Comparison of Lasserre's measure–based bounds for polynomial optimization to bounds obtained by simulated annealing

Etienne de Klerk [*]   Monique Laurent [†]

March 2, 2017

Abstract

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

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

Our goal is to compare two convergent hierarchies of upper bounds on $f_{\min,\mathbf{K}}$, 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 $\mathbf{K}$ with sum-of-squares polynomial density functions with growing degrees, while simulated annealing bounds use Boltzman distributions on $\mathbf{K}$ with decreasing temparature 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 $f_{\min,\mathbf{K}}$ have been proposed, based on sum-of-squares polynomials and semidefinite programming.

For a general compact set $\mathbf{K}$, upper bounds for $f_{\min,\mathbf{K}}$ 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 $\mathbf{K}$. When f is Lipschitz continuous and under some mild assumption on $\mathbf{K}$ (which holds, e.g., when $\mathbf{K}$ 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 $\mathbf{K}$. Related stronger results have been shown for the case when $\mathbf{K}$ 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 $\mathbf{K}$, 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 $f = \sum_{\alpha \in N(n,r)} f_\alpha x^\alpha$, where we set $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$ for $\alpha \in \mathbb{N}^n$ and

$N(n,r) = \{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i \leq r\}$. 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 $f_{\min,\mathbf{K}}$, established by Lasserre [10]:

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

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

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

Clearly, the inequality $f_{\min,\mathbf{K}} \leq \underline{f}_{\mathbf{K}}^{(r)}$ 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 $\underline{f}_{\mathbf{K}}^{(r)}$.

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

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

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

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

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

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

Indeed suppose $f(x) = \sum_{\beta \in N(n,d)} f_\beta x^\beta$ has degree d. Writing $h \in \Sigma[x]_r$ as $h(x) = \sum_{\alpha \in N(n,2r)} h_\alpha x^\alpha$, the parameter $\underline{f}_{\mathbf{K}}^{(r)}$ from (1) can be reformulated as follows:

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

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, $\underline{f}_{\mathbf{K}}^{(r)}$ is equal to the the smallest generalized eigenvalue of the system:

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

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

$$A_{\alpha,\beta} = \sum_{\delta \in N(n,d)} f_\delta \int_{\mathbf{K}} x^{\alpha+\beta+\delta} dx, \quad B_{\alpha,\beta} = \int_{\mathbf{K}} x^{\alpha+\beta} dx \quad \alpha, \beta \in N(n, r). \tag{4}$$

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 $\mathbf{K}$, defined by the density function:

$$P_f(x) := \frac{\exp(-f(x))}{\int_{\mathbf{K}} \exp(-f(x')) \, dx'}.$$

Write $X \sim P_f$ if the random variable X takes values in $\mathbf{K}$ 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

$$f_{\min,\mathbf{K}} \leq \mathbb{E}_{X \sim P_{f/t}}[f(X)]. \tag{5}$$

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

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

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

Example 3.1. *Consider the minimization of the Motzkin polynomial*

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

over $\mathbf{K} = [-1, 1]^2$, *where there are four global minimizers at the points* $\left(\pm\frac{1}{2}, \pm\frac{1}{2}\right)$, *and* $f_{\min,\mathbf{K}} = 0$. *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* $f_{\min,\mathbf{K}} = 0$ *is* $\mathbb{E}_{X \sim P_{f/t}}[f(X)] \approx 0.7257$ $(t = \frac{1}{2})$.

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

When f is linear and $\mathbf{K}$ 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.
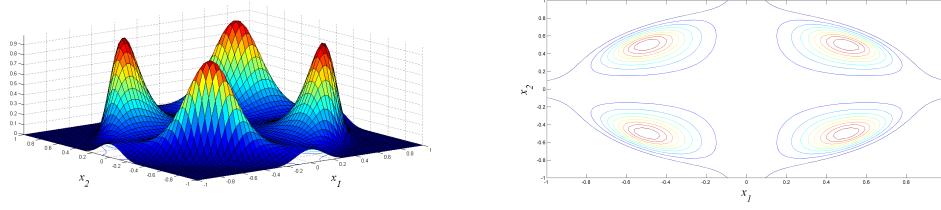
Figure 1: Graph and contours of the Boltzman density with $t = \frac{1}{2}$ for the Motzkin polynomial.

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

$$\mathop{\mathbb{E}}_{X \sim P_{f/t}} [f(X)] - \min_{x \in \mathbf{K}} f(x) \leq nt.$$

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 $\mathbf{K} \subseteq \mathbb{R}^n$ be a convex body. Then, for any $t > 0$, we have*

$$\mathop{\mathbb{E}}_{X \sim P_{f/t}} [f(X)] - \min_{x \in \mathbf{K}} f(x) \leq nt.$$

Proof. Set

$$E_{\mathbf{K}} := \mathop{\mathbb{E}}_{X \sim P_{f/t}} [f(X)] = \frac{\int_{\mathbf{K}} f(x) e^{-f(x)/t} dx}{\int_{\mathbf{K}} e^{\frac{-f(x)}{t}} dx}.$$

Then we have

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

Define the set

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

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

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

Accordingly, define the parameter

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

5

Corollary 3.3 will follow if we show that

$$E_{\widehat{\mathbf{K}}} = E_{\mathbf{K}} + t. \tag{6}$$

To this end set $E_{\mathbf{K}} = \frac{N_{\mathbf{K}}}{D_{\mathbf{K}}}$ and $E_{\widehat{\mathbf{K}}} = \frac{N_{\widehat{\mathbf{K}}}}{D_{\widehat{\mathbf{K}}}}$, where we define

$$N_{\mathbf{K}} := \int_{\mathbf{K}} f(x) e^{-f(x)/t} dx, \quad D_{\mathbf{K}} := \int_{\mathbf{K}} e^{-f(x)/t} dx,$$

$$N_{\widehat{\mathbf{K}}} := \int_{\widehat{\mathbf{K}}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} dx, \quad D_{\widehat{\mathbf{K}}} := \int_{\widehat{\mathbf{K}}} e^{-x_{n+1}/t} dx_{n+1} dx.$$

We work out the parameters $N_{\widehat{\mathbf{K}}}$ and $D_{\widehat{\mathbf{K}}}$ (taking integrations by part):

$$D_{\widehat{\mathbf{K}}} = \int_{\mathbf{K}} \left(\int_{f(x)}^{E_{\mathbf{K}}} e^{-x_{n+1}/t} dx_{n+1} \right) dx = \int_{\mathbf{K}} \left(t e^{-f(x)/t} - t e^{-E_{\mathbf{K}}/t} \right) dx = t D_{\mathbf{K}} - t e^{-E_{\mathbf{K}}/t} \mathrm{vol}(\mathbf{K}),$$

$$\begin{aligned}
N_{\widehat{\mathbf{K}}} &= \int_{\mathbf{K}} \left(\int_{f(x)}^{E_{\mathbf{K}}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} \right) dx \\
&= \int_{\mathbf{K}} \left(-t E_{\mathbf{K}} e^{-E_{\mathbf{K}}/t} + t f(x) e^{-f(x)/t} + t \int_{f(x)}^{E_{\mathbf{K}}} e^{-x_{n+1}/t} dx_{n+1} \right) dx \\
&= -t E_{\mathbf{K}} e^{-E_{\mathbf{K}}/t} \mathrm{vol}(\mathbf{K}) + t N_{\mathbf{K}} + t D_{\widehat{\mathbf{K}}}.
\end{aligned}$$

Then, using the fact that $E_{\mathbf{K}} = \frac{N_{\mathbf{K}}}{D_{\mathbf{K}}}$, we obtain:

$$\frac{N_{\widehat{\mathbf{K}}}}{D_{\widehat{\mathbf{K}}}} = t + \frac{N_{\mathbf{K}} - E_{\mathbf{K}} e^{-E_{\mathbf{K}}/t} \mathrm{vol}(K)}{D_{\mathbf{K}} - e^{-E_{\mathbf{K}}/t} \mathrm{vol}(K)} = t + \frac{N_{\mathbf{K}}}{D_{\mathbf{K}}},$$

which proves relation (6).

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

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

□ □

The bound in the corollary is tight asymptotically, as the following example shows.

Example 3.4. *Consider the univariate problem* $\min_x \{x \mid x \in [0,1]\}$. *Thus, in this case,* $f(x) = x$, $\mathbf{K} = [0,1]$ *and* $\min_{x \in \mathbf{K}} f(x) = 0$. *For given temperature* $t > 0$, *we have*

$$\mathop{\mathbb{E}}_{X \sim P_{f/t}} [f(X)] - \min_{x \in \mathbf{K}} f(x) = \frac{\int_0^1 x e^{-x/t} dx}{\int_0^{\ell} e^{-x/t} dx} - 0 = t - \frac{e^{-1/t}}{1 - e^{-1/t}} \sim t \quad \text{for small } t.$$

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 $\mathbf{K}$ be a compact set and set $\widehat{f}_{\max} = \max_{x \in \mathbf{K}} |f(x)|$. Then we have*

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

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 $\mathbf{K}$ be a convex body. Then for any integer $r \geq 1$ one has*

$$\underline{f}_{\mathbf{K}}^{(rd)} - \min_{x \in \mathbf{K}} f(x) \leq \frac{c}{r},$$

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

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

$$
\begin{aligned}
\underline{f}_{\mathbf{K}}^{(rd)} - \min_{x \in \mathbf{K}} f(x) \ &= \ \left(\underline{f}_{\mathbf{K}}^{(rd)} - \underset{X \sim P_{f/t}}{\mathbb{E}} [f(X)] \right) + \left(\underset{X \sim P_{f/t}}{\mathbb{E}} [f(X)] - f_{\min,\mathbf{K}} \right) \\
&\leq \ \frac{\widehat{f}_{\max}}{2^r} + nt = \frac{\widehat{f}_{\max}}{2^r} + \frac{ne \cdot \widehat{f}_{\max}}{r} \leq \frac{(ne + 1)\widehat{f}_{\max}}{r}.
\end{aligned}
$$

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

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 $\mathbf{K}$ a convex body, then there is a $c > 0$ depending on f and $\mathbf{K}$ only, so that*

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

A suitable value for c is

$$c = (ne + 1) \left(f_{\min,\mathbf{K}} + C_f^1 \cdot diam(\mathbf{K}) + C_f^2 \cdot diam(\mathbf{K})^2 \right),$$

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

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

$$q(x) = f(a) + \nabla f(a)^\top (x - a) + C_f^2 \|x - a\|_2^2,$$

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

By definition of the Lasserre hierarchy,

$$
\begin{aligned}
\underline{f}_{\mathbf{K}}^{(2r)} \quad &:= \quad \inf_{h \in \Sigma[x]_{2r}} \int_{\mathbf{K}} h(x) f(x) dx \quad \text{s.t. } \int_{\mathbf{K}} h(x) dx = 1 \\
&\leq \quad \inf_{h \in \Sigma[x]_{2r}} \int_{\mathbf{K}} h(x) q(x) dx \quad \text{s.t. } \int_{\mathbf{K}} h(x) dx = 1 \\
&\equiv \quad \underline{q}_{\mathbf{K}}^{(2r)}.
\end{aligned}
$$

Invoking Corollary 4.2 and using that the degree of q is 2, we obtain:

$$
\underline{f}_{\mathbf{K}}^{(2r)} \leq \underline{q}_{\mathbf{K}}^{(2r)} \leq f(a) + \frac{(ne+1)\hat{q}_{\max}}{r},
$$

where $\hat{q}_{\max} = \max_{x \in \mathbf{K}} q(x) \leq f_{\min,\mathbf{K}} + C_f^1 \cdot \operatorname{diam}(\mathbf{K}) + C_f^2 \cdot \operatorname{diam}(\mathbf{K})^2$. $\qquad\square$

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,*

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

Then ϕ_{2r} is a sum of squares of polynomials. Moreover, we have

$$
0 \leq \phi_{2r}(\lambda) - e^{-\lambda} \leq \frac{\lambda^{2r+1}}{(2r+1)!} \quad \text{for all } \lambda \geq 0. \tag{7}
$$

We now define the following approximation of the Boltzman density $P_{f/t}$:

$$
\varphi_{2r,t}(x) := \frac{\phi_{2r}(f(x)/t)}{\int_{\mathbf{K}} \phi_{2r}(f(x)/t) dx}. \tag{8}
$$

By construction, $\varphi_{2r,t}$ is a sum-of-squares polynomial probability density function on $\mathbf{K}$, with degree $2rd$ if f is a polynomial of degree d. Moreover, by relation (7) in Lemma 4.4, we obtain

$$
\begin{aligned}
\varphi_{2r,t}(x) \quad &\leq \quad \frac{\phi_{2r}(f(x)/t)}{\int_{\mathbf{K}} \exp(-f(x)/t) dx} \tag{9} \\
&\leq \quad P_{f/t}(x) + \frac{(f(x)/t)^{2r+1}}{(2r+1)! \int_{\mathbf{K}} \exp(-f(x)/t) dx}. \tag{10}
\end{aligned}
$$

From this we can derive the following result.

8

Lemma 4.5. *For any continuous f and scalar $t > 0$ one has*

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

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

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

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

$$
\begin{aligned}
\int_K (f(x) - f_{\min,\mathbf{K}})\varphi_{2r,t}(x)dx &\leq \int_{\mathbf{K}}(f(x) - f_{\min,\mathbf{K}})P_{f/t}(x)dx + \int_{\mathbf{K}}\frac{(f(x) - f_{\min,\mathbf{K}})(f(x)/t)^{2r+1}}{(2r+1)!\int_K \exp(-f(x)/t)dx}dx \\
&= \mathop{\mathbb{E}}_{X \sim P_{f/t}}[f(X)] - f_{\min,\mathbf{K}} + \int_{\mathbf{K}}\frac{(f(x) - f_{\min})(f(x)/t)^{2r+1}}{(2r+1)!\int_K \exp(-f(x)/t)dx}dx.
\end{aligned}
$$

Combining with the inequality (12) gives the desired result. $\qquad\square\qquad\qquad\qquad\square$

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):

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

and to show that $T \leq \frac{\widehat{f}_{\max}}{2^r}$.

By the defininition of $\widehat{f}_{\max}$ we have

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

which implies

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

Combining with the Stirling approximation inequality,

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

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

$$T \leq \frac{2\widehat{f}_{\max}}{\sqrt{2\pi(2r+1)}}\left(\frac{\widehat{f}_{\max}e}{t(2r+1)}\right)^{2r+1}\exp(\widehat{f}_{\max}/t).$$

9

Consider $r \geq \frac{e \cdot \widehat{f}_{\max}}{t}$, so that $\widehat{f}_{\max}/t \leq r/e$. Then, using the fact that $r/(2r+1) \leq 1/2$, we obtain

$$
\begin{aligned}
T &\leq \frac{2\widehat{f}_{\max}}{\sqrt{2\pi}} \frac{\exp(r/e)}{\sqrt{2r+1}} \left(\frac{r}{2r+1}\right)^{2r+1} \\
&\leq \frac{\widehat{f}_{\max}}{\sqrt{2\pi}} \frac{\exp(1/e)^r}{\sqrt{2r+1}} \left(\frac{1}{4}\right)^r \\
&= \frac{\widehat{f}_{\max}}{\sqrt{2\pi}\sqrt{2r+1}} \left(\frac{\exp(1/e)}{4}\right)^r \\
&< \frac{\widehat{f}_{\max}}{2^r}.
\end{aligned}
$$

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 $\underline{f}_{\mathbf{K}}^{(r)}$ and $\mathbb{E}_{X \sim P_{f/t}}[f(X)]$, with $t = \frac{e \cdot d \cdot \widehat{f}_{\max}}{r}$ and d the degree of f. Thus we define, for the purpose of comparison:

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

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

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

Name	$f(x)$	$\widehat{f}_{\max}$	d	Convex?
Booth function	$(10x_1 + 20x_2 - 7)^2 + (20x_1 + 10x_2 - 5)^2$	2594	2	yes
Matyas function	$26(x_1^2 + x_2^2) - 48x_1x_2$	100	2	yes
Motzkin polynomial	$64(x_1^4x_2^2 + x_1^2x_2^4) - 48x_1^2x_2^2 + 1$	81	6	no
Three-Hump Camel function	$\frac{5^6}{6}x_1^6 - 5^4 \cdot 1.05x_1^4 + 50x_1^2 + 25x_1x_2 + 25x_2^2$	2048	6	no

The bounds are shown in Table 2. The bounds $\underline{f}_{\mathbf{K}}^{(r)}$ were taken from [2], while the bounds $SA^{(r)}$ were computed via numerical integration, in particular using the Matlab routine 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 $\underline{f}_{\mathbf{K}}^{(r)}$ and $SA^{(r)}$ are different for convex f. We know that $SA^{(r)} - f_{\min,\mathbf{K}} = \Theta(1/r)$ 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 $\underline{f}_{\mathbf{K}}^{(r)} - f_{\min,\mathbf{K}} = O(1/r^2)$, even for non-convex f. Determining the exact convergence rate $\underline{f}_{\mathbf{K}}^{(r)}$ remains an open problem.

Table 2: Comparison of the upper bounds $SA^{(r)}$ and $\underline{f}_{\mathbf{K}}^{(r)}$ for the test functions.

r	Booth Function		Matyas Function		Three–Hump Camel Function		Motzkin Polynomial	
	$\underline{f}_{\mathbf{K}}^{(r)}$	$SA^{(r)}$	$\underline{f}_{\mathbf{K}}^{(r)}$	$SA^{(r)}$	$\underline{f}_{\mathbf{K}}^{(r)}$	$SA^{(r)}$	$\underline{f}_{\mathbf{K}}^{(r)}$	$SA^{(r)}$
3	118.383	367.834	4.2817	15.4212	29.0005	247.462	1.0614	4.0250
4	97.6473	356.113	3.8942	14.8521	9.5806	241.700	0.8294	3.9697
5	69.8174	345.043	3.6894	14.3143	9.5806	236.102	0.8010	3.9157
6	63.5454	334.585	2.9956	13.8062	4.4398	230.663	0.8010	3.8631
7	47.0467	324.701	2.5469	13.3262	4.4398	225.381	0.7088	3.8118
8	41.6727	315.354	2.0430	12.8726	2.5503	220.251	0.5655	3.7618
9	34.2140	306.510	1.8335	12.4441	2.5503	215.269	0.5655	3.7130
10	28.7248	298.138	1.4784	12.0390	1.7127	210.431	0.5078	3.6654
11	25.6050	290.206	1.3764	11.6560	1.7127	205.734	0.4060	3.6190
12	21.1869	282.687	1.1178	11.2938	1.2775	201.173	0.4060	3.5737
13	19.5588	275.554	1.0686	10.9511	1.2775	196.745	0.3759	3.5296
14	16.5854	268.782	0.8742	10.6267	1.0185	192.446	0.3004	3.4865
15	15.2815	262.348	0.8524	10.3195	1.0185	188.272	0.3004	3.4444
16	13.4626	256.230	0.7020	10.0284	0.8434	184.220	0.2819	3.4034
17	12.2075	250.408	0.6952	9.75250	0.8434	180.287	0.2300	3.3633
18	11.0959	244.863	0.5760	9.49071	0.7113	176.469	0.2300	3.3242
19	9.9938	239.577	0.5760	9.24220	0.7113	172.762	0.2185	3.2860
20	9.2373	234.534	0.4815	9.00615	0.6064	169.164	0.1817	3.2487

Finally, one should point out that it is not really meaningful to compare the computational complexities of computing the two bounds $\underline{f}_{\mathbf{K}}^{(r)}$ and $SA^{(r)}$, as explained below.

For any polynomial f and convex body $\mathbf{K}$, $\underline{f}_{\mathbf{K}}^{(r)}$ may be computed by solving a generalised eigenvalue problem with matrices of order $\binom{n+r}{r}$, as long as the moments of the Lebesgue measure on $\mathbf{K}$ are known. The generalised eigenvalue computation may be done in $O\left(\binom{n+r}{r}^3\right)$ 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 $\mathbb{E}_{X \sim P_{f/t}}[f(X)]$ is not known. When f is linear, it is shown in [1] that $\mathbb{E}_{X \sim P_{rf}}[f(X)]$ with $t = O(1/r)$ may be obtained in $O^*\left(n^{4.5}\log(r)\right)$ oracle membership calls for $\mathbf{K}$, 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

[1] J. Abernethy and E. Hazan. Faster Convex Optimization: Simulated Annealing with an Efficient Universal Barrier. arXiv 1507.02528, July 2015.

[2] E. de Klerk, R. Hess and M. Laurent. Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization. *SIAM J. Optim.* (to appear),

arXiv:1603.03329v1 (2016)

[3] E. de Klerk, J.-B. Lasserre, M. Laurent, and Z. Sun. Bound-constrained polynomial optimization using only elementary calculations. *Mathematics of Operations Research* (to appear), arXiv:1507.04404v2 (2016)

[4] E. de Klerk, M. Laurent, Z. Sun. Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization, *Math. Program. Ser. A*, (2016). doi:10.1007/s10107-016-1043-1.

[5] T. A. Driscoll, N. Hale, and L. N. Trefethen, editors, *Chebfun Guide*, Pafnuty Publications, Oxford, 2014.

[6] A. T. Kalai and S. Vempala. Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2), 253–266 (2006)

[7] S. Kirkpatrick, C.D. Gelatt, Jr., M.P. Vecchi. Optimization by simulated annealing. *Science* 220, 671-680, 1983.

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

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

[10] Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. *SIAM J. Optim.* 21(3), 864–885 (2011)

[11] Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. *In Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 157–270 (2009)

[12] L. Lovász. Hit-and-run mixes fast, *Mathematical Programming*, 86(3), 443–461, 1999.

[13] P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology (2000)

[14] R. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions, *Operations Research* 32(6), 1296-1308, 1984.

[15] J. Spall. *Introduction to Stochastic Search and Optimization.* Wiley, New York, 2003.