On the complexity of optimization over the standard simplex

E. de Klerk D. den Hertog G. Elabwabi

July 14, 2006

Abstract

We review complexity results for minimizing polynomials over the standard simplex and unit hypercube. In addition, we derive new results on the computational complexity of approximating the minimum of some classes of functions (including Lipschitz continuous functions) on the standard simplex. The main tools used in the analysis are Bernstein approximation and Lagrange interpolation on the simplex combined with 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, multivariate Lagrange interpolation, linear programming

AMS classification: 90C60, 90C56, 90C26.

1 Introduction

In this paper we study the computational complexity of approximating the minimum value of a function on the standard simplex. Thus we study the problem:

$$\underline{f} := \min \{ f(x) \ : \ x \in \Delta_m \}, \tag{1.1}$$

where Δ_m is the standard (or unit) simplex

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

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 [5]). In this case the function is Lipschitz continuous, but transcendental. Another example is the minimization of the expected shortfall of a portfolio [1, 7]. A third example is to compute the Lebesgue constant associated with a given set of interpolation nodes on a simplex;

this is important in finite element methods (Hesthaven [23]). Here the function in question is the sum of absolute values of polynomials (see also §5).

We will prove the existence of polynomial-time ϵ-approximation algorithms (in a well-defined sense) for minimizing Lipschitz continuous functions as well as functions with bounded Hessians or suitably bounded higher order derivatives on the simplex. This will be done by approximating the function in question by a polynomial using Bernstein approximation or Lagrange interpolation, and, in the case of Lagrange interpolation, subsequently applying the polynomial-time approximation scheme (PTAS) for minimizing polynomials of fixed degree over the simplex due to De Klerk, Laurent and Parrilo [12]. The Bernstein approximations will in fact only be used to analyze the simple algorithm that evaluates f on a regular grid on the simplex, and returns the smallest value.

Some practical algorithms for optimization of Lipschitz continuous functions on the simplex have been investigated (see e.g. Bagirov and Rubinov [4]), but these algorithms are not known to have the polynomial-time ϵ-approximation property. It is therefore the purpose of this paper to show that it is meaningful to search for practical polynomial-time ϵ-approximation algorithms for problem (1.1).

1.1 Complexity of approximating minima

Consider the generic optimization problem:

$$\underline{f} := \min\left\{f(x) \;:\; x \in K\right\}, \tag{1.2}$$

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

$$\bar{f} := \max\left\{f(x) \;:\; x \in K\right\}.$$

In this paper we will consider the case where K is the standard simplex Δ_m, 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 [3], Bellare and Rogaway [6], Bomze and De Klerk [8], De Klerk, Parrilo and Laurent [12], Nesterov et al. [27], and Vavasis [29].

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

$$\left|\psi_\epsilon - \underline{f}\right| \le \epsilon(\bar{f} - \underline{f}). \tag{1.3}$$

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

If we replace condition (1.3) by the condition

$$\left|\psi_\epsilon - \underline{f}\right| \le \epsilon,$$

then we speak of an ϵ-approximation of $\underline{f}$ in the weak sense.

The following definition is from De Klerk, Laurent and Parrilo [12], and is consistent with the corresponding definition in combinatorial optimization.

Definition 1.2 (PTAS/FPTAS) *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 m and the bit size required to represent f, we say that the problem allows a polynomial time approximation scheme (PTAS).*

If, in addition, the computational time is bounded by a polynomial in $1/\epsilon$, one speaks of a fully polynomial time approximation scheme (FPTAS).

1.2 Known complexity results

If $K = \Delta_m$, then computing $\underline{f}$ is an NP-hard problem, already for quadratic polynomials, as it contains the maximum stable set problem as a special case. 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 [26].

Bomze and De Klerk [8] showed that, for $K = \Delta_m$ and f quadratic, problem (1.2) allows a PTAS. This result was extended to polynomials of fixed degree by De Klerk, Laurent, and Parrilo [12]. On the other hand, this problem cannot have a FPTAS, unless NP=ZPP, due to inapproximability results for the maximum stable set problem by Håstad [21].

Another negative result is due to Bellare and Rogaway [6], who proved that if P $\neq$ NP and $\epsilon \in (0, 1/3)$, there is no polynomial time ϵ-approximation algorithm in the weak sense for the problem of minimizing a polynomial of total degree $d \geq 2$ over the set $K = \{x \in [0,1]^m \mid Ax \leq b\}$.

If $K = [0,1]^m$ and f quadratic, then problem (1.2) 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 [20], and related approximation results for quadratic optimization over a hypercube were given by Nesterov *et al.* [27]. On the negative side, the maximum cut problem cannot be approximated within $\epsilon = 1/17$ (Håstad [22]), and it follows that problem (1.2) does not allow a PTAS for any class of functions that includes the quadratic polynomials if $K = [0,1]^m$. 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 aforementioned papers. In particular, we consider three classes of (multivariate) functions that satisfy suitable ”smoothness” conditions on the standard simplex.

Notation

We equip the space of continuous functions $C(K)$ on on a compact convex set K with the usual supremum norm:

$$\|f\|_{\infty,K} := \max_{x\in K} |f(x)|.$$

We will need the following notation to express bounds on (higher order) derivatives. We denote by $C^k(K)$ the k times continuously differentiable functions on K (or on an open set containing K), and define the following semi-norm on $C^k(K)$:

$$\|D^k f\|_{\infty,\Delta_m} := \sup_{x\in K,\ h\in \mathbb{R}^m,\ \|h\|=1} |D_h^k f(x)|,$$

where $D_h^k f$ is obtained by taking the derivative of f in the direction h successively k times. Finally, recall that the modulus of continuity of $f \in C(K)$ is defined as

$$\omega(f,\delta) := \max_{\substack{x,y\in K \\ \|x-y\|\le\delta}} |f(x) - f(y)|,$$

and satisfies:

$$\omega(f,\kappa\delta) \le \lceil\kappa\rceil\omega(f,\delta) < (1+\kappa)\omega(f,\delta) \qquad \forall\kappa > 0. \tag{1.4}$$

Three function classes

We now introduce three classes of functions that allow polynomial-time ϵ-approximations for the problem $\min_{x\in\Delta_m} f(x)$. We will always make the following assumption.

Assumption 1 *We assume that f is a computable function, and $f(x)$ may be computed in time polynomial in the bit size of x and in the bit size needed to represent f.*

In what follows, we assume that the constants involved in defining the function classes are independent of the number of variables m.

The first class of functions meet the Lipshitz condition of given order $\alpha > 0$ with respect to a given constant $L > 0$:

$$\mathrm{Lip}_L(\alpha) := \{f \in C(\Delta_m) \ : \ \omega(f,\delta) \le \delta^\alpha L\}. \tag{1.5}$$

The second class of functions we consider have second order derivatives bounded by a given $L > 0$ on Δ_m:

$$\left\{f \in C^2(\Delta_m) \ : \ \|D^2 f\|_{\infty,\Delta_m} \le L\right\}. \tag{1.6}$$

The third and last class of functions have suitably bounded higher order derivatives, and is defined in terms of two given constants $L > 0$ and $\bar{k} \in \mathbb{N}$ as follows:

$$\left\{ f \in C^\infty(\Delta_m) \; : \; \|D^{k+1} f\|_{\infty,\Delta_m} \leq 2\pi L \left(\frac{k}{4\sqrt{2}e}\right)^k k^2 \left(\bar{f} - \underline{f}\right) \quad \forall k \geq \bar{k} \right\}. \tag{1.7}$$

Note that this class includes the m-variate polynomials of total degree at most $\bar{k}$, since in this case $\|D^{k+1}f\|_{\infty,\Delta_m} = 0$ if $k \geq \bar{k}$.

Note, however, that polynomials of fixed degree on the simplex do not necessarily belong to the classes (1.5) or (1.6): The Markov inequality for an m-variate polynomial of total degree d defined on a convex body K states that

$$\|Dp\|_{\infty,K} \leq \frac{2d^2}{\omega_{\min}(K)} \|p\|_{\infty,K}, \tag{1.8}$$

where $\omega_{\min}(K)$ is the minimum distance between two distinct parallel supporting hyperplanes of K (see e.g. Kroó [24]). For example, if K is the m-dimensional simplex

$$K = \left\{ x \in \mathbb{R}^m \; : \; \sum_{i=1}^m x_i \leq 1, \; x \geq 0 \right\}, \tag{1.9}$$

then $\omega_{\min}(K) = 1/\sqrt{m}$. The Markov inequality (1.8) is sharp up to a constant, as is clear from the example $p(x) = \sum_{i=1}^m x_i$, and K given by (1.9).

This means that the class of m-variate polynomials of fixed total degree on the simplex is not a subset of the class (1.5) or (1.6).

Main result

We show that, for any $\epsilon > 0$ there exists a polynomial-time ϵ-approximation in the weak sense for the problem of minimizing a function from the classes (1.5) and (1.6) over Δ_m. Moreover, we prove the existence of a PTAS for problem (1.1) for functions from the class (1.7).

We use the properties of multivariate Bernstein operators and Lagrange interpolation on the simplex, as well as the main result in [12] to derive the new results. Our review of multivariate Bernstein operators will be self-contained, since the results we need are somewhat scattered throughout the literature, and even then not always in the required form. More information on Bernstein operators may be found in [2, 13, 14, 15, 16, 25, 30]. We will not give proofs for the results on Lagrange interpolation, as these may be found in the two papers [9, 31]. More information on multivariate Lagrange interpolation is given in [19, 28].

2 Positive linear operators

The results on Bernstein operators presented in the next section 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

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

that in turn implies

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

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

The following result is taken from Waldron [30], but this 'Korovkin-type' analysis is in fact much older (cf. Theorem 4.4 in [13], Chapter 5 in [2], 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 $int(K) = \emptyset$). Let $\phi_{1,i}(x) = x_i$ and $\phi_{2,i}(x) = x_i^2$ $(i = 1, \ldots, m)$.*

One has

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

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

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

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

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

Evaluating the inequality at $x = y$ completes the proof. □

If we do not assume that $f \in C^2(K)$, but merely that f is continous 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 [2]) *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 $\omega(f,\delta)$):

$$\begin{aligned} |f(y) - f(x)| &\leq \omega\left(f, \|x-y\|\right) \\ &\leq \left(1 + \frac{1}{\delta}\|x-y\|\right)\omega(f,\delta) \quad \text{(by (1.4))} \\ &\leq \left(1 + \frac{1}{\delta^2}\|x-y\|^2\right)\omega(f,\delta) \quad (\text{since } \|x-y\| > \delta) \\ &= \left(1 + \frac{1}{\delta^2}(\|x\|^2 - 2x^Ty + \|y\|^2)\right)\omega(f,\delta). \end{aligned}$$

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. □

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

$$\|U_n\phi_{2,i} - \phi_{2,i}\|_{\infty,K} \to 0 \text{ as } n \to \infty \quad (i = 1, \ldots, m). \tag{2.10}$$

By the theorems above, this implies that $U_n f \to f$ uniformly. The rate of convergence is determined by the rate of convergence in (2.10).

3 Bernstein operators on a simplex

3.1 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

$$p_{n,i}(x) = \binom{n}{i} x^i(1-x)^{n-i} \quad (i = 0, \ldots, n). \tag{3.11}$$

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 [13], Chapter 1, for a proof) that B_n is a positive linear operator that preserves constants, and

$$(B_n(\phi_1))(x) = x, \;\; (B_n(\phi_2))(x) = \phi_2(x) + \frac{1}{n}x(1-x), \tag{3.12}$$

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 [25], and for recent developments [18].

3.2 Multivariate Bernstein operators

Let $f \in C^2(\Delta_m)$. The Bernstein approximation of f of order n on Δ_m is the polynomial

$$B_n(f)(x) := \sum_{\alpha\in I(m,n)} f\left(\frac{\alpha}{n}\right)\frac{n!}{\alpha!}\,x^\alpha, \qquad (x \in \Delta_m), \tag{3.13}$$

where

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

Note that this definition coincides with the definition for the univariate case when $m = 2$. We also define a regular grid on Δ_m via

$$\Delta(m,n) := \{x \in \Delta_m \;:\; nx \in \mathbb{N}_0^m\}\,.$$

Thus one has $|\Delta(m,n)| = |I(m,n)| = \binom{n+m}{m}$. Notice that the Bernstein approximation of f in (3.13) only involves the function values that f takes on the grid $\Delta(m,n)$.

The following property of the Bernstein operator is sometimes called the *stability property.*

Lemma 3.1 *One has*

$$\min_{x\in\Delta(m,n)} f(x) \leq B_n(f)(\bar{x}) \leq \max_{x\in\Delta(m,n)} f(x) \quad \forall\, \bar{x} \in \Delta_m.$$

Proof. For any fixed $\bar{x} \in \Delta_m$, $B_n(f)(\bar{x})$ is given as a convex combination of the values $f(x)$ $(x \in \Delta(m,n))$. To see this, one may use the multinomial identity

$$\left(\sum_{j=1}^m x_i\right)^n = \sum_{\alpha\in I(m,n)} \frac{n!}{\alpha!}\,x^\alpha \quad (= 1 \quad \text{if } x \in \Delta_m)\,, \tag{3.14}$$

to find that $\sum_{\alpha\in I(m,n)} \frac{n!}{\alpha!}\,\bar{x}^\alpha = 1$. □

The significance of the stability property is that $\min_{x\in\Delta(m,n)} f(x)$ is at least as good an approximation of $\underline{f}$ as $\min_{x\in\Delta_m} B_n(f)(x)$ is. We therefore need not minimize $B_n(f)$ in practice — we only use $B_n(f)$ in our analysis to prove error bounds for the simple approximation of $\underline{f}$ given by $\min_{x\in\Delta(m,n)} f(x)$.

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 (3.12)). We include a proof for completeness.

Lemma 3.2 *Let $f \in C^2(\Delta_m)$ and $B_n(f)$ as defined in (3.13). Then*

$$B_n(\phi_{1,i}) = \phi_{1,i}, \;\; B_n(\phi_{2,i})(x) = \phi_{2,i}(x) + \frac{1}{n}x_i(1-x_i) \;\; (i=1,\ldots,m), \tag{3.15}$$

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

Proof. By (3.13), 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 ith standard unit vector, and for the last equality we used the multinomial identity (3.14).

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 3.1 (see e.g. Waldron [30]) *Let $f \in C^2(\Delta_m)$ and $B_n(f)$ as defined in (3.13). One has*

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

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

$$\sum_{i=1}^{m} \left(B_n(\phi_{2,i}) - \phi_{2,i}\right)(x) = \frac{1}{n}\sum_{i=1}^{m} x_i(1-x_i). \tag{3.16}$$

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 $x_i = \frac{1}{m}$ $(i = 1, \ldots, m)$. Thus

$$\sum_{i=1}^{m} x_i(1-x_i) \le 1 - \frac{1}{m} \quad \forall x \in \Delta_m. \tag{3.17}$$

□

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

Theorem 3.2 (See e.g. [2], §5.2.11) *Let $f \in C(\Delta_m)$ and $B_n(f)$ as defined in (3.13). One has*

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

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

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

where we have used (3.16) and (3.17) to obtain the last inequality. The required result now follows by setting $\delta = \frac{1}{\sqrt{n}}$. □

4 Complexity analysis using Bernstein approximation

We now use the results of the last section to derive complexity results for optimization over Δ_m. In particular, we derive polynomial-time ϵ-approximation procedures (in the weak sense) for the function classes (1.5) and (1.6).

Theorem 4.1 *Let $\epsilon > 0$ be given and assume $f \in Lip_L(\alpha)$ (see (1.5)) and satisfies Assumption 1. Then for*

$$n = \left\lceil \left(\frac{2L}{\epsilon}\right)^{\frac{1}{2\alpha}} \right\rceil,$$

an implementable polynomial-time approximation (in the weak sense) of $\underline{f}$ is given by

$$\psi_\epsilon := \min_{x \in \Delta(m,n)} f(x).$$

Proof. By Theorem 3.2 one has

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

where the last inequality is due to $f \in \mathrm{Lip}_L(\alpha)$. Using this bound together with the stability property (Lemma 3.1), one has

$$\min_{x \in \Delta(m,n)} f(x) \leq \min_{x \in \Delta_m} B_n(f)(x) \leq \underline{f} + 2L\left(\frac{1}{\sqrt{n}}\right)^\alpha \leq \underline{f} + \epsilon,$$

if $n \geq \left(\frac{2L}{\epsilon}\right)^{1/2\alpha}$.

Finally, note that computing $\min_{x\in\Delta(m,n)} f(x)$ requires $\binom{n+m}{n}$ evaluations of f, and that $\binom{n+m}{n}$ is a polynomial in m if we choose $n = \left\lceil \left(\frac{2L}{\epsilon}\right)^{1/2\alpha} \right\rceil$. □

We have a similar result for functions from the class (1.6) (suitably bounded second derivatives).

Theorem 4.2 *Let $\epsilon > 0$ be given and assume f belongs to the class (1.6) and satisfies Assumption 1. Then for*

$$n = \left\lceil \left(\frac{L}{2\epsilon}\right) \right\rceil,$$

an implementable polynomial-time approximation (in the weak sense) of $\underline{f}$ is given by

$$\psi_\epsilon := \min_{x \in \Delta(m,n)} f(x).$$

Proof. The proof is identical to that of the previous theorem, except that Theorem 3.1 is now required (as opposed to Theorem 3.2). □

Note again that the Bernstein approximations were only used to derive error bounds for the simpler approximation of $\underline{f}$ given by $\min_{x\in\Delta(m,n)} f(x)$. A natural question therefore is whether it is possible to derive the results of this section without using Bernstein approximations. This however does not seem to be straightforward.

5 Lagrange interpolation on the simplex

In addition to Bernstein approximation, we consider Lagrange interpolation on the standard simplex using the interpolation points $\Delta(m,n)$. This interpolation problem is known to be poised, i.e. the Lagrange interpolant of degree n is uniquely determined by the values $f(\theta)$ ($\theta \in \Delta(m,n)$) (see §2.2 in the survey by Gasca and Sauer [19], and the references given there).

The fundamental Lagrange polynomial associated with an interpolation point θ will be denoted by l_θ. In other words, for $x \in \Delta(m,n)$ we have

$$l_\theta(x) = \begin{cases} 1 & \text{if } x = \theta \\ 0 & \text{else.} \end{cases}$$

Note that we may write the Lagrange interpolant as

$$L_{\Delta(m,n)}(f)(x) := \sum_{\theta\in\Delta(m,n)} f(\theta) l_\theta(x).$$

The associated Lebesgue constant $\max_{x\in\Delta_m}\sum_{\theta\in\Delta(m,n)}|l_\theta(x)|$ can be bounded from above as follows.

Theorem 5.1 (Bos [9]) *One has*

$$\max_{x\in\Delta_m}\sum_{\theta\in\Delta(m,n)}|l_\theta(x)| \le \binom{2n-1}{n}. \tag{5.18}$$

The last result may be used to give an error bound for Lagrange interpolation on the simplex using the interpolation points $\theta \in \Delta(m,n)$. To this end, we require the following result.

Theorem 5.2 (Corollary 5.2.8 in [31]) *Let $f \in C^{n+1}(\Delta_m)$ and let Θ denote a finite set of points in $\mathbb{R}^m$. One has*

$$|f(x) - L_\Theta(f)(x)| \le \frac{1}{(n+1)!}\|D^{n+1}f\|_{\infty,conv(\Theta)}\sum_{\theta\in\Theta}\|\theta - x\|^{n+1}|l_\theta(x)| \quad \forall x \in conv(\Theta), \tag{5.19}$$

where $conv(\Theta)$ denotes the convex hull of Θ.

Using the bound (5.18) on the Lebesgue function and the fact that the diameter of Δ_m is $\sqrt{2}$, we obtain the following corollary.

Corollary 5.1 *Let $f \in C^{n+1}(\Delta_m)$ and $\Theta = \Delta(m,n)$. One has*

$$\begin{aligned} \|f - L_{\Delta(m,n)}(f)\|_{\infty,\Delta_m} &\le \frac{1}{(n+1)!}\|D^{n+1}f\|_{\infty,\Delta_m}\left(\sqrt{2}\right)^{n+1}\binom{2n-1}{n} \\ &\le \frac{\left(\frac{4\sqrt{2}e}{n}\right)^n}{2\pi(n+1)n^2}\|D^{n+1}f\|_{\infty,\Delta_m}. \end{aligned}$$

Proof. The first inequality follows from (5.19) by using (5.18), as well as $\mathrm{conv}(\Theta) = \Delta_m$, and the observation that $\|\theta - x\| \leq \sqrt{2}$ if $x \in \Delta_m$ and $\theta \in \Delta(m,n)$. The second inequality of the corollary requires the following version of Sterling's formula:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}},$$

that implies

$$\frac{1}{(n+1)!}\left(\sqrt{2}\right)^{n+1}\binom{2n-1}{n} \leq \frac{\left(\frac{4\sqrt{2}e}{n}\right)^n}{2\pi(n+1)n^2}.$$

□

6 Complexity analysis using Lagrange interpolation

One may use the results of the last section to derive further complexity results for optimization over Δ_m. The idea is as follows:

1. First approximate f on Δ_m via Lagrange interpolation.
2. Minimize the Lagrange interpolant approximately in polynomial time.

First of all we relate the degree of the Lagrange interpolant to the error.

Lemma 6.1 *Assume $\epsilon > 0$ given and that f belongs to the class (1.7) and meets Assumption 1. For*

$$n = \max\left\{\bar{k}, \left\lceil \frac{L}{\epsilon} \right\rceil\right\},$$

one has

$$\|f - L_{\Delta(m,n)}(f)\|_{\infty,\Delta_m} \leq \epsilon(\bar{f} - \underline{f}).$$

Proof. The proof follows immediately from Corollary 5.1 and the definition of the function class (1.7). □

We also need to verify that one may evaluate $L_{\Delta(m,n)}(f)$ in polynomial time.

Lemma 6.2 *For fixed n, one may compute $L_{\Delta(m,n)}(f)(x)$ in time polynomial in m, the bit size of x, and the bit size needed to represent f.*

Proof. We may compute $L_{\Delta(m,n)}(f)(x)$ using

$$L_{\Delta(m,n)}(f)(x) := \sum_{\theta \in \Delta(m,n)} f(\theta) l_\theta(x),$$

if we know the fundamental Lagrange polynomials l_θ ($\theta \in \Delta(m,n)$). Note that the fundamental Lagrange polynomials form an orthogonal basis of the m-variate polynomials of total degree at most n, with respect to the inner product

$$\langle p, q \rangle := \sum_{\theta \in \Delta(m,n)} p(\theta) q(\theta).$$

We may therefore use the Gram-Schmidt procedure to convert the standard monomial basis into the basis l_θ ($\theta \in \Delta(m,n)$). This procedure requires a number of operations polynomial in $\binom{m+n}{n}$. Moreover, the intermediate bit sizes during the Gram-Schmidt procedure remain suitably bounded, since only summation, multiplication, and division is performed.

Also note that the coefficients of $L_{\Delta(m,n)}(f)$ with respect to the basis l_θ ($\theta \in \Delta(m,n)$) are the values $f(\theta)$ ($\theta \in \Delta(m,n)$), which are bounded in absolute value by a polynomial in m, and the bit size required to represent f, by Assumption 1. □

To approximately minimize the Lagrange interpolant, we need the following results due to De Klerk, Laurent, and Parrilo [12] concerning minimization of a form p of degree d on Δ_m. (Related results are given by Faybusovich [17].)

Following [12], we define two sequences of lower bounds on the minimum $\underline{p}$ of p on the Δ_m. The first sequence is given by:

$$p_{\min}^{(r)} := \max\ \lambda \quad \text{s.t. the polynomial } \left(\sum_{i=1}^m x_i\right)^r \left(p(x) - \lambda \left(\sum_{i=1}^m x_i\right)^d\right)$$
$$\text{has nonnegative coefficients} \quad (r = 0, 1, \ldots)$$

Note these bounds may be computed using linear programming, and that this computation is in polynomial time when $r = O(1)$ and $d = O(1)$.

The second sequence is given by

$$p_{\Delta(m,r+d)} := \min\{p(x) \ :\ x \in \Delta(m, r+d)\} \quad (r = 0, 1, \ldots).$$

Obviously, these bounds may again be obtained in polynomial time if $r = O(1)$ and $d = O(1)$.

De Klerk, Laurent, and Parrilo [12] showed that both these sequences of bounds converge to $\underline{p}$ as follows.

Theorem 6.1 (De Klerk, Laurent, and Parrilo [12]) *Let $p(x)$ be a form of degree d and $r \geq 0$ an integer. Then,*

$$\underline{p} - p_{\min}^{(r)} \leq \left(\frac{1}{w_r(d)} - 1\right) \binom{2d-1}{d} d^d (\bar{p} - \underline{p}), \tag{6.20}$$

$$p_{\Delta(m,r+d)} - \underline{p} \leq (1 - w_r(d)) \binom{2d-1}{d} d^d (\bar{p} - \underline{p}), \tag{6.21}$$

where

$$w_r(d) := \frac{(r+d)!}{r!(r+d)^d} = \prod_{i=1}^{d-1}\left(1 - \frac{i}{r+d}\right).$$

One can verify that

$$1 - \binom{d}{2}\frac{1}{r+d} \leq w_r(d) \leq 1,$$

which implies that $\lim_{r\to\infty} w_r(d) = 1$.

We are now in a position to derive an ϵ-approximation algorithm for minimizing f for the class (1.7) over Δ_m.

Theorem 6.2 *Assume that $\epsilon \in (0,1]$ is given and that f belongs to the class (1.7) and meets Assumption 1. Let*

$$n = \max\left\{\bar{k}, \left\lceil \frac{2L}{\epsilon} \right\rceil\right\}, \quad r = \left\lceil \frac{4\binom{n}{2}\binom{2n-1}{n}n^n}{\epsilon} - n \right\rceil. \tag{6.22}$$

Then

$$\min_{x\in\Delta(m,r+n)} L_{\Delta(m,n)}(f)(x) - \underline{f} \leq \epsilon(\bar{f} - \underline{f}). \tag{6.23}$$

Proof. The choice of n in (6.22) guarantees that

$$\|f - L_{\Delta(m,n)}(f)\|_{\infty,\Delta_m} \leq \frac{\epsilon}{2}(\bar{f} - \underline{f}), \tag{6.24}$$

by Lemma 6.1. Moreover, by Theorem 6.1 (see (6.21)), the choice of r in (6.22) implies

$$\begin{aligned}
\min_{x\in\Delta(m,r+n)} L_{\Delta(m,n)}(f)(x) - \min_{x\in\Delta_m} L_{\Delta(m,n)}(f)(x) &\leq \frac{\epsilon}{4}\left(\max_{x\in\Delta_m} L_{\Delta(m,n)}(f)(x) - \min_{x\in\Delta_m} L_{\Delta(m,n)}(f)(x)\right) \\
&\leq \frac{\epsilon}{4}\left((1+\epsilon)(\bar{f} - \underline{f})\right) \\
&\leq \frac{\epsilon}{2}\left(\bar{f} - \underline{f}\right),
\end{aligned}$$

where the last two inequalities follow from (6.24) and $\epsilon \leq 1$ respectively. Using (6.24) once more, we have that $\min_{x\in\Delta_m} L_{\Delta(m,n)}(f)(x) \leq \underline{f} + \frac{\epsilon}{2}(\bar{f} - \underline{f})$, and the required bound (6.23) follows.

Note that the computation of $\min_{x\in\Delta(m,r+n)} L_{\Delta(m,n)}(f)(x)$ requires $\binom{m+r+n}{r+n}$ evaluations of $L_{\Delta(m,n)}(f)$, and this number of evaluations is polynomial in m since $n + r = O(1)$. Moreover, each evaluation my be done in time polynomial in m and in the bit size needed to represent f, by Lemma 6.2.

□

Corollary 6.1 *There exists a PTAS for the minimization of functions from the class 1.7 over Δ_m.*

Remarks

The approximation algorithm in Theorem 6.2 involves evaluating the Lagrange interpolant on a suitable regular grid (see (6.23)). It is also possible (and more practical) to use linear programming (LP) (or semidefinite programming (SDP)) to compute an ϵ-approximation of the minimum of the interpolant; The LP approach requires the bound (6.20) from Theorem 6.1, and stronger SDP-based bounds are discussed in [12].

The definition of the class of functions (1.7) is determined by the error bound for Lagrange interpolation of a regular grid as given in Corollary 5.1. This error bound depends on the choice of interpolation nodes. It is well-known that the regular grid is not a good choice, since the associated Lebesgue constant grows exponentially with the number of nodes n (see (5.18)). For low dimensional simplices, better choices of interpolation nodes are known, like Chebyshev nodes for the one-dimensional case. For Chebyshev nodes the Lebesgue constant grows like $\log n$. Unfortunately, very little is known for general dimension, and the regular grid is the only case where a useful upper bound on the Lebesgue constant is available; see [23] for a detailed discussion of these issues.

7 Conclusion

Our approximation 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 [10] and the references therein.) Thus our results in Theorems 4.1, 4.2, and 6.2 may be viewed as existence proofs for derivative free optimization methods that give polynomial-time approximation guarantees for minimization of certain classes of functions over the simplex.

The algorithms in Theorems 4.1 and 4.2 simply involved evaluating f on a regular grid on the simplex, and are not really practical, since the required number of function evaluations becomes prohibitively large if high accuracy is required. Algorithms based on approximate minimization of a Lagrange interpolant seem more promising from the practical point of view. Indeed, a recent study showed that using Lagrange interpolation at the Chebyshev nodes, in conjunction with semidefinite programming to minimize the interpolant, works well in the one dimensional case [11].

Thus, we hope that our observations will lead to some practical algorithms for optimization over the simplex that retain polynomial-time approximation guarantees.

References

[1] C. Acerbi and D. Tasche. Expected Shortfall: a natural coherent alternative to Value at Risk. *Economic Notes*, 31(2), 379–388, 2002.

[2] F. Altomare and M. Campiti. *Korovkin-type approximation theory and its applications*. De Guyter studies in mathematics 17, Walter de Guyter publishers, Berlin, 1994.

[3] 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.

[4] A.M. Bagirov and A.M. Rubinov. Global minimization of increasing positively homogeneous functions over the unit simplex. *Annals of Operations Research.* 98, 171–188, 2000.

[5] 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.

[6] M. Bellare and P. Rogaway. The complexity of approximating a nonlinear program. *Mathematical Programming*, 69:429–441, 1995.

[7] D. Bertsimas, G. Lauprete and A. Samarov. Shortfall as a risk measure: properties and optimization. *Journal of Economic Dynamics and Control*, 28:7, 1227–1480, 2004.

[8] I. Bomze and E. De Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2), 163–185, 2002.

[9] L.P. Bos. Bounding the Lebesque function for Lagrange interpolation in a simplex. *Journal of Approximation Theory*, 38, 43–59, 1983.

[10] 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`

[11] E. de Klerk, D. den Hertog, and G. Elabwabi. Optimization of univariate functions on bounded intervals by interpolation and semidefinite programming. CentER Discussion paper 2006-26, Tilburg University, The Netherlands, 2006.

[12] 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.

[13] R.A. DeVore and G.G. Lorentz. *Constructive Approximation.* Springer-Verlag, Berlin, 1993.

[14] Z. Ditzian. Inverse theorems for multidimensional Bernstein operators, *Pacific Journal of Mathematics* 121 (2), 293–319, 1986.

[15] Z. Ditzian. Best polynomial approximation and Bernstein polynomial approximation on a simplex, *Indagationes Mathematicae* 92, 243–256, 1989.

[16] Z. Ditzian and X. Zhou. Optimal approximation class for multivariate Bernstein operators, *Pacific Journal of Mathematics* 158, 93–120, 1993.

[17] 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.

[18] M. S. Floater. On the convergence of derivatives of Bernstein approximation. *Journal of Approximation Theory*, 134, 130–135, 2005.

[19] M. Gasca and T. Sauer. Multivariate Polynomial Interpolation. *Advances in Computational Mathematics*, **12**, pp. 377–410, 2000.

[20] 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.

[21] J. Håstad. Clique is hard to approximate within $|V^{1-\epsilon}|$. *Acta Mathematica*, 182, 105–142, 1999.

[22] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48, 798–859, 2001.

[23] J.S. Hesthaven. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. *SIAM Journal on Numerical Analysis*, 35(2), 655–676, 1998.

[24] A. Kroó. On Markov inequality for multivariate polynomials, in Approximation Theory XI (Gatlinburg, 2004), 211-229, Nashboro Press, 2004.

[25] G.G. Lorentz. *Bernstein Polynomials*, 2nd ed., Chelsea, 1986.

[26] T.S. Motzkin and E.G. Straus. Maxima for graphs and a new proof of a theorem of Túran. *Canadian Journal of Mathematics*, 17, 533-540, 1965.

[27] 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*, 361–419. Kluwer Academic Publishers, Norwell, MA, 2000.

[28] T. Sauer and Y. Xu. On multivariate Lagrange interpolation. *Mathematics of Computation*, 64, 1147-1170, 1995.

[29] 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.

[30] 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.

[31] S. Waldron. A multivariate form of Hardy's inequality and L_p-error bounds for multivariate Lagrange interpolation schemes. *SIAM Journal on Mathematical Analysis* 28, no. 1, 233–258, 1997.