**manuscript No.**
(will be inserted by the editor)

# Convergence analysis for Lasserre's measure–based hierarchy of upper bounds for polynomial optimization

**Etienne de Klerk · Monique Laurent · Zhao Sun**

Received: date / Accepted: date

**Abstract** We consider the problem of minimizing a continuous function $f$ over a compact set $\mathbf{K}$. We analyze a hierarchy of upper bounds proposed by Lasserre in [*SIAM J. Optim.* 21(3) (2011), pp. $864-885$], obtained by searching for an optimal probability density function $h$ on $\mathbf{K}$ which is a sum of squares of polynomials, so that the expectation $\int_{\mathbf{K}} f(x)h(x)dx$ is minimized. We show that the rate of convergence is no worse than $O(1/\sqrt{r})$, where $2r$ is the degree bound on the density function. This analysis applies to the case when $f$ is Lipschitz continuous and $\mathbf{K}$ is a full-dimensional compact set satisfying some boundary condition (which is satisfied, e.g., for convex bodies). The $r$th upper bound in the hierarchy may be computed using semidefinite programming if $f$ is a polynomial of degree $d$, and if all moments of order up to $2r+d$ of the Lebesgue measure on $\mathbf{K}$ are known, which holds, for example, if $\mathbf{K}$ is a simplex, hypercube, or a Euclidean ball.

**Keywords** Polynomial optimization · Semidefinite optimization · Lasserre hierarchy

**Mathematics Subject Classification (2000)** 90C22 · 90C26 · 90C30

Etienne de Klerk
Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands
E-mail: E.deKlerk@uvt.nl

Monique Laurent
Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University
CWI, Postbus 94079, 1090 GB Amsterdam, The Netherlands
E-mail: M.Laurent@cwi.nl

Zhao Sun
École Polytechnique de Montréal
GERAD–HEC Montreal 3000, Côte-Sainte-Catherine Rd, Montreal, QC H3T 2A7, Canada
E-mail: Zhao.Sun@polymtl.ca

## 1 Introduction and Preliminaries

### 1.1 Background

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

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

Our main interest will be in the case where $f$ is a polynomial, and $\mathbf{K}$ is defined by polynomial inequalities and equations. For such problems, active research has been done in recent years to construct tractable hierarchies of (upper and lower) bounds for $f_{\min,\mathbf{K}}$, based on using sums of squares of polynomials and semidefinite programming (SDP). The starting point is to reformulate $f_{\min,\mathbf{K}}$ as the problem of finding the largest scalar $\lambda$ for which the polynomial $f - \lambda$ is nonnegative over $\mathbf{K}$ and then to replace the hard positivity condition by a suitable sum of squares decomposition. Alternatively, one may reformulate $f_{\min,\mathbf{K}}$ as the problem of finding a probability measure $\mu$ on $K$ minimizing the integral $\int_{\mathbf{K}} f d\mu$. These two dual points of view form the basis of the approach developed by Lasserre [15] for building hierarchies of semidefinite programming based lower bounds for $f_{\min,\mathbf{K}}$ (see also [16,19] for an overview). Asymptotic convergence to $f_{\min,\mathbf{K}}$ holds (under some mild conditions on the set $\mathbf{K}$). Moreover, error estimates have been shown in [26,24] when $\mathbf{K}$ is a general basic closed semi-algebraic set, and in [4,5,6,7,9,11,27] for simpler sets like the standard simplex, the hypercube and the unit sphere. In particular, [26] shows that the rate of convergence of the hierarchy of lower bounds based on Schmüdgen's Positivstellensatz is in the order $O(1/\sqrt[c]{2r})$, while [24] shows a convergence rate in $O(1/\sqrt[c']{\log(2r/c')})$ for the (weaker) hierarchy of bounds based on Putinar's Positivstellensatz. Here, $c, c'$ are constants (not explicitly known) depending only on $\mathbf{K}$, and $2r$ is the selected degree bound. For the case of the hypercube, [4] shows (using Bernstein approximations) a convergence rate in $O(1/r)$ for the lower bounds based on Schmüdgen's Positivstellensatz.

On the other hand, by selecting suitable probability measures on $\mathbf{K}$, one obtains upper bounds for $f_{\min,\mathbf{K}}$. This approach has been investigated, in particular, for minimization over the standard simplex and when selecting some discrete distributions over the grid points in the simplex. The multinomial distribution is used in [23,6] to show convergence in $O(1/r)$ and the multivariate hypergeometric distribution is used in [7] to show convergence in $O(1/r^2)$ for quadratic minimization over the simplex (and in the general case assuming a rational minimizer exists).

Additionnally, Lasserre [17] shows that, if we fix any measure $\mu$ on $\mathbf{K}$, then it suffices to search for a polynomial density function $h$ which is a sum of squares and minimizes the integral $\int_{\mathbf{K}} fhd\mu$ in order to compute the minimum $f_{\min,\mathbf{K}}$ over $\mathbf{K}$ (see Theorem 1 below). By adding degree constraints on the polynomial density $h$ we get a hierarchy of upper bounds for $f_{\min,\mathbf{K}}$ and our main objective in this paper is to analyze the quality of this hierarchy of upper bounds for $f_{\min,\mathbf{K}}$. Next we will recall this result of Lasserre [17] and then we describe our main results.

### 1.2 Lasserre's hierarchy of upper bounds

Throughout, $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ is the set of polynomials in $n$ variables with real coefficients, and $\mathbb{R}[x]_r$ is the set of polynomials with degree at most $r$. $\Sigma[x]$ is the set of sums of squares of polynomials, and $\Sigma[x]_r = \Sigma[x] \cap \mathbb{R}[x]_{2r}$ consists of all sums of squares of polynomials with degree at most $2r$. We now recall the result of Lasserre [17], which is based on the following characterization for nonnegative continuous functions on a compact set $\mathbf{K}$.

**Theorem 1** *[17, Theorem 3.2] Let $\mathbf{K} \subseteq \mathbb{R}^n$ be compact, let $\mu$ be an arbitrary finite Borel measure supported by $\mathbf{K}$, and let $f$ be a continuous function on $\mathbb{R}^n$. Then, $f$ is nonnegative on $\mathbf{K}$ if and only if*

$$\int_{\mathbf{K}} g^2 f d\mu \geq 0 \quad \forall g \in \mathbb{R}[x].$$

*Therefore, the minimum of $f$ over $\mathbf{K}$ can be expressed as*

$$f_{\min,\mathbf{K}} = \inf_{h \in \Sigma[x]} \int_{\mathbf{K}} h f d\mu \quad s.t. \int_{\mathbf{K}} h d\mu = 1. \tag{1}$$

Note that formula (1) does not appear explicitly in [17, Theorem 3.2], but one can derive it easily from it. Indeed, one can write $f_{\min,\mathbf{K}} = \sup\{\lambda : f(x) - \lambda \geq 0 \text{ over } \mathbf{K}\}$. Then, by the first part of Theorem 1, we have $f_{\min,\mathbf{K}} = \sup\left\{\lambda : \int_{\mathbf{K}} h(f-\lambda)d\mu \geq 0 \ \forall h \in \Sigma[x]\right\}$. As $\int_{\mathbf{K}} h(f-\lambda)d\mu = \int_{\mathbf{K}} hf d\mu - \lambda \int_{\mathbf{K}} h d\mu$, after normalizing $\int_{\mathbf{K}} h d\mu = 1$, we can conclude (1).

If we select the measure $\mu$ to be the Lebesgue measure in Theorem 1, then we obtain the following reformulation for $f_{\min,\mathbf{K}}$, which we will consider in this paper:

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

By bounding the degree of the polynomial $h \in \Sigma[x]$ by $2r$, we can define the parameter:

$$\underline{f}_{\mathbf{K}}^{(r)} := \inf_{h \in \Sigma[x]_r} \int_{\mathbf{K}} h(x) f(x) dx \quad \text{s.t.} \int_{\mathbf{K}} h(x) dx = 1. \tag{2}$$

Clearly, the inequality $f_{\min,\mathbf{K}} \leq \underline{f}_{\mathbf{K}}^{(r)}$ holds for all $r \in \mathbb{N}$. Lasserre [17] gives conditions under which the infimum is attained in the program (2).

**Theorem 2** *[17, Theorems 4.1 and 4.2] Assume $\mathbf{K} \subseteq \mathbb{R}^n$ is compact and has nonempty interior and let $f$ be a polynomial. Then, the program (2) has an optimal solution for every $r \in \mathbb{N}$ and*

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

We now recall how to compute the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ in terms of the moments $m_\alpha(\mathbf{K})$ of the Lebesgue measure on $\mathbf{K}$, where

$$m_\alpha(\mathbf{K}) := \int_{\mathbf{K}} x^\alpha dx \qquad \text{for } \alpha \in \mathbb{N}^n,$$

and $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$.

Let $N(n,r) := \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \le r\}$, and suppose $f(x) = \sum_{\beta \in N(n,d)} f_\beta x^\beta$ has degree $d$. If we write $h \in \Sigma[x]_r$ as $h(x) = \sum_{\alpha \in N(n,2r)} h_\alpha x^\alpha$, then the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ from (2) can be reformulated as follows:

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

Hence, if we know the moments $m_\alpha(\mathbf{K})$ for all $\alpha \in \mathbb{N}^n$ with $|\alpha| := \sum_{i=1}^n \alpha_i \le d+2r$, then we can compute the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ by solving the semidefinite program (3) which involves a LMI of size $\binom{n+2r}{2r}$. So the bound $\underline{f}_{\mathbf{K}}^{(r)}$ can be computed in polynomial time for fixed $d$ and $r$ (to any fixed precision).

When $\mathbf{K}$ is the standard simplex $\Delta_n = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i \le 1\}$, the unit hypercube $\mathbf{Q}_n = [0,1]^n$, or the unit ball $B_1(0) = \{x \in \mathbb{R}^n : \|x\| \le 1\}$, there exist explicit formulas for the moments $m_\alpha(\mathbf{K})$. Namely, for the standard simplex, we have

$$m_\alpha(\Delta_n) = \frac{\prod_{i=1}^n \alpha_i!}{(|\alpha|+n)!}, \tag{4}$$

see e.g., [14, equation (2.4)] or [12, equation (2.2)]. From this one can easily calculate the moments for the hypercube $\mathbf{Q}_n$:

$$m_\alpha(\mathbf{Q}_n) = \int_{\mathbf{Q}_n} x^\alpha dx = \prod_{i=1}^n \int_0^1 x_i^{\alpha_i} dx_i = \prod_{i=1}^n \frac{1}{\alpha_i + 1}.$$

To state the moments for the unit Euclidean ball, we will use the notation $[n] := \{1, \ldots, n\}$, the Euler gamma function $\Gamma(\cdot)$, and the notation for the double factorial of an integer $k$:

$$k!! = \begin{cases} k \cdot (k-2) \cdots 3 \cdot 1, \text{ if } k > 0 \text{ is odd}, \\ k \cdot (k-2) \cdots 4 \cdot 2, \text{ if } k > 0 \text{ is even}, \\ 1 \qquad\qquad\qquad\quad \text{ if } k = 0 \text{ or } k = -1. \end{cases}$$

In terms of this notation, the moments for the unit Euclidean ball are given by:

$$m_\alpha(B_1(0)) = \begin{cases} \frac{\pi^{n/2} \prod_{i=1}^n (\alpha_i - 1)!!}{\Gamma\left(1 + \frac{n+|\alpha|}{2}\right) 2^{|\alpha|/2}} = \frac{\pi^{(n-1)/2} 2^{(n+1)/2} \prod_{i=1}^n (\alpha_i - 1)!!}{(n+|\alpha|)!!} & \text{if } \alpha_i \text{ is even for all } i \in [n], \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

One may prove relation (5) using

$$\int_{B_1(0)} x^\alpha dx = \frac{1}{\Gamma(1 + (n+|\alpha|)/2)} \int_{\mathbb{R}^n} x^\alpha \exp\left(-\|x\|^2\right) dx$$

(see, e.g., [18, Theorem 2.1]), together with the fact (see, e.g., page 872 in [17]) that

$$\int_{-\infty}^{+\infty} t^p \exp\left(-t^2/2\right) dt = \begin{cases} \sqrt{2\pi}(p-1)!! & \text{if } p \text{ is even,} \\ 0 & \text{if } p \text{ is odd,} \end{cases}$$

and the identity $\Gamma(1+\frac{k}{2}) = \frac{k!!}{2^{(k+1)/2}}\sqrt{\pi}$ for all integers $k \in \mathbb{N}$ (see e.g., [1, Section 6.1.12]).
For a general polytope $\mathbf{K} \subseteq \mathbb{R}^n$, it is a hard problem to compute the moments $m_\alpha(\mathbf{K})$. In fact, the problem of computing the volume of polytopes of varying dimensions is already #P-hard [10]. On the other hand, any polytope $\mathbf{K} \subseteq \mathbb{R}^n$ can be triangulated into finitely many simplices (see e.g., [8]) so that one could use (4) to obtain the moments $m_\alpha(\mathbf{K})$ of $\mathbf{K}$. The complexity of this method depends on the number of simplices in the triangulation. However, this number can be exponentially large (e.g., for the hypercube) and the problem of finding the smallest possible triangulation of a polytope is NP-hard, even in fixed dimension $n = 3$ (see e.g., [8]).

*Example*

Consider the minimization of the Motzkin polynomial $f(x_1, x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$ over the hypercube $\mathbf{K} = [-2, 2]^2$, which has four global minimizers at the points $(\pm 1, \pm 1)$, and $f_{\min,\mathbf{K}} = 0$. Figure 1 shows the computed optimal sum of squares density function $h^*$, for $r = 12$, corresponding to $\underline{f}_{\mathbf{K}}^{(12)} = 0.406076$. We observe that the optimal density $h^*$ shows four peaks at the four global minimizers and thus, it appears to approximate the density of a convex combination of the Dirac measures at the four minimizers.

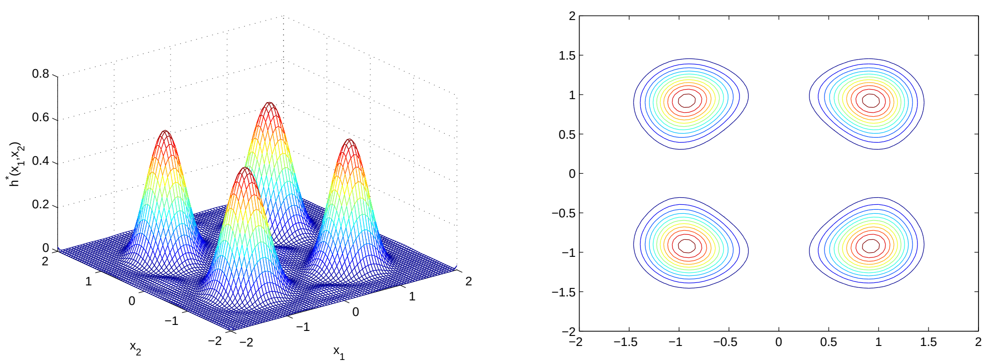


**Fig. 1** Graph and contour plot of $h^*(x)$ on $[-2, 2]^2$ ($r = 12$ and $\deg(h^*) = 24$) for the Motzkin polynomial.

We will present several additional numerical examples in Section 4.

## 1.3 Our main results

In this paper we analyze the quality of the upper bounds $\underline{f}_{\mathbf{K}}^{(r)}$ from (2) for the minimum $f_{\min,\mathbf{K}}$ of $f$ over $K$. Our main result is an upper bound for the range $\underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}}$, which applies to the

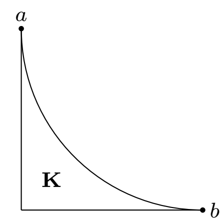


**Fig. 2** This set $\mathbf{K}$ does not satisfy Assumption 1 at the points $a$ and $b$.

case when $f$ is Lipschitz continuous on $\mathbf{K}$ and when $\mathbf{K}$ is a full-dimensional compact set satisfying the additional condition from Assumption 1, see Theorem 3 below. We will use throughout the following notation about the set $\mathbf{K}$.

We let $D(\mathbf{K}) = \max_{x,y\in\mathbf{K}} \|x-y\|^2$ denote the (squared) diameter of the set $\mathbf{K}$, where $\|x\| = \sqrt{\sum_{i=1}^n {x_i}^2}$ is the $\ell_2$-norm. Moreover, $w_{\min}(\mathbf{K})$ is the minimal width of $\mathbf{K}$, which is the minimum distance between two distinct parallel supporting hyperplanes of $\mathbf{K}$. Throughout, $B_\epsilon(a) := \{x \in \mathbb{R}^n : \|x-a\| \leq \epsilon\}$ denotes the Euclidean ball centered at $a \in \mathbb{R}^n$ and with radius $\epsilon > 0$. With $\gamma_n$ denoting the volume of the $n$-dimensional unit ball, the volume of the ball $B_\epsilon(a)$ is given by $\mathrm{vol}B_\epsilon(a) = \epsilon^n\gamma_n$.

We now formulate our geometric assumption about the set $\mathbf{K}$ which says (roughly) that around any point $a \in \mathbf{K}$ there is a ball intersecting a constant fraction of the unit ball.

**Assumption 1** *For all points $a \in \mathbf{K}$ there exist constants $\eta_\mathbf{K} > 0$ and $\epsilon_\mathbf{K} > 0$ such that*

$$\mathrm{vol}(B_\epsilon(a) \cap \mathbf{K}) \geq \eta_\mathbf{K} \mathrm{vol}B_\epsilon(a) = \eta_\mathbf{K}\epsilon^n\gamma_n \quad \textit{for all } 0 < \epsilon \leq \epsilon_\mathbf{K}. \tag{6}$$

Note that Assumption 1 implies that the set $\mathbf{K}$ has positive Lebesgue density at all points $a \in \mathbf{K}$. For all sets $\mathbf{K}$ satisfying Assumption 1, we also define the parameter

$$r_\mathbf{K} := \max\left\{\frac{D(\mathbf{K})e}{2\epsilon_\mathbf{K}^3}, n\right\} \quad \text{if } \epsilon_\mathbf{K} \leq 1, \quad \text{and} \quad r_\mathbf{K} := \frac{D(\mathbf{K})e}{2} \quad \text{if } \epsilon_\mathbf{K} \geq 1. \tag{7}$$

Here, $e = 2.71828...$ denotes the base of the natural logarithm. Note that the parameters $\eta_\mathbf{K}$, $\epsilon_\mathbf{K}$ and $r_\mathbf{K}$ depend not only on the set $\mathbf{K}$ but also on the point $a \in \mathbf{K}$; we omit the dependance on $a$ to simplify notation. Assumption 1 will be used in the case when the point $a$ is a global minimizer in $\mathbf{K}$ of the polynomial to be analyzed.

For instance, convex bodies and, more generally, compact star-shaped sets satisfy Assumption 1 (see Section 5.1). We now give an example of a set $\mathbf{K}$ that does not satisfy Assumption 1 and refer to Section 5.1 for more discussion about Assumption 1.

*Example 1* Consider the following set $\mathbf{K} \subseteq \mathbb{R}^2$, displayed in Figure 2:

$$\mathbf{K} = \{x \in \mathbb{R}^2 \; : \; x \geq 0, (x_1-1)^2 + (x_2-1)^2 \geq 1\}.$$

One can easily check that Assumption 1 is not satisfied, since the condition (6) does not hold for the two points $a$ and $b$.

We now present our main result.

**Theorem 3** *Assume that* $\mathbf{K} \subseteq \mathbb{R}^n$ *is compact and satisfies Assumption 1. Then there exists a constant* $\zeta(\mathbf{K})$ *(depending only on* $\mathbf{K}$*) such that, for all Lipschitz continuous functions* $f$ *with Lipschitz constant* $M_f$ *on* $\mathbf{K}$*, the following inequality holds:*

$$\underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}} \leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{r}} \quad \text{for all } r \geq r_{\mathbf{K}} + 1. \tag{8}$$

*Moreover, if* $f$ *is a polynomial of degree* $d$ *and* $\mathbf{K}$ *is a convex body, then*

$$\underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}} \leq \frac{2d^2\zeta(\mathbf{K})\sup_{x\in\mathbf{K}}|f(x)|}{w_{\min}(\mathbf{K})}\frac{1}{\sqrt{r}} \quad \text{for all } r \geq r_{\mathbf{K}} + 1. \tag{9}$$

The key idea to show this result is to select suitable sums of squares densities which we are able to analyse. For this, we will select a global minimizer $a$ of $f$ over $\mathbf{K}$ and consider the Gaussian distribution with mean $a$ and, as sums of squares densities, we will select the polynomials $H_{r,a}$ obtained by truncating the Taylor series expansion of the Gaussian distribution, see relation (14).

*Remark 1* When the polynomial $f$ has a root in $\mathbf{K}$ (which can be assumed without loss of generality), the parameter $\sup_{x\in\mathbf{K}}|f(x)|$ involved in relation (9) can easily be upper bounded in terms of the range of values of $f$; namely,

$$\sup_{x\in\mathbf{K}}|f(x)| \leq f_{\max,\mathbf{K}} - f_{\min,\mathbf{K}},$$

where $f_{\max,\mathbf{K}}$ denotes the maximum value of $f$ over $\mathbf{K}$. Hence relation (9) also implies an upper bound on $\underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}}$ in terms of the range $f_{\max,\mathbf{K}} - f_{\min,\mathbf{K}}$, as is commonly used in approximation analysis (see, e.g., [3,5]).

### 1.4 Contents of the paper

Our paper is organized as follows. In Section 2, we give a constructive proof for our main result in Theorem 3. In Section 3 we show how to obtain feasible points in $\mathbf{K}$ that correspond to the bounds $\underline{f}_{\mathbf{K}}^{(r)}$ through sampling. This is followed by a section with numerical examples (Section 4). Finally, in the concluding remarks (Section 5), we revisit Assumption 1, and discuss perspectives for future research.

## 2 Proof of our main result in Theorem 3

In this section we prove our main result in Theorem 3. Our analysis holds for Lipschitz continuous functions, so we start by reviewing some relevant properties in Section 2.1. In the next step we indicate in Section 2.2 how to select the polynomial density function $h$ as a special sum of squares that we will be able to analyze. Namely, we let $a$ denote a global minimizer of the function $f$ over the set $\mathbf{K} \subseteq \mathbb{R}^n$. Then we consider the density function $G_a$ in (12) of the Gaussian distribution with mean $a$ (and suitable variance) and the polynomial $H_{r,a}$ in (14), which is obtained from the truncation at degree $2r$ of the Taylor series expansion of the Gaussian density function $G_a$. The final step will be to analyze the quality of the bound obtained by selecting the polynomial $H_{r,a}$ and this will be the most technical part of the proof, carried out in Section 2.3.

### 2.1 Lipschitz continuous functions

A function $f$ is said to be Lipschitz continuous on $\mathbf{K}$, with Lipschitz constant $M_f$, if it satisfies:

$$|f(y) - f(x)| \leq M_f \|y - x\| \quad \text{for all } x, y \in \mathbf{K}.$$

If $f$ is continuous and differentiable on $\mathbf{K}$, then $f$ is Lipschitz continuous on $\mathbf{K}$ with respect to the constant

$$M_f = \max_{x \in \mathbf{K}} \|\nabla f(x)\|. \tag{10}$$

Furthermore, if $f$ is an $n$-variate polynomial with degree $d$, then the Markov inequality for $f$ on a convex body $\mathbf{K}$ reads as

$$\max_{x \in \mathbf{K}} \|\nabla f(x)\| \leq \frac{2d^2}{w_{\min}(\mathbf{K})} \sup_{x \in \mathbf{K}} |f(x)|,$$

see e.g., [3, relation (8)]. Thus, together with (10), we have that $f$ is Lipschitz continuous on $\mathbf{K}$ with respect to the constant

$$M_f \leq \frac{2d^2}{w_{\min}(\mathbf{K})} \sup_{x \in \mathbf{K}} |f(x)|. \tag{11}$$

### 2.2 Choosing the polynomial density function $H_{r,a}$

Consider the function

$$G_a(x) := \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x - a\|^2}{2\sigma^2}\right), \tag{12}$$

which is the probability density function of the Gaussian distribution with mean $a$ and standard variance $\sigma$ (whose value will be defined later). Let the constant $C_{\mathbf{K},a}$ be defined by

$$\int_{\mathbf{K}} C_{\mathbf{K},a} G_a(x) dx = 1. \tag{13}$$

Observe that $G_a(x)$ is equal to the function $\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-t}$ evaluated at the point $t = \frac{\|x-a\|^2}{2\sigma^2}$. Denote by $H_{r,a}$ the Taylor series expansion of $G_a$ truncated at the order $2r$. That is,

$$H_{r,a}(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \sum_{k=0}^{2r} \frac{1}{k!} \left(-\frac{\|x - a\|^2}{2\sigma^2}\right)^k. \tag{14}$$

Moreover consider the constant $c^r_{\mathbf{K},a}$, defined by

$$\int_{\mathbf{K}} c^r_{\mathbf{K},a} H_{r,a}(x) dx = 1. \tag{15}$$

The next step is to show that $H_{r,a}$ is a sum of squares of polynomials and thus $H_{r,a} \in \Sigma[x]_{2r}$. This follows from the next lemma.

**Lemma 1** *Let $\phi_{2r}(t)$ denote the (univariate) polynomial of degree $2r$ obtained by truncating the Taylor series expansion of $e^{-t}$ at the order $2r$. That is,*

$$\phi_{2r}(t) := \sum_{k=0}^{2r} \frac{(-t)^k}{k!}.$$

*Then $\phi_{2r}$ is a sum of squares of polynomials. Moreover, we have*

$$0 \leq \phi_{2r}(t) - e^{-t} \leq \frac{t^{2r+1}}{(2r+1)!} \quad \text{for all } t \geq 0. \tag{16}$$

*Proof* First, we show that $\phi_{2r}$ is a sum of squares. As $\phi_{2r}$ is a univariate polynomial, by Hilbert's Theorem (see e.g., [19, Theorem 3.4]), it suffices to show that $\phi_{2r}(t) \geq 0$ for all $t \in \mathbb{R}$. As $\phi_{2r}(-\infty) = \phi_{2r}(+\infty) = +\infty$, it suffices to show that $\phi_{2r}(t) \geq 0$ at all the stationary points $t$ where $\phi'_{2r}(t) = 0$. For this, observe that $\phi'_{2r}(t) = \sum_{k=1}^{2r}(-1)^k \frac{t^{k-1}}{(k-1)!}$, so that it can be written as $\phi'_{2r}(t) = -\phi_{2r}(t) + \frac{t^{2r}}{(2r)!}$. Hence, for all $t$ with $\phi'_{2r}(t) = 0$, we have $\phi_{2r}(t) = \frac{t^{2r}}{(2r)!} \geq 0$.

Next, we show that $\phi_{2r}(t) \geq e^{-t}$ for all $t \geq 0$. Fix $t \geq 0$. Then, by Taylor Theorem (see e.g., [29]), one has $e^{-t} = \phi_{2r}(t) + \frac{\phi^{(2r+1)}(\xi)t^{2r+1}}{(2r+1)!}$ for some $\xi \in [0, t]$. As $\phi^{(2r+1)}(\xi) = -e^{-\xi}$, one can conclude that $e^{-t} - \phi_{2r}(t) = -\frac{e^{-\xi}t^{2r+1}}{(2r+1)!} \leq 0$ and $e^{-t} - \phi_{2r}(t) \geq -\frac{t^{2r+1}}{(2r+1)!}$. □

We now consider the parameter $f^{(r)}_{\mathbf{K},a}$ defined as

$$f^{(r)}_{\mathbf{K},a} := \int_{\mathbf{K}} f(x) c^r_{\mathbf{K},a} H_{r,a}(x) dx. \tag{17}$$

Our main technical result is the following upper bound for the range $f^{(r)}_{\mathbf{K},a} - f_{\min,\mathbf{K}}$.

**Theorem 4** *Assume $\mathbf{K} \subseteq \mathbb{R}^n$ is compact and satisfies Assumption 1, and consider the parameter $r_{\mathbf{K}}$ from (7). Then there exists a constant $\zeta(\mathbf{K})$ (depending only on $\mathbf{K}$) such that, for all Lipschitz continuous functions $f$ with Lipschitz constant $M_f$ on $\mathbf{K}$, the following inequality holds:*

$$f^{(r)}_{\mathbf{K},a} - f_{\min,\mathbf{K}} \leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{2r+1}}, \quad \text{for all } r \geq \frac{r_{\mathbf{K}}}{2}. \tag{18}$$

*Moreover, if $f$ is a polynomial of degree $d$ and $\mathbf{K}$ is a convex body, then*

$$f^{(r)}_{\mathbf{K},a} - f_{\min,\mathbf{K}} \leq \frac{2d^2\zeta(\mathbf{K}) \sup_{x\in\mathbf{K}} |f(x)|}{w_{\min}(\mathbf{K})\sqrt{2r+1}}, \quad \text{for all } r \geq \frac{r_{\mathbf{K}}}{2}. \tag{19}$$

We will give the proof of Theorem 4, which has lengthy technical details, in Section 2.3 below. We now show how to derive Theorem 3 as a direct application of Theorem 4.

*Proof (of Theorem 3)* Assume $f$ is Lipschitz continuous with Lipschitz constant $M_f$ on $K$ and $a$ is a minimizer of $f$ over the set $\mathbf{K}$. Using the definitions (2) and (17) of the parameters and the fact that $H_{r,a}$ is a sum of squares with degree $4r$, it follows that

$$\underline{f}_{\mathbf{K}}^{(2r+1)} \leq \underline{f}_{\mathbf{K}}^{(2r)} \leq f_{\mathbf{K},a}^{(r)}, \quad \text{for all } r \in \mathbb{N}.$$

Then, from inequality (18) in Theorem 4, one obtains

$$\underline{f}_{\mathbf{K}}^{(2r+1)} - f_{\min,\mathbf{K}} \leq \underline{f}_{\mathbf{K}}^{(2r)} - f_{\min,\mathbf{K}} \leq f_{\mathbf{K},a}^{(r)} - f_{\min,\mathbf{K}} \leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{2r+1}} \quad \text{for all } r \geq \frac{r_{\mathbf{K}}}{2}.$$

Hence, for all $r \geq r_{\mathbf{K}} + 1$,

$$\begin{aligned} \underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}} &\leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{r+1}} \leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{r}} \quad \text{for even } r, \\ \underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}} &\leq \frac{\zeta(\mathbf{K})M_f}{\sqrt{r}} \quad \text{for odd } r. \end{aligned}$$

This concludes the proof for relation (8), and relation (9) follows from (19) in an analogous way. This finishes the proof of Theorem 3. □

### 2.3 Analyzing the polynomial density function $H_{r,a}$

In this section we prove the result of Theorem 4. Recall that $a$ is a global minimizer of $f$ over $\mathbf{K}$. For the proof, we will need the following four technical lemmas.

**Lemma 2** *Assume $\mathbf{K} \subseteq \mathbb{R}^n$ is compact and satisfies Assumption 1. Then, for all $0 < \epsilon \leq \epsilon_{\mathbf{K}}$ and $r \in \mathbb{N}$, we have:*

$$c_{\mathbf{K},a}^r \leq C_{\mathbf{K},a} \leq \frac{(2\pi\sigma^2)^{n/2} \exp\left(\frac{\epsilon^2}{2\sigma^2}\right)}{\eta_{\mathbf{K}}\epsilon^n\gamma_n}. \tag{20}$$

*Proof* By Lemma 1, $\phi_{2r}(t) \geq e^{-t}$ for all $t \geq 0$, which implies $H_{r,a}(x) \geq G_a(x)$ for all $x \in \mathbb{R}^n$. Together with the relations (13) and (15) defining the constants $C_{\mathbf{K},a}$ and $c_{\mathbf{K},a}^r$, we deduce that $c_{\mathbf{K},a}^r \leq C_{\mathbf{K},a}$. Moreover, by the definition (13) of the constant $C_{K,a}$, one has

$$\begin{aligned} \frac{1}{C_{\mathbf{K},a}} &= \int_{\mathbf{K}} G_a(x)dx = \int_{\mathbf{K}} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x-a\|^2}{2\sigma^2}\right) dx \\ &\geq \int_{\mathbf{K}\cap B_\epsilon(a)} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x-a\|^2}{2\sigma^2}\right) dx \\ &\geq \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \text{vol}(\mathbf{K} \cap B_\epsilon(a)). \end{aligned}$$

We now use relation (6) from Assumption 1 in order to conclude that $\text{vol}(\mathbf{K} \cap B_\epsilon(a)) \geq \eta_{\mathbf{K}}\epsilon^n\gamma_n$, which gives the desired upper bound on $C_{K,a}$. □

**Lemma 3** *Given $\tilde{x} \in \mathbb{R}^n$ and a function $F : \mathbb{R}_+ \to \mathbb{R}$, define the function $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = F(\|x - \tilde{x}\|)$ for all $x \in \mathbb{R}^n$. Then, for all $\rho_2 \geq \rho_1 \geq 0$, one has*

$$\int_{B_{\rho_2}(\tilde{x}) \setminus B_{\rho_1}(\tilde{x})} f(x)dx = n\gamma_n \int_{\rho_1}^{\rho_2} z^{n-1} F(z)dz,$$

*where $\gamma_n = \frac{\pi^{(n-1)/2} 2^{(n+1)/2}}{n!!}$ is the volume of the unit Euclidean ball in $\mathbb{R}^n$.*

*Proof* Apply a change of variables using spherical coordinates as explained, e.g., in [2]. $\square$

**Lemma 4** *For all positive integers $r$ and $n$, one has $\left(\frac{1}{2r+1}\right)^{-\frac{n}{4(2r+1)+2n}} < 6n$.*

*Proof* Let $n \in \mathbb{N}$ be given. Denote

$$g(r) := \left(\frac{1}{2r+1}\right)^{-\frac{n}{4(2r+1)+2n}} = (2r+1)^{\frac{n}{4(2r+1)+2n}} \quad (r \geq 0).$$

Observe that, $g(0) = 1$, $g(r) > 0$ for all $r \geq 0$, $\ln(g(r)) = \frac{n}{8r+4+2n} \ln(2r+1)$, and thus $\lim_{r\to\infty} g(r) = 1$. It suffices to show $g(r^*) < 6n$ for all stationary points $r^*$. Since

$$\frac{d \ln(g(r))}{dr} = \frac{-8n \ln(2r+1)}{(8r+4+2n)^2} + \frac{2n}{(2r+1)(8r+4+2n)},$$

and $g'(r) = \frac{1}{g(r)} \frac{d \ln(g(r))}{dr}$, any stationary point $r^*$ satisfies

$$\frac{d \ln(g(r^*))}{dr} = 0 \Longleftrightarrow (2r^*+1)\left[\ln(2r^*+1) - 1\right] = \frac{n}{2}.$$

Since

$$(2r^*+1)(\ln(3) - 1) \leq (2r^*+1)\left[\ln(2r^*+1) - 1\right] = \frac{n}{2},$$

one has $2r^*+1 \leq \frac{n}{2(\ln(3)-1)} < 6n$. Since $g(r) \leq 2r+1$ for all $r \geq 0$, one has $g(r^*) \leq 2r^*+1 < 6n$. $\square$

**Lemma 5** *Assume $\mathbf{K} \subseteq \mathbb{R}^n$ is compact and satisfies Assumption 1. Then, for all $0 < \epsilon \leq \epsilon_{\mathbf{K}}$, one has*

$$\int_{\mathbf{K}} C_{\mathbf{K},a}\|x - a\| G_a(x)dx \leq \epsilon + \frac{n\sigma^{n+1} p(n)}{\epsilon^n \eta_{\mathbf{K}}} e^{\frac{\epsilon^2}{2\sigma^2}},$$

*where $p(n) := \int_0^{+\infty} t^n e^{-t^2/2} dt$ is a constant depending on $n$, given by*

$$p(n) = \begin{cases} 1 & \text{if } n = 1, \\ \sqrt{\frac{\pi}{2}} \prod_{j=1}^{k} (2j-1) & \text{if } n = 2k \text{ and } k \geq 1, \\ \prod_{j=1}^{k} (2j) & \text{if } n = 2k+1 \text{ and } k \geq 1. \end{cases} \tag{21}$$

*Proof* Let $\varphi := \int_{\mathbf{K}} C_{\mathbf{K},a}\|x-a\|G_a(x)dx$ denote the integral that we need to upper bound. We split the integral $\varphi$ as $\varphi = \varphi_1 + \varphi_2$, depending on whether $x$ lies in the ball $B_\epsilon(a)$ or not.

First, we upper bound the term $\varphi_1$ as

$$\varphi_1 := \int_{\mathbf{K}\cap B_\epsilon(a)} \|x-a\| C_{\mathbf{K},a} G_a(x)dx \le \epsilon \int_{\mathbf{K}\cap B_\epsilon(a)} C_{\mathbf{K},a} G_a(x)dx \le \epsilon \int_{\mathbf{K}} C_{\mathbf{K},a} G_a(x)dx = \epsilon.$$

Second, we bound the integral

$$\varphi_2 := C_{\mathbf{K},a} \int_{\mathbf{K}\setminus B_\epsilon(a)} \|x-a\| G_a(x)dx.$$

Since $\mathbf{K} \subseteq B_{\sqrt{D(\mathbf{K})}}(a)$, one has

$$\varphi_2 \le C_{\mathbf{K},a} \int_{B_{\sqrt{D(\mathbf{K})}}(a)\setminus B_\epsilon(a)} \|x-a\| G_a(x)dx,$$

where the right hand side, by Lemma 3, is equal to

$$\frac{C_{\mathbf{K},a} n \gamma_n}{(2\pi\sigma^2)^{n/2}} \int_\epsilon^{\sqrt{D(\mathbf{K})}} z^n \exp\left(-\frac{z^2}{2\sigma^2}\right) dz.$$

By a change of variable $t = \frac{z}{\sigma}$, one obtains

$$\varphi_2 \le \frac{C_{\mathbf{K},a} n \gamma_n \sigma}{(2\pi)^{n/2}} \int_{\epsilon/\sigma}^{\sqrt{D(\mathbf{K})}/\sigma} t^n \exp\left(-\frac{t^2}{2}\right) dt,$$

and thus

$$\varphi_2 \le \frac{C_{\mathbf{K},a} n \gamma_n \sigma}{(2\pi)^{n/2}} \int_0^{+\infty} t^n \exp\left(-\frac{t^2}{2}\right) dt = \frac{C_{\mathbf{K},a} n \gamma_n \sigma}{(2\pi)^{n/2}} p(n).$$

Here we have set $p(n) := \int_0^{+\infty} t^n e^{-\frac{t^2}{2}} dt$ which can be checked to be given by (21) (e.g., using induction on $n$). Now, combining with the upper bound for $C_{\mathbf{K},a}$ from (20), we obtain

$$\varphi_2 \le \frac{n\sigma^{n+1} p(n)}{\epsilon^n \eta_{\mathbf{K}}} e^{\frac{\epsilon^2}{2\sigma^2}}.$$

Therefore, we have shown:

$$\varphi = \varphi_1 + \varphi_2 \le \epsilon + \frac{n\sigma^{n+1} p(n)}{\epsilon^n \eta_{\mathbf{K}}} e^{\frac{\epsilon^2}{2\sigma^2}},$$

which shows the lemma. □

We are now ready to prove Theorem 4.

*Proof (of Theorem 4)* Observe that, if $f$ is a polynomial, then we can use the upper bound (11) for its Lipschitz constant and thus the inequality (19) follows as a direct consequence of the inequality (18). Therefore, it suffices to show the relation (18).

Recall that $a$ is a minimizer of $f$ over $\mathbf{K}$. As $f$ is Lipschitz continuous with Lipschitz constant $M_f$ on $K$, we have

$$f(x) - f(a) \leq M_f \|x - a\| \quad \forall x \in \mathbf{K}.$$

This implies

$$f^{(r)}_{\mathbf{K},a} - f_{\min,\mathbf{K}} = \int_{\mathbf{K}} c^r_{\mathbf{K},a} H_{r,a}(x)(f(x) - f(a))dx \leq M_f \int_{\mathbf{K}} \|x - a\| c^r_{\mathbf{K},a} H_{r,a}(x)dx.$$

Our objective is now to show the existence of a constant $\zeta(\mathbf{K})$ such that

$$\psi := \int_{\mathbf{K}} c^r_{\mathbf{K},a} \|x - a\| H_{r,a}(x)dx \leq \frac{\zeta(\mathbf{K})}{\sqrt{2r+1}}, \quad \text{for all } r \geq r_{\mathbf{K}}, \text{ (see (7))}$$

by which we can then conclude the proof for (18).

For this, we split the integral $\psi$ as the sum of two terms:

$$\psi = \underbrace{\int_{\mathbf{K}} c^r_{\mathbf{K},a} \|x - a\| G_a(x)dx}_{=:\psi_1} + \underbrace{\int_{\mathbf{K}} c^r_{\mathbf{K},a} \|x - a\| (H_{r,a}(x) - G_a(x))dx}_{=:\psi_2}.$$

First, we upper bound the term $\psi_1$. As $c^r_{\mathbf{K},a} \leq C_{\mathbf{K},a}$ (by (20)), we can use Lemma 5 to conclude that, for all $0 < \epsilon \leq \epsilon_{\mathbf{K}}$,

$$\psi_1 \leq \int_{\mathbf{K}} C_{\mathbf{K},a} \|x - a\| G_a(x)dx \leq \epsilon + \frac{n\sigma^{n+1} p(n)}{\epsilon^n \eta_{\mathbf{K}}} e^{\frac{\epsilon^2}{2\sigma^2}} = \epsilon \underbrace{\left[1 + \frac{n\sigma^{n+1} p(n)}{\epsilon^{n+1} \eta_{\mathbf{K}}} e^{\frac{\epsilon^2}{2\sigma^2}}\right]}_{=:\mu_1} = \epsilon \mu_1. \tag{22}$$

Second we bound the integral

$$\psi_2 = \int_{\mathbf{K}} c^r_{\mathbf{K},a} \|x - a\| (H_{r,a}(x) - G_a(x))dx.$$

We can upper bound the function $H_{r,a}(x) - G_a(x)$ using the estimate from (16) and we get

$$H_{r,a}(x) - G_a(x) \leq \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{\|x - a\|^{4r+2}}{(2\sigma^2)^{2r+1}(2r+1)!}.$$

Then we have

$$\psi_2 \leq \frac{1}{(2\pi\sigma^2)^{n/2}} \int_{\mathbf{K}} c^r_{\mathbf{K},a} \frac{\|x - a\|^{4r+3}}{(2\sigma^2)^{2r+1}(2r+1)!} dx = \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{c^r_{\mathbf{K},a}}{(2\sigma^2)^{2r+1}(2r+1)!} \int_{\mathbf{K}} \|x - a\|^{4r+3} dx.$$

Now we upper bound the integral $\int_{\mathbf{K}} \|x-a\|^{4r+3}dx$. Since $\mathbf{K} \subseteq B_{\sqrt{D(\mathbf{K})}}(a)$, one has

$$\int_{\mathbf{K}} \|x-a\|^{4r+3}dx \leq \int_{B_{\sqrt{D(\mathbf{K})}}(a)} \|x-a\|^{4r+3}dx,$$

where the right hand side, by Lemma 3, is equal to

$$n\gamma_n \int_0^{\sqrt{D(\mathbf{K})}} z^{4r+n+2}dz = \frac{n\gamma_n D(\mathbf{K})^{\frac{4r+n+3}{2}}}{4r+n+3} \leq n\gamma_n D(\mathbf{K})^{\frac{4r+n+3}{2}}.$$

Thus, we obtain

$$\psi_2 \leq \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{c^r_{\mathbf{K},a}}{(2\sigma^2)^{2r+1}(2r+1)!} n\gamma_n D(\mathbf{K})^{\frac{4r+n+3}{2}}.$$

We now use the upper bound for $c^r_{\mathbf{K},a}$ from (20):

$$c^r_{\mathbf{K},a} \leq \frac{(2\pi\sigma^2)^{n/2}\exp\left(\frac{\epsilon^2}{2\sigma^2}\right)}{\eta_{\mathbf{K}}\epsilon^n\gamma_n}$$

and we obtain

$$\psi_2 \leq \frac{n\exp\left(\frac{\epsilon^2}{2\sigma^2}\right) D(\mathbf{K})^{\frac{4r+n+3}{2}}}{\eta_{\mathbf{K}}\epsilon^n(2r+1)!(2\sigma^2)^{2r+1}}.$$

Finally we use the Stirling's inequality:

$$(2r+1)! \geq \sqrt{2\pi(2r+1)}\left(\frac{2r+1}{e}\right)^{2r+1},$$

and obtain

$$\begin{aligned}\psi_2 &\leq \underbrace{\frac{n\exp\left(\frac{\epsilon^2}{2\sigma^2}\right) D(\mathbf{K})^{\frac{n+1}{2}}}{\eta_{\mathbf{K}}}}_{=:\mu_2}\left(\frac{D(\mathbf{K})e}{2\sigma^2\epsilon^{n/(2r+1)}(2r+1)}\right)^{2r+1}\frac{1}{\sqrt{2\pi(2r+1)}} \qquad (23)\\ &= \frac{\mu_2}{\sqrt{2\pi(2r+1)}}\left(\frac{D(\mathbf{K})e}{2\sigma^2\epsilon^{n/(2r+1)}(2r+1)}\right)^{2r+1}.\end{aligned}$$

We can now upper bound the quantity $\psi = \psi_1 + \psi_2$, by combining the upper bound for $\psi_1$ in (22) with the above upper bound (23) for $\psi_2$. That is,

$$\psi \leq \epsilon\mu_1 + \frac{\mu_2}{\sqrt{2\pi(2r+1)}}\left(\frac{D(\mathbf{K})e}{2\sigma^2\epsilon^{n/(2r+1)}(2r+1)}\right)^{2r+1}.$$

We now indicate how to select the parameters $\epsilon$ and $\sigma$.

First we select $\sigma = \epsilon$, so that both parameters $\mu_1$ and $\mu_2$ appearing in (22) and (23) are constants depending on $n$ and $\mathbf{K}$, namely

$$\mu_1 = 1 + \frac{np(n)e^{1/2}}{\eta_{\mathbf{K}}} \quad \text{and} \quad \mu_2 = \frac{ne^{1/2}D(\mathbf{K})^{\frac{n+1}{2}}}{\eta_{\mathbf{K}}}.$$

Next we select $\epsilon$ so that $\frac{D(\mathbf{K})e}{2\epsilon^{2+n/(2r+1)}(2r+1)} = 1$, i.e.,

$$\epsilon = \left(\frac{D(\mathbf{K})e}{2(2r+1)}\right)^{\frac{2r+1}{2(2r+1)+n}} = \left(\frac{D(\mathbf{K})e}{2}\right)^{\frac{2r+1}{2(2r+1)+n}} \left(\frac{1}{2r+1}\right)^{\frac{1}{2}-\frac{n}{4(2r+1)+2n}}.$$

Summarizing, we have shown that

$$\begin{aligned}
\psi &\leq \left(\frac{1}{2r+1}\right)^{\frac{1}{2}-\frac{n}{4(2r+1)+2n}} \left[\left(\frac{D(\mathbf{K})e}{2}\right)^{\frac{2r+1}{2(2r+1)+n}} \mu_1 + \frac{\mu_2}{\sqrt{2\pi}} \left(\frac{1}{2r+1}\right)^{\frac{n}{4(2r+1)+2n}}\right] \\
&\leq \left(\frac{1}{2r+1}\right)^{\frac{1}{2}} 6n \left(\mu_1 \max\left\{1, \sqrt{\frac{D(\mathbf{K})e}{2}}\right\} + \frac{\mu_2}{\sqrt{2\pi}}\right). \qquad (24)
\end{aligned}$$

To obtain the last inequality (24), we use the inequality $\left(\frac{1}{2r+1}\right)^{-\frac{n}{4(2r+1)+2n}} < 6n$ (recall Lemma 4), together with the two inequalities $\left(\frac{D(\mathbf{K})e}{2}\right)^{\frac{2r+1}{2(2r+1)+n}} \leq \max\left\{1, \sqrt{\frac{D(\mathbf{K})e}{2}}\right\}$ and $\left(\frac{1}{2r+1}\right)^{\frac{n}{4(2r+1)+2n}} \leq 1$.

Since we have assumed $\epsilon \leq \epsilon_{\mathbf{K}}$ (recall Lemma 2), this implies the condition $r \geq \frac{D(\mathbf{K})e}{4}\epsilon_{\mathbf{K}}^{-\left(2+\frac{n}{2r+1}\right)} - \frac{1}{2}$, i.e., the inequality (24) holds for all $r \geq \frac{D(\mathbf{K})e}{4}\epsilon_{\mathbf{K}}^{-\left(2+\frac{n}{2r+1}\right)} - \frac{1}{2}$. If $\epsilon_{\mathbf{K}} \leq 1$ and $r \geq n/2$, then we have $\epsilon_{\mathbf{K}}^{-(2+\frac{n}{2r+1})} \leq \epsilon_{\mathbf{K}}^{-3}$ and thus the inequality (24) holds for all $r \geq \max\left\{\frac{D(\mathbf{K})e}{4\epsilon_{\mathbf{K}}^3}, \frac{n}{2}\right\}$. If $\epsilon_{\mathbf{K}} \geq 1$ then $\epsilon_{\mathbf{K}}^{-(2+\frac{n}{2r+1})} \leq 1$ and thus (24) holds for all integers $r \geq \frac{D(\mathbf{K})e}{4}$. Hence, the inequality (24) holds for all $r \geq r_{\mathbf{K}}/2$, where $r_{\mathbf{K}}$ is as defined in (7).

Finally, by defining the constant

$$\zeta(\mathbf{K}) := 6n \left(\mu_1 \max\left\{1, \sqrt{\frac{D(\mathbf{K})e}{2}}\right\} + \frac{\mu_2}{\sqrt{2\pi}}\right),$$

which indeed depends only on $\mathbf{K}$ and its dimension $n$, we can conclude the proof for (18). □

*Remark 2* Note that in the proof of Theorem 4, we use Assumption 1 only for the selected minimizer $a \in \mathbf{K}$ (and we use it only in the proof of Lemma 2). Hence, if the selected point $a$ lies in the interior of $\mathbf{K}$, i.e., if there exists $\delta > 0$ such that $B_\delta(a) \subseteq \mathbf{K}$, then the result of Theorem 4 (and thus Theorem 3) holds when selecting $\eta_{\mathbf{K}} = 1$ and $\epsilon_{\mathbf{K}} = \delta$.

Our results extend also to unconstrained global minimization:

$$f^* := \min_{x \in \mathbb{R}^n} f(x),$$

if we know that $f$ has a global minimizer $a$ and we know a ball $B_\delta(0)$ containing $a$. We can then indeed minimize $f$ over a compact set $K$, which can be chosen to be the ball $B_\delta(0)$ or a suitable hypercube containing $a$.

## 3 Obtaining feasible solutions through sampling

In this section we indicate how to sample feasible points in the set $\mathbf{K}$ from the optimal density function obtained by solving the semidefinite program (2).

Let $f \in \mathbb{R}[x]$ be a polynomial. Suppose $h^*(x) \in \Sigma[x]_r$ is an optimal solution of the program (2), i.e., $\underline{f}^{(r)}_{\mathbf{K}} = \int_{\mathbf{K}} f(x)h^*(x)dx$ and $\int_{\mathbf{K}} h^*(x)dx = 1$. Then $h^*$ can be seen as the probability density function of a probability distribution on $\mathbf{K}$, denoted as $\mathcal{T}_{\mathbf{K}}$ and, for all random vector $X = (X_1, \ldots, X_n) \sim \mathcal{T}_{\mathbf{K}}$, the expectation of $f(X)$ is given by:

$$\mathbb{E}\left[f(X)\right] = \int_{\mathbf{K}} f(x)h^*(x)dx = \underline{f}^{(r)}_{\mathbf{K}}. \tag{25}$$

As we now recall one can generate random samples $x \in \mathbf{K}$ from the distribution $\mathcal{T}_{\mathbf{K}}$ using the well known *method of conditional distributions* (see e.g., [20, Section 8.5.1]). Then we will observe that with high probability one of these sample points satisfies (roughly) the inequality $f(x) \leq \underline{f}^{(r)}_{\mathbf{K}}$ (see Theorem 5 for details).

In order to sample a random vector $X = (X_1, \ldots, X_n) \sim \mathcal{T}_{\mathbf{K}}$, we assume that, for each $i = 2, \ldots, n$, we know the cumulative conditional distribution of $X_i$ given that $X_j = x_j$ for $j = 1, \ldots, i-1$, defined in terms of probabilities as

$$F_i(x_i \mid x_1, \ldots, x_{i-1}) := \mathbf{Pr}\left[X_i \leq x_i \ \mid \ X_1 = x_1, \ldots, X_{i-1} = x_{i-1}\right].$$

Additionally, we assume that we know the cumulative marginal distribution function of $X_i$, defined as:

$$F_i(x_i) := \mathbf{Pr}\left[X_i \leq x_i\right].$$

Then one can generate a random sample $x = (x_1, \ldots, x_n) \in \mathbf{K}$ from the distribution $\mathcal{T}_{\mathbf{K}}$ by the following algorithm:

- Generate $x_1$ with cumulative distribution function $F_1(\cdot)$.
- Generate $x_2$ with cumulative distribution function $F_2\left(\cdot|x_1\right)$.

  $\vdots$
- Generate $x_n$ with cumulative distribution function $F_n\left(\cdot|x_1, \ldots, x_{n-1}\right)$.

Then return $x = (x_1, x_2, \ldots, x_n)^T$.

There remains to explain how to generate a (univariate) sample point $x$ with a given cumulative distribution function $F(\cdot)$, since this operation is carried out at each of the $n$ steps of the above algorithm. For this one can use the classical *inverse-transform method* (see e.g., [20, Section 8.2.1]), which reduces to sampling from the uniform distribution on $[0, 1]$ and can be described as follows:

- Generate a sample $u$ from the uniform distribution over $[0, 1]$.
- Return $x = F^{-1}(u)$ (if $F$ is strictly monotone increasing, or $x = \min\{y : F(y) \geq u\}$ otherwise).

Hence, in order to be able to apply the method of conditional distributions for sampling from $\mathbf{K}$ we need to solve the equation $x = F^{-1}(u)$. For instance, when $F(\cdot)$ is a univariate polynomial, solving the equation $x = F^{-1}(u)$ reduces to computing the eigenvalues of the corresponding companion matrix (see, e.g., [19, Section 2.4.1]). This applies, e.g., when $\mathbf{K}$ is the hypercube or the simplex, as we see below.

As an illustration, we first indicate how to compute the cumulative marginal and conditional distributions $F_i(\cdot)$ and $F_i(\cdot \mid x_1 \dots x_{i-1})$ for the case of the hypercube $\mathbf{K} = \mathbf{Q}_n = [0,1]^n$. As before we are given a sum of squares density function $h^*(x)$ on $[0,1]^n$. For $i = 1, \dots, n$, define the polynomial function $f_{1\dots i} \in \mathbb{R}[x_1, \dots, x_i]$ by

$$f_{1\dots i}(x_1, \dots, x_i) = \int_0^1 \cdots \int_0^1 h^*(x_1, \dots, x_n) dx_{i+1} \cdots dx_n. \tag{26}$$

Then the cumulative marginal distribution function $F_1(\cdot)$ is given by

$$F_1(x_1) = \int_0^{x_1} f_1(y) dy$$

and, for $i = 2, \dots, n$, the cumulative conditional distribution function $F_i(\cdot \mid x_1 \dots x_{i-1})$ is given by

$$F_i(x_i \mid x_1 \dots x_{i-1}) = \frac{\int_0^{x_i} f_{1\dots i}(x_1, \dots, x_{i-1}, y) dy}{f_{1\dots(i-1)}(x_1, \dots, x_{i-1})}.$$

The computation of the cumulative marginal and conditional distributions can be carried out in the same way for the simplex $\mathbf{K} = \Delta_n$, after replacing the function $f_{1\dots i} \in \mathbb{R}[x_1, \dots, x_i]$ in (26) by

$$f_{1\dots i}(x_1, \dots, x_i) = \int_0^{1-x_i-x_{i+1}-\dots-x_{n-1}} \int_0^{1-x_i-\dots-x_{n-2}} \cdots \int_0^{1-x_i} h^*(x_1, \dots, x_n) dx_{i+1} \cdots dx_n.$$

Note that in both cases the functions $F_i(x_i \mid x_1 \dots x_{i-1})$ are indeed univariate polynomials. We will apply this sampling method to several examples of polynomial minimization over the hypercube and the simplex in the next section.

We now observe that if we generate sufficiently many samples from the distribution $\mathcal{T}_\mathbf{K}$ then, with high probability, one of these samples is a point $x \in \mathbf{K}$ satisfying (roughly) $f(x) \le \underline{f}_\mathbf{K}^{(r)}$.

**Theorem 5** *Let $X \sim \mathcal{T}_\mathbf{K}$. For all $\epsilon > 0$,*

$$\mathbf{Pr}\left[f(X) \ge \underline{f}_\mathbf{K}^{(r)} + \epsilon\left(\underline{f}_\mathbf{K}^{(r)} - f_{\min,\mathbf{K}}\right)\right] \le \frac{1}{1+\epsilon}.$$

*Proof* Let $X \sim \mathcal{T}_\mathbf{K}$ so that $\mathbb{E}\left[f(X)\right] = \underline{f}_\mathbf{K}^{(r)}$. Define the nonnegative random variable

$$Y := f(X) - f_{\min,\mathbf{K}}.$$

Then, one has $\mathbb{E}\left[Y\right] = \underline{f}_\mathbf{K}^{(r)} - f_{\min,\mathbf{K}}$. Given $\epsilon > 0$, the Markov Inequality (see e.g., [22, Theorem 3.2]) implies

$$\mathbf{Pr}\left[Y \ge (1+\epsilon)\mathbb{E}\left[Y\right]\right] \le \frac{1}{1+\epsilon}.$$

This completes the proof. □

For given $\epsilon > 0$, if one samples $N$ times independently from $\mathcal{T}_\mathbf{K}$, one therefore obtains an $x \in \mathbf{K}$ such that

$$f(x) < \underline{f}_\mathbf{K}^{(r)} + \epsilon\left(\underline{f}_\mathbf{K}^{(r)} - f_{\min,\mathbf{K}}\right)$$

with probability at least $1 - \left(\frac{1}{1+\epsilon}\right)^N$. For example, if $N \ge 1 + \frac{1}{\epsilon}$ then this probability is at least $1 - 1/e$.

## 4 Numerical examples

In this section, we consider several well-known polynomial test functions from global optimization that are listed in Table 1.

**Table 1** Test functions

| Name | Formula | Minimum ($f_{\min,\mathbf{K}}$) | Search domain ($\mathbf{K}$) |
|---|---|---|---|
| Booth Function | $f=(x_1+2x_2-7)^2+(2x_1+x_2-5)^2$ | $f(1,3)=0$ | $[-10,10]^2$ |
| Matyas Function | $f=0.26(x_1^2+x_2^2)-0.48x_1x_2$ | $f(0,0)=0$ | $[-10,10]^2$ |
| Three–Hump Camel Function | $f=2x_1^2-1.05x_1^4+\frac{1}{6}x_1^6+x_1x_2+x_2^2$ | $f(0,0)=0$ | $[-5,5]^2$ |
| Motzkin Polynomial | $f=x_1^4x_2^2+x_1^2x_2^4-3x_1^2x_2^2+1$ | $f(\pm1,\pm1)=0$ | $[-2,2]^2$ |
| Styblinski–Tang Function ($n$-variate) | $f=\sum_{i=1}^n \frac{1}{2}x_i^4-8x_i^2+\frac{5}{2}x_i$ | $f(-2.093534,\dots,-2.093534)=-39.16599n$ | $[-5,5]^n$ |
| Rosenbrock Function ($n$-variate) | $f=\sum_{i=1}^{n-1}100(x_{i+1}-x_i^2)^2+(x_i-1)^2$ | $f(1,\dots,1)=0$ | $[-2.048,2.048]^n$ |
| Matyas Function (Modified-S) | $f=0.26[(20x_1-10)^2+(20x_2-10)^2]-0.48(20x_1-10)(20x_2-10)$ | $f(0.5,0.5)=0$ | $\Delta_2$ |
| Three-Hump Camel Function (Modified-S) | $f=2(10x_1-5)^2-1.05(10x_1-5)^4+\frac{1}{6}(10x_1-5)^6+(10x_1-5)(10x_2-5)+(10x_2-5)^2$ | $f(0.5,0.5)=0$ | $\Delta_2$ |
| Matyas Function (Modified-B) | $f=0.26[(20x_1^2-10)^2+(20x_2^2-10)^2]-0.48(20x_1^2-10)(20x_2^2-10)$ | $f(\pm\frac{\sqrt{2}}{2},\pm\frac{\sqrt{2}}{2})=0$ | $B_1(0)$ |
| Three-Hump Camel Function (Modified-B) | $f=2(10x_1^2-5)^2-1.05(10x_1^2-5)^4+\frac{1}{6}(10x_1^2-5)^6+(10x_1^2-5)(10x_2^2-5)+(10x_2^2-5)^2$ | $f(\pm\frac{\sqrt{2}}{2},\pm\frac{\sqrt{2}}{2})=0$ | $B_1(0)$ |

For these functions, we calculate the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ by solving the SDP (3) for increasing values of the order $r$. As already mentioned by Lasserre [17, Section 4], this computation may be done as a generalised eigenvalue problem — one does not actually have to use an SDP solver. This follows from the fact that the SDP (3) only has one constraint. In particular, $\underline{f}_{\mathbf{K}}^{(r)}$ is equal to the largest scalar $\lambda$ for which $A-\lambda B\succeq 0$, i.e., the smallest generalized eigenvalue of the system:

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

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

$$A_{\alpha,\beta}=\sum_{\delta\in N(n,d)} f_\delta\int_{\mathbf{K}} x^{\alpha+\beta+\delta}dx,\quad B_{\alpha,\beta}=\int_{\mathbf{K}} x^{\alpha+\beta}dx \quad \alpha,\beta\in N(n,r). \tag{27}$$

We performed the computation on a PC with Intel(R) Core(TM) i7-4600U CPU (2.10 GHz) and with 8 GB RAM. The generalized eigenvalue computation was done in Matlab using the `eig` function.

We record the values $\underline{f}_{\mathbf{K}}^{(r)}$ as well as the CPU times in Tables 2, 3, 4, 5, and 6 for minimization over the hypercube, the simplex and the ball. Note that we only list the time for solving the generalised eigenvalue problem, and not for constructing the matrices $A$ and $B$ in (27). In other words, we assume the necessary moments are computed beforehand, and that the time needed to construct the matrices $A$ and $B$ in (27) is negligible if the relevant moments are known.

For instance, in Table 2, we have $n = 2$ and we can compute the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ up to order $r = 20$ for four test functions. Moreover, in Tables 3, 4 and 5, we have $n = 10, 15, 20$, respectively, and the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ can be computed up to order $r = 5$, $r = 4$ and $r = 3$, respectively. Note that in all cases the computation is very fast (at most a few seconds). However, for larger values of $n$ or $r$ we sometimes encountered numerical instability. This may be due to inaccurate calculation of the moments, or to inherent ill-conditioning of the matrices $A$ and $B$ in (27). These issues are of practical importance, but beyond the scope of the present study. Also, one must bear in mind that the order of the matrices $A$ and $B$ grows as $\binom{n+r}{r}$, and this imposes a practical limit on how large the values of $n$ and $r$ may be when computing $\underline{f}_{\mathbf{K}}^{(r)}$.

**Table 2** $\underline{f}_{\mathbf{K}}^{(r)}$ for Booth, Matyas, Three–Hump Camel and Motzkin Functions over the hypercube

| r | Booth Function | | Matyas Function | | Three–Hump Camel Function | | Motzkin Polynomial | |
|---|---|---|---|---|---|---|---|---|
| | Value | Time (sec.) | Value | Time (sec.) | Value | Time (sec.) | Value | Time (sec.) |
| 1 | 244.680 | 0.000666 | 8.26667 | 0.000739 | 265.774 | 0.000742 | 4.2 | 0.000719 |
| 2 | 162.486 | 0.000061 | 5.32223 | 0.000072 | 29.0005 | 0.000062 | 1.06147 | 0.000088 |
| 3 | 118.383 | 0.000083 | 4.28172 | 0.000072 | 29.0005 | 0.000066 | 1.06147 | 0.000080 |
| 4 | 97.6473 | 0.000079 | 3.89427 | 0.000119 | 9.58064 | 0.000117 | 0.829415 | 0.000118 |
| 5 | 69.8174 | 0.000171 | 3.68942 | 0.000208 | 9.58064 | 0.000177 | 0.801069 | 0.000189 |
| 6 | 63.5454 | 0.000277 | 2.99563 | 0.000263 | 4.43983 | 0.000263 | 0.801069 | 0.000208 |
| 7 | 47.0467 | 0.000423 | 2.54698 | 0.000343 | 4.43983 | 0.001146 | 0.708889 | 0.000395 |
| 8 | 41.6727 | 0.000587 | 2.04307 | 0.000417 | 2.55032 | 0.000647 | 0.565553 | 0.000584 |
| 9 | 34.2140 | 0.000657 | 1.83356 | 0.000655 | 2.55032 | 0.000586 | 0.565553 | 0.000766 |
| 10 | 28.7248 | 0.000997 | 1.47840 | 0.000780 | 1.71275 | 0.000782 | 0.507829 | 0.001210 |
| 11 | 25.6050 | 0.001181 | 1.37644 | 0.009241 | 1.71275 | 0.001026 | 0.406076 | 0.001261 |
| 12 | 21.1869 | 0.001942 | 1.11785 | 0.001753 | 1.2775 | 0.001693 | 0.406076 | 0.001712 |
| 13 | 19.5588 | 0.002352 | 1.0686 | 0.001857 | 1.2775 | 0.002031 | 0.3759 | 0.003427 |
| 14 | 16.5854 | 0.002829 | 0.8742 | 0.002253 | 1.0185 | 0.002629 | 0.3004 | 0.003711 |
| 15 | 15.2815 | 0.003618 | 0.8524 | 0.002270 | 1.0185 | 0.002936 | 0.3004 | 0.002351 |
| 16 | 13.4626 | 0.003452 | 0.7020 | 0.003580 | 0.8434 | 0.003452 | 0.2819 | 0.003672 |
| 17 | 12.2075 | 0.004248 | 0.6952 | 0.004662 | 0.8434 | 0.004652 | 0.2300 | 0.004349 |
| 18 | 11.0959 | 0.005217 | 0.5760 | 0.005510 | 0.7113 | 0.004882 | 0.2300 | 0.006060 |
| 19 | 9.9938 | 0.007200 | 0.5760 | 0.005610 | 0.7113 | 0.006752 | 0.2185 | 0.007641 |
| 20 | 9.2373 | 0.009707 | 0.4815 | 0.006975 | 0.6064 | 0.007031 | 0.1817 | 0.007686 |

Furthermore, we use the method described in Section 3 to generate samples that are feasible solutions of (2). We report results for the bivariate Rosenbrock and the Three–Hump Camel functions over the hypercube, and for the Matyas and Three-Hump Camel functions (Modified-S) over the simplex. For each order $r \geq 1$, the sample sizes 20 and 1000 are used. We also generate samples uniformly from the feasible set, for comparison. We give the results in Tables 7, 8, 9 and 10, where we record the mean, variance and the minimum value of these samples together with $\underline{f}_{\mathbf{K}}^{(r)}$ (which equals the sample mean by (25)).

**Table 3** $\underline{f}_{\mathbf{K}}^{(r)}$ for Styblinski–Tang and Rosenbrock Functions (with $n = 10$) over the hypercube

| r | Sty.–Tang ($n = 10$) | | Rosenb. ($n = 10$) | |
|---|---|---|---|---|
| | Value | Time (sec.) | Value | Time (sec.) |
| 1 | −57.1688 | 0.098 | 3649.85 | 0.0005 |
| 2 | −94.5572 | 0.001 | 2813.66 | 0.0009 |
| 3 | −108.873 | 0.011 | 2393.63 | 0.0156 |
| 4 | −132.8810 | 0.349 | 1956.81 | 0.4004 |
| 5 | −146.7906 | 9.245 | 1701.85 | 12.997 |

**Table 4** $\underline{f}_{\mathbf{K}}^{(r)}$ for Styblinski–Tang and Rosenbrock Functions (with $n = 15$) over the hypercube

| r | Sty.–Tang ($n = 15$) | | Rosenb. ($n = 15$) | |
|---|---|---|---|---|
| | Value | Time (sec.) | Value | Time (sec.) |
| 1 | −82.8311 | 0.001071 | 5887.5 | 0.094693 |
| 2 | −130.464 | 0.001707 | 4770.71 | 0.002282 |
| 3 | −148.5594 | 0.170907 | 4160.78 | 0.157897 |
| 4 | −180.9728 | 16.796383 | 3552.04 | 24.696591 |

**Table 5** $\underline{f}_{\mathbf{K}}^{(r)}$ for Styblinski–Tang and Rosenbrock Functions (with $n = 20$) over the hypercube

| r | Sty.–Tang ($n = 20$) | | Rosenb. ($n = 20$) | |
|---|---|---|---|---|
| | Value | Time (sec.) | Value | Time (sec.) |
| 1 | −107.875 | 0.972741 | 8158.36 | 0.000949 |
| 2 | −164.11 | 0.344403 | 6806.74 | 0.011370 |
| 3 | −185.6488 | 2.655447 | 6029.02 | 2.955319 |

**Table 6** $\underline{f}_{\mathbf{K}}^{(r)}$ for Matyas and Three-Hump Camel Functions (Modified) over the Simplex and the Euclidean ball.

| r | Matyas (Modified-S) | | Th.-H. C. (Modified-S) | | Matyas (Modified-B) | | Th.-H. C. (Modified-B) | |
|---|---|---|---|---|---|---|---|---|
| | Value | Time (sec.) | Value | Time (sec.) | Value | Time (sec.) | Value | Time (sec.) |
| 1 | 7.2243 | 0.222604 | 84.354 | 0.000457 | 18.000 | 0.000379 | 146.41 | 0.000454 |
| 2 | 4.6536 | 0.000085 | 22.398 | 0.000081 | 6.3995 | 0.000049 | 138.91 | 0.000052 |
| 3 | 3.9404 | 0.000124 | 12.353 | 0.000115 | 6.3995 | 0.000054 | 48.508 | 0.000069 |
| 4 | 3.7067 | 0.000176 | 3.9153 | 0.000112 | 4.4091 | 0.000133 | 39.673 | 0.000111 |
| 5 | 3.2317 | 0.000696 | 2.9782 | 0.000489 | 4.4091 | 0.000187 | 18.045 | 0.000264 |
| 6 | 2.7328 | 0.000275 | 1.3303 | 0.000255 | 3.9652 | 0.000292 | 13.881 | 0.000309 |
| 7 | 2.2985 | 0.000511 | 1.1773 | 0.000334 | 3.9652 | 0.000323 | 7.7876 | 0.000300 |
| 8 | 1.9536 | 0.001432 | 0.77992 | 0.000560 | 3.8536 | 0.000395 | 5.7685 | 0.000608 |
| 9 | 1.6639 | 0.000709 | 0.73202 | 0.000666 | 3.8536 | 0.000517 | 3.8699 | 0.000636 |
| 10 | 1.4293 | 0.003370 | 0.60846 | 0.001034 | 3.4943 | 0.000687 | 2.8359 | 0.000704 |

Note that the average of the sample function values approximate $\underline{f}_{\mathbf{K}}^{(r)}$ reasonably well for sample size 1000, but poorly for sample size 20. Moreover, the average sample function value for uniform sampling from $\mathbf{K}$ is much higher than $\underline{f}_{\mathbf{K}}^{(r)}$. Also, the minimum function value for sampling from $\mathcal{T}_{\mathbf{K}}$ is significantly lower than the minimum function value obtained by uniform sampling for most values of $r$. In terms of generating "good" feasible solutions, sampling from $\mathcal{T}_{\mathbf{K}}$ therefore outperforms uniform sampling from $\mathbf{K}$ for these examples, as one would expect.

**Table 7** Sampling results for the Rosenbrock Function ($n = 2$) over the hypercube

| r | $f_{\mathbf{K}}^{(r)}$ | Mean | Variance | Minimum | Sample Size |
|---|---|---|---|---|---|
| 1 | 214.648 | 121.125 | 14005.5 | 0.00451826 | 20 |
| | | 209.9 | 80699.0 | 0.0008754 | 1000 |
| 2 | 152.310 | 184.496 | 58423.9 | 4.94265 | 20 |
| | | 149.6 | 54455.0 | 0.02805 | 1000 |
| 3 | 104.889 | 146.618 | 64611.2 | 0.0113339 | 20 |
| | | 110.1 | 26022.0 | 0.0665 | 1000 |
| 4 | 75.6010 | 62.4961 | 5803.21 | 0.0542813 | 20 |
| | | 75.65 | 45777.0 | 0.007285 | 1000 |
| 5 | 51.5037 | 58.4032 | 4397.0 | 0.668679 | 20 |
| | | 50.64 | 6285.0 | 0.01382 | 1000 |
| 6 | 41.7878 | 35.4183 | 2936.24 | 1.16154 | 20 |
| | | 37.64 | 3097.0 | 0.06188 | 1000 |
| 7 | 30.1392 | 29.6545 | 1022.2 | 1.05813 | 20 |
| | | 27.11 | 1332.0 | 0.02044 | 1000 |
| 8 | 25.8329 | 19.5392 | 301.334 | 0.505628 | 20 |
| | | 34.32 | 4106.0 | 0.074 | 1000 |
| 9 | 19.4972 | 20.8982 | 328.475 | 0.564992 | 20 |
| | | 18.65 | 593.6 | 0.07951 | 1000 |
| 10 | 17.3999 | 9.37959 | 146.496 | 0.562473 | 20 |
| | | 15.33 | 685.7 | 0.1448 | 1000 |
| 11 | 13.6289 | 8.74923 | 52.1436 | 0.75774 | 20 |
| | | 15.7 | 7498.0 | 0.1719 | 1000 |
| 12 | 12.5024 | 5.43151 | 66.561 | 0.438172 | 20 |
| | | 12.7 | 764.7 | 0.0945 | 1000 |
| Uniform Sample | | 489.722 | 433549.0 | 9.0754 | 20 |
| | | 465.729 | 361150.0 | 0.0771463 | 1000 |

## 5 Concluding remarks

We conclude with some additional remarks on Assumption 1, and some discussion on perspectives for future work.

### 5.1 Revisiting Assumption 1

In this section we consider in more detail Assumption 1, the geometric assumption which we made about the set $\mathbf{K}$. First we recall another condition, known as the *interior cone condition*, which is classically used in approximation theory (see, e.g., Wendland [28]).

**Definition 1** [28, Definition 3.1] A set $\mathbf{K} \subseteq \mathbb{R}^n$ is said to satisfy an interior cone condition if there exist an angle $\theta \in (0, \pi/2)$ and a radius $\rho > 0$ such that, for every $x \in \mathbf{K}$, a unit vector $\xi(x)$ exists such that the set

$$C(x, \xi(x), \theta, \rho) := \{x + \lambda y : y \in \mathbb{R}^n, \|y\| = 1, y^T\xi(x) \geq \cos\theta, \lambda \in [0, \rho]\} \tag{28}$$

is contained in $\mathbf{K}$.

For instance, as we now recall, Euclidean balls and star-shaped sets satisfy the interior cone condition.

**Table 8** Sampling results for the Three–Hump Camel Function over the hypercube

| r | $\underline{f}_{\mathbf{K}}^{(r)}$ | Mean | Variance | Minimum | Sample Size |
|---|---|---|---|---|---|
| 1 | 265.774 | 216.773 | 177142.0 | 0.106854 | 20 |
| | | 261.23 | 193466.0 | 0.11705 | 1000 |
| 2 | 29.0005 | 28.0344 | 2964.85 | 1.1718 | 20 |
| | | 27.712 | 6712.8 | 0.014255 | 1000 |
| 3 | 29.0005 | 14.9951 | 523.904 | 0.452655 | 20 |
| | | 32.363 | 16681.0 | 0.0088426 | 1000 |
| 4 | 9.58064 | 2.99756 | 14.1201 | 0.175016 | 20 |
| | | 10.364 | 1944.0 | 0.010013 | 1000 |
| 5 | 9.58064 | 4.41907 | 14.1358 | 0.419394 | 20 |
| | | 9.1658 | 643.88 | 0.0015924 | 1000 |
| 6 | 4.43983 | 7.98481 | 245.089 | 0.126147 | 20 |
| | | 4.5791 | 493.12 | 0.0035581 | 1000 |
| 7 | 4.43983 | 3.96711 | 20.3193 | 0.260331 | 20 |
| | | 3.7911 | 57.847 | 0.0076111 | 1000 |
| 8 | 2.55032 | 2.18925 | 3.87943 | 0.0310113 | 20 |
| | | 2.2302 | 8.3767 | 0.0028817 | 1000 |
| 9 | 2.55032 | 1.38102 | 2.27433 | 0.138641 | 20 |
| | | 3.2217 | 812.18 | 0.00014805 | 1000 |
| 10 | 1.71275 | 1.03179 | 0.992636 | 0.0645815 | 20 |
| | | 1.5069 | 3.9581 | 0.0014225 | 1000 |
| 11 | 1.71275 | 1.30757 | 1.90985 | 0.0320489 | 20 |
| | | 1.6379 | 7.2518 | 0.0021144 | 1000 |
| 12 | 1.27749 | 0.841194 | 0.914514 | 0.0369565 | 20 |
| | | 1.2105 | 2.3 | 0.0005154 | 1000 |
| Uniform Sample | | 304.032 | 163021.0 | 1.65885 | 20 |
| | | 243.216 | 183724.0 | 0.00975034 | 1000 |

**Table 9** Sampling results for the Matyas Function (Modified-S) over the simplex

| r | $\underline{f}_{\mathbf{K}}^{(r)}$ | Mean | Variance | Minimum | Sample Size |
|---|---|---|---|---|---|
| 1 | 7.2243 | 6.3018 | 37.373 | 1.2448 | 20 |
| | | 7.0542 | 64.863 | 0.31812 | 1000 |
| 2 | 4.6536 | 5.7252 | 34.964 | 1.8924 | 20 |
| | | 4.5932 | 8.293 | 0.91671 | 1000 |
| 3 | 3.9404 | 3.5187 | 0.31411 | 2.4465 | 20 |
| | | 3.7544 | 1.3576 | 0.071075 | 1000 |
| 4 | 3.7067 | 3.4279 | 1.7187 | 0.92913 | 20 |
| | | 3.8679 | 6.5113 | 0.027508 | 1000 |
| 5 | 3.2317 | 3.8273 | 10.173 | 0.40131 | 20 |
| | | 3.1485 | 6.1263 | 0.035796 | 1000 |
| 6 | 2.7328 | 2.2606 | 3.3343 | 0.2595 | 20 |
| | | 2.5997 | 10.8 | 0.0016761 | 1000 |
| 7 | 2.2985 | 2.4568 | 4.1652 | 0.18947 | 20 |
| | | 2.1541 | 12.868 | 0.002669 | 1000 |
| 8 | 1.9536 | 0.9223 | 0.94139 | 0.064404 | 20 |
| | | 1.9418 | 9.5627 | 0.0000037429 | 1000 |
| 9 | 1.6639 | 1.4446 | 1.9372 | 0.048915 | 20 |
| | | 1.7266 | 16.738 | 0.0019792 | 1000 |
| 10 | 1.4293 | 2.0005 | 2.0226 | 0.016453 | 20 |
| | | 1.4917 | 16.035 | 0.00015252 | 1000 |
| Uniform Sample | | 26.428 | 641.59 | 0.085716 | 20 |
| | | 11.905 | 256.0 | 0.010946 | 1000 |

**Table 10** Sampling results for the Three-Hump Camel Function (Modified-S) over the simplex

| r | $f_{\mathbf{K}}^{(r)}$ | Mean | Variance | Minimum | Sample Size |
|---|---|---|---|---|---|
| 1 | 84.354 | 104.93 | 122488.0 | 0.33441 | 20 |
| | | 89.732 | 48238.0 | 0.0011036 | 1000 |
| 2 | 22.398 | 37.036 | 9864.0 | 0.57012 | 20 |
| | | 22.292 | 10102.0 | 0.0022204 | 1000 |
| 3 | 12.353 | 3.4161 | 49.898 | 0.28108 | 20 |
| | | 11.707 | 1515.9 | 0.00065454 | 1000 |
| 4 | 3.9153 | 2.4193 | 9.0182 | 0.16865 | 20 |
| | | 3.6768 | 592.96 | 0.0016775 | 1000 |
| 5 | 2.9782 | 1.8336 | 6.3414 | 0.11311 | 20 |
| | | 2.5237 | 47.619 | 0.00097905 | 1000 |
| 6 | 1.3303 | 2.355 | 26.176 | 0.0092016 | 20 |
| | | 1.2134 | 8.7253 | 0.00040725 | 1000 |
| 7 | 1.1773 | 1.0385 | 1.0569 | 0.053695 | 20 |
| | | 1.092 | 6.718 | 0.00050329 | 1000 |
| 8 | 0.77992 | 0.9737 | 0.73522 | 0.10604 | 20 |
| | | 0.72927 | 0.73641 | 0.00048517 | 1000 |
| 9 | 0.73202 | 0.69755 | 0.19107 | 0.051634 | 20 |
| | | 0.65302 | 0.28537 | 0.00024601 | 1000 |
| 10 | 0.60846 | 0.67575 | 0.17453 | 0.010351 | 20 |
| | | 0.5616 | 0.17821 | 0.00044175 | 1000 |
| Uniform Sample | | 518.48 | 354855.0 | 0.9165 | 20 |
| | | 485.77 | 391577.0 | 0.32713 | 1000 |

**Lemma 6** *[28, Lemma 3.10] Every Euclidean ball with radius $r > 0$ satisfies an interior cone condition with radius $\rho = r$ and angle $\theta = \pi/3$.*

**Definition 2** [28, Definition 11.25] A set $\mathbf{K}$ is said to be *star-shaped* with respect to a ball $B_r(x_c)$ if, for every $x \in \mathbf{K}$, the closed convex hull of $\{x\} \cup B_r(x_c)$ is contained in $\mathbf{K}$.

**Proposition 1** *[28, Proposition 11.26] If $\mathbf{K}$ is bounded, star-shaped with respect to a ball $B_r(x_c)$, then $\mathbf{K}$ satisfies an interior cone condition with radius $\rho = r$ and angle $\theta = 2\arcsin\left[\frac{r}{2\sqrt{D(\mathbf{K})}}\right]$.*

In fact, any set satisfying the interior cone condition also satisfies the following stronger version of Assumption 1.

**Assumption 2** *There exist constants $\eta_{\mathbf{K}} > 0$ and $\epsilon_{\mathbf{K}} > 0$ such that, for all points $a \in \mathbf{K}$,*

$$\mathrm{vol}(B_\epsilon(a) \cap \mathbf{K}) \geq \eta_{\mathbf{K}} \mathrm{vol} B_\epsilon(a) = \eta_{\mathbf{K}} \epsilon^n \gamma_n \quad \textit{for all } 0 < \epsilon \leq \epsilon_{\mathbf{K}}. \tag{29}$$

Hence the only difference with Assumption 1 is that the constants $\eta_{\mathbf{K}}$ and $\epsilon_{\mathbf{K}}$ now depend only on the set $\mathbf{K}$ and not on the choice of $a \in \mathbf{K}$. Clearly, Assumption 2 implies Assumption 1. Moreover, any set satisfying the interior cone condition satisfies Assumption 2.

**Lemma 7** *If a set $\mathbf{K} \subseteq \mathbb{R}^n$ satisfies the interior cone condition (28) then $\mathbf{K}$ also satisfies Assumption 2 (and thus Assumption 1), where we set*

$$\eta_{\mathbf{K}} = \left[\frac{\sin\theta}{1+\sin\theta}\right]^n \quad \textit{and} \quad \epsilon_{\mathbf{K}} = \rho.$$

*Proof* Assume that $\mathbf{K}$ satisfies the interior cone condition (28). Then, using [28, Lemma 3.7], we know that, for every $x \in \mathbf{K}$ and $h \le \rho/(1+\sin\theta)$, the closed ball $B_{h\sin\theta}(x+h\xi(x))$ is contained in $C(x,\xi(x),\theta,\rho)$ and thus in $\mathbf{K}$. Then, for all $x_0 \in \mathbf{K}$ and $\epsilon \in (0,\rho]$, after setting $h = \epsilon/(1+\sin\theta)$, one can obtain

$$\frac{\mathrm{vol}(B_\epsilon(x_0)\cap\mathbf{K})}{\mathrm{vol}B_\epsilon(x_0)} \ge \frac{\mathrm{vol}C(x_0,\xi(x_0),\theta,\epsilon)}{\mathrm{vol}B_\epsilon(x_0)} \ge \frac{\mathrm{vol}B_{h\sin\theta}(x_0+h\xi(x_0))}{\mathrm{vol}B_\epsilon(x_0)} = \left[\frac{\sin\theta}{1+\sin\theta}\right]^n.$$

Thus, Assumption 2 holds after setting $\eta_{\mathbf{K}} = \left[\frac{\sin\theta}{1+\sin\theta}\right]^n$ and $\epsilon_{\mathbf{K}} = \rho$. □

As any convex body (i.e., full-dimensional convex and compact) is star-shaped with respect to any ball it contains, the next result follows as a direct application of Proposition 1 and Lemma 7.

**Corollary 1** *Any convex body satisfies the interior cone condition and thus Assumptions 1 and 2.*

As an illustration we now consider the parameters $\eta_{\mathbf{K}}$, $\epsilon_{\mathbf{K}}$, and $r_{\mathbf{K}}$ (from relation (7)) when $\mathbf{K}$ is the hypercube, the simplex and the Euclidean ball.

*Remark 3* Consider first the case when $\mathbf{K}$ is the hypercube $\mathbf{Q}_n = [0,1]^n$. By Proposition 1, it satisfies the interior cone condition with radius $\rho = 1/2$ and angle $\theta = 2\arcsin\left[\frac{1}{4\sqrt{n}}\right]$. Hence, Assumption 2 holds with $\epsilon_{\mathbf{K}} = 1/2$ and $\eta_{\mathbf{K}} = \left(\frac{\sqrt{16n-1}}{8n+\sqrt{16n-1}}\right)^n$ (which is $\sim \left(\frac{1}{2\sqrt{n}}\right)^n$ for $n$ large). Moreover, as $D(\mathbf{K}) = n$, it follows that $r_{\mathbf{K}} = 4ne$.
Consider now the case when $\mathbf{K}$ is the full-dimensional simplex $\Delta_n$. By Proposition 1, it satisfies the interior cone condition with radius $\rho = \frac{1}{n+\sqrt{n}}$ and angle $\theta = 2\arcsin\left[\frac{1}{2\sqrt{2}(n+\sqrt{n})}\right]$ (since the ball with center $\rho(1,\ldots,1)^T$ and radius $\rho$ is contained in $\widehat{\Delta}_n$). Hence Assumption 2 holds with $\epsilon_{\mathbf{K}} = \frac{1}{n+\sqrt{n}}$ and $\eta_{\mathbf{K}} = \left(\frac{\sqrt{8(n+\sqrt{n})^2-1}}{4(n+\sqrt{n})^2+\sqrt{8(n+\sqrt{n})^2-1}}\right)^n$ (which is $\sim \left(\frac{1}{\sqrt{2}n}\right)^n$ for $n$ large). As $D(\mathbf{K}) = 2$, it follows that $r_{\mathbf{K}} = e(n+\sqrt{n})^3$.

Finally, for the Euclidean ball $\mathbf{K} = B_1(0)$, we have $\epsilon_{\mathbf{K}} = 1$, $\eta_{\mathbf{K}} = \left(\frac{\sqrt{3}}{2+\sqrt{3}}\right)^n$ and $r_{\mathbf{K}} = \max\{2e, n\}$.

## 5.2 Perspectives

The sampling approach of Section 3 often provides good feasible solutions for the examples in Section 4, even for small values of $r$. One may therefore explore using the sampling technique (for small $r$) as a way of generating starting points for multi-start global optimization algorithms.

Another possibility to enhance computation would be to investigate other sufficient conditions for nonnegativity of $h$ on $\mathbf{K}$, more general than the sum-of-squares condition studied here. This may result in a faster rate of convergence than for $\underline{f}_{\mathbf{K}}^{(r)}$.

Finally, understanding the exact rate of convergence of the upper bounds $\underline{f}_{\mathbf{K}}^{(r)}$ remains an open problem. In particular we do not know whether $1/\sqrt{r}$ is the right rate of convergence.

## Acknowledgements

We thank Jean Bernard Lasserre for bringing our attention to his work [17] and for several valuable suggestions, and Dorota Kurowicka for valuable discussions on multivariate sampling techniques.


## References


1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55 (1972)
2. Blumenson, L.E.: A Derivation of $n$-Dimensional Spherical Coordinates. *The American Mathematical Monthly* 67(1), 63–66 (1960)
3. De Klerk, E., Den Hertog, D., Elabwabi, G.. On the complexity of optimization over the standard simplex. *European journal of operational research,* 191, 773–785 (2008)
4. De Klerk, E., Laurent, M.: Error bounds for some semidefinite programming approaches to polynomial minimization on the hypercube. *SIAM Journal on Optimization,* 20(6), 3104–3120 (2010)
5. De Klerk, E., Laurent, M., Parrilo, P.: A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theory of Computer Science* 361(2-3), 210–225 (2006)
6. De Klerk, E., Laurent, M., Sun, Z.: An alternative proof of a PTAS for fixed-degree polynomial optimization over the simplex. *Mathmatical Programming,* 151(2), 433–457 (2015)
7. De Klerk, E., Laurent, M., Sun, Z.: An error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution. *SIAM Journal on Optimization,* 25(3), 1498–1514 (2015)
8. De Loera, J., Rambau, J., Santos, F.: Triangulations: Structures and algorithms, Book manuscript (2008)
9. Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere. arXiv:1210.5048v2 (2013)
10. Dyer, M.E., Frieze, A.M.: On the Complexity of Computing the Volume of a Polyhedron. *SIAM J. Comput.*, 17(5), 967–974 (1988)
11. Faybusovich, L.: Global optimization of homogeneous polynomials on the simplex and on the sphere. In C. Floudas and P. Pardalos, editors, Frontiers in Global Optimization. Kluwer Academic Publishers (2003)
12. Grundmann, A., Moeller, H.M.: Invariant integration formulas for the n-simplex by combinatorial methods. *SIAM J. Numer. Anal.* 15, 282–290 (1978)
13. Henrion, D., Lasserre, J.B., Loefberg, J.: GloptiPoly 3: moments, optimization and semidefinite programming. *Optim. Method. Softw.* 24(4–5) 761–779 (2009)
14. Lasserre, J.B., Zeron, E.S.: Solving a class of multivariate integration problems via Laplace techniques. *Applicationes Mathematicae*, 28(4), 391–405 (2001)
15. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11, 796–817 (2001)
16. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press (2009)
17. Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* 21(3), 864–885 (2011)
18. Lasserre, J.B.: Unit balls of constant volume: which one has optimal representation? Preprint at arXiv: 1408.1324 (2014)
19. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. *In Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 157–270 (2009)
20. Law, A.M.: Simulation Modeling and Analysis (4th edition). Mc Graw-Hill (2007)
21. Mattila, P.: *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability.* Cambridge University Press (1999)
22. Motwani, R., Raghavan, P.: Randomized Algorithms. Cambridge University Press (1995)
23. Nesterov, Y.: Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper 2003/71, CORE-UCL, Louvain-La-Neuve (2003)
24. Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. *Journal of Complexity*, 23(1), 135–150 (2007)
25. Nie, J.: Certifying convergence of Lasserre's hierarchy via flat truncation. *Math. Program.* 142(1–2), 485–510 (2013)
26. Schweighofer, M.: On the complexity of Schmüdgen's Positivstellensatz. *Journal of Complexity*, 20, 529–543 (2004)

27. Sun, Z.: A refined error analysis for fixed-degree polynomial optimization over the simplex. *Journal of the Operations Research Society of China,* 2(3), 379–393 (2014)
28. Wendland, H.: Scattered Data Approximation. Cambridge University Press (2005)
29. Whittaker, E.T., Watson, G.W.: A course of modern analysis (4ed). Cambridge University Press, New York (1996)