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. 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, 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

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

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

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

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

where the infimum is taken over all probability measures μ with support in $\mathbf{K}$. Using this reformulation one may obtain a sequence of *lower bounds* on $f_{\min,\mathbf{K}}$ that converges to $f_{\min,\mathbf{K}}$, by introducing tractable convex relaxations of the set of probability measures with support in $\mathbf{K}$ (if $\mathbf{K}$ 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

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

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 $f_k^{sos} \to f_{\min,\mathbf{K}}$ as $k \to \infty$ (see Theorem 2.1 below for a precise statement). In principle this approach works for any compact set $\mathbf{K}$ 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 $\mathbf{K}$ 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 $f_k^{sos} - f_{\min,\mathbf{K}} = O(1/\sqrt{k})$ 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 $f_k^H \geq f_{\min,\mathbf{K}}$ such that $f_k^H \to f_{\min,\mathbf{K}}$ as $k \to \infty$. The parameters f_k^H can be effectively computed when the set $\mathbf{K} \subset [0,1]^n$ 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 $\mathbf{K} \subseteq [0,1]^n$ and from the coefficients (f_α) of the polynomial f to minimize. Namely:

$$f_k^H := \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \sum_{\alpha\in\mathbb{N}^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}, \tag{1.2}$$

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

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

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 (1.2) has also a simple interpretation as it reads:

$$f_k^H = \min_{(\eta,\beta)\in\mathbb{N}_k^{2n}} \frac{\displaystyle\int_{\mathbf{K}} f(\mathbf{x})\,\mathbf{x}^\eta(1-\mathbf{x})^\beta\,d\mathbf{x}}{\displaystyle\int_{\mathbf{K}} \mathbf{x}^\eta(1-\mathbf{x})^\beta\,d\mathbf{x}} = \min_{\mu} \left\{ \int_{\mathbf{K}} f\,d\mu :\ \mu \in M(\mathbf{K})_k \right\}, \tag{1.3}$$

where $M(\mathbf{K})_k$ is the set of probability measures on $\mathbf{K}$, absolutely continuous with respect to the Lebesgue measure on $\mathbf{K}$, and whose density is a monomial $\mathbf{x}^\eta(1-\mathbf{x})^\beta$ 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 $d\mu = \mathbf{x}^\eta(1-\mathbf{x})^\beta\,d\mathbf{x}$ (normalized to make then probability measures).

For the case of the hypercube $\mathbf{K}=[0,1]^n$, 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 $\mathbf{K}$ (see Theorem 3.1).

In Section 4, for the case of the hypercube $\mathbf{K}=[0,1]^n$, we analyze the quality of the bounds f_k^H. We show a convergence rate in $O(1/\sqrt{k})$ for the range $f_k^H - f_{\min,\mathbf{K}}$ and a stronger convergence rate in $O(1/k)$ when the polynomial f admits a rational minimizer in $[0,1]^n$ (see Theorem 4.8). This stronger convergence rate applies in particular to quadratic polynomials (since they have a rational minimizer) and Example 4.9 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 $\hat{x}\in\mathbf{K}$ with value $f(\hat{x})\le f_k^H$, once the upper bound f_k^H has been computed, hence also with an error bound guarantee in the case of the box $\mathbf{K}=[0,1]^n$. 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 $\hat{x}$ 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 $\hat{x}$ 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

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 subspace of polynomials of degree at most d, and $\Sigma[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_{2d}$ its subset of sums of squares (SOS) of degree at most $2d$.

We use the convention that $\mathbb{N}$ denotes the nonnegative integers, and let $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \,(=: |\alpha|) = d\}$, 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 $(1-\mathbf{x})^\alpha$ stands for $(1-x_1)^{\alpha_1} \cdots (1-x_n)^{\alpha_n}$, $\alpha \in \mathbb{N}^n$. We will also denote $[n] = \{1, 2, \ldots, n\}$.

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).

In [19], Lasserre proved the following.

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

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

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

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

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

Theorem 2.2. *Let $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\, j = 1, \ldots, m\} \subset \mathbb{R}^n$ 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 $\mathbf{K}$ 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 set $\mathcal{H}_k$ consisting of the polynomials of the form:

$$\sum_{(\eta,\beta) \in \mathbb{N}^{2n}_k} \lambda_{\eta,\beta} \mathbf{x}^\eta (1-\mathbf{x})^\beta \quad \text{where } \lambda_{\eta\beta} \geq 0, \tag{2.4}$$

i.e., all polynomials admitting 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 $\mathbf{x}^\eta (1-\mathbf{x})^\beta$ 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 $\mathbf{x}^\eta (1-\mathbf{x})^\beta$

as a conic combination of terms $\mathbf{x}^{\eta'}(1-\mathbf{x})^{\beta'}$ 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 in an analogous way as f_k^{sos} (see (2.1)), but where the SOS density function $\sigma \in \Sigma_k[\mathbf{x}]$ is replaced by a density $\sigma \in \mathcal{H}_k$.

Lemma 2.4. *Consider the sequence (f_k^H), $k \in \mathbb{N}$, with f_k^H as in* (1.2)*. Then one has:*

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

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

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

where we have used the fact that the penultimate optimization problem is an LP over a simplex that obtains its infimum at one of the vertices. □

Example 2.5. *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 $\mathbf{K}=[0,1]^2$, with minimum $f_{\min,\mathbf{K}} \approx -78.33198$ and minimizer

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

Here one has $f_1^{sos} = -12.9249$, and the corresponding 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 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.

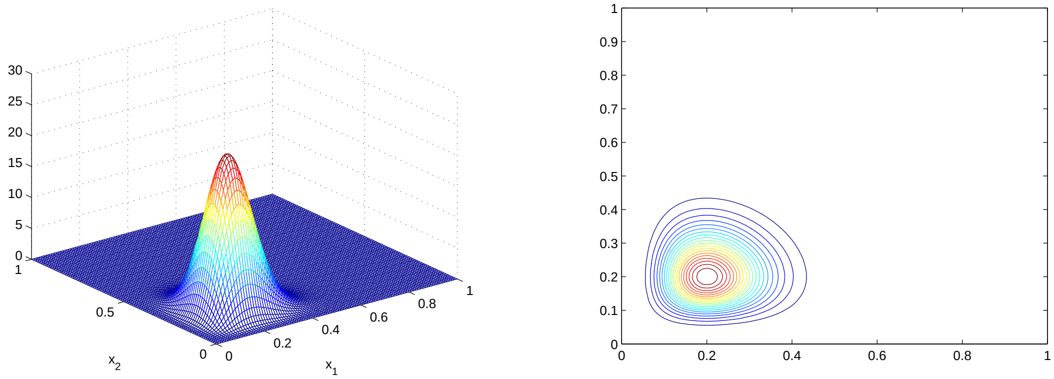

FIGURE 1. Optimal Handelman-type density $\sigma(x)$ of degree 50 on $[0,1]^2$ for the bivariate Styblinski-Tang function.

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

Let $\mathbf{K} \subseteq [0,1]^n$ be a compact set and for every $(\eta, \beta) \in \mathbb{N}^{2n}$, let

$$\gamma_{(\eta,\beta)} := \int_{\mathbf{K}} \mathbf{x}^\eta (1-\mathbf{x})^\beta \, d\mathbf{x}. \tag{3.1}$$

Of course when $\mathbf{K}$ is arbitrary one does not know how to compute such generalized moments. But if $\mathbf{K}$ 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), the hypercube $\{0,1\}^n$ and/or their image by a linear mapping, then such moments are available in closed-form. For instance if $\mathbf{K} = [0,1]^n$ then

$$\int_{\mathbf{K}} \mathbf{x}^\eta \, (1-\mathbf{x})^\beta \, d\mathbf{x} = \prod_{i=1}^n \left(\int_0^1 x_i^{\eta_i} (1-x_i)^{\beta_i} \, dx_i \right), \quad (\eta, \beta) \in \mathbb{N}^{2n},$$

and the univariate integrals may be calculated from

$$\int_0^1 t^i (1-t)^j \, dt = \frac{i!j!}{(i+j+1)!} \qquad i, j \in \mathbb{N}. \tag{3.2}$$

Theorem 3.1. *Let* $f \in \mathbb{R}[\mathbf{x}]_d$ *and let* $\gamma_{(\eta,\beta)}$ *be as in (3.1). Define as before the parameters*

$$f_k^H = \min_{(\eta,\beta) \in \mathbb{N}_k^{2n}} \sum_{\alpha \in \mathbb{N}_{\leq d}^n} f_\alpha \frac{\gamma_{(\eta+\alpha,\beta)}}{\gamma_{(\eta,\beta)}}, \qquad \forall \, k \in \mathbb{N}. \tag{3.3}$$

Then the sequence (f_k^H)*,* $k \in \mathbb{N}$*, is monotone non increasing and* $f_{\min,\mathbf{K}} = \lim_{k\to\infty} f_k^H$*.*

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

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

Also recall from Lemma 2.4 that

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

In view of Lemma 2.3, the sequence (f_k^H) is monotone non-increasing. Moreover, $f_{\min,\mathbf{K}} \le f_k^H$ for all k. Next we show that the sequence (f_k^H) converges to $f_{\min,\mathbf{K}}$.

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

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

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

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

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

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

Hence we obtain:

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

The right most term is equal to

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

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

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

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

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

In this section we analyze the convergence rate of the bounds f_k^H for the hypercube $\mathbf{K} = [0,1]^n$. We prove a convergence rate in $O(1/\sqrt{k})$ for the range $f_k^H - f_{\min,\mathbf{K}}$ 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. **The beta distribution.** By definition, a random variable $Y \in [0,1]$ has the beta distribution with shape parameters $a > 0$ and $b > 0$ (denoted by $Y \sim beta(a,b)$) 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 Y is

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

(see, e.g, [12, Chapter 24]; this also follows using (3.2)).

4.2. **Proof of convergence rate.** Given a polynomial f, consider a global minimizer $\mathbf{x}^*$ of f in $[0,1]^n$. In what follows we indicate how to construct a vector of independent random variables $X = (X_1,\ldots,X_n)$ so that the X_i's have the beta distribution with suitable shape parameters η_i^*, β_i^*, designed to ensure that (roughly) $\mathbb{E}[X] = \mathbf{x}^*$.

We will use the following result about Diophantine approximations.

Theorem 4.1 (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*

$$\left| x - \frac{p}{q} \right| < \frac{\epsilon}{q} \quad \text{and} \quad 1 \leq q \leq \frac{1}{\epsilon}. \tag{4.3}$$

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

If $x_i^* \in (0,1)$ is a rational coordinate of $\mathbf{x}^*$, then we select integers p_i and q_i such that $x_i^* = p_i/q_i$, so that $1 \leq p_i < q_i$. When x_i^* is an irrational coordinate of $\mathbf{x}^*$ we use Theorem 4.1 to construct a pair of suitable integers p_i, q_i. Namely, we consider an integer $r \geq 1$ and apply Theorem 4.1 with $\epsilon = 1/r$. Then, there exist integers p_i and q_i satisfying

$$\left| x_i^* - \frac{p_i}{q_i} \right| < \frac{1}{rq_i}, \ 0 \leq p_i \leq q_i \leq r \text{ and } 1 \leq q_i. \tag{4.4}$$

For convenience, let I_0 (resp., I_1, I) denote the set of indices $i \in [n]$ for which x_i^* is irrational and the integers p_i and q_i in (4.4) satisfy: $p_i = 0$ (resp., $p_i = q_i$, $1 \leq p_i < q_i$). Moreover, define the set J consisting of all indices i for which $x_i^* \in (0,1)$ is rational. Then, $x_i^* \in \{0,1\}$ for all $i \in [n] \setminus (I_0 \cup I_1 \cup I \cup J)$.

We now indicate how to construct the parameters η_i^* and β_i^*.

Definition 4.2. *Let r be a positive integer. For each coordinate $x_i^* \in [0,1]$, consider the integers p_i and q_i defined as above. We define the parameters η_i^* and $\beta_i^* \in \mathbb{N}$ as follows.*

- (i) *Assume $i \in J \cup I$; that is, either $x_i^* \in (0,1)$ is rational of the form $x_i^* = p_i/q_i$, or x_i^* is irrational with $1 \leq p_i < q_i$. Then, we set $\eta_i^* = rp_i$ and $\beta_i^* = r(q_i - p_i)$.*
- (ii) *Assume either $x_i^* = 0$, or $i \in I_0$, i.e., x_i^* is irrational with $p_i = 0$. Then we set $\eta_i^* = 1$ and $\beta_i^* = r$.*
- (iii) *Assume either $x_i^* = 1$, or $i \in I_1$, i.e., x_i^* is irrational with $p_i = q_i$. Then we set $\eta_i^* = r$ and $\beta_i^* = 1$.*

Now we define the vector of independent random variables $X := (X_1, \dots, X_n)$, where $X_i \sim beta(\eta_i^*, \beta_i^*)$ $(i \in [n])$.

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's are independent we have $\mathbb{E}(X^\alpha) = \prod_{i=1}^n \mathbb{E}(X_i^{\alpha_i})$ and the expected value of $f(X) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha X^\alpha$ is given by

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

where $\mathbb{E}(X_i^{\alpha_i})$ can be computed using (4.2). Observe moreover that, by construction,

$$\mathbb{E}(f(X)) = \frac{\int_{[0,1]^n} f(\mathbf{x})\mathbf{x}^{\eta^*-\mathbf{1}}(1-\mathbf{x})^{\beta^*-\mathbf{1}} d\mathbf{x}}{\int_{[0,1]^n} \mathbf{x}^{\eta^*-\mathbf{1}}(1-\mathbf{x})^{\beta^*-\mathbf{1}} d\mathbf{x}} \geq f_{k_r}^H \geq f(\mathbf{x}^*) \ , \tag{4.6}$$

where we set

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

and let $\mathbf{1}$ denote the all-ones vector. We will also use the following estimate on the parameter k_r.

Lemma 4.3. *Consider the parameter $k_r = \sum_{i=1}^n (\eta_i^* - 1 + \beta_i^* - 1)$. Then the following holds:*

(i) *If $x^* \in \mathbb{Q}$ then $k_r \leq ar$ for all $r \geq 1$, where $a > 0$ is a constant (not depending on r).*
(ii) *If $x^* \in \mathbb{R} \setminus \mathbb{Q}$ then $k_r \leq a'r^2$ for all $r \geq 1$, where $a' > 0$ is a constant (not depending on r).*
(iii) *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 (i), $I = \emptyset$ and the constant a does not depend on r. In case (ii), 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 (iii). □

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

Theorem 4.4. *Given a polynomial f, 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^* and random variables X_i in Definition 4.2. Then there exists a constant $C_f > 0$ (depending only on f) such that*

$$\mathbb{E}(f(X)) - f(\mathbf{x}^*) \leq \frac{C_f}{r}.$$

For the proof of Theorem 4.4, we need the following three technical lemmas.

Lemma 4.5. *Let k be a positive integer. There exists a constant $C_k > 0$ (depending only on k) for which the following relation holds:*

$$\frac{rp(rp+1)\cdots(rp+k-1)}{rq(rq+1)\cdots(rq+k-1)} - \frac{p^k}{q^k} \leq \frac{C_k}{r} \tag{4.7}$$

for all integers $1 \leq p < q$ and $r \geq 1$.

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 (4.7), 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:

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

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

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

Thus we get:

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

The first factor is at most 1, since one has: $p(q-p) \le q^2$, 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

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

which depends only on k. We show that

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

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

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

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

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

and thus

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

as desired. □

Lemma 4.6. *Let r be a positive integer. For any $x_i^* \in [0,1]$, we consider the parameters η_i^*, β_i^* and random variables X_i in Definition 4.2. For any integer $k \ge 1$, there exists a constant $C_k' > 0$ (depending only on k) for which the following holds:*

$$\left|\mathbb{E}(X_i^k) - (x_i^*)^k\right| \le \frac{C_k'}{r}.$$

Proof. (i) We consider first the case when $x_i^* \in (0,1)$ with $i \in J \cup I$. Then, by Definition 4.2 (i), one has $\eta_i^* = rp_i$ and $\beta_i^* = r(q_i - p_i)$, where the integers p_i, q_i satisfy $1 \le p_i < q_i$, and either $x_i^* = p_i/q_i$ if x_i^* is rational, or $|x_i^* - p_i/q_i| \le 1/(q_i r) \le 1/r$ if x_i^* is irrational. Then, by (4.2), the k-th moment of X_i is

$$\mathbb{E}(X_i^k) = \frac{\eta_i^*(\eta_i^*+1)\cdots(\eta_i^*+k-1)}{(\eta_i^*+\beta_i^*)(\eta_i^*+\beta_i^*+1)\cdots(\eta_i^*+\beta_i^*+k-1)} = \frac{rp_i(rp_i+1)\cdots(rp_i+k-1)}{rq_i(rq_i+1)\cdots(rq_i+k-1)}$$

and we obtain:

$$|\mathbb{E}(X_i^k) - (x_i^*)^k| \le \underbrace{\left| \frac{rp_i(rp_i+1)\cdots(rp_i+k-1)}{rq_i(rq_i+1)\cdots(rq_i+k-1)} - \frac{p_i^k}{q_i^k} \right|}_{=:T_1} + \underbrace{\left| \frac{p_i^k}{q_i^k} - (x_i^*)^k \right|}_{=:T_2}.$$

For the term T_1, Lemma 4.5 implies:

$$T_1 \le \frac{C_k}{r}.$$

For the term T_2, we have:

$$T_2 = \left| \frac{p_i^k}{q_i^k} - (x_i^*)^k \right| = \left| \frac{p_i}{q_i} - x_i^* \right| \cdot \left(\sum_{h=0}^{k-1} \left(\frac{p_i}{q_i} \right)^h (x_i^*)^{k-h-1} \right) \le \frac{k}{r},$$

since the first factor is at most $1/r$ and, in the second factor, each term in the summation is bounded by 1. Summarizing, we obtain: $|\mathbb{E}(X_i^k) - (x_i^*)^k| \le T_1 + T_2 \le (C_k + k)/r$.

(ii) When $x_i^* = 0$, by Definition 4.2 (ii), one has $\eta_i^* = 1$ and $\beta_i^* = r$. Thus we have:

$$\mathbb{E}(X_i^k) - (x_i^*)^k = E(X_i^k) = \frac{k!}{(r+1)(r+2)\cdots(r+k)} \le \frac{k!}{r}.$$

(iii) When $i \in I_0$, then $x_i^* \le 1/(q_i r) \le 1/r$ and, using the above inequality in (ii), we get:

$$|\mathbb{E}(X_i^k) - (x_i^*)^k| \le E(X_i^k) + (x_i^*)^k \le \frac{k!}{r} + \frac{1}{r^k} \le \frac{k!+1}{r}.$$

(iv) When $x_i^* = 1$, by Definition 4.2 (iii), one has $\eta_i^* = r$ and $\beta_i^* = 1$. Thus we have:

$$|\mathbb{E}(X_i^k) - (x_i^*)^k| = |\mathbb{E}(X_i^k) - 1| = \frac{k}{r+k} \le \frac{k}{r}.$$

(v) Finally, if $i \in I_1$, then $1 - x_i^* \le 1/(q_i r) \le 1/r$ and, using the above inequality in (iv), we get:

$$|\mathbb{E}(X_i^k) - (x_i^*)^k| \le |\mathbb{E}(X_i^k) - 1| + (1 - (x_i^*)^k) \le \frac{k}{r} + \frac{k}{r} \le \frac{2k}{r},$$

where we have used $1 - (x_i^*)^k = (1 - x_i^*) \sum_{h=0}^{k-1} (x_i^*)^{k-h-1} \le k(1 - x^*) \le k/r$.

In all cases (i)-(v), we found $\left|\mathbb{E}(X_i^k) - (x_i^*)^k\right| \le C'_k/r$, after setting $C'_k = \max\{C_k + k, k!+1, 2k\}$. □

Lemma 4.7. *For any $x, y \in \mathbb{R}^n$, one has the following equality:*

$$\prod_{i=1}^n x_i - \prod_{i=1}^n y_i = \sum_{i=1}^n \left[(x_i - y_i) \prod_{j=1}^{i-1} y_j \prod_{j=i+1}^n x_j \right].$$

Proof. Proof by direct verification. □

Now we can prove Theorem 4.4.

Proof. *(of Theorem 4.4)* We write $f(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n} f_\alpha \mathbf{x}^\alpha$. From the definition (4.5), we have

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

By Lemma 4.7, one has

$$A = \sum_{i=1}^n \left((\mathbb{E}(X_i^{\alpha_i}) - (x_i^*)^{\alpha_i}) \underbrace{\prod_{j=1}^{i-1} (x_j^*)^{\alpha_j}}_{=:B} \underbrace{\prod_{j=i+1}^{n} \mathbb{E}(X_j^{\alpha_j})}_{=:C} \right).$$

Since $x_i^* \in [0,1]$ for any $i \in [n]$, then $0 \le B \le 1$. Moreover, by Definition 4.2 and (4.2), one has that $\mathbb{E}(X_i^{\alpha_i}) \in [0,1]$ for any $i \in [n]$, and thus $0 \le C \le 1$. Combining with Lemma 4.6, we can conclude: $|A| \le (\sum_{i=1}^n C'_{\alpha_i})/r$. Therefore, we obtain that

$$\mathbb{E}(f(X)) - f(\mathbf{x}^*) \le \sum_{\alpha\in\mathbb{N}^n} |f_\alpha| \sum_{i=1}^n |\mathbb{E}(X_i^{\alpha_i}) - (x_i^*)^{\alpha_i}|,$$

where the right hand side is at most C_f/r, after setting $C_f := \sum_{\alpha\in\mathbb{N}^n} |f_\alpha| \sum_{i=1}^n C'_{\alpha_i}$. This concludes the proof. □

We can now show the following results for the rate of convergence of the sequence f_k^H.

Theorem 4.8. *Given a polynomial f, let x^* be a global minimizer of f in $[0,1]^n$ and consider as before the parameters*

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

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

$$f_k^H - f(\mathbf{x}^*) \le \frac{M_f}{\sqrt{k}} \quad \text{for all } k \ge k_1, \tag{4.8}$$

where $k_1 = \sum_{i\in J} q_i - 2|J|$ (as in Lemma 4.3 (iii)). Moreover, if f has a rational global minimizer x^*, then there exists a constant M'_f (depending only on f) such that

$$f_k^H - f(\mathbf{x}^*) \le \frac{M'_f}{k} \quad \text{for all } k \ge k_1. \tag{4.9}$$

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

Proof. Consider an integer $r \ge 1$ and, as in Definition 4.2, the parameters η_i^*, β_i^* and the random variables $X_i \sim beta(\eta_i^*, \beta_i^*)$ for $i \in [n]$. Recall also the parameter k_r from Lemma 4.3. Then, as observed earlier in (4.6), by the definition of the parameter $f_{k_r}^H$, we have that $\mathbb{E}(f(X)) \ge f_{k_r}^H$. Moreover, by Theorem 4.4, $\mathbb{E}(f(X)) - f(\mathbf{x}^*) \le C_f/r$ for some constant C_f (depending only on f). Therefore, we have the following inequality:

$$f_{k_r}^H - f(\mathbf{x}^*) \le \frac{C_f}{r}. \tag{4.10}$$

We now 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 $f_k^H - f(\mathbf{x}^*) \leq f_{k_r}^H - f(\mathbf{x}^*)$ and thus, using (4.10), $f_k^H - f(\mathbf{x}^*) \leq \frac{C_f}{r}$. We now bound $1/r$ in terms of k.

If $x^* \in \mathbb{Q}$ then, by Lemma 4.3 (i), $k_{r+1} \leq a(r+1) \leq 2ar$, which 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 (4.9).

If $x^* \notin \mathbb{Q}$ then, by Lemma 4.3 (ii), $k_{r+1} \leq a'(r+1)^2 \leq 4a'r^2$, which 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, $f_k^H - f(\mathbf{x}^*) \leq \frac{C_f}{r} \leq \frac{2\sqrt{a'}C_f}{\sqrt{k}}$, which shows that $f_k^H - f(\mathbf{x}^*) \leq \frac{M_f'}{\sqrt{k}}$ and thus (4.8), after setting $M_f' = 2\sqrt{a'}C_f$.

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 (4.8) and (4.9) hold for all $k \geq k_1$, where k_1 depends only on the rational components in $(0,1)$ of the minimizer 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 4.9. *Consider the polynomial $f = \sum_{i=1}^n x_i$ and the set $\mathbf{K} = [0,1]^n$. Then $f_{min,\mathbf{K}} = 0$ is attained at the zero vector. Using the relations* (3.1), (3.2) *and* (3.3) *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, $f_k^H - f_{\min,\mathbf{K}} = O(1/k^{1+\delta})$. Therefore, when a rational minimizer exists, the convergence rate from Theorem 4.8 in $O(1/k)$ for f_k^H is tight.*

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

As an alternative to computing f_k^H on $\mathbf{K} = Q := [0,1]^n$, 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*

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

Proof. Fix $k \geq 1$. 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 $\|\mathbf{h}\| \leq \frac{\sqrt{n}}{k}$. 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

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

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 $g_{\min,Q(k)} - g(\mathbf{x}_I^*) \leq \frac{C'}{k^2}$ for some constant C' (depending on g and thus on f). As $f_{\min,Q(k)} \leq g_{\min,Q(k)}$ 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 $f(\mathbf{x}) \leq f_k^H$

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

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

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

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

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 $f(\mathbb{E}(X)) \leq f_k^H$ 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 $f_\alpha \geq 0$ for all α. Since $\phi(X_i) = X_i^{\alpha_i}$ is convex on $[0,1]$ ($i \in [n]$), Jensen's inequality yields $\mathbb{E}(X_i^{\alpha_i}) \geq [\mathbb{E}(X_i)]^{\alpha_i}$. Thus

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

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 $\mathbf{x} \in \mathbf{K}$ from the density σ on $\mathbf{K}$ using the well-known *method of conditional distributions* (see e.g., [21, Section 8.5.1]). For $\mathbf{K} = [0,1]^n$, the procedure is described in detail in [6, Section 3]. In this way one may obtain, with high probability, a point $\mathbf{x} \in \mathbf{K}$ with $f(\mathbf{x}) \leq f_k^H + \epsilon$, 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 σ as a candidate solution; *cf.* Example 2.5. 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 behaviour of the sequences of upper bounds, and of the techniques to obtain feasible points.

7.1. **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

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

for all $(\eta, \beta) \in \mathbb{N}_k^{2n}$, and taking the minimum. (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 mentioned before, 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

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

Note that the matrices A and B depend on the moments of the Lebesgue measure on $\mathbf{K} = [0, 1]^n$, and that these moments may be computed beforehand, by assumption. One may compute $A_{\alpha,\beta}$ by taking the inner product of $(f_\delta)_{\delta \in N_f}$ with the vector of moments $\left(\int_{\mathbf{K}} \mathbf{x}^{\alpha+\beta+\delta} d\mathbf{x}\right)_{\delta \in N_f}$. 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.

7.2. **Test functions and results.** We consider several well-known polynomial test functions from global optimization (also used in [6]), that are listed in Table 1. 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).

We start by listing the upper bounds f_k^H 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 (f_k^H bounds) to SOS densities ($f_{k/2}^{sos}$ bounds); see Tables 3 and 4.

[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.

TABLE 1. Test functions

Name	Formula	Minimum ($f_{\min,\mathbf{K}}$)	Search domain ($\mathbf{K}$)
Booth Function	$f=(20x_1+40x_2-37)^2+(40x_1+20x_2-35)^2$	$f(0.55,0.65)=0$	$[0,1]^2$
Matyas Function	$f=0.26[(20x_1-10)^2+(20x_2-10)^2]-0.48(20x_1-10)(20x_2-10)$	$f(0.5,0.5)=0$	$[0,1]^2$
Motzkin Polynomial	$f=(4x_1-2)^4(4x_2-2)^2+(4x_1-2)^2(4x_2-2)^4-3(4x_1-2)^2(4x_2-2)^2+1$	$f(\frac{1}{4},\frac{1}{4})=f(\frac{1}{4},\frac{3}{4})=f(\frac{3}{4},\frac{1}{4})=f(\frac{3}{4},\frac{3}{4})=0$	$[0,1]^2$
Three-Hump Camel Function	$f=2(10x_1-5)^2-1.05(10x_1-5)^4+\frac{1}{6}(10x_1-5)^6+(10x_1-5)(10x_2-5)+(10x_2-5)^2$	$f(0.5,0.5)=0$	$[0,1]^2$
Styblinski-Tang Function	$f=\sum_{i=1}^{n}\frac{1}{2}(10x_i-5)^4-8(10x_i-5)^2+\frac{5}{2}(10x_i-5)$	$f(0.20906466,\dots,0.20906466)=-39.16599n$	$[0,1]^n$
Rosenbrock Function	$f=\sum_{i=1}^{n-1}100(4.096x_{i+1}-2.048-(4.096x_i-2.048)^2)^2+(4.096x_i-3.048)^2$	$f(\frac{3048}{4096},\dots,\frac{3048}{4096})=0$	$[0,1]^n$

TABLE 2. f_k^H for Booth, Matyas, Motzkin, Three–Hump Camel, Styblinski–Tang and Rosenbrock Functions.

k	Booth	Matyas	Motzkin	T-H. Camel	St.-Tang ($n=2$)	Rosen. ($n=2$)	Rosen. ($n=3$)	Rosen. ($n=4$)
1	280.667	17.3333	4.2000	265.77	−12.5	303.16	794.818	1289.9
2	250.667	12.0000	2.1886	86.091	−17.381	235.68	603.931	1097.7
3	214.0	11.0667	2.1886	86.091	−21.548	177.91	536.449	906.76
4	184.0	8.8000	1.2743	40.593	−26.429	148.6	478.673	839.28
5	172.0	8.1333	1.2743	40.593	−28.929	142.2	411.191	781.51
6	151.333	6.9867	1.0218	24.354	−31.429	130.43	343.863	714.02
7	143.905	6.5524	1.0218	24.354	−32.778	120.17	314.559	646.68
8	130.762	5.9048	0.8912	17.322	−34.127	103.43	296.24	579.2
9	125.429	5.6190	0.8912	17.322	−34.921	100.03	266.936	511.86
10	117.571	5.2245	0.8538	13.867	−35.714	91.011	252.003	482.56
11	109.556	5.0317	0.8538	13.867	−36.956	87.425	239.06	460.14
12	106.222	4.7778	0.8384	10.534	−38.305	76.959	225.146	430.83
13	99.4545	4.6444	0.8384	10.534	−39.516	75.033	212.057	406.9
14	94.7407	4.4741	0.8366	8.6752	−40.31	69.148	203.723	377.6
15	90.6667	4.3798	0.8339	8.6752	−41.003	66.266	189.252	362.66
16	85.6364	4.2618	0.8336	7.2466	−42.483	60.434	179.188	349.718
17	83.0909	4.1939	0.8242	7.2466	−43.694	59.243	169.714	334.462
18	78.6434	4.1102	0.8139	6.1763	−44.905	55.276	163.392	321.52
19	75.8648	4.0606	0.8062	6.1763	−45.598	52.947	155.662	309.927
20	73.5152	4.0000	0.8025	5.3826	−46.291	49.381	150.066	294.517
25	61.6535	3.4324	0.7762	4.2267	−49.633	40.704	121.272	242.747
30	53.1228	2.8927	0.7474	3.1892	−52.976	33.338	101.914	205.889
35	46.5982	2.5989	0.7067	2.7367	−55.193	28.72	86.9293	177.821
40	41.6416	2.2609	0.6625	2.2626	−57.411	24.883	75.5008	155.681
45	37.4988	2.0800	0.6254	2.0337	−58.998	21.984	67.1078	138.990
50	34.0573	1.8595	0.5914	1.7768	−60.536	19.739	59.6395	124.115

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, in most cases, $f_{k/2}^{sos} \le f_k^H$, i.e. the SOS densities usually give better bounds for a given degree, but at a higher computational cost.

Next we consider the strategies for generating feasible points corresponding to the bounds f_k^H, as described in Section 6; see Table 5.

In Table 5, 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 bounds than f_k^H, indicating that this heuristic is useful in the non-convex case.

TABLE 3. Comparison of two upper bounds for Booth, Matyas, Three–Hump Camel and Motzkin Functions

degree k	Booth		Matyas		Three–Hump Camel		Motzkin	
	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k
1		280.667		17.3333		265.77		4.2
2	244.680	250.667	8.26667	12.0	265.774	86.091	4.2	2.1886
3		214.0		11.0667		86.091		2.1886
4	162.486	184.0	5.32223	8.8000	29.0005	40.593	1.06147	1.2743
5		172.0		8.1333		40.593		1.2743
6	118.383	151.333	4.28172	6.9867	29.0005	24.354	1.06147	1.0218
7		143.905		6.5524		24.354		1.0218
8	97.6473	130.762	3.89427	5.9048	9.58064	17.322	0.829415	0.8912
9		125.429		5.6190		17.322		0.8912
10	69.8174	117.571	3.68942	5.2245	9.58064	13.867	0.801069	0.8538
11		109.556		5.0317		13.867		0.8538
12	63.5454	106.222	2.99563	4.7778	4.43983	10.534	0.801069	0.8384
13		99.4545		4.6444		10.534		0.8384
14	47.0467	94.7407	2.54698	4.4741	4.43983	8.6752	0.708889	0.8366
15		90.6667		4.3798		8.6752		0.8339
16	41.6727	85.6364	2.04307	4.2618	2.55032	7.2466	0.565553	0.8336
17		83.0909		4.1939		7.2466		0.8242
18	34.2140	78.6434	1.83356	4.1102	2.55032	6.1763	0.565553	0.8139
19		75.8648		4.0606		6.1763		0.8062
20	28.7248	73.5152	1.47840	4.0000	1.71275	5.3826	0.507829	0.8025

TABLE 4. Comparison of two upper bounds for Styblinski–Tang and Rosenbrock Functions

degree k	Sty.–Tang ($n=2$)		Rosenb. ($n=2$)		Rosenb. ($n=3$)		Rosenb. ($n=4$)	
	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k	$f^{sos}_{k/2}$	f^H_k
1		−12.5		303.16		794.818		1289.9
2	−12.9249	−17.381	214.648	235.68	629.086	603.931	1048.19	1097.7
3		−21.548		177.91		536.449		906.76
4	−25.7727	−26.429	152.310	148.6	394.187	478.673	690.332	839.28
5		−28.929		142.2		411.191		781.51
6	−34.4030	−31.429	104.889	130.43	295.811	343.863	536.367	714.02
7		−32.778		120.17		314.559		646.68
8	−41.4436	−34.127	75.6010	103.43	206.903	296.24	382.729	579.2
9		−34.921		100.03		266.936		511.86
10	−45.1032	−35.714	51.5037	91.011	168.135	252.003	314.758	482.56
11		−36.956		87.425		239.06		460.14
12	−51.0509	−38.305	41.7878	76.959	121.558	225.146	236.709	430.83
13		−39.516		75.033		212.057		406.9
14	−56.4050	−40.31	30.1392	69.148	101.953	203.723	202.674	377.6
15		−41.003		66.266		189.252		362.66
16	−58.6004	−42.483	25.8329	60.434	77.4797	179.188	156.295	349.718
17		−43.694		59.243		169.714		334.462
18	−60.7908	−44.905	19.4972	55.276	66.6954	163.392	137.015	321.52

As a final comparison, we also look at the general sampling technique via the method of conditional distributions; see Tables 6 and 7. 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 6 and 7 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 5. 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_k^H	$f(\hat{\mathbf{x}})$	$f(\mathbb{E}(X))$	f_k^H	$f(\hat{\mathbf{x}})$	$f(\mathbb{E}(X))$	f_k^H*	$f(\hat{\mathbf{x}})$	f_k^H	$f(\hat{\mathbf{x}})$
1	280.667	—	2.8889	17.3333	—	0	4.2000	—	265.77	—
2	250.667	—	9.0	12.0000	4.0	0.4444	2.1886	—	86.091	—
3	214.0	194.0	2.8889	11.0667	4.0	1.3889	2.1886	—	86.091	—
4	184.0	194.0	9.0	8.8000	4.0	1.0	1.2743	1.0	40.593	—
5	172.0	96.222	17.0	8.1333	4.0	1.460	1.2743	1.0	40.593	—
6	151.333	96.222	18.0	6.9867	4.0	1.440	1.0218	1.0	24.354	—
7	143.905	96.222	24.222	6.5524	4.0	1.7156	1.0218	1.0	24.354	—
8	130.762	122.0	16.204	5.9048	4.0	1.7778	0.8912	1.0	17.322	—
9	125.429	26.0	2.9796	5.6190	4.0	1.9637	0.8912	1.0	17.322	25.0
10	117.571	96.222	25.806	5.2245	4.0	2.0408	0.8538	1.0	13.867	—
11	109.556	26.0	2.9796	5.0317	4.0	2.1760	0.8538	1.0	13.867	25.0
12	106.222	42.889	9.0	4.7778	4.0	2.2500	0.8384	1.0	10.534	0
13	99.4545	26.0	2.9796	4.6444	4.0	2.3534	0.8384	1.0	10.534	0
14	94.7407	13.592	0.91358	4.4741	4.0	2.4198	0.8366	1.0	8.6752	0
15	90.6667	27.580	7.6777	4.3798	4.0	2.5017	0.8339	1.0	8.6752	0.273
16	85.6364	9.0	2.0	4.2618	4.0	2.5600	0.8336	1.0	7.2466	0
17	83.0909	17.210	4.5785	4.1939	4.0	2.6268	0.8242	1.0	7.2466	0
18	78.6434	9.0	2.0	4.1102	4.0	2.6777	0.8139	1.0	6.1763	0
19	75.8648	5.951	0.35445	4.0606	4.0	2.7332	0.8062	1.0	6.1763	0.209
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

TABLE 6. Sampling results for Motzkin Polynomial

		Sample size 10			Sample size 100		
k	f_k^H	Mean	Variance	Minimum	Mean	Variance	Minimum
1	4.2000	6.2601	66.2605	0.6183	6.5027	188.1445	0.0060
2	2.1886	1.4972	1.6084	0.9158	1.8377	12.5387	0.0657
3	2.1886	1.9658	5.0427	0.0644	2.8413	68.2093	0.0036
4	1.2743	1.1776	1.8501	0.0421	0.8571	0.6764	0.0042
5	1.2743	0.8330	0.0466	0.2790	1.1590	4.2023	0.0525
6	1.0218	1.7002	6.2647	0.3196	0.9336	0.8998	0.0002
7	1.0218	0.8350	0.1672	0.2416	0.9863	1.3777	0.0070
8	0.8912	0.6108	0.1451	0.0218	0.8431	1.4834	0.0070
9	0.8912	0.7545	0.0679	0.1656	0.8879	0.2752	0.0175
10	0.8538	0.7005	0.0800	0.1862	0.8435	0.1448	0.1149
11	0.8538	0.8244	0.0779	0.1123	0.8673	0.2565	0.1100
12	0.8384	0.8912	0.0213	0.5919	0.7835	0.2554	0.0188
13	0.8384	0.8286	0.0412	0.3205	0.7664	0.0714	0.0112
14	0.8366	0.7698	0.0781	0.2083	0.9574	1.2157	0.0778
15	0.8339	0.9063	0.0153	0.6069	0.8465	0.0932	0.0593
16	0.8336	0.7482	0.0750	0.1759	0.7209	0.0875	0.0648
17	0.8242	0.7430	0.0706	0.1500	0.8051	0.0718	0.0984
18	0.8139	0.8546	0.0493	0.4460	0.7749	0.0785	0.0038
19	0.8062	0.6621	0.0892	0.1836	0.7850	0.1273	0.0408
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

8. CONCLUDING REMARKS

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

TABLE 7. Sampling results for Three-Hump Camel function

		Sample size 10			Sample size 100		
k	f_k^H	Mean	Variance	Minimum	Mean	Variance	Minimum
1	265.77	359.98	274477.0	2.4493	300.34	245144.0	0.011095
2	86.091	88.717	24117.0	1.1729	122.12	76646.0	0.082513
3	86.091	14.712	186.23	2.219	58.186	15987.0	0.492
4	40.593	55.091	19297.0	0.10296	44.844	21297.0	0.19439
5	40.593	91.872	27065.0	0.90053	53.656	14575.0	0.58086
6	24.354	12.961	77.377	0.8186	34.115	7862.5	0.019021
7	24.354	33.96	1745.4	0.65266	27.072	10632.0	0.33813
8	17.322	10.029	60.746	1.0931	12.307	314.46	0.074663
9	17.322	9.4932	100.22	0.0027565	20.185	7279.8	0.11239
10	13.867	11.312	45.784	0.8916	14.273	382.98	0.018985
11	13.867	8.3991	87.108	0.0031527	11.928	357.45	0.01384
12	10.534	5.013	52.681	0.30303	12.377	547.42	0.25952
13	10.534	14.281	401.82	0.52373	7.8673	253.02	0.11989
14	8.6752	5.2897	43.81	0.3909	9.4462	362.49	0.051331
15	8.6752	5.6281	31.311	0.21853	10.373	778.32	0.022282
16	7.2466	9.5801	95.901	1.7112	6.465	122.72	0.013084
17	7.2466	5.2511	23.863	2.0409	6.0633	56.495	0.18354
18	6.1763	6.0327	34.298	0.85182	5.2985	35.953	0.071544
19	6.1763	5.3006	52.994	0.6699	5.0383	41.619	0.040785
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

- 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

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

(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 (1.2)), 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

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

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:

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

Thus, the idea is to replace the density $\sigma(x) = \mathbf{x}^\eta(1-\mathbf{x})^\beta / \int_{\mathbf{K}} \mathbf{x}^\eta(1-\mathbf{x})^\beta d\mathbf{x}$ by the density $\sigma(x)^r / \int_{\mathbf{K}} \sigma(r)^r d\mathbf{x}$ 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 $K = [0,1]^n$, one could allow an arbitrary $r > 0$ since the moments are still available as pointed out above.)

In Tables 8, 9, and 10, we show some numerical values for the parameter $f_{r,k}^H$.

TABLE 8. $f_{r,k}^H$ for the Styblinski-Tang function ($n = 2$)

k	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$
1	−12.5	−10.06	−8.3333	−8.3333	−8.3333
2	−17.381	−17.857	−16.919	−15.793	−14.744
3	−21.548	−21.686	−22.582	−23.179	−23.62
4	−26.429	−27.381	−28.256	−30.263	−31.736
5	−28.929	−31.209	−31.167	−32.872	−34.435
6	−31.429	−35.038	−36.842	−38.025	−38.906
7	−32.778	−38.76	−42.505	−45.109	−47.022
8	−34.127	−42.483	−48.179	−52.193	−55.138
9	−34.921	−44.387	−50.577	−54.802	−57.837
10	−35.714	−46.291	−52.976	−57.411	−60.536

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 8. 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 9.

TABLE 9. $f_{r,k}^H$ for the Rosenbrock function $(n = 3)$

k	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$
1	794.818	727.337	698.032	683.822	676.526
2	603.931	512.228	473.974	454.193	443.769
3	536.449	449.625	398.566	367.869	350.671
4	478.673	368.873	294.499	253.135	227.526
5	411.191	274.121	235.89	228.906	232.996
6	343.863	225.146	191.935	151.455	119.98
7	314.559	225.768	166.179	128.62	106.417
8	296.24	198.861	144.94	111.721	88.0661
9	266.936	185.145	133.379	103.162	84.3506
10	252.003	158.448	111.33	87.8805	70.0394

TABLE 10. $f_{r,k}^H$ for the Rosenbrock function $(n = 4)$

k	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$
1	1289.9	1223.8	1194.5	1180.3	1173.0
2	1097.7	1006.9	968.53	948.71	938.29
3	906.76	790.03	742.57	717.15	703.61
4	839.28	727.43	669.06	632.76	612.44
5	781.51	606.15	502.34	446.68	413.72
6	714.02	515.76	397.34	330.93	289.5
7	646.68	421.01	338.74	306.71	294.97
8	579.2	371.11	293.83	229.25	181.95
9	511.86	331.44	269.02	206.42	168.39
10	482.56	323.69	246.84	189.36	149.9

On the other hand, it is clear from Tables 8, 9, and 10 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 $\mathbf{K} = [0,1]^n$ 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 4.2: $\eta_i^* = rp_i + 1$ 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 (4.6) and Theorem 4.4, we can conclude that the following inequality holds:

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

where C_f is a constant that depends on f only.

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

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

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 the second author was funded by by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 666981 TAMING).

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; 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`

Tilburg University; PO Box 90153, 5000 LE Tilburg, The Netherlands.
E-mail address: `Z.Sun@uvt.nl`