Comparison of Lasserre's Measure-based Bounds for Polynomial Optimization to Bounds Obtained by Simulated Annealing

de Klerk, Etienne; Laurent, Monique

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Comparison of Lasserre’s measure–based bounds for polynomial optimization to bounds obtained by simulated annealing

Etienne de Klerk * Monique Laurent †

March 3, 2017

Abstract

We consider the problem of minimizing a continuous function $f$ over a compact set $K$. We compare the hierarchy of upper bounds proposed by Lasserre in [SIAM J. Optim. 21(3) (2011), pp. 864 – 885] to bounds that may be obtained from simulated annealing.

We show that, when $f$ is a polynomial and $K$ a convex body, this comparison yields a faster rate of convergence of the Lasserre hierarchy than what was previously known in the literature.

Keywords: Polynomial optimization; Semidefinite optimization; Lasserre hierarchy; simulated annealing

AMS classification: 90C22; 90C26; 90C30

1 Introduction

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

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

Our goal is to compare two convergent hierarchies of upper bounds on $f_{\min,K}$, namely measure-based bounds introduced by Lasserre [10], and simulated annealing bounds, as studied by Kalai and Vempala [6]. The bounds of Lasserre are obtained by minimizing over measures on $K$ with sum-of-squares polynomial density functions with growing degrees, while simulated annealing bounds use Boltzman distributions on $K$ with decreasing temperature parameters.

In this note we establish a relationship between these two approaches, linking the degree and temperature parameters in the two bounds (see Theorem 4.1 for a precise statement). As an application, when $f$ is a polynomial and $K$ is a convex body, we can show a faster convergence rate for the measure-based bounds of Lasserre. The new convergence rate is in $O(1/r)$ (see Corollary 4.3),

*Tilburg University and Delft University of Technology, E.deKlerk@uvt.nl
†Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University, monique@cwi.nl
where $2r$ is the degree of the sum-of-squares polynomial density function, while the dependence was in $O(1/\sqrt{r})$ in the previously best known result from [4].

Polynomial optimization is a very active research area in the recent years since the seminal works of Lasserre [8] and Parrilo [13] (see also, e.g., the book [9] and the survey [11]). In particular, hierarchies of (lower and upper) bounds for the parameter $f_{\min,K}$ have been proposed, based on sum-of-squares polynomials and semidefinite programming.

For a general compact set $K$, upper bounds for $f_{\min,K}$ have been introduced by Lasserre [10], obtained by searching for a sum-of-squares polynomial density function of given maximum degree $2r$, so as to minimize the integration of $f$ with respect to the corresponding probability measure on $K$. When $f$ is Lipschitz continuous and under some mild assumption on $K$ (which holds, e.g., when $K$ is a convex body), estimates for the convergence rate of these bounds have been proved in [4] that are in order $O(1/\sqrt{r})$. Improved rates have been subsequently shown when restricting to special sets $K$. Related stronger results have been shown for the case when $K$ is the hypercube $[0,1]^n$ or $[-1,1]^n$. In [3] the authors show a hierarchy of upper bounds using the Beta distribution, with the same convergence rate in $O(1/\sqrt{r})$, but whose computation needs only elementary operations; moreover an improved convergence in $O(1/r)$ can be shown, e.g., when $f$ is quadratic. In addition, a convergence rate in $O(1/r^2)$ is shown in [2], using distributions based on Jackson kernels and a larger class of sum-of-squares density functions.

In this paper we investigate the hierarchy of measure-based upper bounds of [10] and show that when $K$ is a convex body, convexity can be exploited to show an improved convergence rate in $O(1/r)$, even for nonconvex functions. The key ingredient for this is to establish a relationship with upper bounds based on simulated annealing and to use a known convergence rate result from [6] for simulated annealing bounds in the convex case.

Simulated annealing was introduced by Kirkpatrick et al. [7] as a randomized search procedure for general optimization problems. It has enjoyed renewed interest for convex optimization problems since it was shown by Kalai and Vempala [6] that a polynomial-time implementation is possible. This requires so-called hit-and-run sampling from $K$, as introduced by Smith [14], that was shown to be a polynomial-time procedure by Lovász [12]. Most recently, Abernethy and Hazan [1] showed formal equivalence with a certain interior point method for convex optimization.

This unexpected equivalence between seemingly different methods has motivated this current work to relate the bounds by Lasserre [10] to the simulating annealing bounds as well.

In what follows, we first introduce the measure-based upper bounds of Lasserre [10]. Then we recall the bounds based on simulated annealing and the known convergence results for a linear objective function $f$, and we give an explicit proof of their extension to the case of a general convex function $f$. After that we state our main result and the next section is devoted to its proof. In the last section we conclude with numerical examples showing the quality of the two types of bounds and some final remarks.

## 2 Lasserre’s hierarchy of upper bounds

Throughout, $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ is the set of polynomials in $n$ variables with real coefficients and, for an integer $r \in \mathbb{N}$, $\mathbb{R}[x]_r$ is the set of polynomials with degree at most $r$. Any polynomial $f \in \mathbb{R}[x]_r$ can be written $f = \sum_{\alpha \in \mathbb{N}(n,r)} f_\alpha x^\alpha$, where we set $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$ for $\alpha \in \mathbb{N}^n$ and
\( \{ \alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \leq r \} \). We let \( \Sigma[x] \) denote the set of sums of squares of polynomials, and \( \Sigma[x]_r = \Sigma[x] \cap \mathbb{R}[x]_{2r} \) consists of all sums of squares of polynomials with degree at most \( 2r \).

We recall the following reformulation for \( f_{\min,K} \), established by Lasserre [10]:

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

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

\[
f^{(r)}_K := \inf_{h \in \Sigma[x], r} \int_K h(x)f(x)dx \quad \text{s.t.} \quad \int_K h(x)dx = 1. \tag{1}
\]

Clearly, the inequality \( f_{\min,K} \leq f^{(r)}_K \) holds for all \( r \in \mathbb{N} \). Lasserre [10] gave conditions under which the infimum is attained in the program (1). De Klerk, Laurent and Sun [4, Theorem 3] established the following rate of convergence for the bounds \( f^{(r)}_K \).

**Theorem 2.1** (De Klerk, Laurent, and Sun [4]). Let \( f \in \mathbb{R}[x] \) and \( K \) a convex body. There exist constants \( C_{f,K} \) (depending only on \( f \) and \( K \)) and \( r_K \) (depending only on \( K \)) such that

\[
f_K^{(r)} - f_{\min,K} \leq C_{f,K} \frac{1}{\sqrt{r}} \quad \text{for all} \quad r \geq r_K. \tag{2}
\]

That is, the following asymptotic convergence rate holds: \( f_K^{(r)} - f_{\min,K} \simeq O \left( \frac{1}{\sqrt{r}} \right) \).

This result of [4] holds in fact under more general assumptions, namely when \( f \) is Lipschitz continuous and \( K \) satisfies a technical assumption (Assumption 1 in [4]), which says (roughly) that around any point in \( K \) there is a ball whose intersection with \( K \) is at least a constant fraction of the unit ball.

As explained in [10] the parameter \( f_K^{(r)} \) can be computed using semidefinite programming, assuming one knows the moments \( m_\alpha(K) \) of the Lebesgue measure on \( K \), where

\[
m_\alpha(K) := \int_K x^\alpha dx \quad \text{for} \quad \alpha \in \mathbb{N}^n.
\]

Indeed suppose \( f(x) = \sum_{\beta \in N(n,d)} f_\beta x^\beta \) has degree \( d \). Writing \( h(x) = \sum_{\alpha \in N(n,2r)} h_\alpha x^\alpha \), the parameter \( f_K^{(r)} \) from (1) can be reformulated as follows:

\[
f_K^{(r)} = \min_{\beta \in N(n,d)} \sum_{\alpha \in N(n,2r)} h_\alpha m_{\alpha+\beta}(K) \quad \text{s.t.} \quad \sum_{\alpha \in N(n,2r)} h_\alpha m_\alpha(K) = 1, \quad \sum_{\alpha \in N(n,2r)} h_\alpha x^\alpha \in \Sigma[x]_r.
\]

Since the sum-of-squares condition on \( h \) may be written as a linear matrix inequality, this is a semidefinite program. In fact, since it only has one linear equality constraint, it may even be
rewritten as a generalised eigenvalue problem. In particular, $f^{(r)}_{\mathbb{K}}$ is equal to the smallest generalised eigenvalue of the system:

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

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

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

(4)

For more details, see [10, 4, 3].

3 Bounds from simulated annealing

Given a continuous function $f$, consider the associated Boltzman distribution over the set $\mathbb{K}$, defined by the density function:

$$P_{f}(x) := \frac{\exp(-f(x))}{\int_{\mathbb{K}} \exp(-f(x')) dx'}.$$

Write $X \sim P_{f}$ if the random variable $X$ takes values in $\mathbb{K}$ according to the Boltzman distribution.

The idea of simulated annealing is to sample $X \sim P_{f/t}$ where $t > 0$ is a fixed ‘temperature’ parameter, that is subsequently decreased. Clearly, for any $t > 0$, we have

$$f_{\min,\mathbb{K}} \leq E_{X \sim P_{f/t}}[f(X)].$$

(5)

The point is that, under mild assumptions, these bounds converge to the minimum of $f$ over $\mathbb{K}$ (see, e.g., [15]):

$$\lim_{t \downarrow 0} E_{X \sim P_{f/t}}[f(X)] = f_{\min,\mathbb{K}}.$$

The key step in the practical utilization of these bounds is therefore to perform the sampling of $X \sim P_{f/t}$.

Example 3.1. Consider the minimization of the Motzkin polynomial

$$f(x_1, x_2) = 64(x_1^4x_2^2 + x_2^4x_1^2) - 48x_1^2x_2^2 + 1$$

over $\mathbb{K} = [-1,1]^2$, where there are four global minimizers at the points $(\pm \frac{1}{2}, \pm \frac{1}{2})$, and $f_{\min,\mathbb{K}} = 0$. Figure 1 shows the corresponding Boltzman density function for $t = \frac{1}{2}$. Note that this density has four modes, roughly positioned at the four global minimizers of $f$ in $[-1,1]^2$. The corresponding upper bound on $f_{\min,\mathbb{K}} = 0$ is $E_{X \sim P_{f/t}}[f(X)] \approx 0.7257$ ($t = \frac{1}{2}$).

To obtain a better upper bound on $f_{\min,\mathbb{K}}$ from the Lasserre hierarchy, one needs to use a degree 14 s.o.s. polynomial density; in particular, one has $f^{(6)}_{\mathbb{K}} = 0.8010$ (degree 12) and $f^{(7)}_{\mathbb{K}} = 0.7088$ (degree 14). More detailed numerical results are given in Section 5.

When $f$ is linear and $\mathbb{K}$ a convex body, Kalai and Vempala [6, Lemma 4.1] show that the rate of convergence of the bounds in (5) is linear in the temperature $t$. 
Figure 1: Graph and contours of the Boltzman density with $t = \frac{1}{2}$ for the Motzkin polynomial.

**Theorem 3.2** (Kalai and Vempala [6]). Let $f(x) = c^T x$ where $c$ is a unit vector, and let $K$ be a convex body. Then, for any $t > 0$, we have

$$E_{X \sim \mathcal{P}_{f/t}} [f(X)] - \min_{x \in K} f(x) \leq nt.$$  

We indicate how to extend the result of Kalai and Vempala in Theorem 3.2 to the case of an arbitrary convex function $f$. This more general result is hinted at in §6 of [6], where the authors write

“... a statement analogous to [Theorem 2] holds also for general convex functions ...”

but no precise statement is given there. In any event, as we will now show, the more general result may readily be derived from Theorem 3.2 (in fact, from the special case of a linear coordinate function $f(x) = x_i$ for some $i$).

**Corollary 3.3.** Let $f$ be a convex function and let $K \subseteq \mathbb{R}^n$ be a convex body. Then, for any $t > 0$, we have

$$E_{X \sim \mathcal{P}_{f/t}} [f(X)] - \min_{x \in K} f(x) \leq nt.$$  

Proof. Set

$$E_K := E_{X \sim \mathcal{P}_{f/t}} [f(X)] = \frac{\int_{K} f(x) e^{-f(x)/t} dx}{\int_{K} e^{-f(x)/t} dx}.$$  

Then we have

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

Define the set

$$\tilde{K} := \{(x, x_{n+1}) \in \mathbb{R}^{n+1} : x \in K, f(x) \leq x_{n+1} \leq E_K\}.$$  

Then $\tilde{K}$ is a convex body and we have

$$\min_{x \in K} f(x) = \min_{(x, x_{n+1}) \in \tilde{K}} x_{n+1}.$$  

Accordingly, define the parameter

$$E_{\tilde{K}} := \frac{\int_{\tilde{K}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} dx}{\int_{\tilde{K}} e^{-x_{n+1}/t} dx_{n+1} dx}.$$  

5
Corollary 3.3 will follow if we show that

$$E_{\hat{K}} = E_K + t. \quad (6)$$

To this end set \( E_K = \frac{N_K}{D_K} \) and \( E_{\hat{K}} = \frac{N_{\hat{K}}}{D_{\hat{K}}} \), where we define

$$N_K := \int_K f(x)e^{-f(x)/t}dx, \quad D_K := \int_K e^{-f(x)/t}dx,$$

$$N_{\hat{K}} := \int_{\hat{K}} x_{n+1}e^{-x_{n+1}/t}dx_{n+1}dx, \quad D_{\hat{K}} := \int_{\hat{K}} e^{-x_{n+1}/t}dx_{n+1}dx.$$ We work out the parameters \( N_{\hat{K}} \) and \( D_{\hat{K}} \) (taking integrations by part):

$$D_{\hat{K}} = \int_{\hat{K}} \left( \int_{f(x)}^{E_K} e^{-x_{n+1}/t}dx_{n+1} \right) dx = \int_{\hat{K}} \left( te^{-f(x)/t} - te^{-E_K/t} \right) dx = tD_K - te^{-E_K/t}\text{vol}(K),$$

$$N_{\hat{K}} = \int \left( \int_{f(x)}^{E_K} x_{n+1}e^{-x_{n+1}/t}dx_{n+1} \right) dx = \int \left( -te^{-E_K/t} + tf(x)e^{-f(x)/t} + t \int_{f(x)}^{E_K} e^{-x_{n+1}/t}dx_{n+1} \right) dx = -te^{-E_K/t}\text{vol}(K) + tN_K + tD_{\hat{K}}.$$

Then, using the fact that \( E_K = \frac{N_K}{D_K} \), we obtain:

$$\frac{N_{\hat{K}}}{D_{\hat{K}}} = t + \frac{N_K - E_K e^{-E_K/t}\text{vol}(K)}{D_K - e^{-E_K/t}\text{vol}(K)} = t + \frac{N_K}{D_K},$$

which proves relation (6).

We can now derive the result of Corollary 3.3. Indeed, using Theorem 2 applied to \( \hat{K} \) and the linear function \( x_{n+1} \), we get

$$E_{X \sim P_{f/t}} [f(X)] - \min_{x \in K} f(x) = E_K - \min_{x \in K} f(x) = (E_{\hat{K}} - \min_{x,x_{n+1} \in \hat{K}} x_{n+1}) + (E_K - E_{\hat{K}}) \leq t(n+1) - t = tn.$$

We can now derive the result of Corollary 3.3. Indeed, using Theorem 2 applied to \( \hat{K} \) and the linear function \( x_{n+1} \), we get

$$E_{X \sim P_{f/t}} [f(X)] - \min_{x \in K} f(x) = E_K - \min_{x \in K} f(x) = (E_{\hat{K}} - \min_{x,x_{n+1} \in \hat{K}} x_{n+1}) + (E_K - E_{\hat{K}}) \leq t(n+1) - t = tn.$$
4 Main results

We will prove the following relationship between the sum-of-squares based upper bound (1) of Lasserre and the bound (5) based on simulated annealing.

**Theorem 4.1.** Let $f$ be a polynomial of degree $d$, let $K$ be a compact set and set $\hat{f}_{\text{max}} = \max_{x \in K} |f(x)|$. Then we have

$$f^{(rd)}/K \leq \mathbb{E}_{X \sim P_{f/t}} [f(X)] + \frac{\hat{f}_{\text{max}}}{2^r} \quad \text{for any integer } r \geq e \cdot \frac{\hat{f}_{\text{max}}}{t} \text{ and any } t > 0.$$  

For the problem of minimizing a convex polynomial function over a convex body, we obtain the following improved convergence rate for the sum-of-squares based bounds of Lasserre.

**Corollary 4.2.** Let $f \in \mathbb{R}[x]$ be a convex polynomial of degree $d$ and let $K$ be a convex body. Then for any integer $r \geq 1$ one has

$$f^{(rd)}/K - \min_{x \in K} f(x) \leq \frac{cr}{r},$$

for some constant $c > 0$ that does not depend on $r$. (For instance, $c = (ne + 1)\hat{f}_{\text{max}}$.)

**Proof.** Let $r \geq 1$ and set $t = \frac{e \cdot \hat{f}_{\text{max}}}{r}$. Combining Theorems 3.2 and 4.1, we get

$$f^{(rd)}/K - \min_{x \in K} f(x) = \left( f^{(rd)}/K - \mathbb{E}_{X \sim P_{f/t}} [f(X)] \right) + \left( \mathbb{E}_{X \sim P_{f/t}} [f(X)] - f_{\text{min},K} \right) \leq \frac{\hat{f}_{\text{max}}}{2^r} + \frac{ne \cdot \hat{f}_{\text{max}}}{r} \leq \frac{(ne + 1)\hat{f}_{\text{max}}}{r}.$$  

For convex polynomials $f$, this improves on the known $O(1/\sqrt{r})$ result from Theorem 2.1. One may in fact use the last corollary to obtain the same rate of convergence in terms of $r$ for all polynomials, without the convexity assumption, as we will now show.

**Corollary 4.3.** If $f$ be a polynomial and $K$ a convex body, then there is a $c > 0$ depending on $f$ and $K$ only, so that

$$f^{(2r)}/K - \min_{x \in K} f(x) \leq \frac{c}{r}.$$  

A suitable value for $c$ is

$$c = (ne + 1) \left( f_{\text{min},K} + C^{1}_{f} \cdot \text{diam}(K) + C^{2}_{f} \cdot \text{diam}(K)^2 \right),$$

where $C^{1}_{f} = \max_{x \in K} \|\nabla f(x)\|_2$ and $C^{2}_{f} = \max_{x \in K} \|\nabla^2 f(x)\|_2$.

We first define a convex quadratic function $q$ that upper bounds $f$ on $K$ as follows:

$$q(x) = f(a) + \nabla f(a)^\top (x - a) + C^{2}_{f} \|x - a\|_2^2,$$

where $C^{2}_{f} = \max_{x \in K} \|\nabla^2 f(x)\|_2$, and $a$ is the minimizer of $f$ on $K$. Note that $q(x) \geq f(x)$ for all $x \in K$ by Taylor’s theorem, and $\min_{x \in K} q(x) = f(a)$. 

7
By definition of the Lasserre hierarchy,
\[
\inf_{h \in \Sigma_{2r}^n} \int_K h(x)f(x)dx \quad \text{s.t.} \quad \int_K h(x)dx = 1
\leq \inf_{h \in \Sigma_{2r}^n} \int_K h(x)q(x)dx \quad \text{s.t.} \quad \int_K h(x)dx = 1
\equiv \varphi^{(2r)}_K.
\]

Invoking Corollary 4.2 and using that the degree of \( q \) is 2, we obtain:
\[
\inf_{h \in \Sigma_{2r}^n} \int_K h(x)q(x)dx \quad \text{s.t.} \quad \int_K h(x)dx = 1
\equiv \varphi^{(2r)}_K.
\]

The last result improves on the known \( O\left(\frac{1}{\sqrt{r}}\right) \) rate in Theorem 2.1.

**Proof of Theorem 4.1**

The key idea in the proof of Theorem 4.1 is to replace the Boltzman density function by a polynomial approximation.

To this end, we first recall a basic result on approximating the exponential function by its truncated Taylor series.

**Lemma 4.4** (De Klerk, Laurent and Sun [4]). Let \( \phi_{2r}(\lambda) \) denote the (univariate) polynomial of degree \( 2r \) obtained by truncating the Taylor series expansion of \( e^{-\lambda} \) at the order \( 2r \). That is,
\[
\phi_{2r}(\lambda) := \sum_{k=0}^{2r} \frac{(-\lambda)^k}{k!}.
\]

Then \( \phi_{2r} \) is a sum of squares of polynomials. Moreover, we have
\[
0 \leq \phi_{2r}(\lambda) - e^{-\lambda} \leq \frac{\lambda^{2r+1}}{(2r + 1)!} \quad \text{for all} \quad \lambda \geq 0. \tag{7}
\]

We now define the following approximation of the Boltzman density \( P_{f/t} \):
\[
\varphi_{2r,t}(x) := \frac{\phi_{2r}(f(x)/t)}{\int_K \phi_{2r}(f(x)/t)dx} \tag{8}
\]

By construction, \( \varphi_{2r,t} \) is a sum-of-squares polynomial probability density function on \( K \), with degree \( 2rd \) if \( f \) is a polynomial of degree \( d \). Moreover, by relation (7) in Lemma 4.4, we obtain
\[
\varphi_{2r,t}(x) \leq \frac{\phi_{2r}(f(x)/t)}{\int_K \exp(-f(x)/t)dx} \leq P_{f/t}(x) + \frac{(f(x)/t)^{2r+1}}{(2r + 1)! \int_K \exp(-f(x)/t)dx}. \tag{9}
\]

From this we can derive the following result.
Lemma 4.5. For any continuous $f$ and scalar $t > 0$ one has

$$
\frac{f^{(rd)}}{\mathbf{K}} \leq \int_{\mathbf{K}} f(x) \varphi_{2r,t}(x) \, dx \leq \mathbb{E}_{X \sim P_{f/t}} [f(X)] + \frac{\int_{\mathbf{K}} (f(x) - f_{\min,K})(f(x))^{2r+1} \, dx}{t^{2r+1}(2r+1)! \int_{\mathbf{K}} \exp(-f(x)/t) \, dx}. \tag{11}
$$

Proof. As $\varphi_{2r,t}(x)$ is a polynomial of degree $2rd$ and a probability density function on $\mathbf{K}$ (by (8)), we have:

$$
\frac{f^{(rd)}}{\mathbf{K}} \leq \int_{\mathbf{K}} f(x) \varphi_{2r,t}(x) \, dx = \int_{\mathbf{K}} (f(x) - f_{\min,K}) \varphi_{2r,t}(x) \, dx + f_{\min,K}. \tag{12}
$$

Using the above inequality (10) for $\varphi_{2r,t}(x)$ we can upper bound the integral on the right hand side:

$$
\int_{\mathbf{K}} (f(x) - f_{\min,K}) \varphi_{2r,t}(x) \, dx \leq \int_{\mathbf{K}} (f(x) - f_{\min,K}) P_{f/t}(x) \, dx + \int_{\mathbf{K}} (f(x) - f_{\min,K})(f(x)/t)^{2r+1} \, dx
\leq \mathbb{E}_{X \sim P_{f/t}} [f(X)] - f_{\min,K} + \int_{\mathbf{K}} (f(x) - f_{\min,K})(f(x)/t)^{2r+1} \, dx.
$$

Combining with the inequality (12) gives the desired result. \qed

We now proceed to the proof of Theorem 4.1. In view of Lemma 4.5, we only need to bound the last right-hand-side term in (11):

$$
T := \frac{\int_{\mathbf{K}} (f(x) - f_{\min,K})(f(x))^{2r+1} \, dx}{t^{2r+1}(2r+1)! \int_{\mathbf{K}} \exp(-f(x)/t) \, dx}
$$

and to show that $T \leq \hat{\max}_{f_{\max}}$.

By the definition of $\hat{\max}_{f_{\max}}$ we have

$$(f(x) - f_{\min})(f(x))^{2r+1} \leq 2 \hat{\max}_{f_{\max}}^{2(r+1)} \text{ and } \exp(-f(x)/t) \geq \exp(\hat{\max}_{f_{\max}}/t) \text{ on } \mathbf{K},$$

which implies

$$
T \leq \frac{2 \hat{\max}_{f_{\max}}^{2(r+1)} \exp(\hat{\max}_{f_{\max}}/t)}{t^{2r+1}(2r+1)!}.
$$

Combining with the Stirling approximation inequality,

$$
r! \geq \sqrt{2\pi r} \left( \frac{r}{e} \right)^r \quad (r \in \mathbb{N}),
$$

applied to $(2r + 1)!$, we obtain:

$$
T \leq \frac{2 \hat{\max}_{f_{\max}}}{\sqrt{2\pi(2r + 1)}} \left( \frac{\hat{\max}_{f_{\max}} e}{t(2r + 1)} \right)^{2r+1} \exp(\hat{\max}_{f_{\max}}/t).
$$
Consider \( r \geq \frac{e}{2} \hat{f}_{\text{max}} \), so that \( \hat{f}_{\text{max}} / t \leq \frac{r}{e} \). Then, using the fact that \( r / (2r + 1) \leq 1/2 \), we obtain

\[
T \leq \frac{2 \hat{f}_{\text{max}} \exp(r/e)}{\sqrt{2\pi}} \left( \frac{r}{2r + 1} \right)^{2r+1} \leq \frac{\exp(1/e)}{\sqrt{2\pi}} \left( \frac{r}{2r + 1} \right)^{r} \leq \frac{\hat{f}_{\text{max}}}{\sqrt{2\pi}} \left( \frac{\exp(1/e)}{4} \right)^{r} \leq \frac{\hat{f}_{\text{max}}}{2r}.
\]

This concludes the proof of Theorem 4.1.

5 Concluding remarks

We conclude with a numerical comparison of the two hierarchies of bounds. By Theorem 4.1, it is reasonable to compare the bounds \( f^{(r)}_{\text{K}} \) and \( \mathbb{E}_{X \sim P_{f,t}}[f(X)] \), with \( t = \frac{e d \hat{f}_{\text{max}}}{r} \) and \( d \) the degree of \( f \). Thus we define, for the purpose of comparison:

\[
SA^{(r)} = \mathbb{E}_{X \sim P_{f,t}}[f(X)], \quad \text{with } t = \frac{e d \hat{f}_{\text{max}}}{r}.
\]

We calculated the bounds for the polynomial test functions listed in Table 1.

Table 1: Test functions, all with \( n = 2 \), domain \( K = [-1,1]^2 \), and minimum \( f_{\text{min,K}} = 0 \).

<table>
<thead>
<tr>
<th>Name</th>
<th>( f(x) )</th>
<th>( \hat{f}_{\text{max}} )</th>
<th>( d )</th>
<th>Convex?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booth function</td>
<td>((10x_1 + 20x_2 - 7)^2 + (20x_1 + 10x_2 - 5)^2)</td>
<td>2594</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>Matyas function</td>
<td>(26(x_1^2 + x_2^2) - 48x_1 x_2)</td>
<td>100</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>Motzkin polynomial</td>
<td>(64(x_1^3 x_2^2 + x_1^2 x_2^3) - 48x_1^2 x_2^2 + 1)</td>
<td>81</td>
<td>6</td>
<td>no</td>
</tr>
<tr>
<td>Three-Hump Camel</td>
<td>(\frac{5}{4} x_1^6 - 5^4 \cdot 1.05x_1^4 + 50x_1^2 + 25x_1 x_2 + 25x_2^2)</td>
<td>2048</td>
<td>6</td>
<td>no</td>
</tr>
</tbody>
</table>

The bounds are shown in Table 2. The bounds \( f^{(r)}_{\text{K}} \) were taken from [2], while the bounds \( SA^{(r)} \) were computed via numerical integration, in particular using the Matlab routine \texttt{sum2} of the package Chebfun [5].

The results in the table show that the bound in Theorem 4.1 is far from tight for these examples. In fact, it may well be that the convergence rates of \( f^{(r)}_{\text{K}} \) and \( SA^{(r)} \) are different for convex \( f \). We know that \( SA^{(r)} - f_{\text{min,K}} = \Theta(1/r) \) is the exact convergence rate for the simulated annealing bounds for convex \( f \) (cf. Example 3.4), but it was speculated in [2] that one may in fact have \( f^{(r)}_{\text{K}} - f_{\text{min,K}} = O(1/r^2) \), even for non-convex \( f \). Determining the exact convergence rate \( f^{(r)}_{\text{K}} \) remains an open problem.
Finally, one should point out that it is not really meaningful to compare the computational complexities of computing the two bounds $f^{(r)}_K$ and $SA^{(r)}$, as explained below.

For any polynomial $f$ and convex body $K$, $f^{(r)}_K$ may be computed by solving a generalised eigenvalue problem with matrices of order $(n + r)^3$, as long as the moments of the Lebesgue measure on $K$ are known. The generalised eigenvalue computation may be done in $O((n + r)^3)$ operations; see [3] for details. Thus this is a polynomial-time procedure for fixed values of $r$.

For non-convex $f$, the complexity of computing $E_{X \sim P_{t/r}}[f(X)]$ is not known. When $f$ is linear, it is shown in [1] that $E_{X \sim P_{t/r}}[f(X)]$ with $t = O(1/r)$ may be obtained in $O^*(n^{4.5} \log(r))$ oracle membership calls for $K$, where the $O^*(\cdot)$ notation suppresses logarithmic factors.

Since the assumptions on the available information is different for the two types of bounds, there is no simple way to compare these respective complexities.

References


