R}^n, \mathbb{R})$ such that $K = \Psi^{-1}((-\infty, 0])$ and $\partial K = \Psi^{-1}(0)$. Assume $\nabla \Psi(a) \neq 0$ for all $a \in \partial K$. The set K is said to be *2-strictly convex* if the following holds:

$$x^T \nabla^2 \Psi(a) x > 0 \quad \text{for all } x \in T_a K \backslash \{0\} \text{ and } a \in \partial K.$$

In other words, the Hessian of Ψ at any boundary point should be positive definite, when restricted to the tangent space.

Example 1 Consider the unit ball for the ℓ_4-norm:

$$K = \left\{ (x_1, x_2) : \Psi(x_1, x_2) := x_1^4 + x_2^4 \leq 1 \right\} \subseteq \mathbb{R}^2.$$

Then, K is strictly convex, but not 2-strictly convex. Indeed, at any of the points $a = (0, \pm 1)$ and $(\pm 1, 0)$, the Hessian of Ψ is not positive definite on the tangent space. For instance, for $a = (0, -1)$, we have $\nabla \Psi(a) = (0, -4)$ and $x^T \Psi^2(a) x = 12 x_2^2$, which vanishes at $x = (1, 0) \in T_a K$. In fact, one can verify that K does not have a circumscribed tangent ball at any of the points $(0, \pm 1)$, $(\pm 1, 0)$.

It is shown in [4] that the set of 2-strictly convex bodies lies dense in the set of all convex bodies. For K with C^2-boundary, it turns out that 2-strict convexity is equivalent to the existence of circumscribed tangent balls at all boundary points.

Theorem 8 ([5, Corollary 3.3]) *Let K be a convex body with C^2-boundary. Then K is 2-strictly convex if and only if there exists $\epsilon_{circ} > 0$ such that K has a circumscribed tangent ball of radius ϵ_{circ} at all boundary points $a \in \partial K$.*

Combining Theorems 7 and 8 then gives a full classification of the ball-like convex bodies K with C^2-boundary.

Corollary 3 *Let $K \subseteq \mathbb{R}^n$ be a convex body with C^2-boundary. Then K is ball-like if and only if it is 2-strictly convex.*
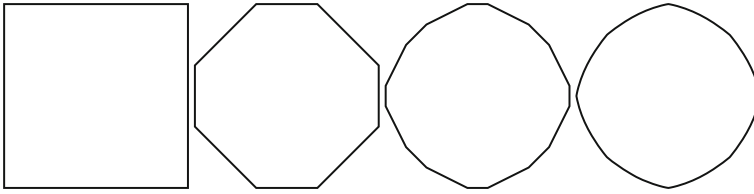
Fig. 3 From left to right: the curve C_k for $k = 0, 1, 2, 8$

A convex body without inscribed tangent balls We now give an example of a convex body K which does not have inscribed tangent balls, going back to de Rham [14]. The idea is to construct a curve by starting with a polygon, and then successively 'cutting corners'. Let C_0 be the polygon in $\mathbb{R}^2$ with vertices $(-1, -1)$, $(1, -1)$, $(1, 1)$ and $(-1, 1)$, i.e., a square. For $k \geq 1$, we obtain C_k by subdividing each edge of C_{k-1} into three equal parts and taking the convex hull of the resulting subdivision points (see Fig. 3). We then let C be the limiting curve obtained by letting k tend to ∞. Then, C is a continuously differentiable, convex curve (see [6] for details). It is not, however, C^2 everywhere. We indicate below some point where no inscribed tangent ball exists for the convex body with boundary C.

Consider the point $m = (0, -1) \in C$, which is an element of C_k for all k. Fix $k \geq 1$. If we walk anti-clockwise along C_k starting at m, the first corner point encountered is $s_k = (1/3^k, -1)$, the slope of the edge starting at s_k is $l_k = 1/k$ and its end point is

$$e_k = \left((2k + 1)/3^k, 2/3^k - 1\right).$$

Now suppose that there exists an inscribed tangent ball $B_\epsilon(c)$ at the point m. Then, $\epsilon > 0, c = (0, \epsilon - 1)$ and any point $(x, y) \in C$ lies outside of the ball $B_\epsilon(c)$, so that

$$x^2 + (y + 1)^2 - 2\epsilon(y + 1) \geq 0 \quad \text{for all } (x, y) \in C.$$

As C is contained in the polygonal region delimited by any C_k, also $e_k \notin B_\epsilon(c)$ and thus $\left(\frac{2k+1}{3^k}\right)^2 + \left(\frac{2}{3^k}\right)^2 - \frac{4\epsilon}{3^k} \geq 0$. Letting $k \to \infty$, we get $\epsilon = 0$, a contradiction.

3.6 The simplex

We now consider a full-dimensional simplex $\Delta^n := \text{conv}(\{v_0, v_1, v_2, \ldots, v_n\}) \subseteq \mathbb{R}^n$, equipped with the Lebesgue measure. We show the following.

Theorem 9 *Let $K = \Delta^n$ be a simplex, equipped with the Lebesgue measure. Then*

$$E_{\Delta^n}^{(r)}(f) = O\left(\frac{1}{r^2}\right).$$

Proof Let $a \in \Delta^n$ be a global minimizer of f on Δ^n. The idea is to apply an affine transformation ϕ to Δ^n whose image $\phi(\Delta^n)$ is locally similar to $[0, 1]^n$ at the global
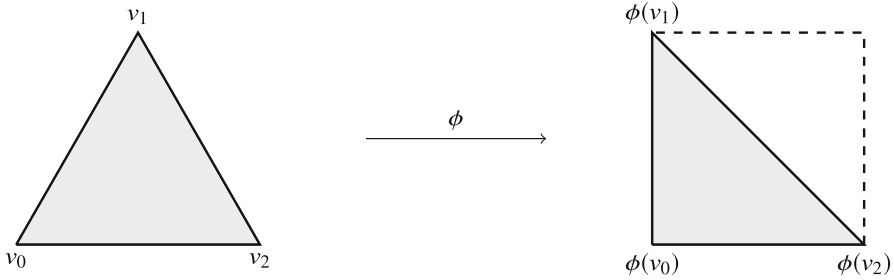
Fig. 4 The map ϕ from the proof of Theorem 9 for $n = 2$

minimizer $\phi(a)$ of $g := f \circ \phi^{-1}$, after which we may 'transport' the $O(1/r^2)$ rate from the hypercube to the simplex.

Let $F := \mathrm{conv}(v_1, v_2, \ldots, v_n)$ be the facet of Δ^n which does not contain v_0. By reindexing, we may assume w.l.o.g. that $a \notin F$. Consider the map ϕ determined by $\phi(v_0) = 0$ and $\phi(v_i) = e_i$ for all $i \in [n]$, where e_i is the i-th standard basis vector of $\mathbb{R}^n$. See Fig. 4. Clearly, ϕ is nonsingular, and $\phi(\Delta^n) \subseteq [0, 1]^n$. $\qquad\square$

Lemma 10 *We have* $\phi(\Delta^n) \subseteq_{\phi(x)} [0, 1]^n$ *for all* $x \in \Delta^n \backslash F$.

Proof By definition of F, we have

$$\Delta^n \backslash F = \left\{ \sum_{i=0}^{n} \lambda_i v_i : \sum_{i=1}^{n} \lambda_i < 1, \lambda \geq 0 \right\},$$

and so

$$\phi(\Delta^n \backslash F) = \left\{ y \in [0, 1]^n : \sum_{i=1}^{n} y_i < 1 \right\},$$

which is an open subset of $[0, 1]^n$. But this means that for each $y = \phi(x) \in \phi(\Delta^n \backslash F)$ there exists $\epsilon > 0$ such that

$$B_{\epsilon}^n(y) \cap [0, 1]^n \subseteq B_{\epsilon}^n(y) \cap \phi(\Delta^n \backslash F),$$

which concludes the proof of the lemma. $\qquad\square$

The above lemma tells us in particular that $\phi(\Delta^n) \subseteq_{\phi(a)} [0, 1]^n$. We now apply Corollary 1 with $K = \phi(\Delta^n)$, $\widehat{K} = [0, 1]^n$ and weight functions $w = \widehat{w} = 1$ on $K, \widehat{K}$, respectively. This yields a polynomial upper estimator h of g on $[0, 1]^n$ having

$$E_{\phi(\Delta^n)}^{(r)}(g) \leq 2E_{[0,1]^n}^{(r)}(h) = O(1/r^2),$$

for $r \in \mathbb{N}$ large enough, using Theorem 3 for the right most equality. It remains to apply Lemma 2 to obtain:

$$E_{\Delta^n}^{(r)}(f) = E_{\phi(\Delta^n)}^{(r)}(g) = O\left(1/r^2\right),$$

which concludes the proof of Theorem 9. □

4 General compact sets

In this section we analyze the error $E^{(r)}(f)$ for a general compact set K equipped with the Lebesgue measure. We will show the following two results: when K satisfies a mild assumption (Assumption 1) we prove a convergence rate in $O(\log r/r)$ (Theorem 10), which improves on the previous rate in $O(1/\sqrt{r})$ from [12], and when K is a convex body we prove a convergence rate in $O((\log r/r)^2)$ (Theorem 11), improving the previous rate in $O(1/r)$ from [8]. As a byproduct of our analysis, we can show the stronger bound $O((\log r/r)^\beta)$ when all partial derivatives of f of order at most $\beta - 1$ vanish at a global minimizer (see Theorem 14). We begin with introducing Assumption 1.

Assumption 1 There exist constants $\epsilon_K, \eta_K > 0$ such that

$$\mathrm{vol}\left(B_\delta^n(x) \cap K\right) \geq \eta_K \, \mathrm{vol}\left(B_\delta^n(x)\right) = \delta^n \eta_K \, \mathrm{vol}(B^n) \text{ for all } x \in K \text{ and } 0 < \delta \leq \epsilon_K.$$

In other words, Assumption 1 claims that K contains a constant fraction η_K of the full ball $B_\delta^n(x)$ around x for any radius $\delta > 0$ small enough. This rather mild assumption is discussed in some detail in [12]. In particular, it is implied by the so-called *interior cone condition* used in approximation theory; it is satisfied by convex bodies and, more generally, by sets that are star-shaped with respect to a ball.

Theorem 10 *Let $K \subseteq \mathbb{R}^n$ be a compact set satisfying Assumption 1. Then we have*

$$E^{(r)}(f) = O\left(\frac{\log r}{r}\right).$$

Theorem 11 *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then we have*

$$E^{(r)}(f) = O\left(\frac{\log^2 r}{r^2}\right).$$

Outline of the proofs First of all, if f has a global minimizer which lies in the interior of K, then we may apply Theorem 5 to obtain a convergence rate in $O(1/r^2) = O((\log r/r)^2)$ and so there is nothing to prove. Hence, in the rest of the section, we assume that f has a global minimizer which lies on the boundary of K.

The basic proof strategy for both theorems is to construct explicit sum-of-squares polynomials q_r giving good feasible solutions to the program (3). The building blocks for these polynomials q_r will be provided by the *needle polynomials* from [19]; these are degree r univariate polynomials $v_r^h, \widehat{v}_r^h$, parameterized by a constant $h \in (0, 1)$, that approximate well the Dirac delta at 0 on $[-1, 1]$ and $[0, 1]$, respectively.

For Theorem 10, we are able to use the needle polynomials v_r^h directly after applying the transform $x \mapsto \|x\|$ and selecting the value $h = h(r)$ carefully. We then make use of Lipschitz continuity of f to bound the integral in the objective of (3).

For Theorem 11, a more complicated analysis is needed. We then construct q_r as a product of n univariate well-selected needle polynomials, exploiting geometric properties of the boundary of K in the neighbourhood of a global minimizer.

Simplifying assumptions In order to simplify notation in the subsequent proofs we assume throughout this section that $0 \in K \subseteq B^n \subseteq \mathbb{R}^n$, and $f_{\min,K} = f(0) = 0$, so $a = 0$ is a global minimizer of f over K. As K is compact, and in light of Lemma 2, this is without loss of generality.

We now introduce needle polynomials and their main properties in Sect. 4.1, and then give the proofs of Theorems 10 and 11 in Sects. 4.2 and 4.3, respectively.

4.1 Needle polynomials

We begin by recalling some of the basic properties of the Chebyshev polynomials. The Chebyshev polynomials $T_r \in \mathbb{R}[t]_r$ can be defined by the recurrence relation (4), and also by the following explicit expression:

$$T_r(t) = \begin{cases} \cos(r \arccos t) & \text{for } |t| \leq 1, \\ \frac{1}{2}(t + \sqrt{t^2 - 1})^r + \frac{1}{2}(t - \sqrt{t^2 - 1})^r & \text{for } |t| \geq 1. \end{cases} \tag{15}$$

From this definition, it can be seen that $|T_r(t)| \leq 1$ on the interval $[-1, 1]$, and that $T_r(t)$ is nonnegative and monotone nondecreasing on $[1, \infty)$. The Chebyshev polynomials form an orthogonal basis of $\mathbb{R}[t]$ with respect to the Chebyshev measure (with weight $(1 - t^2)^{-1/2}$ on $[-1, 1]$ and they are used extensively in approximation theory. For instance, they are the polynomials attaining equality in the Markov brother's inequality on $[-1, 1]$, recalled below.

Lemma 11 (Markov Brothers' Inequality; see, e.g., [27]) *Let $p \in \mathbb{R}[t]$ be a univariate polynomial of degree at most r. Then, for any scalars $a < b$, we have*

$$\max_{t \in [a,b]} |p'(t)| \leq \frac{2r^2}{b - a} \cdot \max_{t \in [a,b]} |p(t)|.$$
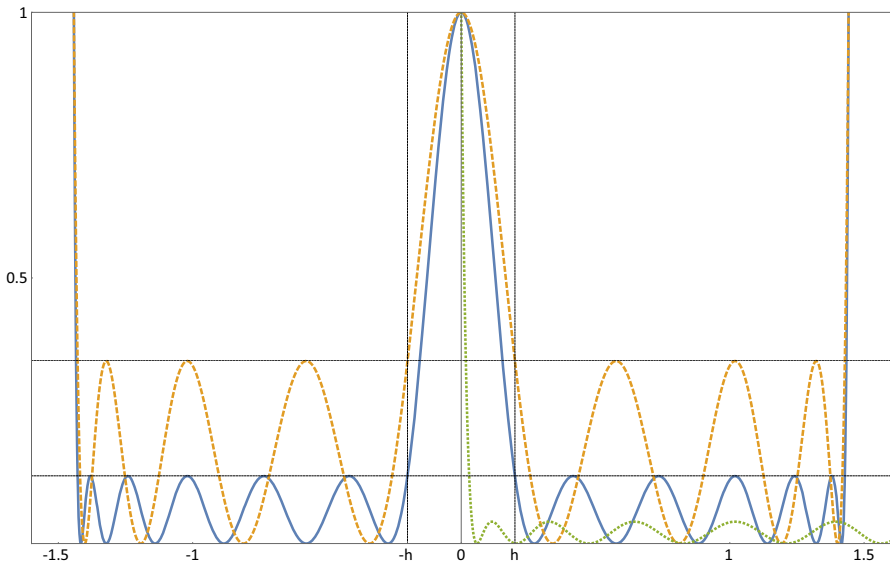
Kroó and Swetits [19] use the Chebyshev polynomials to construct the so-called (univariate) *needle polynomials*.

Definition 5 For $r \in \mathbb{N}$, $h \in (0, 1)$, we define the *needle polynomial* $v_r^h \in \mathbb{R}[t]_{4r}$ by

$$v_r^h(t) = \frac{T_r^2 \left(1 + h^2 - t^2\right)}{T_r^2 \left(1 + h^2\right)}.$$

Additionally, we define the $\frac{1}{2}$-*needle polynomial* $\widehat{v}_r^h \in \mathbb{R}[t]_{4r}$ by

$$\widehat{v}_r^h(t) = T_{2r}^2 \left(\frac{2 + h - 2t}{2 - h}\right) \cdot T_{2r}^{-2} \left(\frac{2 + h}{2 - h}\right).$$

Fig. 5 The needle polynomials v_4^h (dashed), v_6^h (solid) and the $\frac{1}{2}$-needle $\widehat{v}_4^{h2}$ (dotted) for $h = 1/5$

By construction, the needle polynomials v_r^h and $\widehat{v}_r^h$ are squares and have degree $4r$. They approximate well the Dirac delta function at 0 on $[-1, 1]$ and $[0, 1]$, respectively. In [26], a construction similar to the needles presented here is used to obtain the best polynomial approximation of the Dirac delta in terms of the Hausdorff distance.

The needle polynomials satisfy the following bounds (see Fig. 5 for an illustration).

Theorem 12 (cf. [18–20]) *For any $r \in \mathbb{N}$ and $h \in (0, 1)$, the following properties hold for the polynomials v_r^h and $\widehat{v}_r^h$:*

$$v_r^h(0) = 1, \tag{16}$$

$$0 \leq v_r^h(t) \leq 1 \quad for\ t \in [-1, 1], \tag{17}$$

$$v_r^h(t) \leq 4e^{-\frac{1}{2}rh} \quad for\ t \in [-1, 1]\ with\ |t| \geq h, \tag{18}$$

$$\widehat{v}_r^h(0) = 1, \tag{19}$$

$$0 \leq \widehat{v}_r^h(t) \leq 1 \quad for\ t \in [0, 1], \tag{20}$$

$$\widehat{v}_r^h(t) \leq 4e^{-\frac{1}{2}r\sqrt{h}} \quad for\ t \in [0, 1]\ with\ t \geq h. \tag{21}$$

As this result plays a central role in our treatment we give a short proof, following the argument given in [22]. We need the following lemma.

Lemma 12 *For any $r \in \mathbb{N}$, $t \in [0, 1)$ we have $T_r(1 + t) \geq \frac{1}{2}e^{r\sqrt{t}\log(1+\sqrt{2})} \geq \frac{1}{2}e^{\frac{1}{4}r\sqrt{t}}$.*

Proof Using the explicit expression (15) for T_r, we have

$$2T_r(1+t) \geq \left(1+t+\sqrt{(1+t)^2-1}\right)^r = \left(1+t+\sqrt{2t+t^2}\right)^r$$
$$\geq (1+\sqrt{2t})^r = e^{r\log\left(1+\sqrt{2}\cdot\sqrt{t}\right)}.$$

By concavity of the logarithm, we have

$$\log\left(1+\sqrt{2}\sqrt{t}\right) = \log\left(\sqrt{t}\cdot\left(1+\sqrt{2}\right) + \left(1-\sqrt{t}\right)\cdot 1\right)$$
$$\geq \sqrt{t}\cdot\log\left(1+\sqrt{2}\right) + \left(1-\sqrt{t}\right)\log(1) = \sqrt{t}\cdot\log\left(1+\sqrt{2}\right) \geq \frac{1}{4}\sqrt{t},$$

and so, using the above lower bound on $T_r(1+t)$, we obtain

$$T_r(1+t) \geq \frac{1}{2}e^{r\sqrt{t}\log\left(1+\sqrt{2}\right)} \geq \frac{1}{2}e^{\frac{1}{4}r\sqrt{t}}.$$

$\square$

Proof (of Theorem 12) Properties (16), (19) are clear. We first check (17)–(18). If $|t| \leq h$ then $1+h^2 \geq 1+h^2-t^2 \geq 1$, giving $v_r^h(t) \leq v_r^h(0) = 1$ by monotonicity of $T_r(t)$ on $[1,\infty)$. Assume now $h \leq |t| \leq 1$. Then $T_r^2(1+h^2-t^2) \leq 1$ as $1+h^2-t^2 \in [-1,1]$, and $T_r^2(1+h^2) \geq 1$ (again by monotonicity), which implies $v_r^h(t) \leq 1$. In addition, since $T_r(1+h^2) \geq \frac{1}{2}e^{\frac{1}{4}rh}$ by Lemma 12, we obtain $v_r^h(t) \leq T_r^{-2}(1+h^2) \leq 4e^{-\frac{1}{2}rh}$.

We now check (20)–(21). If $t \in [0,h]$ then $\widehat{v}_r^h(t) \leq \widehat{v}_r^h(0) = 1$ follows by monotonicity of $T_{2r}(t)$ on $[1,\infty)$. Assume now $h \leq t \leq 1$. Then, $\frac{2+h-2t}{2-h} \in [-1,1]$ and thus $T_{2r}^2\left(\frac{2+h-2t}{2-h}\right) \leq 1$. On the other hand, we have $T_{2r}^2\left(\frac{2+h}{2-h}\right) \geq 1$, which gives $\widehat{v}_r^h(t) \leq 1$. In addition, as $\frac{2+h}{2-h} \geq 1+h \geq 1$, using again monotonicity of T_{2r} and Lemma 12, we get $T_{2r}^2\left(\frac{2+h}{2-h}\right) \geq T_{2r}^2(1+h) \geq \frac{1}{4}e^{\frac{1}{2}r\sqrt{h}}$, which implies (21). $\square$

We now give a simple lower estimator for a nonnegative polynomial p with $p(0) = 1$. This lower estimator will be useful later to lower bound the integral of the needle and $\frac{1}{2}$-needle polynomials on small intervals $[-h,h]$ and $[0,h]$, respectively.

Lemma 13 *Let $p \in \mathbb{R}[t]_r$ be a polynomial, which is nonnegative over $\mathbb{R}_{\geq 0}$ and satisfies $p(0) = 1$, $p(t) \leq 1$ for all $t \in [0,1]$. Let $\Lambda_r : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be defined by*

$$\Lambda_r(t) = \begin{cases} 1 - 2r^2t & \text{if } t \leq \frac{1}{2r^2}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Lambda_r(t) \leq p(t)$ for all $t \in \mathbb{R}_{\geq 0}$.

Proof Suppose not. Then there exists $s \in \mathbb{R}_{\geq 0}$ such that $\Lambda_r(s) > p(s)$. As $p \geq 0$ on $\mathbb{R}_{\geq 0}$, $p(0) = 1$ and $\Lambda_r(t) = 0$ for $t \geq \frac{1}{2r^2}$, we have $0 < s < \frac{1}{2r^2}$. We find that $p(s) - p(0) < \Lambda_r(s) - 1 = -2r^2s$. Now, by the mean value theorem, there exists an element $z \in (0,s)$ such that $p'(z) = \frac{p(s)-p(0)}{s} < \frac{-2r^2s}{s} = -2r^2$. But this is in contradiction with Lemma 11, which implies that $\max_{t\in[0,1]}|p'(t)| \leq 2r^2$. $\square$

Corollary 4 *Let $h \in (0, 1)$, and let $v_r^h, \widehat{v}_r^h$ as above. Then $\Lambda_{4r}(t) \leq v_r^h(t) = v_r^h(-t)$ and $\Lambda_{4r}(t) \leq \widehat{v}_r^h(t)$ for all $t \in [0, 1]$.*

4.2 Compact sets satisfying Assumption 1

In this section we prove Theorem 10. Recall we assume that K satisfies Assumption 1 with constants ϵ_K and η_K. We also assume that $0 \in \partial K$ is a global minimizer of f over K, $f(0) = 0$, and $K \subseteq B^n$, so that $\epsilon_K < 1$. By Lemma 6, we have $f(x) \leq_0 \beta_{K,f} \|x\|$ on K. Hence, in view of Lemma 3, it suffices to find a polynomial $q_r \in \Sigma_{2r}$ for each $r \in \mathbb{N}$ such that $\int_K q_r(x) dx = 1$ and

$$\int_K q_r(x) \|x\| dx = O\left(\frac{\log r}{r}\right).$$

The idea is to set $q_r(x) \sim \sigma_r^h(x) := v_r^h(\|x\|)$ and then select carefully the constant $h = h(r)$. The main technical component of the proof is the following lemma, which bounds the normalized integral $\int_K \sigma_r^h(x) \|x\|^\beta dx$ in terms of r, h and $\beta \geq 1$. For Theorem 10 we only need the case $\beta = 1$, but allowing $\beta \geq 1$ permits to show a sharper convergence rate when the polynomial f has special properties at the minimizer (see Theorem 14).

Lemma 14 *Let $r \in \mathbb{N}$ and $h \in (0, 1)$ with $\epsilon_K \geq h \geq 1/64r^2$. Let $\beta \geq 1$. Then*

$$\frac{1}{\int_K \sigma_r^h(x) dx} \int_K \sigma_r^h(x) \|x\|^\beta dx \leq h^\beta + Cr^{2n} e^{-\frac{1}{2}hr}, \tag{22}$$

where $C > 0$ is a constant depending only on K.

Proof Set $\rho = 1/64r^2$, so that $\rho \leq h \leq \epsilon_K$. We define the sets

$$B_h := B_h^n(0) \cap K \text{ and } B_\rho := B_\rho^n(0) \cap K \subseteq B_h.$$

Note that $\text{vol}(B_h) \geq \text{vol}(B_\rho) \geq \eta_K \rho^n \text{vol}(B^n)$ by Assumption 1. For $x \in B_h$, we have the bounds $\sigma_r^h(x) \leq 1$ (by (17), since $\|x\| \leq 1$ as $K \subseteq B^n$) and $\|x\|^\beta \leq h^\beta$. On the other hand, for $x \in K \backslash B_h$, we have the bound $\|x\|^\beta \leq 1$, but now $\sigma_r^h(x)$ is exponentially small (by (18)). We exploit this for bounding the integral in (22):

$$\int_K \sigma_r^h(x) \|x\|^\beta dx = \int_{B_h} \sigma_r^h(x) \|x\|^\beta dx + \int_{K \backslash B_h} \sigma_r^h(x) \|x\|^\beta dx$$

$$\leq h^\beta \int_{B_h} \sigma_r^h(x) dx + \int_{K \backslash B_h} \sigma_r^h(x) dx.$$

Combining with the following lower bound on the denominator:

$$\int_K \sigma_r^h(x) dx \geq \int_{B_h} \sigma_r^h(x) dx \geq \int_{B_\rho} \sigma_r^h(x) dx,$$

we get

$$\frac{1}{\int_K \sigma_r^h(x)dx} \int_K \sigma_r^h(x)\|x\|^\beta dx \ \leq \ h^\beta + \frac{\int_{K\setminus B_h} \sigma_r^h(x)dx}{\int_{B_\rho} \sigma_r^h(x)dx}.$$

It remains to upper bound the last term in the above expression. By (18) we have $\sigma_r^h(x) \leq 4e^{-\frac{1}{2}hr}$ for any $x \in K\setminus B_h$ and so

$$\int_{K\setminus B_h} \sigma_r^h(x)dx \leq 4e^{-\frac{1}{2}hr} \cdot \mathrm{vol}(K\setminus B_h) \leq 4e^{-\frac{1}{2}hr} \cdot \mathrm{vol}(B^n).$$

Furthermore, by Lemma 13, we have $\sigma_r^h(x) \geq \Lambda_{4r}(\|x\|) = 1 - 32r^2\|x\| \geq \frac{1}{2}$ for all $x \in B_\rho$. Using Assumption 1 we obtain

$$\int_{B_\rho} \sigma_r^h(x)dx \geq \frac{1}{2}\mathrm{vol}(B_\rho) \geq \frac{1}{2}\eta_K \rho^n \mathrm{vol}(B^n) = \frac{\eta_K \mathrm{vol}(B^n)}{2\cdot 64^n r^{2n}}. \tag{23}$$

Putting things together yields

$$\frac{\int_{K\setminus B_h} \sigma_r^h(x)dx}{\int_{B_h} \sigma_r^h(x)dx} \leq 4e^{-\frac{1}{2}hr} \cdot \mathrm{vol}(B^n)\frac{2\cdot 64^n r^{2n}}{\eta_K \mathrm{vol}(B^n)} = \frac{8\cdot 64^n}{\eta_K}r^{2n}e^{-\frac{1}{2}hr}.$$

This shows the lemma with the constant $C = \frac{8\cdot 64^n}{\eta_K}$. $\qquad\square$

It remains to choose $h = h(r)$ to obtain the polynomials q_r. Our choice here is essentially the same as the one used in [18,26]. With the next result (applied with $\beta = 1$) the proof of Theorem 10 is now complete.

Proposition 2 *For $r \in \mathbb{N}$ and $\beta \geq 1$, set $h(r) = 2(2n + \beta)\log r/r$ and define the polynomial $q_r := \sigma_r^{h(r)}/\int_K \sigma_r^{h(r)}(x)dx$. Then q_r is a sum-of-squares polynomial of degree $4r$ with $\int_K q_r(x)dx = 1$ and*

$$\int_K q_r(x)\|x\|^\beta dx = O\left(\frac{\log^\beta r}{r^\beta}\right).$$

Proof For r sufficiently large, we have $h(r) < \epsilon_K$ and $h(r) \geq 1/64r^2$ and so we may use Lemma 14 to obtain

$$\int_K q_r(x)\|x\|^\beta dx \leq h(r)^\beta + Cr^{2n}e^{-\frac{1}{2}h(r)r}$$

$$= \left(2(2n + \beta)\frac{\log r}{r}\right)^\beta + \frac{C}{r^\beta} = O\left(\frac{\log^\beta r}{r^\beta}\right).$$

$$\square$$

4.3 Convex bodies

We now prove Theorem 11. Here, K is assumed to be a convex body, hence it still satisfies Assumption 1 for certain constants ϵ_K, η_K. As before we also assume that $0 \in \partial K$ is a global minimizer of f in K, $f(0) = 0$ and $K \subseteq B^n$.

If $\nabla f(0) = 0$, then in view of Taylor's theorem (Theorem 1) we know that $f(x) \leq_0 \gamma_{K,f} \|x\|^2$ on K. Hence we may apply Proposition 2 (with $\beta = 2$) to this quadratic upper estimator of f to obtain $E^{(r)}(f) = O(\log^2 r/r^2)$ (recall Lemma 3).

In the rest of this section, we will therefore assume that $\nabla f(0) \neq 0$. In this case, we cannot get a better upper estimator than $f(x) \leq_0 \beta_{K,f} \|x\|$ on K, and so the choice of q_r in Proposition 2 is not sufficient. Instead we will need to make use of the sharper $\frac{1}{2}$-needles $\widehat{v}_r^h$. We will show how to do this in the univariate case first.

The univariate case If $K \subseteq [-1, 1]$ is convex with 0 on its boundary, we may assume w.l.o.g. that $K = [0, b]$ for some $b \in (0, 1]$ (in which case we may choose $\epsilon_K = b$). By using the $\frac{1}{2}$-needle $\widehat{v}_r^h$ instead of the regular needle v_r^h, we immediately get the following analog of Lemma 14.

Lemma 15 *Let $b \in (0, 1]$ and $K = [0, b]$. Let $r \in \mathbb{N}$ and $h \in (0, 1)$ with $b \geq h \geq \frac{1}{64r^2}$. Then we have*

$$\frac{1}{\int_K \widehat{v}_r^h(x)dx} \int_K \widehat{v}_r^h(x)|x|dx \leq h + Cr^2 e^{-\frac{1}{2}\sqrt{hr}}, \tag{24}$$

where $C > 0$ is a universal constant.

Proof Same proof as for Lemma 14, using now the fact that $\widehat{v}_r^h(x) \leq 1$ on K and $\widehat{v}_r^h(x) \leq 4e^{-\frac{1}{2}\sqrt{hr}}$ on $K \backslash B_h$ from (20) and (21). □

Since the exponent in (24) now contains the term '$\sqrt{h}$' instead of 'h', we may square our previous choice of $h(r)$ in Proposition 2 to obtain the following result.

Proposition 3 *Assume $K = [0, b]$. Set $h(r) = \left(2\frac{\log(r^4)}{r}\right)^2 = \left(8\frac{\log r}{r}\right)^2$ and define the polynomial $q_r := \widehat{v}_r^{h(r)} / \int_K \widehat{v}_r^{h(r)}(x)dx$. Then q_r is a sum-of-squares polynomial of degree $4r$ satisfying $\int_K q_r(x)dx = 1$ and*

$$\int_K q_r(x)xdx = O\left(\frac{\log^2 r}{r^2}\right).$$

Proof For r sufficiently large, we have $h(r) < b$ and $h(r) \geq 1/64r^2$ and so we may use Lemma 15 to obtain

$$\int_K q_r(x)xdx \leq h(r) + Cr^2 e^{-\frac{1}{2}r\sqrt{h(r)}} = \left(8\frac{\log r}{r}\right)^2 + \frac{C}{r^2} = O\left(\frac{\log^2 r}{r^2}\right).$$

□

Since $f(x) \leq_0 \beta_{K,f} \cdot x$ on K we obtain $E^{(r)}(f) = O((\log r/r)^2)$, the desired result.

The multivariate case Let $v := \nabla f(0)/\|\nabla f(0)\|$ and let $w_1, w_2, \ldots w_{n-1}$ be an orthonormal basis of $v^\perp$. Then

$$U = U(f) := \{v, w_1, w_2, \ldots, w_{n-1}\} \tag{25}$$

is an orthonormal basis, which we will use as basis of $\mathbb{R}^n$.

The basic idea of the proof is as follows. For any $j \in [n-1]$, if we minimize f in the direction of w_j then we minimize the univariate polynomial $\tilde{f}(t) = f(tw_j)$, which satisfies: $\tilde{f}'(0) = \langle \nabla f(0), w_j \rangle = 0$. Hence, by Taylor's theorem, there is a quadratic upper estimator when minimizing in the direction w_j, so that using a regular needle polynomial will suffice for the analysis. On the other hand, if we minimize f in the direction v, then $\min_{tv \in K} f(tv) = \min_{t \in [0,1]} f(tv)$, since $K \subseteq B^n$ and $v \in N_K(0)$. As explained above this univariate minimization problem can be dealt with using $\frac{1}{2}$-needle polynomials to get the desired convergence rate. This motivates defining the following sum-of-squares polynomials.

Definition 6 For $r \in \mathbb{N}, h \in (0, 1)$ we define the polynomial $\sigma_r^h \in \Sigma_{2nr}$ by

$$\sigma_r^h(x) = \widehat{v}_r^{h^2}(\langle x, v \rangle) \cdot \prod_{j=1}^{n-1} v_r^h(\langle x, w_j \rangle).$$

This construction is similar to the one used by Kroó in [18] to obtain sharp multivariate needle polynomials at boundary points of K.

Proposition 4 *We have* $\sigma_r^h(0) = 1$ *and*

$$\sigma_r^h(x) \in [0, 1] \quad \text{for } x \in K,$$

$$\sigma_r^h(x) \leq 4e^{-\frac{1}{2}hr} \quad \text{for } x \in K \text{ with } \langle x, v \rangle \geq h^2, \tag{26}$$

$$\sigma_r^h(x) \leq 4e^{-\frac{1}{2}hr} \quad \text{for } x \in K \text{ with } \max_{j \in [n-1]} |\langle x, w_j \rangle| \geq h. \tag{27}$$

Proof Note that for any $x \in K$ we have $0 \leq \langle x, v \rangle \leq \|x\| \leq 1$ and $|\langle x, w_j \rangle| \leq \|x\| \leq 1$ for $j \in [n-1]$. The required properties then follow immediately from those of the needle and $\frac{1}{2}$-needle polynomials discussed in Theorem 12. $\qquad \square$

It remains to formulate and prove an analog of Lemma 14 for the polynomial σ_r^h. Before we are able to do so, we first need a few technical statements. For $h > 0$ we define the polytope

$$P_h := \left\{ x \in \mathbb{R}^n : 0 \leq \langle x, v \rangle \leq h^2, |\langle x, w_j \rangle| \leq h \text{ for all } j \in [n-1] \right\}.$$

Note that for $h \in (0, 1)$, the inequalities (26) and (27) can be summarized as

$$\sigma_r^h(x) \leq 4e^{-\frac{1}{2}hr} \text{ for } x \in K \backslash P_h,$$

which means $\sigma_r^h(x)$ is exponentially small for $x \in K$ outside of P_h. When instead $x \in K \cap P_h$, the following two lemmas show that the function value $f(x)$ is small.

Lemma 16 *Let $h \in (0, 1)$. Then $\|x\| \leq \sqrt{n}h$ for all $x \in P_h$.*

Proof Let $x \in P_h$. By expressing x in the orthonormal basis U from (25), we obtain

$$\|x\|^2 = \langle x, v \rangle^2 + \sum_{i=1}^{n-1} \langle x, w_i \rangle^2 \leq nh^2,$$

using the definition of P_h for the second inequality. □

Lemma 17 *Let $h \in (0, 1)$. Then $f(x) \leq (\beta_{K,f} + n\gamma_{K,f})h^2$ for all $x \in K \cap P_h$.*

Proof Using Taylor's Theorem 1, Lemma 16 and $\langle x, v \rangle \leq h^2$ for $x \in P_h$, we obtain

$$f(x) \leq \langle \nabla f(0), x \rangle + \gamma_{K,f} \|x\|^2 \leq \|\nabla f(0)\| \langle x, v \rangle + n\gamma_{K,f}h^2 \leq (\beta_{K,f} + n\gamma_{K,f})h^2.$$

□

We now give a lower bound on $\int_{K \cap P_h} \sigma_r^h(x)dx$ [compare to (23)]. First we need the following bound on $\text{vol}(K \cap P_h)$.

Lemma 18 *Let $h \in (0, 1)$. If $h < \epsilon_K$ then we have: $\text{vol}(K \cap P_h) \geq \eta_K h^{2n} \text{vol}(B^n)$.*

Proof Consider the halfspace $H_v := \{x \in \mathbb{R}^n : \langle v, x \rangle \geq 0\}$. As $v \in N_K(0)$, we have the inclusion $K \subseteq H_v$. We show that $B_{h^2}^n(0) \cap H_v \subseteq P_h$, implying that $B_{h^2}^n(0) \cap K \subseteq B_{h^2}^n(0) \cap H_v \subseteq P_h$. Let $x \in B_{h^2}^n(0) \cap H_v$. By expressing x in the orthonormal basis $U(f)$ from (25), we get $\|x\|^2 = \langle v, x \rangle^2 + \sum_{j=1}^{n-1} \langle w_j, x \rangle^2 \leq h^4$. Since $x \in H_v$ and $0 < h < 1$, we get $0 \leq \langle v, x \rangle \leq h^2$ and $|\langle w_j, x \rangle| \leq h^2 \leq h$, thus showing $x \in P_h$. See Fig. 6 for an illustration. We may now apply Assumption 1 to find $\text{vol}(P_h) \geq \text{vol}(B_{h^2}^n(0) \cap K) \geq \eta_K h^{2n} \text{vol}(B^n)$. □

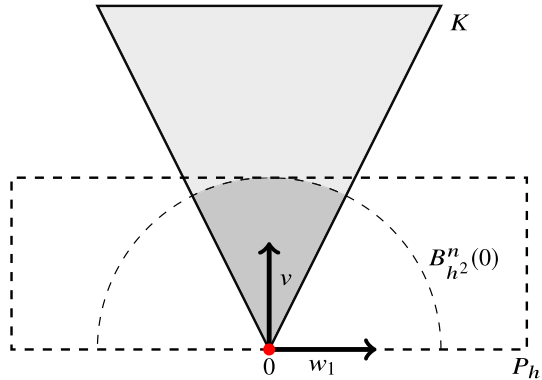**Lemma 19** *Let $r \in \mathbb{N}$, $h \in (0, 1)$. Assume that $\epsilon_K > h > \rho = 1/64r^2$. Then*

$$\int_{K \cap P_h} \sigma_r^h(x)dx \geq \frac{1}{2^n}\eta_K \rho^{2n} \text{vol}(B^n).$$

Proof The integral $\int_{K \cap P_h} \sigma_r^h(x)dx$ is equal to

$$\int_{K \cap P_h} \widehat{v}_r^{h^2}(\langle x, v \rangle) \cdot \prod_{i=1}^{n-1} v_r^h(\langle x, w_i \rangle)dx \qquad \text{[using Definition 6]}$$

$$\geq \int_{K \cap P_h} \Lambda_{4r}(\langle x, v \rangle) \cdot \prod_{i=1}^{n-1} \Lambda_{4r}(|\langle x, w_i \rangle|)dx \qquad \text{[using Corollary 4]}$$

Fig. 6 Overview of the situation in the proof of Lemma 18. Note that as long as $v \in N_K(0)$, the entire region $B_{h^2}^n(0) \cap K$ (in dark gray) is contained in P_h

$$\geq \int_{K \cap P_\rho} \Lambda_{4r}(\langle x, v \rangle) \cdot \prod_{i=1}^{n-1} \Lambda_{4r}(|\langle x, w_i \rangle|) dx \qquad \text{[since } P_\rho \subseteq P_h\text{]}$$

$$\geq \int_{K \cap P_\rho} \Lambda_{4r}(\rho) \cdot \prod_{i=1}^{n-1} \Lambda_{4r}(\rho) dx \qquad \text{[since } \Lambda_{4r}(t) \geq \Lambda_{4r}(\rho) \text{ if } t \in [0, \rho]\text{]}$$

$$\geq \int_{K \cap P_\rho} \frac{1}{2^n} dx = \frac{1}{2^n} \text{vol}(K \cap P_\rho) \qquad \left[\text{as } \Lambda_{4r}(\rho) = \frac{1}{2}\right]$$

$$\geq \frac{1}{2^n} \eta_K \rho^{2n} \text{vol}(B^n) \qquad \text{[using Lemma 18]}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$$

We are now able to prove an analog of Lemma 14.

Lemma 20 *Let $r \in \mathbb{N}$ and $h \in (0, 1)$. If $\epsilon_K > h > 1/64r^2$ then we have*

$$\frac{1}{\int_K \sigma_r^h(x) dx} \int_K \sigma_r^h(x) f(x) dx \leq (\beta_{K,f} + n\gamma_{K,f}) h^2 + C' r^{4n} e^{-\frac{1}{2}hr},$$

where C' is a constant depending only on K.

Proof Set $\rho = 1/64r^2$. By Lemma 17, $f(x) \leq (\beta_{K,f} + n\gamma_{K,f}) h^2$ for all $x \in K \cap P_h$. Moreover, by Proposition 4, we have $\sigma_r^h(x) \leq 4e^{-\frac{1}{2}hr}$ for all $x \in K \setminus P_h$. Hence,

$$\int_K \sigma_r^h(x) f(x) dx = \int_{K \cap P_h} \sigma_r^h(x) f(x) dx + \int_{K \setminus P_h} \sigma_r^h(x) f(x) dx$$

$$\leq (\beta_{K,f} + n\gamma_{K,f}) h^2 \int_{K \cap P_h} \sigma_r^h(x) dx + 4e^{-\frac{1}{2}hr} f_{\max,K} \text{vol}(B^n),$$

where $f_{\max,K} = \max_{x \in K} f(x)$. Combining with

$$\int_K \sigma_r^h(x) dx \geq \int_{K \cap P_h} \sigma_r^h(x) dx \geq \frac{1}{2^n} \eta_K \rho^{2n} \text{vol}(B^n),$$

where we use Lemma 19 for the last inequality, we obtain

$$\frac{1}{\int_K \sigma_r^h(x)dx} \int_K \sigma_r^h(x)f(x)dx \le (\beta_{K,f} + n\gamma_{K,f})h^2 + \frac{4 \cdot 2^n \cdot 64^{2n} f_{\max,K}}{\eta_K}r^{4n}e^{-\frac{1}{2}hr}.$$

This shows the lemma, with the constant $C' = \frac{4 \cdot 2^n \cdot 64^{2n} f_{\max,K}}{\eta_K}$. □

From the preceding lemma we get the following corollary, which immediately implies Theorem 11.

Corollary 5 *For any $r \in \mathbb{N}$, set $h(r) = (8n + 4)\frac{\log r}{r}$ and consider the polynomial $q_r := \sigma_r^{h(r)}/\int_K \sigma_r^{h(r)}(x)dx$. Then q_r is a sum-of-squares polynomial of degree $4nr$, which satisfies $\int_K q_r(x)dx = 1$ and*

$$\int_K q_r(x)f(x)dx = O\left(\frac{\log^2 r}{r^2}\right).$$

Proof For r sufficiently large, we have $\epsilon_K > h(r) > 1/64r^2$ and so we may apply Lemma 20, which implies directly

$$\int_K q_r(x)f(x)dx \le (\beta_{K,f} + n\gamma_{K,f})h(r)^2 + C'r^{4n}e^{-\frac{1}{2}rh(r)} = O\left(\frac{\log^2 r}{r^2}\right).$$

□

5 Numerical experiments

In this section, we illustrate some of the results in this paper with numerical examples. We consider the test functions listed below in Table 2, the latter four of which are well-known in global optimization and also used for this purpose in [12].

We compare the behaviour of the error $E_K^{(r)}(f)$ for these functions on different sets K, namely the hypercube, the unit ball, and a regular octagon in $\mathbb{R}^2$. On the unit ball and the regular octagon, we consider the Lebesgue measure. On the hypercube, we

Table 2 Polynomial test functions

Name	Formula	$f_{\min,[-1,1]^2}$
Linear	$f_{li}(x) = x_1$	$f_{li}(-1, 0) = -1$
Quadratic	$f_{qu}(x) = x_1 + x_2^2$	$f_{qu}(-1, 0) = -1$
Booth	$f_{bo}(x) = (10x_1 + 20x_2 - 7)^2 + (20x_1 + 10x_2 - 5)^2$	$f_{bo}(\frac{1}{10}, \frac{3}{10}) = 0$
Matyas	$f_{ma}(x) = 26(x_1^2 + x_2^2) - 48x_1x_2$	$f_{ma}(0, 0) = 0$
Camel	$f_{ca}(x) = 50x_1^2 - \frac{2625}{4}x_1^4 + \frac{15625}{6}x_1^6 + 25x_1x_2 + 25x_2^2$	$f_{ca}(0, 0) = 0$
Motzkin	$f_{mo}(x) = 64x_1^4x_2^2 + 64x_1^2x_2^4 - 48x_1^2x_2^2 + 1$	$f_{mo}(\pm\frac{1}{2}, \pm\frac{1}{2}) = 0$

consider both the Lebesgue measure and the Chebyshev measure. In each case, we compute the Lasserre bounds of order r in the range $1 \leq r \leq 20$, corresponding to sos-densities of degree up to 40.

Computing the bounds As explained in Sect. 1, it is possible to compute the degree $2r$ Lasserre bound $f_{K,\mu}^{(r)}$ by finding the smallest eigenvalue of the truncated moment matrix $M_{r,f}$ of f, defined by

$$M_{r,f}(\alpha, \beta) = \int_K f p_\alpha p_\beta d\mu(x) \quad (\alpha, \beta \in \mathbb{N}_r^n),$$

assuming that one has an orthonormal basis $\{p_\alpha : \alpha \in \mathbb{N}_r^n\}$ of $\mathbb{R}[x]_r$ w.r.t. the inner product induced by the measure μ, i.e., such that $\int_K p_\alpha p_\beta d\mu(x) = \delta_{\alpha,\beta}$.

More generally, if we use an arbitrary linear basis $\{p_\alpha\}$ of $\mathbb{R}[x]_r$ then the bound $f_{K,\mu}^{(r)}$ is equal to the smallest *generalized eigenvalue* of the system:

$$M_{r,f} v = \lambda B_r v, \tag{28}$$

where $B_r := M_{r,1}$ is the matrix with entries $B_r(\alpha, \beta) = \int_K p_\alpha p_\beta d\mu(x)$. Note that if the p_α are orthonormal, then B_r is the identity matrix and one recovers the eigenvalue formulation of Sect. 1. For details, see, e.g., [21].

This formulation in terms of generalized eigenvalues allows us to work with the standard monomial basis of $\mathbb{R}[x]_r$. To compute the entries of the matrices $M_{r,f}$ and B_r, we therefore only require knowledge of the moments:

$$\int_K x^\alpha d\mu(x) \quad (\alpha \in \mathbb{N}^n).$$

For the hypercube, simplex and unit ball, closed form expressions for these moments are known (see, e.g., Table 1 in [9]). For the octagon, they can then be computed by triangulation. We solve the generalized eigenvalue problem (28) using the `eig` function of the SciPy software package.

The linear case We consider first the linear case $f(x) = f_{li}(x) = x_1$ and $K = [-1, 1]^2$ equipped with the Lebesgue measure. Figure 7 shows the values of the parameters $E_K^{(r)}(f_{li})$ and $E_K^{(r)}(f_{li}) \cdot r^2$. In accordance with Theorem 3 [and 2(ii)], it appears indeed that $E_K^{(r)}(f_{li}) = O(1/r^2)$, as suggested by the fact that the parameter $E_K^{(r)}(f_{li}) \cdot r^2$ approaches a constant value as r grows.

The unit ball Next, we consider the unit ball B^2, again equipped with the Lebesgue measure. Figure 8 shows the values of the ratio

$$E_{B^2}^{(r)}(f_*)/E_{[-1,1]^2}^{(r)}(f_*) \tag{29}$$

for $* \in \{li, qu, bo, ma, ca, mo\}$. In each case, the ratio (29) appears to tend to a constant value, suggesting that the error $E_K^{(r)}(f_*)$ has similar asymptotic behaviour
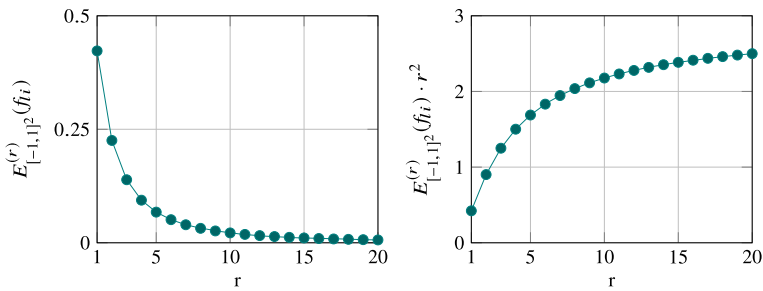
Fig. 7 The error of upper bounds for $f(x) = x_1$ computed on $[-1, 1]^2$ w.r.t. the Lebesgue measure
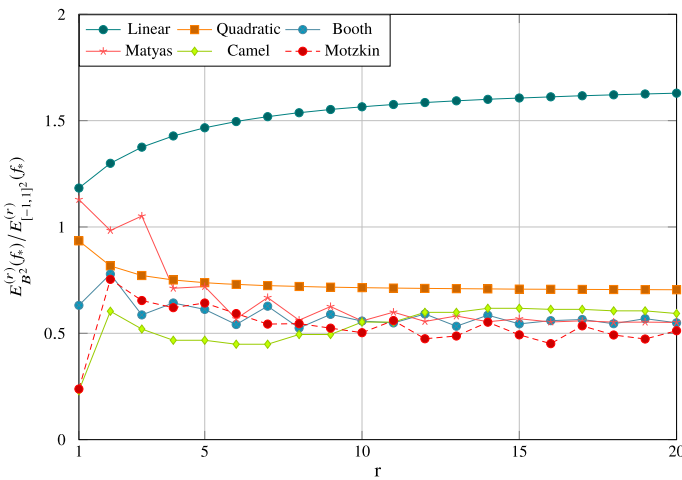

Fig. 8 Comparison of the errors of upper bounds for the functions in Table 2 computed on $[-1, 1]^2$ and the unit ball B^2 w.r.t. the Lebesgue measure

for $K = [-1, 1]^2$ and $K = B^2$. This matches the result of Theorem 4 both in the case of a minimizer on the boundary ($* \in \{li, qu\}$) and in the case of a minimizer in the interior ($* \in \{bo, ma, ca, mo\}$).

The regular octagon Consider now the regular octagon (with the Lebesgue measure)

$$O = \text{conv}\left\{(\pm 1, 0), (0, \pm 1), \left(\pm\frac{1}{2}\sqrt{2}, \pm\frac{1}{2}\sqrt{2}\right)\right\} \subseteq [-1, 1]^2, \qquad (30)$$

which is an example of a convex body that is not ball-like (see Definition 3). Note that as a result, the strongest theoretical guarantee we have shown for the convergence rate of the Lasserre bounds on O is in $O(\log^2 r / r^2)$ (see Theorem 11). Figure 9 shows the values of the ratio

$$E_O^{(r)}(f_*) / E_{[-1,1]^2}^{(r)}(f_*) \qquad (31)$$
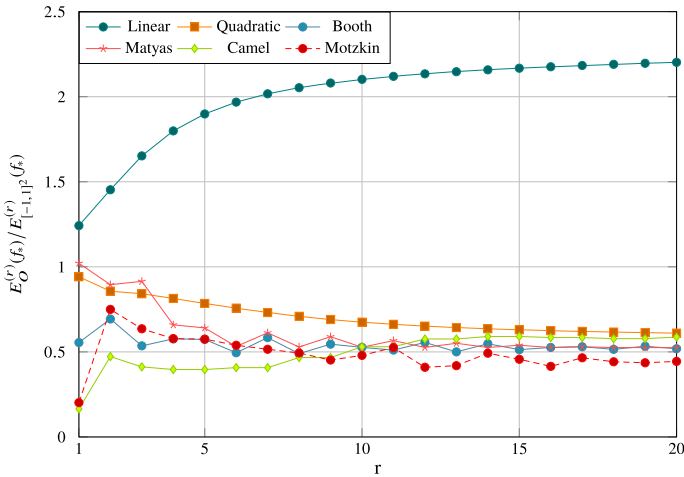
Fig. 9 Comparison of the errors of upper bounds for the functions in Table 2 computed on $[-1, 1]^2$ and the regular octagon O [see (30)] w.r.t. the Lebesgue measure

for $* \in \{li, qu, bo, ma, ca, mo\}$. As for the unit ball, the ratio (31) seemingly tends to a constant value for each of the test polynomials. This indicates a similar asymptotic behaviour of the error $E_K^{(r)}(f_*)$ for $K = [-1, 1]^2$ and $K = O$ and suggests that the convergence rate guaranteed by Theorem 11 might not be tight in this instance.

The Chebyshev measure Finally, we consider the Chebyshev measure $d\mu(x) = (1 - x_1^2)^{-1/2}(1 - x_2^2)^{-1/2}dx$ on $[-1, 1]^2$, which we compare to the Lebesgue measure. Figure 10 shows the values of the fraction

$$E_{[-1,1]^2,\mu}^{(r)}(f_*)/E_{[-1,1]^2}^{(r)}(f_*) \tag{32}$$

for $* \in \{li, qu, bo, ma, ca, mo\}$. Again, we observe that the fraction (32) appears to tend to a constant value in each case, matching the result of Theorem 3.

6 Concluding remarks

Extension to non-polynomial functions Throughout, we have assumed that the function f is a polynomial. Strictly speaking, this assumption is not necessary to obtain our results. For the results in Sect. 3 and in Theorem 11, it suffices that f has an upper estimator, exact at one of its global minimizers on K, and satisfying the properties given in Lemma 4. In light of Taylor's Theorem, such an upper estimator exists for all $f \in C^2(\mathbb{R}^n, \mathbb{R})$. For Theorem 10, it is even sufficient that f satisfies $f(x) \leq f(a) + M_f ||x - a||$ for all $x \in f$, where $M_f > 0$ is a constant. That is, it suffices that f is Lipschitz continuous on K. Finally, as shown in [11, Theorem 10], results on the convergence rate of the bounds $f^{(r)}$ for polynomials f extend directly to the case of rational functions f.
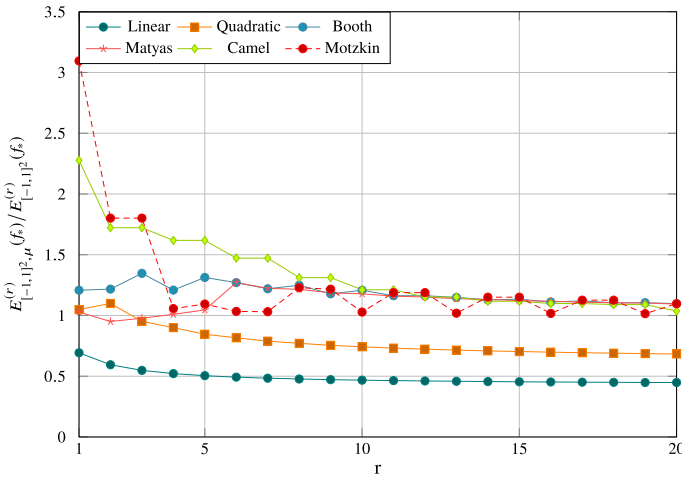
Fig. 10 Comparison of the errors of upper bounds for the functions in Table 2 computed on $[-1, 1]^2$ w.r.t. the Lebesgue and Chebyshev measures

Accelerated convergence results For the minimization of linear polynomials the convergence rate of the bounds $f^{(r)}$ is shown to be in the order $\Theta(1/r^2)$ for the hypercube [10] and the unit sphere [11]. Hence, for arbitrary polynomials, a quadratic rate is the best we can hope for. On the other hand, if we restrict to a class of functions with additional properties, then a better convergence rate can be shown. Indeed, a faster convergence rate can be achieved when the function f has many vanishing derivatives at a global minimizer. We will make use of the following consequence of Taylor's theorem.

Theorem 13 (Taylor's theorem) *Assume* $f \in C^\beta(\mathbb{R}^n, \mathbb{R})$ *with* $\beta \geq 1$. *Then we have*

$$f(x) \leq \sum_{\alpha \in \mathbb{N}^n, |\alpha| \leq \beta - 1} \frac{1}{\alpha!} (D^\alpha)(f)(a)(x-a)^\alpha + \delta_{K,f} \|x - a\|^\beta \quad \text{for all } x \in K$$

for some constant $\delta_{K,f} > 0$.

Theorem 14 *Let* $f \in C^\beta(\mathbb{R}^n, \mathbb{R})$ ($\beta \geq 1$) *and let* a *be a global minimizer of* f *on* K. *Assume that all partial derivatives* $(D^\alpha f)(a)$ *vanish for* $1 \leq |\alpha| \leq \beta - 1$. *Then, given any* $\epsilon > 0$ *we have*

$$E^{(r)}(f) = O\left(\frac{\log^\beta r}{r^\beta}\right) = o\left(\frac{1}{r^{\beta - \epsilon}}\right).$$

Proof This follows as a direct application of Proposition 2. ☐

This applies, e.g., for the univariate polynomial $f(x) = x^\beta$ on the interval $K = [0, 1]$. As an application we can answer in the negative a question posed in [10], where the authors asked about the existence of a 'saturation result' for the convergence rate of the Lasserre upper bounds, namely whether

$$E^{(r)}(f) = o\left(\frac{1}{r^2}\right) \overset{?}{\Longleftrightarrow} \ f \text{ is a constant polynomial.}$$

Application to the generalized problem of moments (GPM) and cubature rules As shown in [7] results on the convergence analysis of the bounds $E^{(r)}(f)$ have direct implications for the following generalized moment problem (GMP):

$$\text{val} := \inf\left\{\int_K f_0(x)d\nu(x) : \int_K f_i(x)d\nu(x) = b_i \ (i \in [m])\right\},$$

where $b_i \in \mathbb{R}$ and $f_i \in \mathbb{R}[x]$ are given, and the variable ν is a Borel measure on K. Bounds can be obtained by searching for measures of the form $q_r d\mu$ with μ a given Borel measure on K and $q_r \in \Sigma_r$. Their quality can be analyzed via the parameter

$$\Delta(r) = \min_{q_r \in \Sigma_r} \max_{i=0}^{m} \left|\int_K f_i(x)q_r(x)d\mu(x) - b_i\right|,$$

setting $b_0 = \text{val}$. It is shown in [7] (see also [9]) that, if $E_{K,\mu}^{(r)}(f) = O(\epsilon(r))$ for all polynomials f, then $\Delta(r) = O(\sqrt{\epsilon(r)})$. Hence, our results in this paper imply directly that $\Delta(r) = O(\log r / r)$ for general convex bodies and $O(1/r)$ for hypercubes, balls and simplices (recall Table 1 for exact details). An important instance of (GMP) is finding cubature schemes for numerical integration on K (see, e.g., [9] and references therein). If $\{x^{(j)}, \lambda_j : j \in [N]\}$ form a cubature scheme with positive weights $\lambda_j > 0$ that permits to integrate any polynomial of degree at most $d + 2r$ on K w.r.t. measure μ, then, as shown in [23], we have

$$f_{K,\mu}^{(r)} \geq f_{cub}^{(r)} := \min_{j=1}^{N} f\left(x^{(j)}\right) \geq f_{\min,K}.$$

Hence any upper bound on $f_{K,\mu}^{(r)}$ directly gives an upper bound on the parameter $f_{cub}^{(r)}$. Conversely, any lower bound on $f_{cub}^{(r)}$ implies a lower bound on $f_{K,\mu}^{(r)}$, which is the fact used in [10,11] to show the lower bound $\Omega(1/r^2)$ for the hypercube and the sphere.

Finally, let us mention that the needle polynomials are used already in [18] to study cubature rules. There, the author considers degree r cubature rules for which the sum $\sum_{j\in[N]} |\lambda_j|$ is polynomially bounded in r. For all $x \in K$, $A \subseteq K$, define the parameters

$$\rho(x, A) := \sup\left\{h > 0 : B_h^n(x) \cap A = \emptyset\right\}, \ \rho(K, A) := \sup_{x \in K} \rho(x, A),$$

which indicate how densely A is distributed at x or in K, respectively. Kroó [18] shows that if X_r is the set of nodes of a degree r cubature rule on a convex body K, we then have

$$\rho(K, X_r) = O(\log r / r)$$

and that, if $x_0 \in K$ is a vertex of K, we even have

$$\rho(x_0, X_r) = O\left(\log^2 r/r^2\right).$$

Although the asymptotic rates here are the same as the ones we find in Theorems 10 and 11, we are not aware of any direct link between the density of cubature points and the convergence rate of the Lasserre upper bounds.

Some open questions There are several natural questions left open by this work. The first natural question is whether the convergence rate in $O(1/r^2)$ can be proved for all convex bodies. So far we can only prove a rate in $O(\log^2 r/r^2)$, but we suspect that the $\log r$ term is just a consequence of the analysis technique used here. The computational results for the octagon in Sect. 5 seem to support this. Another question is whether this also applies to general compact sets under Assumption 1, since we know of no example showing this is not possible.

In particular, it is interesting to determine the exact rate of convergence for polytopes. We could so far only deal with hypercubes and simplices. The main tool we used was the 'local similarity' of the simplex with the hypercube. For a general polytope K, if the minimum is attained at a point lying in the interior of K or of one of its facets, then we can still apply the 'local similarity' tool (and deduce the $O(1/r^2)$ rate). However, at other points (like its vertices) K is in general not locally similar to the hypercube, so another proof technique seems needed. A possible strategy could be splitting K into simplices and using the known convergence rate for the simplex containing a global minimizer; however, a difficulty there is keeping track of the distribution of mass of an optimal sum-of-squares on the other simplices.

References

1. Bonnesen, T., Fenchel, W.: Theory of Convex Bodies. BCS Associates, Mountain View (1987)
2. Borwein, J., Lewis, A.: Convex Analysis and Nonlinear Optimization. Springer, Berlin (2006)
3. Dai, F., Xu, Y.: Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer, Berlin (2013)
4. Dalla, L., Hatziafratis, T.: Strict convexity of sets in analytic terms. J. Aust. Math. Soc. **81**(1), 49–61 (2006)
5. Dalla, L., Samiou, E.: Curvature and q-strict convexity. Beitr. Algebra Geom. **48**, 83–93 (2007)
6. de Boor, C.: Cutting corners always works. Comput. Aided Geom. Des. **4**(1–2), 125–131 (1987)
7. de Klerk, E., Kuhn, D., Postek, K.: Distributionally robust optimization with polynomial densities: theory, models and algorithms. Math. Program. (2019). https://doi.org/10.1007/s10107-019-01429-5

8. de Klerk, E., Laurent, M.: Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing. Math. Oper. Res. **43**, 1317–1325 (2017)
9. de Klerk, E., Laurent, M.: A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In: Araujo, C., Benkart, G., Praeger, C., Tanbay, B. (eds.) World Women in Mathematics 2018. Association for Women in Mathematics Series, vol. 20, pp. 17–56. Springer, Cham (2019) https://doi.org/10.1007/978-3-030-21170-7_1
10. de Klerk, E., Laurent, M.: Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube. Math. Oper. Res. (2019). https://doi.org/10.1287/moor.2018.0983
11. de Klerk, E., Laurent, M.: Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. Math. Program. (2020) (**accepted/in press**)
12. de Klerk, E., Laurent, M., Sun, Z.: Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization. Math. Program. **162**(1), 363–392 (2017)
13. de Klerk, E., Pasechnik, D.: Approximation of the stability number of a graph via copositive programming. SIAM J. Optim. **12**(4), 875–892 (2002)
14. de Rham, G.: Un peu de mathématiques à propos d'une courbe plane. Elem. Math. **2**(73–76), 89–97 (1947)
15. Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere. arXiv:1210.5048 (2012)
16. Dunkl, C.F., Xu, Y.: Orthogonal Polynomials of Several Variables, Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge (2014)
17. Koutroufiotis, D.: On Blaschke's rolling theorems. Arch. Math. **23**(1), 655–670 (1972)
18. Kroó, A.: Multivariate "needle" polynomials with application to norming sets and cubature formulas. Acta Math. Hung. **147**(1), 46–72 (2015)
19. Kroó, A., Swetits, J.J.: On density of interpolation points, a Kadec-type theorem, and Saff's principle of contamination in Lp-approximation. Constr. Approx. **8**(1), 87–103 (1992)
20. Kroó, A., Lubinsky, D.: Christoffel functions and universality in the bulk for multivariate orthogonal polynomials. Can. J. Math. **65**, 600–620 (2012)
21. Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. SIAM J. Optim. **21**(3), 864–885 (2010)
22. Lasserre, J.B., Pauwels, E.: The empirical Christoffel function with applications in data analysis. Adv. Comput. Math. **45**(3), 1439–1468 (2019)
23. Martinez, A., Piazzon, F., Sommariva, A., Vianello, M.: Quadrature-based polynomial optimization. Optim. Lett. (2019). https://doi.org/10.1007/s11590-019-01416-x
24. Motzkin, T.S., Straus, E.G.: Maxima for graphs and a new proof of a theorem of Turán. Can. J. Math. **17**, 533–540 (1965)
25. Rivlin, T.J.: The Chebyshev Polynomials. Wiley-Interscience, New York (1990)
26. Sendov, B.: Hausdorff Approximations. Kluwer Academic Publishers, Dordrecht (1990)
27. Shadrin, A.: Twelve Proofs of the Markov Inequality. Approximation Theory—A Volume Dedicated to B. Bojanov, pp. 233–298. Professor Marin Drinov Publishing House, Sofia (2004)