

Tilburg University

On the Complexity of Optimization over the Standard Simplex

de Klerk, E.; den Hertog, D.; Elfadul, G.E.E.

Publication date:
2005

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

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
de Klerk, E., den Hertog, D., & Elfadul, G. E. E. (2005). *On the Complexity of Optimization over the Standard Simplex*. (CentER Discussion Paper; Vol. 2005-125). Operations research.

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.

CentER



Discussion Paper

No. 2005–125

ON THE COMPLEXITY OF OPTIMIZATION OVER THE STANDARD SIMPLEX

By Etienne de Klerk, Dick den Hertog, Gamal Elfadul Elabwabi

December 2005

ISSN 0924-7815

On the complexity of optimization over the standard simplex

E. de Klerk

D. den Hertog

G. Elabwabi

December 6, 2005

Abstract

We review complexity results for minimizing polynomials over the standard simplex and unit hypercube.

In addition, we show that there exists a polynomial time approximation scheme (PTAS) for minimizing Lipschitz continuous functions and functions with uniformly bounded Hessians over the standard simplex. This extends an earlier result by De Klerk, Laurent and Parrilo [A PTAS for the minimization of polynomials of fixed degree over the simplex, *Theoretical Computer Science*, to appear.]

Keywords: global optimization, standard simplex, PTAS, multivariate Bernstein approximation, semidefinite programming

JEL code: C61

1 Introduction

In this paper we study the computational complexity of approximating the minimum value of a function on the standard simplex. This relatively simple optimization problem has several applications. If the function is quadratic, the applications already include portfolio optimization, population dynamics, genetics, finding maximum stable sets in graphs, and lower bounding the crossing number of certain classes of graphs.

There are not as many applications yet for more general functions, but one example is the training of neural networks (Beliakov and Abraham [4]). In this case the function is Lipschitz continuous, but transcendental.

We will prove the existence of a polynomial-time approximation scheme (PTAS) for minimizing Lipschitz continuous functions as well as functions with bounded Hessians over the simplex.

Some practical algorithms for optimization of Lipschitz continuous functions over the simplex have been investigated (see e.g. Bagirov and Rubinov [3]), but these algorithms are not known to have the PTAS property. It is therefore the purpose of this paper to show that it is meaningful to search for practical PTAS algorithms.

1.1 Complexity of approximating minima

Consider the generic optimization problem:

$$\underline{f} := \min \{f(x) : x \in S\}. \quad (1.1)$$

for some continuous $f : \mathbb{R}^m \mapsto \mathbb{R}$ and compact convex set S , and let

$$\bar{f} := \max \{f(x) : x \in S\},$$

In this paper we will consider the cases where S is the standard simplex

$$\Delta_m := \left\{ x \in \mathbb{R}^m : \sum_i x_i = 1, x \geq 0 \right\},$$

but we will also review known results for the hypercube $[0, 1]^m$.

The next definition has been used by several authors, including Ausiello, d'Atri and Protasi [2], Bellare and Rogaway [5], Bomze and De Klerk [6], De Klerk, Parrilo and Laurent [8], Nesterov et al. [21], and Vavasis [22].

Definition 1.1 *A value ψ_ϵ approximates \underline{f} with relative accuracy $\epsilon \in [0, 1]$ if*

$$|\psi_\epsilon - \underline{f}| \leq \epsilon(\bar{f} - \underline{f}).$$

Then one also says that ψ_ϵ is an ϵ -approximation of \underline{f} . The approximation is called implementable if $\psi_\epsilon = f(x)$ for some $x \in S$.

The following definition is from De Klerk, Laurent and Parrilo [8], and is consistent with the corresponding definition in combinatorial optimization.

Definition 1.2 (PTAS) *If a problem allows an implementable approximation $\psi_\epsilon = f(x_\epsilon)$ for each $\epsilon \in (0, 1]$, such that $x_\epsilon \in S$ can be computed in time polynomial in n and the bit size required to represent f , we say that the problem allows a polynomial time approximation scheme (PTAS).*

Since we will consider more general functions than polynomials, we need to elaborate on the notion of the bit size required to represent f . We will only assume that $f(x)$ may be computed in time polynomial in the bit size of x .

Since, in practice, the evaluation of f can be costly, we will explicitly state the required number of function evaluations when stating results on the complexity of approximating \underline{f} .

1.2 Known complexity results

If S is the standard m -simplex, then computing \underline{f} is an NP-hard problem, already for quadratic polynomials, as it contains the maximum stable set problem. Indeed, let G be a graph with adjacency matrix A and let

I denote the identity matrix; then the maximum size $\alpha(G)$ of a stable set in G can be expressed as

$$\frac{1}{\alpha(G)} = \min_{x \in \Delta} x^T (I + A)x$$

by the theorem of Motzkin and Straus [20].

Bomze and De Klerk [6] showed that, for $S = \Delta_m$ and f quadratic, problem (1.1) allows a PTAS. This result was extended to polynomials of fixed degree by De Klerk, Laurent, and Parrilo [8]. On the other hand, this problem cannot have a FPTAS, unless $\text{NP}=\text{ZPP}$, due to inapproximability results for the maximum stable set problem by Håstad [16].

Another negative result is due to Bellare and Rogaway [5], who proved that if $\text{P} \neq \text{NP}$ and $\epsilon \in (0, 1/3)$, there is no polynomial time ϵ -approximation algorithm for the problem of minimizing a polynomial of total degree $d \geq 2$ over the set $S = \{x \in [0, 1]^m \mid Ax \leq b\}$.

If $S = [0, 1]^m$ and f quadratic, then problem (1.1) contains the maximum cut problem in graphs as a special case. Indeed, for a graph $G = (V, E)$ with Laplacian matrix L , the size of the maximum cut is given by

$$|\text{maximum cut}| = \max_{x \in [-1, 1]^{|V|}} \frac{1}{4} x^T L x = \max_{x \in [0, 1]^{|V|}} \frac{1}{4} (2x - e)^T L (2x - e),$$

where e is the all-ones vector.

For the maximum cut problem there is a celebrated (1-0.878)-approximation result due to Goemans and Williamson [15], and related approximation results for quadratic optimization over a hypercube were given by Nesterov [21]. On the negative side, the maximum cut problem cannot be approximated within $\epsilon = 1/17$ (Håstad [17]), and it follows that problem (1.1) does not allow a PTAS for any class of functions that includes the quadratic polynomials if $S = [0, 1]^m$. So, in a well-defined sense optimization over the unit hypercube is much harder than over the simplex.

1.3 New results

We prove some new approximability results in the spirit of the abovementioned papers. In particular, we consider two classes of (multivariate) functions on the standard simplex that satisfy suitable "smoothness" conditions.

The first class is given by

$$\left\{ f \in C(\Delta_m) : \omega\left(f, \frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}(\bar{f} - \underline{f})\right) \right\}, \quad (1.2)$$

where ω is the usual modulus of continuity, i.e.

$$\omega(f, \delta) = \max_{\substack{x, y \in \Delta_m \\ \|x - y\| \leq \delta}} |f(x) - f(y)|.$$

For the purposes of optimization, we may assume that

$$\max_{x \in \Delta_m} |f(x)| =: \|f\|_{\infty, \Delta_m} \leq \bar{f} - \underline{f} \leq 2\|f\|_{\infty, \Delta_m},$$

by replacing f by $f - f(x_0)$ for any fixed $x_0 \in \Delta_m$. Thus the class of functions (1.2) may also be defined as

$$\left\{ f \in C(\Delta_m) : \omega\left(f, \frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\|f\|_{\infty, \Delta_m}\right) \right\}.$$

This class contains the Lipschitz continuous functions on Δ_m in the sense that it contains

$$\text{cone}\{f \in C(\Delta_m), \|f\|_{\infty, \Delta_m} = 1 : |f(x) - f(y)| \leq L\|x - y\| \ \forall x, y \in \Delta_m\}, \quad (1.3)$$

where L denotes a (Lipschitz) constant. In other words, each constant L defines a function class via (1.3), that is contained in the class (1.2).

The second class of functions we consider is

$$\{f \in C^2(\Delta_m) : \|\nabla^2 f\|_{\infty, S} = O(\bar{f} - \underline{f})\}, \quad (1.4)$$

where

$$\|\nabla^2 f\|_{\infty, \Delta_m} := \max_{x \in \Delta_m} \rho(\nabla^2 f(x)),$$

and ρ denotes the spectral radius of a matrix. As before, an equivalent definition is

$$\{f \in C^2(\Delta_m) : \|\nabla^2 f\|_{\infty, \Delta_m} = O(\|f\|_{\infty, \Delta_m})\}. \quad (1.5)$$

As before, we can view this set as the cone generated by functions of unit supremum norm and bounded Hessian norm on Δ_m .

We show that there exists a PTAS for the problem of minimizing a function from the classes (1.2) and (1.4) over Δ_m . Thus we will extend the main result in [8] to a larger class of nonlinear functions.

We use the properties of multivariate Bernstein operators as well as the main result in [8] to derive the new PTAS result. In particular, we approximate f using multivariate Bernstein polynomials, and subsequently apply the main result from [8]. Our review of multivariate Bernstein operators will be self-contained, but more information may be found in [1, 9, 10, 11, 12, 19, 23].

2 Positive linear operators

The results on Bernstein operators presented in the following sections may be derived in a simple way by using the framework of positive linear operators, and we review their basic properties here.

Definition 2.1 *A linear operator U acting on $C(K)$, where $K \subset \mathbb{R}^m$ is convex and compact, is called positive if $f(x) \geq 0 \ \forall x \in K$ implies $(Uf)(x) \geq 0 \ \forall x \in K$.*

We write $f \geq 0$ as shorthand for $f(x) \geq 0 \ \forall x \in K$.

Note that

$$g \geq f \implies Ug \geq Uf,$$

that in turn implies

$$U(|f|) \geq U(f) \text{ and } U(|f|) \geq U(-f) = -U(f),$$

i.e. $U(|f|) \geq |U(f)|$.

If U preserves constants, then its operator norm satisfies $\|U\| = \sup_{f \in C(K)} \frac{\|Uf\|_{\infty, K}}{\|f\|_{\infty, K}} = 1$, due to

$$|U(f)| \leq U(|f|) \leq U(\|f\|_{\infty, K} 1) = \|f\|_{\infty, K} U(1) = \|f\|_{\infty, K}.$$

The following result is taken from Waldron [23], but this type of analysis is in fact much older (cf. Theorem 4.4 in [9], Chapter 5 in [1], and the references therein).

Theorem 2.1 *Let $U : C(K) \mapsto C(K)$ be a positive linear operator that preserves constants, and let $f \in C^2(K)$ (or twice continuously differentiable on an open set containing K if $\text{int}(K) = \emptyset$). Let $\phi_{1,i}(x) = x_i$ and $\phi_{2,i}(x) = x_i^2$ ($i = 1, \dots, m$).*

One has

$$|Uf - f| \leq \|\nabla f\|_{\infty, K} \sum_{i=1}^m |\phi_{1,i} - U(\phi_{1,i})| + \frac{1}{2} \|\nabla^2 f\|_{\infty, K} \left| \sum_{i=1}^m (U(\phi_{2,i}) + \phi_{2,i} - 2\phi_{1,i}U(\phi_{1,i})) \right|.$$

In particular, if U also preserves linear functions, i.e. $U\phi_{1,i} = \phi_{1,i}$ then one has

$$|Uf - f| \leq \frac{1}{2} \|\nabla^2 f\|_{\infty, K} \left| \sum_{i=1}^m (U(\phi_{2,i}) - \phi_{2,i}) \right|.$$

Proof. Fix $y \in K$. By Taylor's theorem, for every $x \in K$ one has

$$f(x) = f(y + (x - y)) = f(y) + \nabla f(y)^T (x - y) + \frac{1}{2} (x - y)^T \nabla^2 f(\zeta(x)) (x - y),$$

where $\zeta(x) = \alpha(x)x + (1 - \alpha(x))y$ for some $\alpha(x) \in [0, 1]$.

Applying the operator U on both sides we get

$$\begin{aligned} |Uf - f(y)| &= \left| U(\nabla f(y)^T (\cdot - y)) + U\left(\frac{1}{2}(\cdot - y)^T \nabla^2 f(\zeta(\cdot))(\cdot - y)\right) \right| \\ &\leq \|\nabla f\|_{\infty, K} \sum_{i=1}^m |y_i - U(\phi_{1,i})| + \frac{1}{2} \|\nabla^2 f\|_{\infty, K} \left| \sum_{i=1}^m (U(\phi_{2,i}) + y_i^2 - 2y_i U(\phi_{1,i})) \right|. \end{aligned}$$

Evaluating the inequality at $x = y$ completes the proof. \square

If we do not assume that $f \in C^2(K)$, but merely that f is continuous on K , then we have the following result in terms of the modulus of continuity of f (as opposed to the norm of the Hessian). The proof is simple, and we again include it for completeness.

Theorem 2.2 (Proposition 5.1.5 in [1]) *Let $K \in \mathbb{R}^m$ be convex and compact, and let $U : C(K) \mapsto C(K)$ be a positive linear operator that preserves affine functions.*

Then for every $f \in C(K)$, $x \in K$, and $\delta > 0$ one has

$$|(Uf)(x) - f(x)| \leq \left(1 + \frac{1}{\delta^2} \sum_{i=1}^m (U\phi_{2,i}(x) - \phi_{2,i}(x))\right) \omega(f, \delta).$$

Proof. Fix $f \in C(K)$, $x \in K$ and $\delta > 0$. For any $y \in K$ such that $\|x - y\| > \delta$ one has (by the definition of $\omega(f, \delta)$):

$$\begin{aligned} |f(y) - f(x)| &\leq \left(1 + \frac{1}{\delta} \|x - y\|\right) \omega(f, \delta) \\ &\leq \left(1 + \frac{1}{\delta^2} \|x - y\|^2\right) \omega(f, \delta) \\ &= \left(1 + \frac{1}{\delta^2} (\|x\|^2 - 2x^T y + \|y\|^2)\right) \omega(f, \delta). \end{aligned}$$

Obviously, the same inequality holds if $\|x - y\| \leq \delta$.

Applying U on both sides and evaluating the resulting inequality at $y = x$ yields the required result. \square

These theorems are useful in the following setting. We will study a sequence of positive operators U_n that preserve affine functions, and such that

$$\|U_n \phi_{2,i} - \phi_{2,i}\|_{\infty, K} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (i = 1, \dots, m). \quad (2.6)$$

By the theorems above, this implies that $U_n f \rightarrow f$ uniformly. The rate of convergence is determined by the rate of convergence in (2.6).

3 Univariate Bernstein operators

Let $f \in C^2[0, 1]$, and define the Bernstein basis for the univariate polynomials of degree at most n by

$$p_{n,i}(x) = \binom{n}{i} x^i (1-x)^{n-i} \quad (i = 0, \dots, n). \quad (3.7)$$

Consider the Bernstein approximation of f with respect to this basis:

$$B_n(f)(x) := \sum_{k=0}^n f\left(\frac{k}{n}\right) p_{n,k}(x).$$

One can use Theorem 2.1 to show that $B_n(f)$ converges uniformly to f . To this end, it is simple to verify (or see [9], Chapter 1, for a proof) that B_n is a positive linear operator that preserves constants, and

$$(B_n(\phi_1))(x) = x, \quad (B_n(\phi_2))(x) = \phi_2(x) + \frac{1}{n} x(1-x), \quad (3.8)$$

where $\phi_1(x) = x$, $\phi_2(x) = x^2$, as before.

In other words, B_n preserves linear functions as well. By Theorem 2.1, an error bound is therefore given by

$$\|B_n(f) - f\|_{\infty, [0,1]} \leq \frac{x(1-x)}{2n} \|f^{(2)}\|_{\infty, [0,1]} \leq \frac{1}{8n} \|f^{(2)}\|_{\infty, [0,1]}.$$

For a discussion of these historical results see [19], and for recent developments [14].

4 Multivariate Bernstein operators on a simplex

Consider a twice continuously differentiable function f defined on the standard m -simplex

$$\Delta_m := \left\{ x \in \mathbb{R}^m : \sum_i x_i = 1, x_i \geq 0 \right\}.$$

The Bernstein approximation of f of order n on Δ_m is the polynomial

$$B_n(f)(x) := \sum_{\alpha \in I(m,n)} f\left(\frac{\alpha}{n}\right) \frac{n!}{\alpha!} x^\alpha, \quad (4.9)$$

where

$$I(m,n) := \left\{ \alpha \in \mathbb{N}_0^m \mid \sum_{i=1}^m \alpha_i = n \right\}, \quad x^\alpha := x_1^{\alpha_1} \dots x_m^{\alpha_m}, \quad \alpha! := \prod_i \alpha_i!$$

Note that this definition coincides with the definition for the univariate case when $m = 2$. Also note that finding the coefficients of $B_n(f)$ requires $|I(m,n)| = \binom{n+m}{m}$ evaluations of f .

Similarly to the univariate case, one can show that B_n preserves linear functions and gives an $O(1/n)$ -error for quadratic functions (see (3.8)). We include a proof for completeness.

Lemma 4.1 *Let $f \in C^2(\Delta_m)$ and $B_n(f)$ as defined in (4.9). Then*

$$B_n(\phi_{1,i}) = \phi_{1,i}, \quad B_n(\phi_{2,i})(x) = \phi_{2,i}(x) + \frac{1}{n} x_i (1 - x_i) \quad (i = 1, \dots, m), \quad (4.10)$$

where $\phi_{1,i}(x) = x_i$ and $\phi_{2,i}(x) = x_i^2$ ($i = 1, \dots, m$) as before.

Proof. By (4.9), we have

$$\begin{aligned} B_n(\phi_{1,i})(x) &= \sum_{\alpha \in I(m,n)} \frac{\alpha_i}{n} \frac{n!}{\alpha!} x^\alpha \\ &= x_i \sum_{\substack{\alpha \in I(m,n) \\ \alpha_i \neq 0}} \frac{(n-1)!}{(\alpha - e_i)!} x^{\alpha - e_i} \\ &= x_i \sum_{\beta \in I(m,n-1)} \frac{(n-1)!}{\beta!} x^\beta = x_i, \end{aligned}$$

where e_i is the i th standard unit vector, and for the last equality we used the multinomial identity

$$\left(\sum_{j=1}^m x_j\right)^k = \sum_{\beta \in I(m,k)} \frac{(k)!}{\beta!} x^\beta \quad (= 1 \text{ if } x \in \Delta_m).$$

Similarly,

$$\begin{aligned} B_n(\phi_{2,i})(x) &= \sum_{\alpha \in I(m,n)} \left(\frac{\alpha_i}{n}\right)^2 \frac{n!}{\alpha!} x^\alpha \\ &= \frac{n-1}{n} x_i^2 \sum_{\substack{\alpha \in I(m,n) \\ \alpha_i \neq 0,1}} \frac{(n-2)!}{(\alpha - 2e_i)!} x^{\alpha - 2e_i} + \frac{1}{n} x_i \sum_{\substack{\alpha \in I(m,n) \\ \alpha_i = 1}} \frac{(n-1)!}{(\alpha - e_i)!} x^{\alpha - e_i} \\ &= \frac{n-1}{n} x_i^2 + \frac{1}{n} x_i = x_i^2 + \frac{1}{n} x_i (1 - x_i). \end{aligned}$$

□

One therefore has the following approximation result, by Theorem 2.1, and we include a proof for completeness.

Theorem 4.1 (see e.g. Waldron [23]) *Let $f \in C^2(\Delta_m)$ and $B_n(f)$ as defined in (4.9). One has*

$$\|B_n(f) - f\|_{\infty, \Delta_m} \leq \frac{1}{2n} \|\nabla^2 f\|_{\infty, \Delta_m}.$$

Proof. By Theorem 2.1, we only have to give an upper bound on

$$\sum_{i=1}^m (B_n(\phi_{2,i}) - \phi_{2,i})(x) = \frac{1}{n} \sum_{i=1}^m x_i (1 - x_i). \quad (4.11)$$

Consider therefore the convex optimization problem

$$\max_{x \in \Delta_m} \frac{1}{n} \sum_{i=1}^m x_i (1 - x_i).$$

The KKT conditions for this problem are necessary and sufficient for optimality, and it is easy to verify that a KKT point is given by $x_i = \frac{1}{m}$ ($i = 1, \dots, m$). Thus

$$\sum_{i=1}^m x_i (1 - x_i) \leq 1 - \frac{1}{m} \quad \forall x \in \Delta_m. \quad (4.12)$$

□

If we only assume that f is continuous on Δ_m then we can use Theorem 2.2 to derive the following result.

Theorem 4.2 (See e.g. [1], §5.2.11) *Let $f \in C(\Delta_m)$ and $B_n(f)$ as defined in (4.9). One has*

$$\|B_n(f) - f\|_{\infty, \Delta_m} \leq 2\omega\left(f, \frac{1}{\sqrt{n}}\right).$$

Proof. By Theorem 2.2, we have, for any $\delta > 0$,

$$\begin{aligned} |(B_n f)(x) - f(x)| &\leq \left(1 + \frac{1}{\delta^2} \sum_{i=1}^m (B_n \phi_{2,i}(x) - \phi_{2,i}(x))\right) \omega(f, \delta) \\ &= \left(1 + \frac{1}{\delta^2 n} \sum_{i=1}^m x_i(1 - x_i)\right) \omega(f, \delta) \\ &\leq \left(1 + \frac{1}{\delta^2 n}\right) \omega(f, \delta), \end{aligned}$$

where we have used (4.11) and (4.12) to obtain the last inequality. The required result now follows by setting $\delta = \frac{1}{\sqrt{n}}$. \square

5 Complexity analysis

One may use the results of the last section to derive complexity results for optimization over Δ_m . We use the following result due to De Klerk, Laurent, and Parrilo [8] (see also Faybusovich [13]).

Theorem 5.1 *Let p be a polynomial of total degree d on Δ_m . Then one may compute an $x \in \Delta_m$ such that*

$$p(x) - \underline{p} \leq \epsilon(\bar{p} - \underline{p})$$

in time polynomial in m and the bit size of the coefficients of p , and exponential in d and in $1/\epsilon$.

Combining Theorem 5.1 with Theorem 4.1 and Theorem 4.2 yields the following result.

Theorem 5.2 *Assume that $\epsilon > 0$ is given and that f belongs to the class (1.2) or (1.4). Then one may compute an $x \in \Delta_m$ such that*

$$f(x) - \underline{f} \leq \epsilon(\bar{f} - \underline{f})$$

where the computation complexity is polynomial in m and in the bit size required to represent f , and exponential in $1/\epsilon$, and in addition requires $\binom{m+n}{n}$ evaluations of f , where $n = O(1/\epsilon)$.

Proof. We only give the proof for the function class (1.4). The proof when f belongs to the class (1.2) is

similar and requires Theorem 4.2.

For given $\epsilon > 0$, set $n = M/\epsilon = O(1/\epsilon)$ where we assume $\|\nabla^2 f\|_{\infty, \Delta_m} \leq M(\bar{f} - \underline{f})$. By Theorem 4.1, we have

$$\|B_n(f) - f\|_{\infty, \Delta_m} \leq \frac{\epsilon}{2}(\bar{f} - \underline{f}). \quad (5.13)$$

Now use Theorem 5.1 to conclude that one can obtain an $\frac{\epsilon}{2}$ -approximation to the minimum of $B_n(f)$ on Δ_m in time polynomial in m and the bit sizes of the coefficients of $B_n(f)$, and exponential in $\frac{1}{\epsilon}$ and M .

Note that the coefficients of $B_n(f)$ in the Bernstein basis are $f(\alpha/n)$ ($\alpha \in I(m, n)$). By our general assumptions on f in Section 1.1, these numbers have bit sizes bounded by a polynomial in the bit size required to represent f and in n and m .

Using (5.13) completes the proof. □

Corollary 5.1 *There exists a PTAS for the problem $\min_{x \in \Delta} f(x)$ if f belongs to the class (1.2) or (1.4).*

6 Discussion

Relation to derivative free optimization

Our PTAS algorithms fall in the category of derivative free optimization methods, i.e. methods for optimization of functions where the gradient does not exist or is too expensive to compute in practice. (For a review of these methods see [7] and the references therein.)

Thus our PTAS result in Corollary 5.1 also implies that there exists a derivative free optimization algorithm for minimizing functions from the classes (1.2) and (1.4) over the simplex, that is also a PTAS.

Having said that, the algorithms presented here are probably not practical, due to the following reasons:

- The Bernstein approximation $B_n f$ of f converges only linearly to f , and this convergence is known to be slow in practice as well.
- The PTAS for minimizing $B_n f$ over the simplex from [8] has complexity that depends on n via $O(n^n)$. For the theoretical analysis, this was fine, but in practice the slow convergence of $B_n f$ to f would obviously lead to prohibitive computational requirements.

However, we hope that our observations will lead to some practical algorithms for optimization over the simplex that retain the PTAS guarantee. For example, one could replace the Bernstein approximations by suitable splines to obtain faster convergence.

Which class of functions allows a PTAS for $\min_{x \in \Delta_m} f(x)$?

Our main result was to extend the result by De Klerk, Laurent, and Parrilo [8] (Theorem 5.1) on the existence of a PTAS for minimizing polynomials of fixed degree over the simplex to a larger class of nonlinear functions (Corollary 5.1).

Note however, that Corollary 5.1 does not imply Theorem 5.1, since polynomials of fixed degree on the simplex do not necessarily belong to the classes (1.2) or (1.4). The Markov inequality for an m -variate polynomial of total degree d defined on a convex body S states that

$$\|\nabla p(\cdot)^T h\|_{\infty, S} \leq \frac{2d^2}{\omega(S)} \|p\|_{\infty, S} \quad \forall \|h\| = 1, \quad (6.14)$$

where $\omega(S)$ is the minimum distance between two distinct parallel supporting hyperplanes of S (see e.g. Kroó [18]).

For example, if S is the m -dimensional simplex

$$S = \left\{ x \in \mathbb{R}^m : \sum_i x_i \leq 1, x_i \geq 0 \right\},$$

then $\omega(S) = 1/\sqrt{m}$. The Markov inequality (6.14) is sharp, which means that the class of n -variate polynomials of fixed total degree on the simplex is not a subset of the class (1.2).

Thus we do not yet have a complete understanding of the class of functions that allow a PTAS for the problem $\min_{x \in \Delta_m} f(x)$.

References

- [1] F. Altomare and M. Campiti. *Korovkin-type approximation theory and its applications*. De Gruyter studies in mathematics 17, Walter de Gruyter publishers, Berlin, 1994.
- [2] G. Ausiello, A. D’Atri, and M. Protasi. Structure preserving reductions among convex optimization problems. *Journal of Computer and System Sciences*, 21:136–153, 1980.
- [3] A.M. Bagirov and A.M. Rubinov. Global minimization of increasing positively homogeneous functions over the unit simplex, *Annals of Operation Research*, 98, 171–188, 2000.
- [4] G. Beliakov and A. Abraham. Global optimisation of neural networks using a deterministic hybrid approach. In *Proceedings of Hybrid Information Systems-2001*, Physica-Verlag, 79–92, 2002.
- [5] M. Bellare and P. Rogaway. The complexity of approximating a nonlinear program. *Mathematical Programming*, 69:429–441, 1995.
- [6] I. Bomze and E. De Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Global Optimization*, 24(2):163–185, 2002.

- [7] A.R. Conn, K. Scheinberg, and L.N. Vicente. Geometry of sample sets in derivative free optimization. Part I: Polynomial interpolation, Manuscript, April 2003, available at www.mat.uc.pt/~lnv/papers/reports.html
- [8] E. de Klerk, M. Laurent, P. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex *Theoretical Computer Science*, to appear.
- [9] R.A. DeVore and G.G. Lorentz. *Constructive Approximation*. Springer-Verlag, Berlin, 1993.
- [10] Z. Ditzian. Inverse theorems for multidimensional Bernstein operators, *Pacific J. Math.* 121 (2), 293–319, 1986.
- [11] Z. Ditzian. Best polynomial approximation and Bernstein polynomial approximation on a simplex, *Indagationes Math.* 92, 243–256, 1989.
- [12] Z. Ditzian and X. Zhou. Optimal approximation class for multivariate Bernstein operators, *Pacific Jour. of Math.* 158, 93–120, 1993.
- [13] L. Faybusovich. Global optimization of homogeneous polynomials on the simplex and on the sphere. In C. Floudas and P. Pardalos, editors, *Frontiers in Global Optimization*. Kluwer Academic Publishers, 2003.
- [14] M. S. Floater. On the convergence of derivatives of Bernstein approximation. *J. Approx. Theory*, 134, 130–135, 2005.
- [15] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.
- [16] J. Håstad. Clique is hard to approximate within $|V^{1-\epsilon}|$. *Acta Mathematica*, 182:105–142, 1999.
- [17] J. Håstad. Some optimal inapproximability results. *Journal of ACM*, 48, 798–859, 2001.
- [18] A. Kroó. On Markov inequality for multivariate polynomials, in *Approximation Theory XI (Gatlinburg, 2004)*, 211–229, Nashboro Press.
- [19] G.G. Lorentz. *Bernstein Polynomials*, 2nd ed., Chelsea, 1986.
- [20] T.S. Motzkin and E.G. Straus. Maxima for graphs and a new proof of a theorem of Túrán. *Canadian J. Math.*, 17:533–540, 1965.
- [21] Yu. Nesterov, H. Wolkowicz, and Y. Ye. Semidefinite programming relaxations of nonconvex quadratic optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of semidefinite programming*, pages 361–419. Kluwer Academic Publishers, Norwell, MA, 2000.
- [22] S. Vavasis. Approximation algorithms for concave quadratic programming. In C.A. Floudas and P. Pardalos, eds., *Recent Advances in Global Optimization*, Princeton University Press, 3–18, 1992.
- [23] S. Waldron. Sharp error estimates for multivariate positive linear operators which reproduce the linear polynomials. In *Approximation Theory IX - Vol. 1*, 339–346, (edited by C. K. Chui and L. L. Schumaker), Vanderbilt University Press, 1998.