Bound-constrained polynomial optimization using only elementary calculations

De Klerk, Etienne; Lasserre, Jean B.; Laurent, Monique; Sun, Zhao

DOI
10.1287/moor.2016.0829

Publication date
2017

Document Version
Accepted author manuscript

Published in
Mathematics of Operations Research

Citation (APA)
De Klerk, E., Lasserre, J. B., Laurent, M., & Sun, Z. (2017). Bound-constrained polynomial optimization using only elementary calculations. *Mathematics of Operations Research*, *42*(3), 834-853. https://doi.org/10.1287/moor.2016.0829

BOUND-CONSTRAINED POLYNOMIAL OPTIMIZATION USING ONLY ELEMENTARY CALCULATIONS

ETIENNE DE KLERK, JEAN B. LASSERRE, MONIQUE LAURENT, AND ZHAO SUN

Abstract. We provide a monotone non-increasing sequence of upper bounds f_k^H ($k \geq 1$) converging to the global minimum of a polynomial f on simple sets like the unit hypercube in $\mathbb{R}^n$. The novelty with respect to the converging sequence of upper bounds in [J.B. Lasserre, A new look at nonnegativity on closed sets and polynomial optimization, *SIAM J. Optim.* **21**, pp. 864–885, 2010] is that only elementary computations are required. For optimization over the hypercube $[0,1]^n$, we show that the new bounds f_k^H have a rate of convergence in $O(1/\sqrt{k})$. Moreover we show a stronger convergence rate in $O(1/k)$ for quadratic polynomials and more generally for polynomials having a rational minimizer in the hypercube. In comparison, evaluation of all rational grid points with denominator k produces bounds with a rate of convergence in $O(1/k^2)$, but at the cost of $O(k^n)$ function evaluations, while the new bound f_k^H needs only $O(n^k)$ elementary calculations.

1. Introduction

Consider the problem of computing the global minimum

$$f_{\min,\mathcal{K}} = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{K}\}, \tag{1.1}$$

of a polynomial f on a compact set $\mathcal{K} \subset \mathbb{R}^n$. (We will mainly deal with the case where $\mathcal{K}$ is a basic semi-algebraic set.)

A fruitful perspective, introduced by Lasserre [16], is to reformulate problem (1.1) as

$$f_{\min,\mathcal{K}} = \inf_{\mu} \int_{\mathcal{K}} f d\mu,$$

where the infimum is taken over all probability measures μ with support in $\mathcal{K}$. Using this reformulation one may obtain a sequence of *lower bounds* on $f_{\min,\mathcal{K}}$ that converges to $f_{\min,\mathcal{K}}$, by introducing tractable convex relaxations of the set of probability measures with support in $\mathcal{K}$ (if $\mathcal{K}$ is semi-algebraic). For more details on this approach the interested reader is referred to Lasserre [15, 16, 18], and [20, 17] for a comparison between linear programming (LP) and semidefinite programming (SDP) relaxations.

As an alternative, one may obtain a sequence of *upper bounds* by optimizing over specific classes of probability distributions. In particular, Lasserre [19] defined the sequence (also called hierarchy) of upper bounds

$$f_k^{sos} := \min_{\sigma \in \Sigma_k[\mathbf{x}]} \left\{ \int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} \ : \ \int_{\mathcal{K}} \sigma(\mathbf{x})d\mathbf{x} = 1 \right\}, \quad (k = 1, 2, \ldots), \tag{1.2}$$

where $\Sigma_k[\mathbf{x}]$ denotes the cone of sums of squares (SOS) of polynomials of degree at most $2k$. Thus the optimization is restricted to probability distributions where the probability density function is

2000 *Mathematics Subject Classification.* 90C22 and 90C26 and 90C30.

Key words and phrases. Polynomial optimization and bound-constrained optimization and Lasserre hierarchy.

an SOS polynomial of degree at most $2k$. Lasserre [19] showed that $f_k^{sos} \to f_{\min,\mathcal{K}}$ as $k \to \infty$ (see Theorem 2.1 below for a precise statement). In principle this approach works for any compact set $\mathcal{K}$ and any polynomial but for practical implementation it requires knowledge of moments of the measure $\sigma(\mathbf{x})d\mathbf{x}$. So in practice the approach is limited to *simple* sets $\mathcal{K}$ like the Euclidean ball, the hypersphere, the simplex, the hypercube and/or their image by a linear transformation.

In fact computing such upper bounds reduces to computing the smallest generalized eigenvalue associated with two real symmetric matrices whose size increases in the hierarchy. For more details the interested reader is referred to Lasserre [19]. In a recent paper, De Klerk et al. [6] have provided the first convergence analysis for this hierarchy and shown a bound $f_k^{sos} - f_{\min,\mathcal{K}} = O(1/\sqrt{k})$ on the rate of convergence. In a related analysis of convergence Romero and Velasco [23] provide a bound on the rate at which one may approximate from outside the cone of nonnegative homogeneous polynomials (of fixed degree) by the hierarchy of spectrahedra defined in [19].

It should be emphasized that it is a difficult challenge in optimization to obtain a sequence of upper bounds converging to the global minimum and having a known estimate on the rate of convergence. So even if the convergence to the global minimum of the hierarchy of upper bounds obtained in [19] is rather slow, and even though it applies to the restricted context of "simple sets", to the best of our knowledge it provides one of the first results of this kind. A notable earlier result was obtained for polynomial optimization over the simplex, where it has been shown that brute force grid search leads to a polynomial time approximation scheme for minimizing polynomials of fixed degree [1, 4]. When minimizing over the set of grid points in the standard simplex with given denominator k, the rate of convergence is in $O(1/k)$ [1, 4] and, for quadratic polynomials (and for general polynomials having a rational minimizer), in $O(1/k^2)$ [5]. Grid search over the hypercube was also shown to have a rate of convergence in $O(1/k)$ [3] and, as we will indicate in this paper, a stronger rate of convergence in $O(1/k^2)$ can be shown. Note however that computing the best grid point in the hypercube $[0,1]^n$ with denominator k requires $O(k^n)$ computations, thus exponential in the dimension.

Contribution. As our main contribution we provide a monotone non-increasing converging sequence $(f_k^H)_{k\in\mathbb{N}}$, of upper bounds $f_k^H \geq f_{\min,\mathcal{K}}$ such that $f_k^H \to f_{\min,\mathcal{K}}$ as $k \to \infty$. The parameters f_k^H can be effectively computed when the set $\mathcal{K} \subseteq [0,1]^n$ is a "simple set" like, for example, a Euclidean ball, sphere, simplex, hypercube, or any linear transformation of them.

This "hierarchy" of upper bounds is inspired from the one defined by Lasserre in [19], but with the novelty that:

Computing the upper bounds (f_k^H) does not require solving an SDP or computing the smallest generalized eigenvalue of some pair of matrices (as is the case in [19]*). It only requires elementary calculations (but possibly many of them for good quality bounds).*

Indeed, computing the upper bound f_k^H only requires finding the minimum in a list of $O(n^k)$ scalars $(\gamma_{(\eta,\beta)})$, formed from the moments $\boldsymbol{\gamma}$ of the Lebesgue measure on the set $\mathcal{K} \subseteq [0,1]^n$ and from the coefficients (f_α) of the polynomial f to minimize. Namely:

$$f_k^H := \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}, \tag{1.3}$$

where $\mathbb{N}$ denotes the nonnegative integers, $f(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \mathbf{x}^\alpha$, $\mathbb{N}_k^{2n} = \{(\eta,\beta) \in \mathbb{N}^{2n} : |\eta+\beta| = k\}$, and the scalars

$$\gamma_{(\eta,\beta)} := \int_{\mathcal{K}} x_1^{\eta_1} \cdots x_n^{\eta_n} (1-x_1)^{\beta_1} \cdots (1-x_n)^{\beta_n} \, d\mathbf{x}, \quad (\eta,\beta) \in \mathbb{N}^{2n},$$

are available in closed-form. (Our informal notion of "simple set" therefore means that the moments $\gamma_{(\eta,\beta)}$ are known a priori.)

The upper bound (1.3) has also a simple interpretation as it reads:

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \frac{\int_{\mathcal{K}} f(\mathbf{x})\,\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\,d\mathbf{x}}{\int_{\mathcal{K}} \mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\,d\mathbf{x}} = \min_{\mu} \left\{ \int_{\mathcal{K}} f\,d\mu : \mu \in M(\mathcal{K})_k \right\}, \tag{1.4}$$

where $M(\mathcal{K})_k$ is the set of probability measures on $\mathcal{K}$, absolutely continuous with respect to the Lebesgue measure on $\mathcal{K}$, and whose density is a monomial $\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta$ with $(\eta,\beta)\in\mathbb{N}_k^{2n}$. (Such measures are in fact products of (univariate) beta distributions, see Section 4.1.) This also proves that at any point $\mathbf{a}\in[0,1]^n$ one may approximate the Dirac measure $\delta_{\mathbf{a}}$ with measures of the form $d\mu = \mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\,d\mathbf{x}$ (normalized to make then probability measures).

For the case of the hypercube $\mathcal{K}=[0,1]^n$, we analyze the rate of convergence of the bounds f_k^H and show a rate of convergence in $O(1/\sqrt{k})$ for general polynomials, and in $O(1/k)$ for quadratic polynomials (and general polynomials having a rational minimizer). As a second minor contribution, we revisit grid search over the rational points with given denominator k in the hypercube and observe that its convergence rate is in $O(1/k^2)$ (which follows as an easy application of Taylor's theorem). However as observed earlier the computation of the best grid point with denominator k requires $O(k^n)$ function evaluations while the computation of the parameter f_k^H requires only $O(n^k)$ elementary calculations.

Organization of the paper. We start with some basic facts about the bounds f_k^H in Section 2 and in Section 3 we show their convergence to the minimum of f over the set $\mathcal{K}$ (see Theorem 3.1).

In Section 4, for the case of the hypercube $\mathcal{K}=[0,1]^n$, we analyze the quality of the bounds f_k^H. We show a convergence rate in $O(1/\sqrt{k})$ for the range $f_k^H - f_{\min,\mathcal{K}}$ and a stronger convergence rate in $O(1/k)$ when the polynomial f admits a rational minimizer in $[0,1]^n$ (see Theorem 4.9). This stronger convergence rate applies in particular to quadratic polynomials (since they have a rational minimizer) and Example 4.10 shows that this bound is tight. When no rational minimizer exists the weaker rate follows using Diophantine approximations. So again the main message of this paper is that one may obtain non-trivial upper bounds with error guarantees (and converging to the global minimum) via elementary calculations and without invoking a sophisticated algorithm.

In Section 5 we revisit the simple technique which consists of evaluating the polynomial f at all rational points in $[0,1]^n$ with given denominator k. By a simple application of Taylor's theorem we can show a convergence rate in $O(1/k^2)$. However, in terms of computational complexity, the parameters f_k^H are easier to compute. Indeed, for fixed k, computing f_k^H requires $O(n^k)$ computations (similar to function evaluations), while computing the minimum of f over all grid points with given denominator k requires an exponential number k^n of function evaluations.

In Section 6 we present some additional (simple) techniques to provide a feasible point $\hat{\mathbf{x}}\in\mathcal{K}$ with value $f(\hat{\mathbf{x}}) \le f_k^H$, once the upper bound f_k^H has been computed, hence also with an error bound guarantee in the case of the box $\mathcal{K}=[0,1]^n$. This includes, in the case when f is convex, getting a feasible point using Jensen inequality (Section 6.1) and, in the general case, taking the mode $\hat{\mathbf{x}}$ of the optimal density function (i.e., its global maximizer) (see Section 6.2).

In Section 7, we present some numerical experiments, carried out on several test functions on the box $[0,1]^n$. In particular, we compare the values of the new bound f_k^H with the bound $f_{k/2}^{sos}$ (whose definition uses a sum of squares density), and we apply the proposed techniques to find a feasible point in the box. As expected the sos based bound is tighter in most cases but the bound

f_k^H can be computed for much larger values of k. Moreover, the feasible points $\hat{\mathbf{x}}$ returned by the proposed mode heuristic are often of very good quality for sufficiently large k. Finally, in Section 8 we conclude with some remarks on variants of the bound f_k^H that may offer better results in practice.

2. Notation, definitions and preliminary results

Throughout we let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbb{R}[\mathbf{x}]_d$ is the subspace of polynomials of degree at most d, and $\Sigma[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_{2d}$ is its subset of sums of squares (SOS) of degree at most $2d$.

We use the convention that $\mathbb{N}$ denotes the set of nonnegative integers, and set $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \,(=: |\alpha|) = d\}$, and similarly $\mathbb{N}^n_{\leq d} := \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \leq d\}$. The notation $\mathbf{x}^\alpha$ stands for the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, while $(\mathbf{1} - \mathbf{x})^\alpha$ stands for $(1 - x_1)^{\alpha_1} \cdots (1 - x_n)^{\alpha_n}$, $\alpha \in \mathbb{N}^n$. We will also denote $[n] = \{1, 2, \ldots, n\}$ and let $\mathbf{1}$ denote the all-ones vector (of suitable size).

One may write every polynomial $f \in \mathbb{R}[\mathbf{x}]_d$ in the monomial basis

$$\mathbf{x} \mapsto f(\mathbf{x}) \;=\; \sum_{\alpha \in \mathbb{N}^n_{\leq d}} f_\alpha \, \mathbf{x}^\alpha,$$

with vector of (finitely many) coefficients (f_α).

2.1. The bounds f_k^{sos} and f_k^H.

In [19], Lasserre introduced the parameters f_k^{sos} as upper bounds for the minimum $f_{\min,\mathcal{K}}$ of f over $\mathcal{K}$ and he proved the following result.

Theorem 2.1 (Lasserre [19])**.** *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be compact, let $f_{\min,\mathcal{K}}$ be as in (1.1), and let*

$$f_k^{sos} \;:=\; \inf_\sigma \left\{ \int_{\mathcal{K}} f(\mathbf{x})\, \sigma(\mathbf{x})\, d\mathbf{x} : \; \int_{\mathcal{K}} \sigma(\mathbf{x})\, d\mathbf{x} = 1, \sigma \in \Sigma[\mathbf{x}]_k \right\}, \quad k \in \mathbb{N}. \tag{2.1}$$

Then $f_{\min,\mathcal{K}} \;\leq\; f_{k+1}^{sos} \;\leq\; f_k^{sos}$ for all k and

$$f_{\min,\mathcal{K}} \;=\; \lim_{k\to\infty} f_k^{sos}. \tag{2.2}$$

We will also use the following important result due to Krivine [13, 14] and Handelman [10].

Theorem 2.2. *Let $\mathcal{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \, j = 1, \ldots, m\} \subset \mathbb{R}^n$ be a polytope with a nonempty interior and where each g_j is an affine polynomial, $j = 1, \ldots, m$. If $f \in \mathbb{R}[\mathbf{x}]$ is strictly positive on $\mathcal{K}$ then*

$$f(\mathbf{x}) \;=\; \sum_{\alpha \in \mathbb{N}^m} \lambda_\alpha \, g_1(\mathbf{x})^{\alpha_1} \cdots g_m(\mathbf{x})^{\alpha_m}, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \tag{2.3}$$

for finitely many positive scalars λ_α.

We will call the expression in (2.3) the *Handelman representation* of f, and call any f that allows a Handelman representation to be *of the Handelman type*. Throughout we consider the following set of polynomials:

$$\mathcal{H}_k := \left\{ p \in \mathbb{R}[\mathbf{x}] : p(\mathbf{x}) = \sum_{(\eta,\beta) \in \mathbb{N}^{2n}_k} \lambda_{\eta,\beta} \mathbf{x}^\eta (\mathbf{1} - \mathbf{x})^\beta \quad \text{where } \lambda_{\eta\beta} \geq 0 \right\}, \tag{2.4}$$

i.e., all polynomials that admit a Handelman representation of degree at most k in terms of the polynomials $x_i, 1 - x_i$ defining the hypercube $[0, 1]^n$.

Observe that any term $\mathbf{x}^\eta(\mathbf{1} - \mathbf{x})^\beta$ with degree $|\eta + \beta| < k$ also belongs to the set $\mathcal{H}_k$. This follows by iteratively applying the identity: $1 = x_i + (1 - x_i)$, which permits to rewrite $\mathbf{x}^\eta(\mathbf{1} - \mathbf{x})^\beta$

as a conic combination of terms $\mathbf{x}^{\eta'}(\mathbf{1}-\mathbf{x})^{\beta'}$ with degree $|\eta'+\beta'|=k$. The next claim follows then as a direct application.

Lemma 2.3. *We have the inclusion: $\mathcal{H}_k \subseteq \mathcal{H}_{k+1}$ for all k.*

We may now interpret the new upper bounds f_k^H from (1.3) in an analogous way as the bounds f_k^{sos} from (2.1), but where the SOS density function $\sigma \in \Sigma_k[\mathbf{x}]$ is now replaced by a density $\sigma \in \mathcal{H}_k$.

For clarity we first repeat the definition (1.3) of the parameters f_k^H below:

$$f_k^H := \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}},$$

where the scalars

$$\gamma_{(\eta,\beta)} = \int_{\mathcal{K}} \mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta d\mathbf{x} = \int_{\mathcal{K}} x_1^{\eta_1}\cdots x_n^{\eta_n}(1-x_1)^{\beta_1}\cdots(1-x_n)^{\beta_n}\, dx_1\cdots dx_n, \quad (\eta,\beta)\in\mathbb{N}^{2n},$$

denote the moments of the Lebesgue measure on the set $\mathcal{K}$. Using the fact that

$$\sum_{\alpha\in\mathbb{N}^n} f_\alpha \gamma_{(\eta+\alpha,\beta)} = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \int_{\mathcal{K}} \mathbf{x}^{\eta+\alpha}(\mathbf{1}-\mathbf{x})^\beta d\mathbf{x} = \int_{\mathcal{K}} f(\mathbf{x})\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta d\mathbf{x},$$

we can rewrite the parameter f_k^H as in (1.4):

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \frac{\int_{\mathcal{K}} f(\mathbf{x})\,\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\, d\mathbf{x}}{\int_{\mathcal{K}} \mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\, d\mathbf{x}}.$$

We now give yet another reformulation for the parameter f_k^H, where we optimize over density functions in the set $\mathcal{H}_k$, which turn out to be convex combinations of density functions of the form $\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta$ (after suitable scaling).

Lemma 2.4. *Let $\mathcal{K}\subseteq[0,1]^n$, let f be a polynomial, and consider the parameters f_k^H, $k\in\mathbb{N}$, from (1.3). Then one has:*

$$f_k^H = \inf_{\sigma\in\mathcal{H}_k} \left\{ \int_{\mathcal{K}} f(\mathbf{x})\,\sigma(\mathbf{x})\,d\mathbf{x} : \int_{\mathcal{K}} \sigma(\mathbf{x})\,d\mathbf{x} = 1 \right\} \quad \textit{for all } k\in\mathbb{N},$$

and the sequence $(f_k^H)_k$ is monotonically non-increasing: $f_{k+1}^H \le f_k^H$.

Proof. Note that, for given $k\in\mathbb{N}$,

$$\begin{aligned}
&\inf_{\sigma} \left\{ \int_{\mathcal{K}} f(\mathbf{x})\,\sigma(\mathbf{x})\,d\mathbf{x} : \int_{\mathcal{K}} \sigma(\mathbf{x})\,d\mathbf{x} = 1,\ \sigma\in\mathcal{H}_k \right\} \\
=\ &\inf_{\lambda\ge 0} \left\{ \sum_{\alpha\in\mathbb{N}_d^n} f_\alpha \left(\sum_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \lambda_{\eta\beta} \underbrace{\int_{\mathcal{K}} \mathbf{x}^{\eta+\alpha}(\mathbf{1}-\mathbf{x})^\beta\, d\mathbf{x}}_{\gamma_{(\eta+\alpha,\beta)}} \right) : \sum_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \lambda_{\eta\beta} \underbrace{\int_{\mathcal{K}} \mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\, d\mathbf{x}}_{\gamma_{(\eta,\beta)}} = 1 \right\} \\
=\ &\inf_{\lambda\ge 0} \left\{ \sum_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \lambda_{\eta\beta} \left(\sum_{\alpha\in\mathbb{N}_d^n} f_\alpha\, \gamma_{(\eta+\alpha,\beta)} \right) : \sum_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \lambda_{\eta\beta}\,\gamma_{(\eta,\beta)} = 1 \right\} \\
=\ &\min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}_d^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}} = f_k^H,
\end{aligned}$$

where we have used the fact that the penultimate optimization problem is an LP over a simplex that attains its infimum at one of the vertices. The monotonicity of the sequence $(f_k^H)_{k\in\mathbb{N}}$ now follows from Lemma 2.3. $\square$

2.2. **Calculating moments on $\mathcal{K}$.** For $\mathcal{K} \subseteq [0,1]^n$ a compact set and for every $(\eta,\beta) \in \mathbb{N}^{2n}$, we need to calculate the parameters

$$\gamma_{(\eta,\beta)} := \int_{\mathcal{K}} \mathbf{x}^{\eta}(\mathbf{1}-\mathbf{x})^{\beta}\, d\mathbf{x}, \tag{2.5}$$

in order to compute f_k^H. When $\mathcal{K}$ is arbitrary one does not know how to compute such generalized moments. But if $\mathcal{K}$ is the unit hypercube $[0,1]^n$, the simplex $\Delta := \{\mathbf{x} : \mathbf{x} \geq 0;\, \sum_{i=1}^n x_i \leq 1\}$, a Euclidean ball (or sphere), and/or their image by a linear mapping, then such moments are available in closed-form; see e.g. [19]. We give the moments for the hypercube $\mathcal{K} = [0,1]^n$, which we will treat in detail in this paper. Namely,

$$\int_{[0,1]^n} \mathbf{x}^{\eta}\,(\mathbf{1}-\mathbf{x})^{\beta}\, d\mathbf{x} = \prod_{i=1}^{n} \left(\int_0^1 x_i^{\eta_i} (1-x_i)^{\beta_i}\, dx_i \right), \qquad \text{for any } (\eta,\beta) \in \mathbb{N}^{2n},$$

and the univariate integrals may be calculated from

$$\int_0^1 t^i (1-t)^j\, dt = \frac{i!j!}{(i+j+1)!}, \qquad \text{for any } i,j \in \mathbb{N}. \tag{2.6}$$

2.3. **The complexity of computing f_k^H and f_k^{sos}.** We let N_f denote the set of indices $\alpha \in \mathbb{N}^n$ for which $f_\alpha \neq 0$; note that $|N_f| \leq \binom{n+d}{d}$ if d is the total degree of f. The computation of f_k^H is done by computing the summations:

$$\sum_{\alpha \in N_f} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}$$

for all $(\eta,\beta) \in \mathbb{N}_k^{2n}$, and taking the minimum one. (We assume that the values $\gamma_{(\eta,\beta)}$ are pre-computed for all $(\eta,\beta) \in \mathbb{N}_{k+d}^{2n}$.)

Thus, for fixed $(\eta,\beta) \in \mathbb{N}_k^{2n}$, one may first compute the inner product of the vectors with components f_α and $\gamma_{(\eta+\alpha,\beta)}$ (indexed by α). Note that these vectors are of size $|N_f|$. Since there are $\binom{2n+k-1}{k}$ pairs $(\eta,\beta) \in \mathbb{N}_k^{2n}$, the entire computation requires $(2|N_f|+1)\binom{2n+k-1}{k}$ flops[1].

As explained in [19], the computation of the upper bounds f_k^{sos} may be done by finding the smallest generalized eigenvalue λ of the system:

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

for suitable symmetric matrices A and B of order $\binom{n+k}{k}$. In particular, the rows and columns of the two matrices are indexed by $\mathbb{N}_{\leq k}^n$, and

$$A_{\alpha,\beta} = \sum_{\delta \in N_f} f_\delta \int_{\mathcal{K}} \mathbf{x}^{\alpha+\beta+\delta} d\mathbf{x}, \quad B_{\alpha,\beta} = \int_{\mathcal{K}} \mathbf{x}^{\alpha+\beta} d\mathbf{x} \quad \alpha,\beta \in \mathbb{N}_{\leq k}^n.$$

Note that the matrices A and B depend on the moments of the Lebesgue measure on $\mathcal{K}$, and that these moments may be computed beforehand, by assumption. One may compute $A_{\alpha,\beta}$ by taking

[1]We define floating point operations (flops) as in [9, p. 18]; in particular, by this definition the inner product of two n-vectors requires $2n$ flops.

the inner product of $(f_\delta)_{\delta\in N_f}$ with the vector of moments $\left(\int_{\mathcal{K}} \mathbf{x}^{\alpha+\beta+\delta} d\mathbf{x}\right)_{\delta\in N_f}$. Thus computation of the elements of A require a total of $|N_f|\left(\binom{n+k}{k}+1\right)^2$ flops.

Also note that the matrix B is a positive definite (Gram) matrix. Thus one has to solve a so-called symmetric-definite generalized eigenvalue problem, and this may be done in $14\binom{n+k}{k}^3$ flops; see e.g. [9, Section 8.7.2]. Thus one may compute f_k^{sos} in at most $14\binom{n+k}{k}^3 + |N_f|\left(\binom{n+k}{k}+1\right)^2$ flops.

2.4. **An illustrating example.** We give an example to illustrate the behaviour of the bounds f_k^{sos} and f_k^H. More examples will be given in Section 7.

Example 2.5. *As an example we consider the bivariate Styblinski-Tang function*

$$f(x_1,x_2) = \sum_{i=1}^{2} \frac{1}{2}(10x_i-5)^4 - 8(10x_i-5)^2 + \frac{5}{2}(10x_i-5)$$

over the square $\mathcal{K}=[0,1]^2$, with minimum $f_{\min,\mathcal{K}} \approx -78.33198$ and minimizer

$$\mathbf{x}^* \approx (0.20906466, 0.20906466).$$

Using a SOS density function, the upper bound of degree 2 is $f_1^{sos} = -12.9249$, and the corresponding optimal SOS density of degree 2 is (roughly)

$$\sigma(x_1,x_2) = (1.9169 - 1.005x_1 - 1.005x_2)^2.$$

Using a Handelman-type density function, the upper bound of degree 2 is $f_2^H = -17.3810$, with corresponding optimal density

$$\sigma(x_1,x_2) = 6x_2(1-x_2).$$

On the other hand, if we consider densities of degree 6 then we get $f_3^{sos} = -34.403$ and $f_6^H = -31.429$.

Thus there is no general ordering between the bounds f_k^{sos} and f_{2k}^H. Having said that, we will show in Section 7 that, for most of the examples we have considered, one has $f_k^{sos} \le f_{2k}^H$ for all k, as one may expect from the relative computational efforts. As a final illustration, Figure 1 shows the plot and contour plot of the Handelman-type density corresponding to the bound $f_{50}^H = -60.536$ (i.e. degree 50).

The figure illustrates the earlier assertion that the optimal density approximates the Dirac delta measure at the minimizer $\mathbf{x}^* \approx (0.20906466, 0.20906466)$. Indeed, it is clear from the contour plot that the mode of the optimal density is close to $\mathbf{x}^*$.

3. Convergence proof for the bounds f_k^H on $\mathcal{K}\subseteq[0,1]^n$

In this section we prove the convergence of the sequence $(f_k^H)_{k\in\mathbb{N}}$ to the minimum of f over any compact set $\mathcal{K}\subseteq[0,1]^n$.

Theorem 3.1. *Let $\mathcal{K}\subseteq[0,1]^n$, let $f\in\mathbb{R}[\mathbf{x}]_d$ and let $\gamma_{(\eta,\beta)}$ be as in (2.5). Define as in (1.3) the parameters*

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}_{\le d}^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}, \qquad \forall\, k\in\mathbb{N}. \tag{3.1}$$

Then, $f_{\min,\mathcal{K}} = \lim_{k\to\infty} f_k^H$.

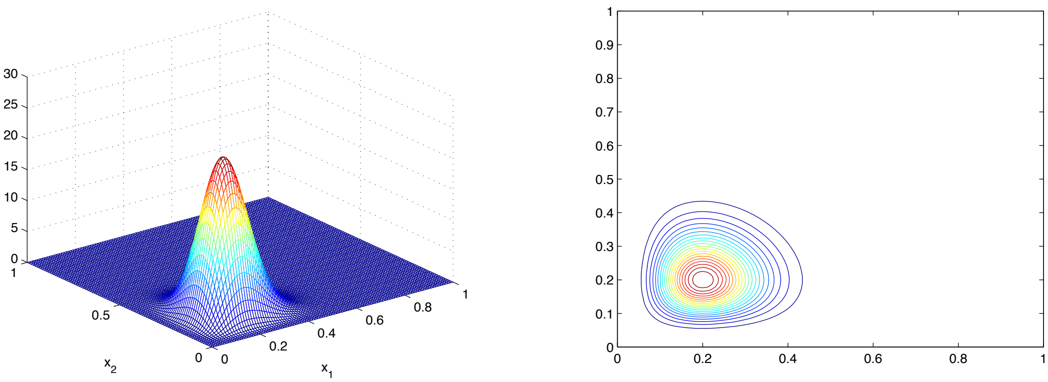

FIGURE 1. Optimal Handelman-type density σ of degree 50 on $[0,1]^2$ for the bivariate Styblinski-Tang function.

Proof. As in (1.2), let f_k^{sos} denote the bound obtained by searching over an SOS density σ of degree at most $2k$:

$$f_k^{sos} = \min \int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} \;\text{ such that }\; \int_{\mathcal{K}} \sigma(\mathbf{x})d\mathbf{x} = 1,\; \sigma \in \Sigma[\mathbf{x}]_k.$$

Also recall from Lemma 2.4 that

$$f_k^{H} = \min \int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} \;\text{ such that }\; \int_{\mathcal{K}} \sigma(\mathbf{x})d\mathbf{x} = 1,\; \sigma \in \mathcal{H}_k.$$

By Lemma 2.4, the sequence (f_k^H) is monotone non-increasing, with $f_{\min,\mathcal{K}} \le f_k^H$ for all k. Hence it has a limit which is at least $f_{\min,\mathcal{K}}$, we now show that the limit is equal to $f_{\min,\mathcal{K}}$.

To this end, let $\epsilon > 0$. As the sequence (f_k^{sos}) converges to $f_{\min,\mathcal{K}}$ (Theorem 2.1), there exists an integer k such that

$$f_{\min,\mathcal{K}} \le f_k^{sos} \le f_{\min,\mathcal{K}} + \epsilon.$$

Next, there exists a polynomial $\sigma \in \Sigma_k$ such that $\int_{\mathcal{K}} \sigma(\mathbf{x})d\mathbf{x} = 1$ and

$$f_k^{sos} \le \int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} \le f_k^{sos} + \epsilon.$$

Define now the polynomial $\hat{\sigma}(\mathbf{x}) = \sigma(\mathbf{x}) + \epsilon$. Then $\hat{\sigma}$ is strictly positive on $[0,1]^n$ and thus, by Theorem 2.2 applied to the hypercube $[0,1]^n$, $\hat{\sigma} \in \mathcal{H}_{j_k}$ for some integer j_k. Observe that

$$\int_{\mathcal{K}} \hat{\sigma}(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{K}} (\sigma(\mathbf{x}) + \epsilon) d\mathbf{x} \geq \int_{\mathcal{K}} \sigma(\mathbf{x}) d\mathbf{x} = 1.$$

Hence we obtain:

$$f_{j_k}^H - f_{\min,\mathcal{K}} \leq \frac{\int_{\mathcal{K}} f(\mathbf{x})\hat{\sigma}(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{K}} \hat{\sigma}(\mathbf{x})d\mathbf{x}} - f_{\min,\mathcal{K}} = \frac{\int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})\hat{\sigma}(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{K}} \hat{\sigma}(\mathbf{x})d\mathbf{x}} \leq \int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})\hat{\sigma}(\mathbf{x})d\mathbf{x}.$$

The right most term is equal to

$$\int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})\sigma(\mathbf{x})d\mathbf{x} + \epsilon \int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})d\mathbf{x} = \int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} - f_{\min,\mathcal{K}} + \epsilon \int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})d\mathbf{x},$$

where we used the fact that $\int_{\mathcal{K}} \sigma(\mathbf{x})d\mathbf{x} = 1$. Finally, combining with the fact that $\int_{\mathcal{K}} f(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x} \leq f_k^{sos} + \epsilon \leq f_{\min,\mathcal{K}} + 2\epsilon$, we can derive that

$$f_{j_k}^H - f_{\min,\mathcal{K}} \leq \epsilon \left(2 + \int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})d\mathbf{x}\right) = \epsilon C,$$

where $C := 2 + \int_{\mathcal{K}} (f(\mathbf{x}) - f_{\min,\mathcal{K}})d\mathbf{x}$ is a constant. This concludes the proof. □

Note that, in the proof, it was essential to have $\hat{\sigma}$ strictly positive on all of $[0,1]^n$, for the application of Handelman's theorem. The fact that $\hat{\sigma}(\mathbf{x}) = \sigma(\mathbf{x}) + \epsilon$ with σ SOS and $\epsilon > 0$ guaranteed this strict positivity.

4. Bounding the rate of convergence for the parameters f_k^H on $\mathcal{K} = [0,1]^n$

In this section we analyze the rate of convergence of the bounds f_k^H for the hypercube $\mathcal{K} = [0,1]^n$. We prove a convergence rate in $O(1/\sqrt{k})$ for the range $f_k^H - f_{\min,\mathcal{K}}$ in general, and a stronger convergence rate in $O(1/k)$ when f has a rational global minimizer in $[0,1]^n$, which is the case, for instance, when f is quadratic.

Our main tool will be exploiting some properties of the moments $\gamma_{(\eta,\beta)}$ which, as we recall below, arise from the moments of the beta distribution.

4.1. **Properties of the beta distribution.** By definition, a random variable $X \in [0,1]$ has the beta distribution with shape parameters $a > 0$ and $b > 0$, which is denoted by $X \sim \mathit{beta}(a,b)$, if its probability density function is given by

$$y \mapsto \frac{y^{a-1}(1-y)^{b-1}}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}.$$

If $a > 1$ and $b > 1$, then the (unique) mode of the distribution (i.e., the maximizer of the density function) is

$$y = (a-1)/(a+b-2). \tag{4.1}$$

Moreover, the k-th moment of X is given by

$$\mathbb{E}(X^k) = \frac{a(a+1)\cdots(a+k-1)}{(a+b)(a+b+1)\cdots(a+b+k-1)}, \quad (k = 1,2,3,\ldots) \tag{4.2}$$

(see, e.g, [12, Chapter 24]; this also follows using (2.6)).

In what follows we will consider families of random random variables with the beta distribution of the form $X \sim beta(ar, br)$, where a and b are positive real numbers and $r \geq 1$ is an integer. By (4.2), any such random variable has mean

$$\mathbb{E}(X) = \frac{ar}{ar+br} = \frac{a}{a+b}.$$

In Lemma 4.2 below we show how the moments of such random variables relate to powers of the mean. The proof relies on the following technical lemma.

Lemma 4.1. *Let k be a positive integer. There exists a constant $C_k > 0$ (depending only on k) for which the following relation holds:*

$$\frac{rp(rp+1)\cdots(rp+k-1)}{rq(rq+1)\cdots(rq+k-1)} - \frac{p^k}{q^k} \leq \frac{C_k}{r} \tag{4.3}$$

for all integers $r \geq 1$, and real numbers $0 < p < q$.

Proof. Consider the univariate polynomial $\phi(t) = (t+1)\cdots(t+k-1) = \sum_{i=0}^{k-1} a_i t^i$, where the scalars $a_i > 0$ depend only on k and $a_{k-1} = 1$. Denote by Δ the left hand side in (4.3), which can be written as $\Delta = N/D$, where we set

$$N := rpq^k\phi(rp) - rqp^k\phi(rq), \quad D := rq^{k+1}\phi(rq).$$

We first work out the term N:

$$N = rpq\left(\sum_{i=0}^{k-2} a_i r^i p^i q^{k-1} - \sum_{i=0}^{k-2} a_i r^i q^i p^{k-1}\right) = rpq\sum_{i=0}^{k-2} a_i r^i p^i q^i (q^{k-1-i} - p^{k-1-i}).$$

Write: $q^{k-1-i} - p^{k-1-i} = (q-p)\sum_{j=0}^{k-2-i} q^j p^{k-2-i-j} \leq (q-p)q^{k-2-i}(k-1-i)$, where we use the fact that $p < q$. This implies:

$$N \leq rpq(q-p)\sum_{i=0}^{k-2} a_i r^i p^i q^{k-2}(k-1-i) = rpq^{k-1}(q-p)\sum_{i=0}^{k-2} a_i(k-1-i)r^i p^i =: N'.$$

Thus we get:

$$\Delta \leq \frac{N'}{D} = \frac{p(q-p)}{q^2} \cdot \frac{\sum_{i=0}^{k-2} a_i(k-1-i)r^i p^i}{\phi(rq)}.$$

The first factor is at most 1, since one has: $p(q-p) \leq q^2$, as $q^2 - p(q-p) = (q-p)^2 + pq$. Second, we bound the sum $\sum_{i=0}^{k-2} a_i(k-1-i)r^i p^i$ in terms of $\phi(rq) = \sum_{j=0}^{k-1} a_j r^j q^j$. Namely, define the constant

$$C_k := \max_{0\leq i\leq k-2} \frac{a_i(k-1-i)}{a_{i+1}},$$

which depends only on k. We show that

$$a_i(k-1-i)r^i p^i \leq \frac{C_k}{r}.$$

Indeed, for each $0 \leq i \leq k-2$, using $p^i \leq q^{i+1}$ and the definition of C_k, we get:

$$r \cdot a_i(k-1-i)r^i p^i \leq a_i(k-1-i)r^{i+1}q^{i+1} \leq C_k a_{i+1} r^{i+1} q^{i+1}.$$

Summing over $i = 0, 1, \ldots, k-2$ gives:

$$r\sum_{i=0}^{k-2} a_i(k-1-i)r^i p^i \leq C_k \sum_{i=0}^{k-2} a_{i+1} r^{i+1} q^{i+1} \leq C_k\phi(rq),$$

and thus

$$\Delta \leq \frac{N'}{D} \leq \frac{C_k}{r}$$

as desired. □

Lemma 4.2. *For any integer $k \geq 1$, there exists a constant $C'_k > 0$ (depending only on k) for which the following holds:*

$$\left|\mathbb{E}(X^k) - (\mathbb{E}(X))^k\right| \leq \frac{C'_k}{r},$$

for all integers $r \geq 1$, real numbers $a, b > 0$, and where $X \sim beta(ar, br)$.

Proof. Directly using (4.2), $\mathbb{E}(X) = \frac{a}{a+b}$, and Lemma 4.1 applied to $p = a$ and $q = a + b$. □

Now we consider i.i.d. random variables $X_1, \ldots, X_n$ such that

$$X_i \sim beta(a_i r, b_i r) \quad a_i, b_i > 0 \; (i \in [n]),\; r \geq 1, r \in \mathbb{N}, \tag{4.4}$$

and denote $X = (X_1, \ldots, X_n)$. For given $\alpha \in \mathbb{N}^n$, we denote $X^\alpha = \prod_{i=1}^n X_i^{\alpha_i}$. Since the random variables X_i are independent we have $\mathbb{E}(X^\alpha) = \prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i})$ and, for a polynomial $f = \sum f_\alpha \mathbf{x}^\alpha$, the expected value of $f(X) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha X^\alpha$ is given by

$$\mathbb{E}(f(X)) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \mathbb{E}(X^\alpha) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i}). \tag{4.5}$$

Recall that the explicit value of $\mathbb{E}(X_i^{\alpha_i})$ is given by (4.2). The next result relates $\mathbb{E}(f(X))$ (the expected value of $f(X)$) and $f(\mathbb{E}(X))$ (the value of f evaluated at the mean of X).

Lemma 4.3. *Let $f(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \mathbf{x}^\alpha$ and $X = (X_1, \ldots, X_n)$, where the i.i.d. random variables X_i $(i \in [n])$ are as in (4.4). Then there is a constant $\hat{C}_f > 0$ (depending on f only) such that*

$$|\mathbb{E}(f(X)) - f(\mathbb{E}(X))| \leq \frac{\hat{C}_f}{r}.$$

Proof. We have

$$\mathbb{E}(f(X)) - f(\mathbb{E}(X)) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \left(\prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i}) - \prod_{i=1}^n (\mathbb{E}(X_i))^{\alpha_i} \right).$$

By the identity:

$$\prod_{i=1}^n x_i - \prod_{i=1}^n y_i = \sum_{i=1}^n \left[(x_i - y_i) \prod_{j=1}^{i-1} y_j \prod_{j=i+1}^n x_j \right] \qquad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^n), \tag{4.6}$$

one has

$$\prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i}) - \prod_{i=1}^n (\mathbb{E}(X_i))^{\alpha_i} = \sum_{i=1}^n \left((\mathbb{E}(X_i^{\alpha_i}) - (\mathbb{E}(X_i))^{\alpha_i}) \prod_{j=1}^{i-1} (\mathbb{E}(X_j))^{\alpha_j} \prod_{j=i+1}^n \mathbb{E}(X_j^{\alpha_j}) \right).$$

Since $\mathbb{E}(X_i) \in [0,1]$ and $\mathbb{E}(X_i^{\alpha_i}) \in [0,1]$ for any $i \in [n]$, it follows that

$$\begin{aligned} |\mathbb{E}(f(X)) - f(\mathbb{E}(X))| &\leq \sum_{\alpha \in \mathbb{N}^n} |f_\alpha| \sum_{i=1}^n |\mathbb{E}(X_i^{\alpha_i}) - (\mathbb{E}(X_i))^{\alpha_i}| \\ &\leq \sum_{\alpha \in \mathbb{N}^n} |f_\alpha| \sum_{i=1}^n \frac{C'_{\alpha_i}}{r}, \end{aligned}$$

where the second inequality is from Lemma 4.2, and the constant $C'_{\alpha_i} > 0$ only depends on α_i. Setting $\hat{C}_f := \sum_{\alpha \in \mathbb{N}^n} |f_\alpha| \sum_{i=1}^n C'_{\alpha_i}$ concludes the proof. □

4.2. **Proof of the convergence rate.** Let $\mathbf{x}^*$ be a global minimizer of f in $[0,1]^n$. Our objective is to analyze the rate of convergence of the sequence $(f_k^H - f(\mathbf{x}^*))_k$. Our strategy is to define suitable shape parameters η_i^*, β_i^* from the components x_i^* of the global minimizer $\mathbf{x}^*$ so that, if we choose a vector $X = (X_1, \ldots, X_n)$ of i.i.d. random variables with $X_i \sim beta(\eta_i^*, \beta_i^*)$, then (roughly) $\mathbb{E}(X) \approx \mathbf{x}^*$ and $\mathbb{E}(f(X)) \approx f_k^H$ (so that we can use the result of Lemma 4.3 to estimate $f_k^H - f(\mathbf{x}^*)$.

In a first step we indicate how to define the shape parameters η_i^*, β_i^*. For any given integer $r \geq 1$ we will select them of the form $\eta_i^* = ra_i$, $\beta_i^* = rb_i$, where a_i, b_i are constructed from the coordinates of $\mathbf{x}^*$. As we want η_i^*, β_i^* to be integer valued we need to discuss whether a coordinate x_i is rational or not, and to deal with irrational coordinates we will use the following result about Diophantine approximations.

Theorem 4.4 (Dirichlet's theorem)**.** *(See e.g.* [24, Chapter 6.1]*) Consider a real number $x \in \mathbb{R}$ and $0 < \epsilon \leq 1$. Then there exist integers p and q satisfying*

$$\left| x - \frac{p}{q} \right| < \frac{\epsilon}{q} \quad \text{and} \quad 1 \leq q \leq \frac{1}{\epsilon}.$$

If $x \in (0,1)$, then one may moreover assume $0 \leq p \leq q$.

Definition 4.5 (Shape parameters for rational components)**.** *Fix an integer $r \geq 1$. For rational coordinates $x_i^* \in \mathbb{Q}$ define η_i^*, β_i^* as follows:*

(i) *If $x_i^* = 0$ then set $\eta_i^* = 1$ and $\beta_i^* = r$.*
(ii) *If $x_i^* = 1$ then set $\eta_i^* = r$ and $\beta_i^* = 1$.*
(iii) *If $x_i^* \in \mathbb{Q} \setminus \{0,1\}$ then write $x_i^* = p_i/q_i$ where $1 \leq p_i < q_i$ are integers, and set $\eta_i^* = rp_i$ and $\beta_i^* = r(q_i - p_i)$.*

Definition 4.6 (Shape parameters for irrational components)**.** *Fix an integer $r \geq 1$. For each irrational coordinate $x_i^* \in \mathbb{R} \setminus \mathbb{Q}$, apply Theorem 4.4 with $\epsilon = 1/r$ to obtain integers p_i, q_i satisfying*

$$\left| x_i^* - \frac{p_i}{q_i} \right| < \frac{1}{rq_i}, \ 0 \leq p_i \leq q_i \leq r, \ \text{and} \ 1 \leq q_i.$$

Define the sets $I_0 = \{i \in [n] : x_i^* \in \mathbb{R} \setminus \mathbb{Q},\ p_i = 0\}$, $I_1 = \{i \in [n] : x_i^* \in \mathbb{R} \setminus \mathbb{Q},\ p_i = q_i\}$, and $I = \{i \in [n] : x_i^* \in \mathbb{R} \setminus \mathbb{Q},\ 1 \leq p_i < q_i\}$, and define η_i^*, β_i^* as follows:

(iv) *If $i \in I_0$ then set $\eta_i^* = 1$ and $\beta_i^* = r$.*
(v) *If $i \in I_1$ then set $\eta_i^* = r$ and $\beta_i^* = 1$.*
(vi) *If $i \in I$ then set $\eta_i^* = rp_i$ and $\beta_i^* = r(q_i - p_i)$.*

As above consider i.i.d. $X = (X_1, \dots, X_n)$, where $X_i \sim beta(\eta_i^*, \beta_i^*)$. Then, by construction, for all $i \in [n]$, one has

$$\mathbb{E}(X_i) = \frac{\eta_i^*}{\eta_i^* + \beta_i^*} = \begin{cases} \frac{1}{r+1} & \text{in cases (i), (iv),} \\ \frac{r}{r+1} & \text{in cases (ii), (v),} \\ \frac{p_i}{q_i} & \text{in cases (iii), (vi).} \end{cases}$$

One can verify that in all cases one has

$$|\mathbb{E}(X_i) - x_i^*| \leq 1/r \quad \text{for all } i \in [n]. \tag{4.7}$$

Observe morever that, again by construction,

$$\mathbb{E}(f(X)) = \frac{\int_{[0,1]^n} f(\mathbf{x})\mathbf{x}^{\eta^*-\mathbf{1}}(\mathbf{1}-\mathbf{x})^{\beta^*-\mathbf{1}} d\mathbf{x}}{\int_{[0,1]^n} \mathbf{x}^{\eta^*-\mathbf{1}}(\mathbf{1}-\mathbf{x})^{\beta^*-\mathbf{1}} d\mathbf{x}} \geq f_{k_r}^H \geq f(\mathbf{x}^*) \ , \tag{4.8}$$

where we let $\mathbf{1}$ denote the all-ones vector and we define the parameter

$$k_r := \sum_{i=1}^n (\eta_i^* - 1 + \beta_i^* - 1). \tag{4.9}$$

We will use the following estimate on the parameter k_r.

Lemma 4.7. *Consider the parameter* $k_r = \sum_{i=1}^n (\eta_i^* - 1 + \beta_i^* - 1)$ *and* $J = \{i \in [n] : x_i^* \in \mathbb{Q} \setminus \{0,1\}\}$. *Then the following holds:*

(a) *If* $\mathbf{x}^* \in \mathbb{Q}^n$ *then* $k_r \leq ar$ *for all* $r \geq 1$, *where* $a > 0$ *is a constant (not depending on* r*).*
(b) *If* $\mathbf{x}^* \notin \mathbb{Q}^n$ *then* $k_r \leq a'r^2$ *for all* $r \geq 1$, *where* $a' > 0$ *is a constant (not depending on* r*).*
(c) *For* $r = 1$, *we have that* $k_1 = \sum_{i \in J} q_i - 2|J|$.

Proof. By construction, $\eta_i^* + \beta_i^* - 2 = rq_i - 2$ for each $i \in I \cup J$, and $\eta_i^* + \beta_i^* - 2 = r - 1$ otherwise. From this one gets $k_r = r(\sum_{i \in I \cup J} q_i + n - |I \cup J|) - n - |I \cup J| =: ar - b$, after setting $b := n + |I \cup J|$ and $a := \sum_{i \in I \cup J} q_i + n - |I \cup J|$, so that $a, b \geq 0$. Thus, $k_r \leq ar$ holds.

Next, note that $q_i \leq r$ for each $i \in I$, while q_i does not depend on r for $i \in J$ (since then $x_i^* = p_i/q_i$). Hence, in case (a), $I = \emptyset$ and the constant a does not depend on r. In case (b), we obtain: $a \leq r|I| + \sum_{i \in J} q_i + n - |I \cup J| \leq a'r$, after setting $a' := |I| + \sum_{i \in J} q_i + n - |I \cup J|$, which is thus a constant not depending on r. Then, $k_r \leq ar \leq a'r^2$.

In the case $r = 1$, the set I is empty and thus $k_1 = \sum_{i \in J} q_i - 2|J|$, showing (c). ∎

We can now prove the following upper bound for the range $\mathbb{E}(f(X)) - f(\mathbf{x}^*)$ (thus also for the range $f_{k_r}^H - f(\mathbf{x}^*)$) which will be crucial for establishing the rate of convergence of the parameters f_k^H.

Theorem 4.8. *Given a polynomial* f *of total degree* d, *consider a global minimizer* $\mathbf{x}^*$ *of* f *in* $[0,1]^n$. *Let* r *be a positive integer. For any* $x_i^* \in [0,1]$ *(*$i \in [n]$*), consider the parameters* η_i^*, β_i^* *as in Definitions 4.5 and 4.6, and i.i.d. random variables* $X_i \sim beta(\eta_i^*, \beta_i^*)$. *Then there exists a constant* $C_f > 0$ *(depending only on* f*) such that*

$$f_{k_r}^H - f(\mathbf{x}^*) \leq \mathbb{E}(f(X)) - f(\mathbf{x}^*) \leq \frac{C_f}{r},$$

where k_r *is as in (4.9).*

Proof. The leftmost inequality follows using (4.8), we show the rightmost one. By Lemma 4.3 one has:

$$\begin{aligned} \mathbb{E}(f(X)) - f(\mathbf{x}^*) &= \mathbb{E}(f(X)) - f(\mathbb{E}(X)) + f(\mathbb{E}(X)) - f(\mathbf{x}^*) \\ &\leq \hat{C}_f/r + f(\mathbb{E}(X)) - f(\mathbf{x}^*), \end{aligned}$$

where $\hat{C}_f > 0$ is a constant that depends on f only. Thus we need only bound $f(\mathbb{E}(X)) - f(\mathbf{x}^*)$. To this end, note that

$$f(\mathbb{E}(X)) - f(\mathbf{x}^*) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \left(\prod_{i=1}^n \mathbb{E}(X_i)^{\alpha_i} - \prod_{i=1}^n (x_i^*)^{\alpha_i} \right).$$

Using again the identity (4.6) one has

$$\left| \left(\prod_{i=1}^n \mathbb{E}(X_i)^{\alpha_i} - \prod_{i=1}^n (x_i^*)^{\alpha_i} \right) \right| \leq \sum_{i:\alpha_i>0} |\mathbb{E}(X_i) - x_i^*| \leq \frac{d}{r},$$

where d is the degree of f, and we have used $|\mathbb{E}(X_i) - x_i^*| \leq 1/r$, $x_i^* \in [0,1]$ and $\mathbb{E}(X_i) \in [0,1]$ for all $i \in [n]$. Setting

$$C_f = \hat{C}_f + d \sum_{\alpha\in\mathbb{N}^n} |f_\alpha|$$

completes the proof. □

Finally we can now show the following for the rate of convergence of the sequence f_k^H, which is our main result.

Theorem 4.9. *Let f be a polynomial, let $\mathbf{x}^*$ be a global minimizer of f in $[0,1]^n$, and consider as before the parameters*

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \frac{\int_{[0,1]^n} f(\mathbf{x})\, \mathbf{x}^\eta (\mathbf{1}-\mathbf{x})^\beta \, d\mathbf{x}}{\int_{[0,1]^n} \mathbf{x}^\eta (\mathbf{1}-\mathbf{x})^\beta \, d\mathbf{x}} \quad (k = 1, 2, \ldots).$$

There exists a constant M_f (depending only on f) such that

$$f_k^H - f(\mathbf{x}^*) \leq \frac{M_f}{\sqrt{k}} \quad \text{for all } k \geq k_1, \tag{4.10}$$

where $k_1 = \sum_{i\in J} q_i - 2|J|$, with $J = \{i \in [n] : x_i^* \in \mathbb{Q} \setminus \{0,1\}\}$ and $x_i^* = p_i/q_i$ for integers $1 \leq p_i < q_i$ if $i \in J$. Moreover, if f has at least one rational global minimizer $\mathbf{x}^*$, then there exists a constant M_f' (depending only on f) such that

$$f_k^H - f(\mathbf{x}^*) \leq \frac{M_f'}{k} \quad \text{for all } k \geq k_1. \tag{4.11}$$

In particular, the convergence rate is in $O(1/k)$ when f is a quadratic polynomial.

Proof. Consider an arbitrary integer $k \geq k_1$. Let $r \geq 1$ be the largest integer for which $k \geq k_r$. Then we have $k_r \leq k < k_{r+1}$. As $k_r \leq k$, we have the inequality $f_k^H \leq f_{k_r}^H$ and thus, by Theorem 4.8, we obtain

$$f_k^H - f(\mathbf{x}^*) \leq f_{k_r}^H - f(\mathbf{x}^*) \leq \frac{C_f}{r},$$

where the constant C_f depends only on f.

If $\mathbf{x}^* \in \mathbb{Q}^n$ then, by Lemma 4.7 (a), $k_{r+1} \leq a(r+1) \leq 2ar$. This implies $k \leq k_{r+1} \leq 2ar$, where the constant a does not depend on r. Thus,

$$f_k^H - f(\mathbf{x}^*) \leq \frac{C_f}{r} \leq \frac{2aC_f}{k} = \frac{M_f}{k},$$

where the constant $M_f = 2aC_f$ depends only on f. This shows (4.11).

If $\mathbf{x}^* \notin \mathbb{Q}^n$ then, by Lemma 4.7 (b), $k_{r+1} \leq a'(r+1)^2 \leq 4a'r^2$. This implies $k \leq k_{r+1} \leq 4a'r^2$ and thus $\frac{1}{r} \leq \frac{2\sqrt{a'}}{\sqrt{k}}$, where the constant a' does not depend on r. Therefore,

$$f_k^H - f(\mathbf{x}^*) \leq \frac{C_f}{r} \leq \frac{2\sqrt{a'}C_f}{\sqrt{k}} = \frac{M'_f}{\sqrt{k}},$$

where the constant $M'_f = 2\sqrt{a'}C_f$ depends only on f. This shows (4.10).

Finally, if f is quadratic then, by a result of Vavasis [25], f has a rational minimizer over the hypercube and thus the rate of convergence is $O(1/k)$. □

Note that the inequalities (4.10) and (4.11) hold for all $k \geq k_1$, where k_1 depends only on the rational components in $(0,1)$ of the minimizer $\mathbf{x}^*$. The constant k_1 can be in $O(1)$, e.g., when all but $O(1)$ of these rational components have a small denominator (say, equal to 2). Thus we can, for some problem classes, get a bound with an error estimate in polynomial time.

Example 4.10. *Consider the polynomial $f = \sum_{i=1}^n x_i$ and the set $\mathcal{K} = [0,1]^n$. Then $f_{min,\mathcal{K}} = 0$ is attained at $\mathbf{x}^* = 0$. Using the relations* (2.5)*,* (2.6) *and* (3.1)*, it follows that*

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{i=1}^n \frac{\eta_i + 1}{\eta_i + \beta_i + 2}.$$

Since $\eta_i + \beta_i \leq k$ and $\eta_i \geq 0$ (for any $i \in [n]$), we have $f_k^H \geq \frac{n}{k+2}$.
By this example, there does not exist any $\delta > 0$ such that, for any f, $f_k^H - f_{\min,\mathcal{K}} = O(1/k^{1+\delta})$. Therefore, when a rational minimizer exists, the convergence rate from Theorem 4.9 in $O(1/k)$ for f_k^H is tight.

5. Bounding the rate of convergence for grid search over $\mathcal{K} = [0,1]^n$

As an alternative to computing f_k^H on $\mathcal{K} = Q := [0,1]^n$, one may minimize f over the regular grid:

$$Q(k) := \{\mathbf{x} \in Q = [0,1]^n \mid k\mathbf{x} \in \mathbb{N}^n\},$$

i.e., the set of rational points in $[0,1]^n$ with denominator k. Thus we get the upper bound

$$f_{\min,Q(k)} := \min_{\mathbf{x}\in Q(k)} f(x) \geq f_{\min,Q} \quad k = 1,2,\ldots$$

De Klerk and Laurent [3] showed a rate of convergence in $O(1/k)$ for this sequence of upper bounds:

$$f_{\min,Q(k)} - f_{\min,Q} \leq \frac{L(f)}{k}\binom{d+1}{3}n^d \text{ for any } k \geq d, \tag{5.1}$$

where d is the degree of f and $L(f)$ is the constant

$$L(f) = \max_\alpha |f_\alpha| \frac{\prod_{i=1}^n \alpha_i!}{|\alpha|!}.$$

We can in fact show a stronger convergence rate in $O(1/k^2)$.

Theorem 5.1. *Let f be a polynomial and let $\mathbf{x}^*$ be a global minimizer of f in $[0,1]^n$. Then there exists a constant C_f (depending on f) such that*

$$f_{\min,Q(k)} - f(\mathbf{x}^*) \le \frac{C_f}{k^2} \quad \text{for all } k \ge 1.$$

Proof. Fix $k \ge 1$. By looking at the grid point in $Q(k)$ closest to $\mathbf{x}^*$, there exists $\mathbf{h} \in [0,1]^n$ such that $\mathbf{x}^* + \mathbf{h} \in Q(k)$ and $\|\mathbf{h}\| \le \frac{\sqrt{n}}{k}$. Then, by Taylor's theorem, we have that

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \mathbf{h}^T \nabla f(\mathbf{x}^*) + \frac{1}{2}\mathbf{h}^T \nabla^2 f(\zeta)\mathbf{h}, \tag{5.2}$$

for some point ζ lying in the segment $[\mathbf{x}^*, \mathbf{x}^* + \mathbf{h}] \subseteq [0,1]^n$.

Assume first that the global minimizer $\mathbf{x}^*$ lies in the interior of $[0,1]^n$. Then $\nabla f(\mathbf{x}^*) = 0$ and thus

$$f_{\min,Q(k)} - f(\mathbf{x}^*) \le f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \le C\|\mathbf{h}\|^2 \le \frac{nC}{k^2},$$

after setting $C := \max_{\zeta \in [0,1]^n} \|\nabla^2 f(\zeta)\|/2$.

Assume now that $\mathbf{x}^*$ lies on the boundary of $[0,1]^n$ and let I_0 (resp., I_1, I) denote the set of indices $i \in [n]$ for which $x_i^* = 0$ (resp., $x_i^* = 1$, $x_i^* \in (0,1)$). Define the polynomial $g(y) = f(y, 0, \ldots, 0, 1, \ldots, 1)$ (with 0 at the positions $i \in I_0$ and 1 at the positions $i \in I_1$) in the variable $y \in \mathbb{R}^{|I|}$. Then $\mathbf{x}_I^* = (x_i^*)_{i \in I}$ is a global minimizer of g over $[0,1]^{|I|}$ which lies in the interior. So we may apply the preceding reasoning to the polynomial g and conclude that $g_{\min,Q(k)} - g(\mathbf{x}_I^*) \le \frac{C'}{k^2}$ for some constant C' (depending on g and thus on f). As $f_{\min,Q(k)} \le g_{\min,Q(k)}$ and $f(\mathbf{x}^*) = g(\mathbf{x}_I^*)$ the result follows. □

Therefore the bounds $f_{\min,Q(k)}$ obtained through grid search have a faster convergence rate than the bounds f_k^H. However, for any fixed value of k, for the bound f_k^H one needs a polynomial number $O(n^k)$ of computations (similar to function evaluations), while computing the bound $f_{\min,Q(k)}$ requires an exponential number k^n of function evaluations. Hence the 'measure-based' guided search producing the bounds f_k^H is superior to the brute force grid search technique in terms of complexity.

6. Obtaining feasible points $\mathbf{x}$ with $f(\mathbf{x}) \le f_k^H$

In this section we describe how to generate a point $\mathbf{x} \in \mathcal{K} \subseteq [0,1]^n$ such that $f(\mathbf{x}) \le f_k^H$ (or such that $f(\mathbf{x}) \le f_k^H + \epsilon$ for some small $\epsilon > 0$).

We will discuss in turn:

- the convex case (and related cases), and
- the general case.

6.1. The convex case (and related cases): using the Jensen inequality.

Our main tool for treating the convex case (and related cases) will be the Jensen inequality.

Lemma 6.1 (Jensen inequality)**.** *If $\mathcal{C} \subseteq \mathbb{R}^n$ is convex, $\phi : \mathcal{C} \to \mathbb{R}$ is a convex function, and $X \in \mathcal{C}$ a random variable, then*

$$\phi(\mathbb{E}(X)) \le \mathbb{E}(\phi(X)).$$

Theorem 6.2. *Assume that $\mathcal{K} \subseteq [0,1]^n$ is closed and convex, and $(\eta, \beta) \in \mathbb{N}_k^{2n}$ is such that*

$$f_k^H = \frac{\int_{\mathcal{K}} f(\mathbf{x})\, \mathbf{x}^\eta (\mathbf{1} - \mathbf{x})^\beta \, d\mathbf{x}}{\int_{\mathcal{K}} \mathbf{x}^\eta (\mathbf{1} - \mathbf{x})^\beta \, d\mathbf{x}}.$$

Let $X = (X_1, \ldots, X_n)$ be a vector of random variables with $X_i \sim beta(\eta_i + 1, \beta_i + 1)$ $(i \in [n])$.

Then one has $f(\mathbb{E}(X)) \le f_k^H$ in the following cases:

(1) *f is convex;*
(2) *f has only nonnegative coefficients;*
(3) *f is square-free, i.e., $f(\mathbf{x}) = \sum_{\alpha \in \{0,1\}^n} f_\alpha \mathbf{x}^\alpha$.*

Proof. The proof uses the fact that, by construction,

$$f_k^H = \mathbb{E}(f(X)).$$

Thus the first item follows immediately from Jensen's inequality. For the proof of the second item, recall that

$$f_k^H = \mathbb{E}(f(X)) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i})$$

where we now assume $f_\alpha \ge 0$ for all α. Since $\phi(X_i) = X_i^{\alpha_i}$ is convex on $[0,1]$ $(i \in [n])$, Jensen's inequality yields $\mathbb{E}(X_i^{\alpha_i}) \ge [\mathbb{E}(X_i)]^{\alpha_i}$. Thus

$$f_k^H \ge \sum_{\alpha \in \mathbb{N}^n} f_\alpha \mathbb{E}(X)^\alpha,$$

as required. For the third item, where f is assumed square-free, one has

$$f_k^H = \mathbb{E}(f(X)) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i})$$

where all $\alpha \in \{0,1\}^n$ so that $\mathbb{E}(X_i^{\alpha_i}) = [\mathbb{E}(X_i)]^{\alpha_i}$, and consequently

$$f_k^H = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \mathbb{E}(X)^\alpha.$$

This completes the proof. □

6.2. **The general case.**

Sampling. One may generate random samples $\mathbf{x} \in \mathcal{K}$ from the density σ on $\mathcal{K}$ using the well-known *method of conditional distributions* (see e.g., [21, Section 8.5.1]). For $\mathcal{K} = [0,1]^n$, the procedure is described in detail in [6, Section 3]. In this way one may obtain, with high probability, a point $\mathbf{x} \in \mathcal{K}$ with $f(\mathbf{x}) \le f_k^H + \epsilon$, for any given $\epsilon > 0$. (The size of the sample depends on ϵ.) Here we only mention that this procedure may be done in time polynomial in n and $1/\epsilon$; for details the reader is referred to [6, Section 3].

A heuristic based on the mode. As an alternative, one may consider the heuristic that returns the mode (i.e., maximizer) of the density function σ as a candidate solution. By way of illustration, recall that in Example 2.5 the mode was a good approximation of the global minimizer for σ of degree 50; see Figure 1. The mode may be calculated one variable at a time using (4.1).

In Section 7 below, we will illustrate the performance of all the strategies described in this section on numerical examples.

7. Numerical examples

In this section we will present numerical examples to illustrate the behavior of the sequences of upper bounds, and of the techniques to obtain feasible points.

We consider several well-known polynomial test functions from global optimization (also used in [6]), that are listed in Table 1, where we set

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

Note that the Booth and Matyas functions are convex. Note also that the functions have a rational minimizer in the hypercube (except the Styblinski-Tang function).

Table 1. Test functions

Name	Formula	Minimum ($f_{\min,\mathcal{K}}$)	Maximum ($f_{\max,\mathcal{K}}$)	Search domain ($\mathcal{K}$)
Booth Function	$f = (20x_1 + 40x_2 - 37)^2 + (40x_1 + 20x_2 - 35)^2$	$f(0.55, 0.65) = 0$	$f(0, 0) = 2594$	$[0, 1]^2$
Matyas Function	$f = 0.26[(20x_1 - 10)^2 + (20x_2 - 10)^2] - 0.48(20x_1 - 10)(20x_2 - 10)$	$f(0.5, 0.5) = 0$	$f(0, 1) = 100$	$[0, 1]^2$
Motzkin Polynomial	$f = (4x_1 - 2)^4(4x_2 - 2)^2 + (4x_1 - 2)^2(4x_2 - 2)^4 - 3(4x_1 - 2)^2(4x_2 - 2)^2 + 1$	$f(\frac{1}{4}, \frac{1}{4}) = f(\frac{1}{4}, \frac{3}{4}) = f(\frac{3}{4}, \frac{1}{4}) = f(\frac{3}{4}, \frac{3}{4}) = 0$	$f(1, 1) = 81$	$[0, 1]^2$
Three-Hump Camel Function	$f = 2(10x_1 - 5)^2 - 1.05(10x_1 - 5)^4 + \frac{1}{6}(10x_1 - 5)^6 + (10x_1 - 5)(10x_2 - 5) + (10x_2 - 5)^2$	$f(0.5, 0.5) = 0$	$f(1, 1) = 2047.92$	$[0, 1]^2$
Styblinski-Tang Function	$f = \sum_{i=1}^{n} \frac{1}{2}(10x_i - 5)^4 - 8(10x_i - 5)^2 + \frac{5}{2}(10x_i - 5)$	$f(0.209, \ldots, 0.209) = -39.16599n$	$f(1, \ldots, 1) = 125n$	$[0, 1]^n$
Rosenbrock Function	$f = \sum_{i=1}^{n-1} 100(4.096x_{i+1} - 2.048 - (4.096x_i - 2.048)^2)^2 + (4.096x_i - 3.048)^2$	$f(\frac{3048}{4096}, \ldots, \frac{3048}{4096}) = 0$	$f(0, \ldots, 0) = 3905.93(n - 1)$	$[0, 1]^n$

We start by listing the relative gaps $RG(\%) = \frac{f_k^H - f_{\min,\mathcal{K}}}{f_{\max,\mathcal{K}} - f_{\min,\mathcal{K}}} \times 100$ for these test functions in Table 2 for densities with degree up to $k = 50$.

One notices that the observed convergence rate is more-or-less in line with the $O(1/k)$ bound.

In a next experiment, we compare the Handelman-type densities ($RG(\%)$ by f_k^H bounds) to SOS densities (we still use the notation $RG(\%) = (f_k^{sos} - f_{\min,\mathcal{K}})/(f_{\max,\mathcal{K}} - f_{\min,\mathcal{K}}) \times 100$); we also compare their computation times (in seconds), for which we use the approaches described in Section 2.3, and we assume that the values $\gamma_{(\eta,\beta)}$ for all $(\eta, \beta) \in \mathbb{N}^{2n}_{k+d}$ and the moments of the Lebesgue measure on $\mathcal{K} = [0, 1]^n$ are computed beforehand; see Tables 3, 4 and 5. We performed the computation using Matlab on a Laptop with Intel Core i7-4600U CPU (2.10 GHz) and 8 GB RAM. The generalized eigenvalue computation was done in Matlab using the eig function.

As described in Example 2.5, there is no ordering possible in general between $f_{k/2}^{sos}$ and f_k^H, but one observes that $f_{k/2}^{sos} \leq f_k^H$ holds in most cases, i.e., the SOS densities usually give better bounds for a given degree. One should bear in mind though, that the $f_{k/2}^{sos}$ are in general much

TABLE 2. Relative gaps of f_k^H for test functions in Table 1.

k	Booth	Matyas	Motzkin	T-H. Camel	St.-Tang ($n=2$)	Rosen. ($n=2$)	Rosen. ($n=3$)	Rosen. ($n=4$)
1	10.8199	17.3333	5.1852	12.9776	20.0499	7.7615	10.1745	11.0081
2	9.6633	12.0000	2.7020	4.2038	18.5633	6.0339	7.7310	9.3678
3	8.2498	11.0667	2.7020	4.2038	17.2942	4.5549	6.8671	7.7383
4	7.0933	8.8000	1.5732	1.9822	15.8076	3.8045	6.1275	7.1624
5	6.6307	8.1333	1.5732	1.9822	15.0461	3.6406	5.2637	6.6694
6	5.8340	6.9867	1.2615	1.1892	14.2847	3.3393	4.4018	6.0935
7	5.5476	6.5524	1.2615	1.1892	13.8738	3.0766	4.0267	5.5188
8	5.0409	5.9048	1.1002	0.8458	13.4630	2.6480	3.7922	4.9429
9	4.8354	5.6190	1.1002	0.8458	13.2211	2.5610	3.4171	4.3682
10	4.5324	5.2245	1.0541	0.6771	12.9796	2.3301	3.2259	4.1182
11	4.2234	5.0317	1.0541	0.6771	12.6013	2.2383	3.0602	3.9269
12	4.0949	4.7778	1.0351	0.5144	12.1905	1.9703	2.8821	3.6767
13	3.8340	4.6444	1.0351	0.5144	11.8216	1.9210	2.7146	3.4725
14	3.6523	4.4741	1.0328	0.4236	11.5798	1.7703	2.6079	3.2225
15	3.4952	4.3798	1.0295	0.4236	11.3687	1.6965	2.4226	3.0950
16	3.3013	4.2618	1.0291	0.3539	10.9180	1.5472	2.2938	2.9845
17	3.2032	4.1939	1.0175	0.3539	10.5491	1.5167	2.1725	2.8543
18	3.0317	4.1102	1.0048	0.3016	10.1803	1.4152	2.0916	2.7439
19	2.9246	4.0606	0.9953	0.3016	9.9692	1.3556	1.9926	2.6449
20	2.8340	4.0000	0.9907	0.2628	9.7582	1.2643	1.9210	2.5134
25	2.3768	3.4324	0.9583	0.2064	8.7403	1.0421	1.5524	2.0716
30	2.0479	2.8927	0.9227	0.1557	7.7221	0.8535	1.3046	1.7571
35	1.7964	2.5989	0.8725	0.1336	7.0469	0.7353	1.1128	1.5175
40	1.6053	2.2609	0.8179	0.1105	6.3713	0.6371	0.9665	1.3286
45	1.4456	2.0800	0.7721	0.0993	5.8880	0.5628	0.8591	1.1861
50	1.3129	1.8595	0.7301	0.0868	5.4195	0.5054	0.7634	1.0592

TABLE 3. Comparison of two upper bounds for Booth, Matyas and Three–Hump Camel functions in relative gaps and computation times (sec.)

k	Booth				Matyas				Three–Hump Camel			
	$f_{k/2}^{sos}$		f_k^H		$f_{k/2}^{sos}$		f_k^H		$f_{k/2}^{sos}$		f_k^H	
	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time
2	9.433	0.0007	9.663	0.0001	8.267	0.0009	12.0	0.0001	12.98	0.0008	4.204	0.0001
4	6.264	0.0006	7.093	0.0003	5.322	0.0005	8.8	0.0003	1.416	0.0006	1.982	0.0002
6	4.564	0.0008	5.834	0.0008	4.282	0.0009	6.987	0.0007	1.416	0.0011	1.189	0.0007
8	3.764	0.0015	5.041	0.0025	3.894	0.0017	5.905	0.0018	0.4678	0.002	0.8458	0.0017
10	2.691	0.0025	4.532	0.0038	3.689	0.0033	5.224	0.0039	0.4678	0.0035	0.6771	0.0037
12	2.45	0.0047	4.095	0.0065	2.996	0.0056	4.778	0.0074	0.2168	0.0086	0.5144	0.0063
14	1.814	0.0072	3.652	0.0109	2.547	0.0102	4.474	0.0112	0.2168	0.0128	0.4236	0.0117
16	1.607	0.0097	3.301	0.0177	2.043	0.0131	4.262	0.0178	0.1245	0.0139	0.3539	0.0179
18	1.319	0.0146	3.032	0.0276	1.834	0.0226	4.11	0.0266	0.1245	0.0377	0.3016	0.027
20	1.107	0.0242	2.834	0.0391	1.478	0.0329	4.0	0.0384	0.08363	0.0312	0.2628	0.0397

TABLE 4. Comparison of two upper bounds for Motzkin, Styblinski-Tang ($n = 2$) and Rosenbrock ($n = 2$) functions in relative gaps and computation times (sec.)

k	Motzkin				Sty.–Tang ($n=2$)				Rosenb. ($n=2$)			
	$f_{k/2}^{sos}$		f_k^H		$f_{k/2}^{sos}$		f_k^H		$f_{k/2}^{sos}$		f_k^H	
	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time
2	5.185	0.0008	2.702	0.0001	19.92	0.0008	18.56	0.0001	5.495	0.001	6.034	0.0001
4	1.31	0.0005	1.573	0.0003	16.01	0.0005	15.81	0.0002	3.899	0.0009	3.804	0.0003
6	1.31	0.0009	1.261	0.0009	13.38	0.0009	14.28	0.0008	2.685	0.0018	3.339	0.0013
8	1.024	0.0016	1.1	0.002	11.23	0.0016	13.46	0.0021	1.936	0.0031	2.648	0.0034
10	0.989	0.0034	1.054	0.0043	10.12	0.0028	12.98	0.0037	1.319	0.0031	2.33	0.0057
12	0.989	0.0062	1.035	0.006	8.308	0.0063	12.19	0.0078	1.07	0.0049	1.97	0.008
14	0.8752	0.0096	1.033	0.0168	6.678	0.0097	11.58	0.0177	0.7716	0.0083	1.77	0.012
16	0.6982	0.0216	1.029	0.0179	6.009	0.014	10.92	0.0214	0.6614	0.0119	1.547	0.0237
18	0.6982	0.0242	1.005	0.0266	5.342	0.0231	10.18	0.0358	0.4992	0.0198	1.415	0.0264
20	0.6269	0.0298	0.9907	0.046	4.36	0.0286	9.758	0.042	0.4455	0.0324	1.264	0.0383

more expensive to compute than f_k^H, as discussed in Section 2.3. This is not really visible in the computational times presented here, since the values of n in the examples are too small.

TABLE 5. Comparison of two upper bounds for Rosenbrock functions ($n = 3, 4$) in relative gaps and computation times (sec.)

k	Rosenb. ($n = 3$)				Rosenb. ($n = 4$)			
	$f^{sos}_{k/2}$		f^H_k		$f^{sos}_{k/2}$		f^H_k	
	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time	$RG(\%)$	time
2	8.053	0.0033	7.731	0.0001	8.945	0.0204	9.368	0.0002
4	5.046	0.0009	6.128	0.0007	5.891	0.0243	7.162	0.0017
6	3.787	0.0024	4.402	0.0021	4.577	0.0111	6.093	0.0062
8	2.649	0.0078	3.792	0.0054	3.266	0.0442	4.943	0.0228
10	2.152	0.016	3.226	0.0135	2.686	0.2087	4.118	0.0699
12	1.556	0.0355	2.882	0.0244	2.02	0.3774	3.677	0.1837
14	1.305	0.0811	2.608	0.041	1.73	0.9121	3.222	0.431
16	0.9918	0.1324	2.294	0.0684	1.334	1.986	2.985	1.099
18	0.8538	0.2272	2.092	0.1139	1.169	4.279	2.744	1.92

Next we consider the strategies for generating feasible points corresponding to the bounds f^H_k, as described in Section 6; see Table 6.

TABLE 6. Comparing strategies for generating feasible points for Booth, Matyas, Motzkin, and Three–Hump Camel functions. Here, $\hat{\mathbf{x}}$ denotes the mode of the optimal density.

k	Booth			Matyas			Motzkin		Three-H. Camel	
	f^H_k	$f(\hat{\mathbf{x}})$	$f(\mathbb{E}(X))$	f^H_k	$f(\hat{\mathbf{x}})$	$f(\mathbb{E}(X))$	f^H_k	$f(\hat{\mathbf{x}})$	f^H_k	$f(\hat{\mathbf{x}})$
5	172.0	96.222	17.0	8.1333	4.0	1.460	1.2743	1.0	40.593	—
10	117.571	96.222	25.806	5.2245	4.0	2.0408	0.8538	1.0	13.867	—
15	90.6667	27.580	7.6777	4.3798	4.0	2.5017	0.8339	1.0	8.6752	0.273
20	73.5152	9.0	2.0	4.0000	0.16	0.1111	0.8025	1.0	5.3826	0
25	61.6535	4.5785	1.8107	3.4324	0.3161	0.2404	0.7762	1.0	4.2267	0.1653
30	53.1228	1.6403	0.41428	2.8927	0.0178	0.0138	0.7474	1.0	3.1892	0
35	46.5982	1.0923	0.53061	2.5989	0.1071	0.0897	0.7067	0.4214	2.7367	0.110
40	41.6416	0.8454	0.64566	2.2609	0	0	0.6625	0.2955	2.2626	0
45	37.4988	2.0	0.80157	2.0800	0	0	0.6254	0.1985	2.0337	0.0783
50	34.0573	0.9784	0.22222	1.8595	0	0	0.5914	0.1297	1.7768	0

In Table 6, the columns marked $f(\mathbb{E}(X))$ refer to the convex case in Theorem 6.2. The columns marked $f(\hat{\mathbf{x}})$ correspond to the mode $\hat{\mathbf{x}}$ of the optimal density; an entry '—' in these columns means that the mode of the optimal density was not unique.

For the convex Booth and Matyas functions $f(\mathbb{E}(X))$ gives the best upper bound. For sufficiently large k the mode $\hat{\mathbf{x}}$ gives a better bound than f^H_k, indicating that this heuristic is useful in the non-convex case.

As a final comparison, we also look at the general sampling technique via the method of conditional distributions; see Tables 7 and 8. We present results for the Motzkin polynomial and the Three hump camel function.

For each degree k, we use the sample sizes 10 and 100. In Tables 7 and 8 we record the mean, variance and the minimum value of these samples. (Recall that the expected value of the sample mean equals f^H_k.) We also generate samples uniformly from $[0, 1]^n$, for comparison.

The mean of the sample function values approximates f^H_k reasonably well for sample size 100, but less so for sample size 10. Moreover, the mean sample function value for uniform sampling from $[0, 1]^n$ is much higher than f^H_k. Also, the minimum function value for sampling is significantly lower than the minimum function value obtained by uniform sampling for most values of k.

TABLE 7. Sampling results for Motzkin polynomial

		Sample size 10			Sample size 100		
k	f_k^H	Mean	Variance	Minimum	Mean	Variance	Minimum
5	1.2743	0.8330	0.0466	0.2790	1.1590	4.2023	0.0525
10	0.8538	0.7005	0.0800	0.1862	0.8435	0.1448	0.1149
15	0.8339	0.9063	0.0153	0.6069	0.8465	0.0932	0.0593
20	0.8025	0.7704	0.0336	0.3826	0.9326	1.6454	0.0040
25	0.7762	0.7995	0.1014	0.2433	0.7493	0.0717	0.0722
30	0.7474	1.0104	1.2852	0.1091	0.8290	0.8620	0.0522
35	0.7067	0.5930	0.0981	0.1940	0.7647	1.3012	0.0016
40	0.6625	0.6967	0.0497	0.2867	0.6028	0.1371	0.0021
45	0.6254	0.6258	0.0500	0.3548	0.7007	0.2242	0.0090
50	0.5914	0.6244	0.0718	0.3000	0.5782	0.1406	0.0154
Uniform Sample		4.2888	37.4427	0.5290	3.7397	53.8833	0.0492

TABLE 8. Sampling results for Three-Hump Camel function

		Sample size 10			Sample size 100		
k	f_k^H	Mean	Variance	Minimum	Mean	Variance	Minimum
5	40.593	91.872	27065.0	0.90053	53.656	14575.0	0.58086
10	13.867	11.312	45.784	0.8916	14.273	382.98	0.018985
15	8.6752	5.6281	31.311	0.21853	10.373	778.32	0.022282
20	5.3826	3.5174	16.053	0.43269	9.4178	653.27	0.041752
25	4.2267	10.741	776.55	0.59616	5.0642	112.61	0.039463
30	3.1892	2.2515	8.6915	0.063265	2.2096	6.2611	0.040845
35	2.7367	1.5032	1.4626	0.0085016	3.0679	16.47	0.24175
40	2.2626	1.3941	1.1995	0.21653	2.3431	17.735	0.069473
45	2.0337	2.3904	10.934	0.57818	1.8928	3.6581	0.050042
50	1.7768	1.664	3.3983	0.061995	1.6301	1.6966	0.048476
Uniform Sample		306.96	275366.0	0.15602	368.28	296055.0	0.59281

8. Concluding remarks

One may consider several strategies to improve the upper bounds f_k^H, and we list some in turn.

- A natural idea is to use density functions that are convex combinations of SOS and Handelman-type densities, i.e., that belong to $\mathcal{H}_k + \Sigma[x]_r$ for some nonnegative integers k, r. Unfortunately one may show that this does not yield a better upper bound than $\min\{f_r^{sos}, f_k^H\}$, namely

$$\min\{f_r^{sos}, f_k^H\} = \inf_{\sigma \in \mathcal{H}_k + \Sigma[x]_r} \left\{ \int_{\mathcal{K}} f(\mathbf{x})\,\sigma(\mathbf{x})\,d\mathbf{x} : \int_{\mathcal{K}} \sigma(\mathbf{x})\,d\mathbf{x} = 1 \right\}, \quad k, r \in \mathbb{N}.$$

(We omit the proof since it is straightforward, and of limited interest.)

- For optimization over the hypercube, a second idea is to replace the integer exponents in Handelman representations of the density by more general positive real exponents. (This is amenable to analysis since the beta distribution is defined for arbitrary positive shape parameters and with its moments available via relation (4.2).) If we drop the integrality requirement for (η, β) in the definition of f_k^H (see (1.3)), we obtain the bound:

$$f_k^H \geq f_k^{beta} := \min_{(\eta,\beta) \in \Delta_k^{2n}} \sum_{\alpha \in \mathbb{N}^n_{\leq d}} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}, \quad k \in \mathbb{N},$$

where Δ_k^{2n} is the simplex $\Delta_k^{2n} := \{(\eta, \beta) \in \mathbb{R}^{2n}_+ \,:\, \sum_{i=1}^n (\eta_i + \beta_i) = k\}$.

As with f_k^H, when (η, β) is such that $f_k^{beta} = \sum_{\alpha \in \mathbb{N}^n_{\leq d}} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}$, one has that $f_k^{beta} = \mathbb{E}(f(X))$ where $X = (X_1, \ldots, X_n)$ and $X_i \sim beta(\eta_i+1, \beta_i+1)$ $(i \in [n])$. Using the moments of the beta distribution in (4.2), we obtain

$$f_k^{beta} = \min_{(\eta,\beta)\in\Delta_k^{2n}} \sum_{\alpha\in\mathbb{N}_d^n} f_\alpha \prod_{i=1}^n \frac{(\eta_i+1)\cdots(\eta_i+\alpha_i)}{(\eta_i+\beta_i+2)\cdots(\eta_i+\beta_i+\alpha_i+1)}, \quad k\in\mathbb{N}. \tag{8.1}$$

Thus one may obtain the bounds f_k^{beta} by minimizing a rational function over a simplex. A question for future research is whether one may approximate f_k^{beta} to any fixed accuracy in time polynomial in k and n. (This may be possible, since the minimization of fixed-degree polynomials over a simplex allows a PTAS [4], and the relevant algorithmic techniques have been extended to rational objective functions [11].)

One may also use the value of $(\eta,\beta)\in\Delta_k^{2n}$ that gives f_k^H as a starting point in the minimization problem (8.1), and employ any iterative method to obtain a better upper bound heuristically. Subsequently, one may use the resulting density function to obtain 'good' feasible points as described in Section 6. Of course, one may also use the feasible points (generated by sampling) as starting points for iterative methods. Suitable iterative methods for bound-constrained optimization are described in the books [2, 7, 8], and the latest algorithmic developments for bound constrained global optimization are surveyed in the recent thesis [22].

- Perhaps the most promising practical variant of the f_k^H bound is the following parameter:

$$\begin{aligned} f_{r,k}^H &= \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \frac{\int_{\mathcal{K}} f(\mathbf{x}) \left(\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta\right)^r d\mathbf{x}}{\int_{\mathcal{K}} (\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta)^r d\mathbf{x}} \\ &= \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}^n} f_\alpha \frac{\gamma_{(r\eta+\alpha,r\beta)}}{\gamma_{(r\eta,r\beta)}} \qquad \text{for } r,k\in\mathbb{N}. \end{aligned}$$

Thus, the idea is to replace the density $\sigma(x)=\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta/\int_{\mathcal{K}}\mathbf{x}^\eta(\mathbf{1}-\mathbf{x})^\beta d\mathbf{x}$ by the density $\sigma(x)^r/\int_{\mathcal{K}}\sigma(r)^r d\mathbf{x}$ for some power $r\in\mathbb{N}$. Hence, for $r=1$, $f_{1,k}^H=f_k^H$. Note that the calculation of $f_{r,k}^H$ requires exactly the same number of elementary operations as the calculation of f_k^H, provided all the required moments are available. (Also note that, for $\mathcal{K}=[0,1]^n$, one could allow an arbitrary $r>0$ since the moments are still available as pointed out above.)

In Tables 9, 10, and 11, we show some relative gaps for the parameter $f_{r,k}^H$, defined as $(f_{r,k}^H - f_{\min,\mathcal{K}})/(f_{\max,\mathcal{K}}-f_{\min,\mathcal{K}})\times 100$.

TABLE 9. Relative gaps of $f_{r,k}^H$ for the Styblinski-Tang function ($n=2$)

k	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$
1	20.0499	20.7931	21.3190	21.3190	21.3190
2	18.5633	18.4184	18.7040	19.0470	19.3665
3	17.2942	17.2522	16.9793	16.7974	16.6631
4	15.8076	15.5176	15.2511	14.6398	14.1912
5	15.0461	14.3517	14.3645	13.8452	13.3692
6	14.2847	13.1855	12.6361	12.2758	12.0074
7	13.8738	12.0519	10.9113	10.1182	9.5355
8	13.4630	10.9180	9.1831	7.9606	7.0636
9	13.2211	10.3381	8.4528	7.1660	6.2416
10	12.9796	9.7582	7.7221	6.3713	5.4195

TABLE 10. Relative gaps of $f_{r,k}^H$ for the Rosenbrock function ($n = 3$)

k	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
1	10.1745	9.3107	8.9356	8.7536	8.6603
2	7.7310	6.5571	6.0674	5.8142	5.6807
3	6.8671	5.7557	5.1021	4.7091	4.4890
4	6.1275	4.7220	3.7699	3.2404	2.9126
5	5.2637	3.5090	3.0196	2.9302	2.9826
6	4.4018	2.8821	2.4570	1.9388	1.5359
7	4.0267	2.8901	2.1273	1.6465	1.3623
8	3.7922	2.5456	1.8554	1.4301	1.1273
9	3.4171	2.3701	1.7074	1.3206	1.0798
10	3.2259	2.0283	1.4251	1.1250	0.8966

TABLE 11. Relative gaps of $f_{r,k}^H$ for the Rosenbrock function ($n = 4$)

k	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
1	11.0081	10.4440	10.1939	10.0727	10.0104
2	9.3678	8.5929	8.2655	8.0963	8.0074
3	7.7383	6.7421	6.3371	6.1202	6.0046
4	7.1624	6.2079	5.7098	5.4000	5.2266
5	6.6694	5.1729	4.2870	3.8120	3.5307
6	6.0935	4.4015	3.3909	2.8242	2.4706
7	5.5188	3.5929	2.8908	2.6175	2.5173
8	4.9429	3.1671	2.5076	1.9564	1.5528
9	4.3682	2.8285	2.2958	1.7616	1.4370
10	4.1182	2.7624	2.1065	1.6160	1.2793

A first important observation is that, for fixed k, the values of $f_{r,k}^H$ are not monotonically decreasing in r; see e.g. the row $k = 2$ in Table 9. Likewise, the sequence $f_{r,k}^H$ is not monotonically decreasing in k for fixed r; see, e.g., the column $r = 5$ in Table 10.

On the other hand, it is clear from Tables 9, 10, and 11 that $f_{r,k}^H$ can provide a much better bound than f_k^H for $r > 1$.

Since $f_{r,k}^H$ is not monotonically decreasing in r (for fixed k), or in k (for fixed r), one has to consider the convergence question. An easy case is when $\mathcal{K} = [0,1]^n$ and the global minimizer $\mathbf{x}^*$ is rational. Say $x_i^* = \frac{p_i}{q_i}$ ($i \in [n]$), setting $q_i = 1$ and $p_i = x_i^*$ when $x_i^* \in \{0,1\}$. Consider the following variation of the parameters η_i^*, β_i^* from Definition 4.5: $\eta_i^* = rp_i + 1$ and $\beta_i^* = r(q_i - p_i) + 1$ for $i \in [n]$, so that $\sum_{i=1}^n \eta_i^* + \beta_i^* - 2 = r(\sum_{i=1}^n q_i)$. Combining relation (4.8) and Theorem 4.8, we can conclude that the following inequality holds:

$$f_{r,k}^H - f(\mathbf{x}^*) \leq \frac{C_f}{r} \quad \text{for all} \quad k \geq \textstyle\sum_{i=1}^n q_i \quad \text{and} \quad r \geq 1,$$

where C_f is a constant that depends on f only.

For more general sets $\mathcal{K}$, one may ensure convergence by considering instead the following parameter (for fixed $R \in \mathbb{N}$):

$$\min_{r \in [R]} f_{k,r}^H \leq f_k^H \quad (k \in \mathbb{N}).$$

Then convergence follows from the convergence results for $f_{k,r}^H$. Moreover, this last parameter may be computed in polynomial time if k is fixed, and R is bounded by a polynomial in n.

Acknowledgements. Etienne de Klerk would like to thank Dorota Kurowicka for valuable discussions on the beta distribution. The research of Jean B. Lasserre was funded by by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 666981 TAMING). We thank two anonymous referees for their useful suggestions that helped improve the presentation of the paper.

References

[1] Bomze, I.M., Klerk, E. de.: Solving standard quadratic optimization problems via semidefinite and copositive programming. *J. Global Optim.* 24(2), 163–185 (2002)

[2] Bertsekas, D.P.: *Constrained Optimization and Lagrange Multiplier Methods.* Athena Scientific, Belmont, MA (1996)

[3] De Klerk, E., Laurent, M.: Error bounds for some semidefinite programming approaches to polynomial optimization on the hypercube. *SIAM J. Optim.* 20(6), 3104–3120 (2010)

[4] De Klerk, E., Laurent, M., Parrilo, P.: A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoret. Comput. Sci.* 361(2–3), 210–225 (2006)

[5] De Klerk, E., Laurent, M., Sun, Z.: An error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution. *SIAM J. Optim.* (to appear) (2015)

[6] De Klerk, E., Laurent, M., Sun, Z.: Error analysis for Lasserre hierarchy of upper bounds for continuous optimization. *arXiv: 1411.6867* (2014)

[7] Fletcher, R.: *Practical Methods of Optimization*, 2nd ed., John Wiley & Sons, Inc., New York (1987)

[8] Gill, P.E., Murray, W., Wright, M.H.: *Practical Optimization*, Academic Press, New York (1981)

[9] Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edition. The John Hopkins University Press, Baltimore and London (1996)

[10] Handelman, D.: Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific J. Math.* 132(1), 35–62 (1988)

[11] Jibetean, D., De Klerk, E.: Global optimization of rational functions: a semidefinite programming approach. *Math. Program.* 106(1), 93–109 (2006)

[12] Johnson, N.L., Kotz, S.: *Continuous univariate distributions – 2.* John Wiley & Sons (1970)

[13] Krivine, J.L.: Anneaux préordonnés, *J. Anal. Math.* 12, 307–326 (1964)

[14] Krivine, J.L.: Quelques propriétés des préordres dans les anneaux commutatifs unitaires. *Comptes Rendus de l'Académie des Sciences de Paris*, 258, 3417–3418 (1964)

[15] Lasserre, J.B.: Optimisation globale et théorie des moments. *C. R. Acad. Sci.* Paris 331, Série 1, 929–934 (2000)

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

[17] Lasserre, J.B.: Semidefinite programming vs. LP relaxations for polynomial programming. *Math. Oper. Res.* 27, 347-C360 (2002)

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

[19] Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* 21, 864–885 (2011)

[20] Laurent, M.: A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxation for 0-1 programming. *Math. Oper. Res.* 28(3), 470–498 (2003)

[21] Law, A.M.: *Simulation Modeling and Analysis* (4th edition). Mc Graw-Hill (2007)

[22] Pál, L.: Global optimization algorithms for bound constrained problems. PhD thesis, University of Szeged (2010) Available at `http://www2.sci.u-szeged.hu/fokozatok/PDF/Pal_Laszlo/Diszertacio_PalLaszlo.pdf`

[23] Romero, J., Velasco, M.: Semidefinite approximations of conical hulls of measured sets, *arXiv:1409.9272v2* (2014)

[24] Schrijver, A.: *Theory of Linear and Integer Programming.* Wiley (1986)

[25] Vavasis, S.A.: Quadratic programming is in NP. *Inform. Process. Lett.* 36, 73–77 (1990)

Tilburg University and Delft University of Technology; PO Box 90153, 5000 LE Tilburg, The Netherlands.

E-mail address: `E.deKlerk@uvt.nl`

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4, France, Tel: +33561336415

E-mail address: `lasserre@laas.fr`

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

École Polytechnique de Montréal; Canada Excellence Research Chair in Data Science for Real-Time Decision-Making, C.P. 6079, Succ. Centre-ville, Montral, H3C 3A7, Canada.
E-mail address: Zhao.Sun@polymtl.ca