An alternative proof of a PTAS for fixed-degree polynomial optimization over the simplex

Etienne de Klerk · Monique Laurent · Zhao Sun

Abstract The problem of minimizing a polynomial over the standard simplex is one of the basic NP-hard nonlinear optimization problems — it contains the maximum clique problem in graphs as a special case. It is known that the problem allows a polynomial-time approximation scheme (PTAS) for polynomials of fixed degree, which is based on polynomial evaluations at the points of a sequence of regular grids. In this paper, we provide an alternative proof of the PTAS property. The proof relies on the properties of Bernstein approximation on the simplex. We also refine a known error bound for the scheme for polynomials of degree three. The main contribution of the paper is to provide new insight into the PTAS by establishing precise links with Bernstein approximation and the multinomial distribution.

Etienne de Klerk
Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands
E-mail: E.deKlerk@uvt.nl

Monique Laurent
Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University
CWI, Postbus 94079, 1090 GB Amsterdam, The Netherlands
E-mail: monique@cwi.nl

Zhao Sun
Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands
E-mail: z.sun@uvt.nl

1 Introduction and preliminaries

1.1 Polynomial optimization over the simplex

Let $\mathcal{H}_{n,d}$ denote the set of all homogeneous polynomials of degree d in n variables. We consider the problem of minimizing a polynomial $f \in \mathcal{H}_{n,d}$ on the standard simplex

$$\Delta_n = \left\{ x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1 \right\}.$$

That is, the problem of computing

$$\underline{f} = \min_{x \in \Delta_n} f(x), \quad \text{or} \quad \overline{f} = \max_{x \in \Delta_n} f(x).$$

This problem is known to be NP-hard, even if f is a quadratic function. Indeed, if G denotes a graph with vertex set V and adjacency matrix A, and I denotes the identity matrix, then the maximum cardinality $\alpha(G)$ of a stable set in G can be obtained via

$$\frac{1}{\alpha(G)} = \min_{x \in \Delta_{|V|}} x^T (I + A) x,$$

by a theorem of Motzkin and Straus [15].

On the other hand, the problem does allow a polynomial-time approximation scheme (PTAS), as was shown by Bomze and De Klerk (for quadratic f) [3], and by De Klerk, Laurent and Parrilo (for more general, fixed-degree f) [11]. The PTAS is particularly simple, and takes the minimum of f on the regular grid:

$$\Delta(n, r) = \{x \in \Delta_n : rx \in \mathbb{N}^n\}$$

for increasing values of r. We denote the minimum over the grid by

$$f_{\Delta(n,r)} = \min_{x \in \Delta(n,r)} f(x),$$

and observe that the computation of $f_{\Delta(n,r)}$ requires $|\Delta(n, r)| = \binom{n+r-1}{r}$ evaluations of f.

Several properties of the regular grid $\Delta(n, r)$ have been studied in the literature. In Bos [5], the Lebesgue constant of $\Delta(n, r)$ is studied in the context of Lagrange interpolation and finite element methods. Given a point $x \in \Delta_n$, Bomze, Gollowitzer and Yildirim [4] study a scheme to find the closest point to x on $\Delta(n, r)$ with respect to certain norms (including ℓ_p-norms for finite p). Furthermore, for any quadratic polynomial $f \in \mathcal{H}_{n,2}$ and $r \geq 2$, Sagol and Yildirim [18] and Yildirim [20] consider the upper bound on $\underline{f}$ defined by $\min_{x \in \cup_{k=2}^r \Delta(n,k)} f(x)$ $(r = 2, 3, \ldots)$, and analyze the error bound. The following error bounds are known for the approximation $f_{\Delta(n,r)}$ of $\underline{f}$.

Theorem 1 ((i) Bomze-De Klerk [3] and (ii) De Klerk-Laurent-Parrilo [11])

(i) For any quadratic polynomial $f \in \mathcal{H}_{n,2}$ and $r \geq 2$, one has

$$f_{\Delta(n,r)} - \underline{f} \leq \frac{\overline{f} - \underline{f}}{r}.$$

(ii) For any polynomial $f \in \mathcal{H}_{n,d}$ and $r \geq d$, one has

$$f_{\Delta(n,r)} - \underline{f} \leq \left(1 - \frac{r^{\underline{d}}}{r^d}\right)\binom{2d-1}{d}d^d(\overline{f} - \underline{f}),$$

where $r^{\underline{d}} = r(r-1)\cdots(r-d+1)$.

Note that these results indeed imply the existence of a PTAS in the sense of the following definition, that has been used by several authors (see e.g. [2,10,11,17,19]).

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

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

The approximation is called implementable if $\psi_\epsilon = f(x_\epsilon)$ for some feasible x_ϵ. If a problem allows an implementable approximation $\psi_\epsilon = f(x_\epsilon)$ for each $\epsilon \in (0,1]$, such that the feasible x_ϵ 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).

Indeed, Theorem 1 clearly implies that $f_{\Delta(n,r)}$ yields a PTAS for polynomials of fixed degree d. In this paper we give alternative proofs of this result, and also refine the relevant error bound in the special case of degree three polynomials. The proof of the PTAS in the quadratic case is completely elementary, and much simpler than the proof given in [3]. It is in fact closely related to a proof given by Nesterov [16]; see Section 6. In fact, the main contribution of our paper is to provide new insight into the PTAS by establishing precise connections with Bernstein approximation, and the approach of Nesterov [16], which in turn requires an understanding of the precise connection to the multinomial distribution. We also prove, by giving an example, that the error bound in Theorem 1(i) is tight in terms of its dependence on r.

Our main tool will be Bernstein approximation on the simplex (which is similar to the approach used by De Klerk and Laurent [12] for polynomial optimization over the hypercube).

We start by reviewing the necessary background material on Bernstein approximation.

1.2 Bernstein approximation on the simplex

Given an integer $r \geq 0$, we define

$$I(n,r) := \left\{\alpha \in \mathbb{N}^n : \sum_{i=1}^n \alpha_i = r\right\} = r\Delta(n,r).$$

The Bernstein approximation of order $r \geq 1$ on the simplex of a polynomial $f \in \mathcal{H}_{n,d}$ is the polynomial $B_r(f) \in \mathcal{H}_{n,r}$ defined by

$$B_r(f)(x) = \sum_{\alpha \in I(n,r)} f\left(\frac{\alpha}{r}\right)\frac{r!}{\alpha!}x^\alpha, \tag{1}$$

where $\alpha! := \prod_{i=1}^n \alpha_i!$ and $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$. For instance, for the constant polynomial $f \equiv 1$, its Bernstein approximation of any order r is $\sum_{\alpha \in I(n,r)} \frac{r!}{\alpha!}x^\alpha$, which is equal to $(\sum_{i=1}^n x_i)^r$ by the multinomial theorem, and thus to 1 for any $x \in \Delta_n$.

There is a vast literature on Bernstein approximation, and the interested reader may consult e.g. the papers by Ditzian [6,7], Ditzian and Zhou [8], the book by Altomare and Campiti [1], and the references therein for more details than given here.

To motivate our use of Bernstein approximation, we state one well-known result that shows uniform convergence.

Theorem 2 (See e.g. [1], §5.2.11) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be any continuous function defined on Δ_n, and $B_r(f)$ as defined in (1). One has*

$$|B_r(f)(x) - f(x)| \leq 2\omega\left(f, \frac{1}{\sqrt{r}}\right) \qquad \forall x \in \Delta_n,$$

where ω denotes the modulus of continuity:

$$\omega(f, \delta) := \max_{\substack{x,y \in \Delta_n \\ \|x-y\| \leq \delta}} |f(x) - f(y)| \qquad (\delta \geq 0).$$

Next we state some simple inequalities relating a polynomial, its Bernstein approximation and their minimum over the set $\Delta(n,r)$ of grid points.

Lemma 1 *Given a polynomial $f \in \mathcal{H}_{n,d}$ and $r \geq 1$, one has*

$$\min_{x \in \Delta_n} B_r(f)(x) \geq f_{\Delta(n,r)}, \tag{2}$$

$$f_{\Delta(n,r)} - \underline{f} \leq \min_{x \in \Delta(n,r)} B_r(f)(x) - \underline{f} \leq \max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\}. \tag{3}$$

Proof We first show (2). Fix $x \in \Delta_n$. By the multinomial theorem, $1 = (\sum_{i=1}^{n} x_i)^r = \sum_{\alpha \in I(n,r)} \frac{r!}{\alpha!} x^\alpha$. Hence, $B_r(f)(x)$ is a convex combination of the values $f(\frac{\alpha}{r})$ ($\alpha \in I(n,r)$), which implies that $B_r(f)(x) \geq \min_{\alpha \in I(n,r)} f(\frac{\alpha}{r}) = f_{\Delta(n,r)}$.

The left most inequality in (3) follows directly from (2). To show the right most inequality, let x^* be a global minimizer of f over Δ_n, so that $f(x^*) = \underline{f}$. Then, $\min_{x \in \Delta_n} B_r(f)(x) - \underline{f}$ is at most $B_r(f)(x^*) - \underline{f} = B_r(f)(x^*) - f(x^*)$, which concludes the proof. □

The motivation for using Bernstein approximation to study the quantity $f_{\Delta(n,r)}$ is now clear from Theorem 2 and relation (2). Indeed, the Bernstein approximation $B_r(f)$ converges uniformly to f as $r \to \infty$, and the minimum of $B_r(f)$ on Δ_n is lower bounded by $f_{\Delta(n,r)}$.

Our strategy for upper bounding the range $f_{\Delta(n,r)} - \underline{f}$ will be to upper bound the (possibly larger) range $\max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\}$ – see Theorems 4, 5, 6 and 8. Hence our results can be seen as refinements of the previously known results quoted in Theorem 1 above.

The following example shows that all inequalities can be strict in relation (3).

Example 1 Consider the quadratic polynomial $f = 2x_1^2 + x_2^2 - 5x_1x_2 \in \mathcal{H}_{2,2}$. Then, $B_2(f)(x) = x_1^2 + \frac{1}{2}x_2^2 - \frac{5}{2}x_1x_2 + x_1 + \frac{1}{2}x_2$. One can easily check that $\underline{f} = -\frac{17}{32}$ (attained at the unique minimizer $(\frac{7}{16}, \frac{9}{16})$), $\min_{x \in \Delta_2} B_2(f)(x) = \frac{7}{16}$ (attained at the unique minimizer $x = (\frac{3}{8}, \frac{5}{8})$), and $f_{\Delta(2,2)} = -\frac{1}{2}$ (attained at the unique minimizer $(\frac{1}{2}, \frac{1}{2})$). In this example, the polynomial f and its Bernstein approximation $B_2(f)(x)$ do not have a common minimizer over the simplex.

Moreover, we note that $\overline{f} = 2$ and $\max_{x \in \Delta_2}\{B_2(f)(x) - f(x)\} = 1$, so that we have the following chain of strict inequalities:

$$f_{\Delta(2,2)} - \underline{f} \, (= \frac{1}{32}) < \min_{x \in \Delta_2} B_2(f)(x) - \underline{f} \, (= \frac{31}{32}) < \max_{x \in \Delta_2}\{B_2(f)(x) - f(x)\} \, (= 1) < \frac{1}{2}(\overline{f} - \underline{f}) \, (= \frac{81}{64}),$$

which shows that all the inequalities can be strict in (3).

1.3 Bernstein coefficients

For any polynomial $f = \sum_{\beta \in I(n,d)} f_\beta x^\beta \in \mathcal{H}_{n,d}$, one can write

$$f = \sum_{\beta \in I(n,d)} f_\beta x^\beta = \sum_{\beta \in I(n,d)} \left(f_\beta \frac{\beta!}{d!} \right) \frac{d!}{\beta!} x^\beta.$$

We call $f_\beta \frac{\beta!}{d!}$ ($\beta \in I(n,d)$) the *Bernstein coefficients* of f, since they are the coefficients of the polynomial f when it is expressed in the Bernstein basis $\{ \frac{d!}{\beta!} x^\beta : \beta \in I(n,d)\}$ of $\mathcal{H}_{n,d}$. Using the multinomial theorem (as in the proof of Lemma 1), one can see that, for $x \in \Delta_n$, $f(x)$ is a convex combination of its Bernstein coefficients $f_\beta \frac{\beta!}{d!}$ ($\beta \in I(n,d)$). Therefore, one has

$$\min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} \leq \underline{f} \leq \overline{f} \leq \max_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!}. \tag{4}$$

We will use the following result of [11], which bounds the range of the Bernstein coefficients in terms of the range of function values.

Theorem 3 *[11, Theorem 2.2] For any polynomial $f = \sum_{\beta \in I(n,d)} f_\beta x^\beta \in \mathcal{H}_{n,d}$ and $x \in \Delta_n$, one has*

$$\overline{f} - \underline{f} \leq \max_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} - \min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} \leq \binom{2d-1}{d} d^d (\overline{f} - \underline{f}).$$

1.4 Structure of the paper

The paper is organized as follows. In Section 2, we give an elementary proof of the PTAS for quadratic polynomial optimization over the simplex, that is closely related to a proof given by Nesterov [16]. In Section 3, we refine the known PTAS result for cubic polynomial optimization over the simplex. In addition, we give an elementary proof of the PTAS result for square-free polynomial optimization over the simplex in Section 4. Moreover, in Section 5, we provide an alternative proof of the PTAS for general (fixed-degree) polynomial optimization over the simplex. We conclude with a discussion of the exact relation between our analysis and that by Nesterov [16] in Section 6. In the Appendix we provide a self-contained proof for an explicit description of the moments of the multinomial distribution in terms of the Stirling numbers of the second kind.

1.5 Notation

Throughout we use the notation $[n] = \{1, 2, \ldots, n\}$ and $\mathbb{N}^n$ is the set of all nonnegative integral vectors. For $\alpha \in \mathbb{N}^n$, we define $|\alpha| = \sum_{i=1}^n \alpha_i$ and $\alpha! = \alpha_1! \alpha_2! \cdots \alpha_n!$. For two vectors $\alpha, \beta \in \mathbb{N}^n$, the inequality $\alpha \leq \beta$ is coordinate-wise and means that $\alpha_i \leq \beta_i$ for any $i = 1, \ldots, n$. We let $e_1, \ldots, e_n$ denote the standard unit vectors in $\mathbb{R}^n$. Moreover, for $I \subseteq [n]$ we set $e_I = \sum_{i \in I} e_i$ and we let e denote the all-ones vector in $\mathbb{R}^n$. As before, $I(n,d) = \{\alpha \in \mathbb{N}^n : |\alpha| = d\}$ and $\mathcal{H}_{n,d}$ denotes the set of all multivariate homogeneous polynomials in n variables with degree d. Monomials in $\mathcal{H}_{n,d}$ are denoted as $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$ for $\alpha \in I(n,d)$, while for $I \subseteq [n]$, we use the notation $x^I = \prod_{i \in I} x_i$. Finally, for $\beta \in \mathbb{N}^n$, we also use ϕ_β to denote the monomial x^β, i.e., we set $\phi_\beta(x) = x^\beta$.

2 PTAS for quadratic polynomial optimization over the simplex

We first recall the explicit Bernstein approximation of the monomials of degree at most two, i.e., we compute $B_r(\phi_{e_i})$, $B_r(\phi_{2e_i})$ and $B_r(\phi_{e_i+e_j})$. We give a proof for clarity.

Lemma 2 *For $r \geq 1$ one has $B_r(\phi_{e_i})(x) = x_i$, $B_r(\phi_{2e_i})(x) = \frac{1}{r}x_i(1-x_i)+x_i^2$, and $B_r(\phi_{e_i+e_j})(x) = \frac{r-1}{r}x_ix_j$ for all $x \in \Delta_n$.*

Proof By the definition (1), one has:

$$B_r(\phi_{e_i})(x) = \sum_{\alpha \in I(n,r)} \frac{\alpha_i}{r}\frac{r!}{\alpha!}x^\alpha = x_i \sum_{\beta \in I(n,r-1)} \frac{(r-1)!}{\beta!}x^\beta = x_i(\sum_{i=1}^n x_i)^{r-1} = x_i,$$

$$B_r(\phi_{2e_i})(x) = \sum_{\alpha \in I(n,r)} \frac{\alpha_i^2}{r^2}\frac{r!}{\alpha!}x^\alpha = \frac{r-1}{r}x_i^2 \sum_{\beta \in I(n,r-2)} \frac{(r-2)!}{\beta!}x^\beta + \frac{1}{r}x_i \sum_{\beta \in I(n,r-1)} \frac{(r-1)!}{\beta!}x^\beta$$

$$= \frac{r-1}{r}x_i^2 + \frac{1}{r}x_i = \frac{1}{r}x_i(1-x_i) + x_i^2,$$

$$B_r(\phi_{e_i+e_j})(x) = \sum_{\alpha \in I(n,r)} \frac{\alpha_i\alpha_j}{r^2}\frac{r!}{\alpha!}x^\alpha = \frac{r-1}{r}x_ix_j \sum_{\beta \in I(n,r-2)} \frac{(r-2)!}{\beta!}x^\beta = \frac{r-1}{r}x_ix_j,$$

where we have used at several places the multinomial theorem (and the fact that an empty summation is equal to 0). □

Consider now a quadratic polynomial $f = x^T Q x \in \mathcal{H}_{n,2}$. By Lemma 2, its Bernstein approximation on the simplex is given by

$$B_r(f)(x) = \frac{1}{r}\sum_{i=1}^n Q_{ii}x_i + (1 - \frac{1}{r})f(x) \quad \forall x \in \Delta_n. \tag{5}$$

Theorem 4 *For any polynomial $f = x^T Q x \in \mathcal{H}_{n,2}$ and $r \geq 1$, one has*

$$\max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\} \leq \frac{Q_{\max} - \underline{f}}{r} \leq \frac{\overline{f} - \underline{f}}{r}.$$

setting $Q_{\max} = \max_{i \in [n]} Q_{ii}$.

Proof Using (5), one obtains that

$$rB_r(f)(x) = \sum_{i=1}^n Q_{ii}x_i + (r-1)f(x)$$

$$\leq \max_{x \in \Delta_n}\sum_{i=1}^n Q_{ii}x_i + rf(x) - \min_{x \in \Delta_n} f(x)$$

$$= \max_i Q_{ii} - \underline{f} + rf(x)$$

$$\leq \overline{f} - \underline{f} + rf(x),$$

where in the last inequality we have used the fact that $\max_i Q_{ii} \leq \overline{f}$, since $Q_{ii} = f(e_i) \leq \overline{f}$ for $i \in [n]$. This gives the two right-most inequalities in the theorem. □

Combining Theorem 4 with Lemma 1, we obtain the following corollary, which gives the PTAS result by Bomze and de Klerk [3, Theorem 3.2].

Corollary 1 *For any polynomial $f = x^T Q x \in \mathcal{H}_{n,2}$ and $r \geq 1$, one has*

$$f_{\Delta(n,r)} - \underline{f} \leq \frac{Q_{\max} - \underline{f}}{r} \leq \frac{\overline{f} - \underline{f}}{r}.$$

We note that the proof given here is completely elementary and much simpler than the original one in [3]. Our proof is, however, closely related to another proof by Nesterov [16], we will give the precise relation in Section 6.

Example 2 Consider the quadratic polynomial $f = \sum_{i=1}^{n} x_i^2 \in \mathcal{H}_{n,2}$. As f is convex, it is easy to check that $\underline{f} = \frac{1}{n}$ (attained at $x = \frac{1}{n}e$) and $\overline{f} = 1$ (attained at any standard unit vector).

For the computation of $f_{\Delta(n,r)}$, it is convenient to write r as $r = kn + s$, where $k \geq 0$ and $0 \leq s < n$. Then we have

$$f_{\Delta(n,r)} = \frac{1}{n} + \frac{1}{r^2}\frac{s(n-s)}{n},$$

which is attained at any point $x \in \Delta(n,r)$ having $n-s$ coordinates equal to $\frac{k}{r}$ and s coordinates equal to $\frac{k+1}{r}$. To see this, pick a minimizer $x \in \Delta(n,r)$. First we claim that $x_i - x_j \leq \frac{1}{r}$ for any $i \neq j \in [n]$. Indeed, assume (say) that $x_2 - x_1 > \frac{1}{r}$. Then define the new point $x' \in \Delta(n,r)$ by $x_1' = x_1 + \frac{1}{r}$, $x_2' = x_2 - \frac{1}{r}$ and $x_i' = x_i$ for all $i \neq 1,2$ and observe that $f(x') < f(x)$, which contradicts the optimality of x. Therefore, the coordinates of x can take at most two possible values $\frac{h}{r}, \frac{h+1}{r}$ for some $0 \leq h \leq r-1$ and it is easy to see these two values belong to $\{\frac{k}{r}, \frac{k+1}{r}\}$. Hence we obtain that

$$f_{\Delta(n,r)} - \underline{f} = \frac{1}{r^2}\frac{s(n-s)}{n} \quad \text{and} \quad \frac{f_{\Delta(n,r)} - \underline{f}}{\overline{f} - \underline{f}} = \frac{1}{r^2}\frac{s(n-s)}{n-1}.$$

We observe that this latter ratio might be in the order $\frac{1}{r}$, thus matching the upper bound in Corollary 1 in terms of the dependence of the error bound on r. For instance, for $r = \frac{3n}{2}$ (i.e., $k = 1$, $s = \frac{n}{2}$), we have that

$$f_{\Delta(n,r)} - \underline{f} = \frac{1}{6r - 9}(\overline{f} - \underline{f}). \tag{6}$$

Moreover, we have $B_r(f)(x) = \frac{1}{r} + (1 - \frac{1}{r})f(x)$, so that

$$\min_{x \in \Delta_n} B_r(f)(x) - \underline{f} = \max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\} = \frac{1}{r}(\overline{f} - \underline{f}).$$

Hence, equality holds throughout in the inequalities of Theorem 4, which shows that the upper bound is tight on this example.

By Example 2, there does not exist any $\epsilon > 0$ such that, for any quadratic form f,

$$f_{\Delta(n,r)} - \underline{f} \leq \frac{1}{r^{1+\epsilon}}(\overline{f} - \underline{f}) \quad \forall r \geq 1,$$

and thus the error bound in Corollary 1 is tight in terms of its dependence on r. On the other hand, one may easily show that, for the polynomial f in Example 2,

$$\rho_r(f) := \frac{f_{\Delta(n,r)} - \underline{f}}{\overline{f} - \underline{f}} \leq \frac{n}{4r^2} = O(1/r^2).$$

Thus, $\limsup_{r\to\infty}(r^2\rho_r(f)) < \infty$, i.e. the asymptotic convergence rate of the sequence $\{\rho_r(f)\}$ for the example is $O(1/r^2)$. It turns out that this will be the case also for the other polynomials considered in Examples 3, 4 and 5.

3 PTAS for cubic polynomial optimization over the simplex

Using similar arguments as for Lemma 2, one can compute the Bernstein approximations of the monomials of degree three. Namely, for distinct $i, j, k \in [n]$ and $x \in \Delta_n$,

$$B_r(\phi_{3e_i})(x) = \frac{1}{r^2}x_i + \frac{3(r-1)}{r^2}x_i^2 + \frac{(r-1)(r-2)}{r^2}x_i^3,$$

$$B_r(\phi_{2e_i+e_j})(x) = \frac{(r-1)}{r^2}x_ix_j + \frac{(r-1)(r-2)}{r^2}x_i^2x_j,$$

$$B_r(\phi_{e_i+e_j+e_k})(x) = \frac{(r-1)(r-2)}{r^2}x_ix_jx_k.$$

We show the following result.

Theorem 5 *For any polynomial $f \in \mathcal{H}_{n,3}$ and $r \geq 2$, one has*

$$\max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\} \leq \left(\frac{4}{r} - \frac{4}{r^2}\right)(\overline{f} - \underline{f}).$$

Proof Consider a cubic polynomial $f \in \mathcal{H}_{n,3}$ of the form

$$f = \sum_{i=1}^{n}f_ix_i^3 + \sum_{1\leq i<j\leq n}(f_{ij}x_ix_j^2 + g_{ij}x_i^2x_j) + \sum_{1\leq i<j<k\leq n}f_{ijk}x_ix_jx_k.$$

Applying the above description for the Bernstein approximation of degree 3 monomials, the Bernstein approximation of f at any $x \in \Delta_n$ reads

$$B_r(f)(x) = \frac{(r-1)(r-2)}{r^2}f(x) + \frac{1}{r^2}\left[\sum_{i=1}^{n}f_ix_i + (r-1)\left(\sum_{i=1}^{n}3f_ix_i^2 + \sum_{i<j}(f_{ij} + g_{ij})x_ix_j\right)\right]. \quad (7)$$

Evaluating f at e_i and at $(e_i + e_j)/2$ yields, respectively, the relations:

$$\underline{f} \leq f_i \leq \overline{f}, \quad (8)$$

$$f_i + f_j + f_{ij} + g_{ij} \leq 8\overline{f}. \quad (9)$$

Using (9) and the fact that $\sum_{i=1}^{n}x_i = 1$, one can obtain

$$\sum_{i<j}(f_{ij} + g_{ij})x_ix_j \leq \sum_{i<j}(8\overline{f} - f_i - f_j)x_ix_j = 8\overline{f}\sum_{i<j}x_ix_j - \sum_{i=1}^{n}f_ix_i(1 - x_i). \quad (10)$$

Combining (7) and (10), one obtains that, for any $x \in \Delta_n$,

$$r^2 B_r(f)(x) = (r-1)(r-2)f(x) + \sum_{i=1}^{n} f_i x_i + (r-1) \left(\sum_{i=1}^{n} 3 f_i x_i^2 + \sum_{i<j} (f_{ij} + g_{ij}) x_i x_j \right)$$

$$\leq (r-1)(r-2)f(x) - (r-2) \sum_{i=1}^{n} f_i x_i + (r-1) \left(\sum_{i=1}^{n} 4 f_i x_i^2 + 8\overline{f} \sum_{i<j} x_i x_j \right).$$

We now use (8) to bound the two inner summations as follows:

$$- \sum_i f_i x_i \leq -\underline{f} \sum_i x_i = -\underline{f} \quad \text{and} \quad \sum_{i=1}^{n} 4 f_i x_i^2 + 8\overline{f} \sum_{i<j} x_i x_j \leq 4\overline{f} (\sum_i x_i)^2 = 4\overline{f}.$$

This implies:

$$r^2 (B_r(f)(x) - f(x)) \leq -(3r-2)\underline{f} - (r-2)\underline{f} + 4(r-1)\overline{f} = 4(r-1)(\overline{f} - \underline{f}),$$

which concludes the proof. □

Combining Theorem 5 with Lemma 1, we obtain the following error bound.

Corollary 2 *For any polynomial $f \in \mathcal{H}_{n,3}$ and $r \geq 2$, one has*

$$f_{\Delta(n,r)} - \underline{f} \leq \left(\frac{4}{r} - \frac{4}{r^2} \right) (\overline{f} - \underline{f}).$$

This result is a bit stronger than the result by de Klerk et al. [11, Theorem 3.3], which states that $f_{\Delta(n,r)} - \underline{f} \leq \frac{4}{r}(\overline{f} - \underline{f})$.

Example 3 Consider the cubic polynomial $f = x_1^3 + x_2^3 \in \mathcal{H}_{2,3}$. One can check that $\overline{f} = 1, \underline{f} = \frac{1}{4}$,

$$f_{\Delta(2,r)} = \begin{cases} 1/4 & \text{if } r \text{ is even,} \\ \frac{1}{4} + \frac{3}{4r^2} & \text{if } r \text{ is odd.} \end{cases}$$

Moreover, one can check that $B_r(f)(x) = 1 + \left(\frac{3}{r} - 3 \right) x_1 x_2$ and $\min_{x \in \Delta_2} B_r(f)(x) = \frac{1}{4} + \frac{3}{4r}$. Hence, for any integer $r \geq 2$, one has strict inequality $\min_{x \in \Delta_2} B_r(f)(x) > f_{\Delta(2,r)}$. Moreover, for $r \geq 2$,

$$\min_{x \in \Delta_2} B_r(f)(x) - \underline{f} = \max_{x \in \Delta_2} \{ B_r(f)(x) - f(x) \} = \frac{3}{4r} = \frac{1}{r}(\overline{f} - \underline{f}) < (\frac{4}{r} - \frac{4}{r^2})(\overline{f} - \underline{f}).$$

On the other hand, for odd r, the range $f_{\Delta(2,r)} - \underline{f}$ is equal to $\frac{3}{4r^2} = \frac{1}{r^2}(\overline{f} - \underline{f})$ and thus grows proportionally to $\frac{1}{r^2}$.

4 PTAS for square-free polynomial optimization over the simplex

Here we consider square-free (aka multilinear) polynomials, involving only monomials x^I for $I \subseteq [n]$. The Bernstein approximation of the square-free monomial $\phi_{e_I}(x) := x^I$, with $d = |I|$, is given by

$$B_r(\phi_{e_I})(x) = \sum_{\alpha \in I(n,r)} \frac{\alpha^I}{r^{\underline{d}}} \frac{r!}{\alpha!} x^\alpha = \frac{r^{\underline{d}}}{r^d} x^I \sum_{\alpha \in I(n,r-d)} \frac{(r-d)!}{\alpha!} x^\alpha = \frac{r^{\underline{d}}}{r^d} x^I (\sum_i x_i)^{r-d} = \frac{r^{\underline{d}}}{r^d} x^I$$

for $x \in \Delta_n$. Recall that, for an integer $r \geq 1$, $r^{\underline{d}} = r(r-1) \cdots (r-d+1)$ and observe that $r^{\underline{d}} = 0$ if $r < d$. Hence the Bernstein approximation of the square-free polynomial $f = \sum_{I \subseteq [n], |I|=d} f_I x^I$ satisfies

$$B_r(f)(x) = \frac{r^{\underline{d}}}{r^d} f(x) \quad \forall x \in \Delta_n,$$

which implies the following identities:

$$\min_{x \in \Delta_n} B_r(f)(x) - \underline{f} = \max_{x \in \Delta_n} \{B_r(f)(x) - f(x)\} = -\left(1 - \frac{r^{\underline{d}}}{r^d}\right) \underline{f}. \tag{11}$$

Theorem 6 *For any square-free polynomial $f \in \mathcal{H}_{n,d}$ and $r \geq 1$, one has*

$$\max_{x \in \Delta_n} \{B_r(f)(x) - f(x)\} \leq \left(1 - \frac{r^{\underline{d}}}{r^d}\right) (\overline{f} - \underline{f}) \quad \forall x \in \Delta_n.$$

Proof We use (11). For degree $d = 1$ the result is clear and, for $d \geq 2$, we use the fact that $\overline{f} \geq 0$ since $f(e_i) = 0$ for any $i \in [n]$. $\qquad \square$

Combining with Lemma 1 we obtain the following error bound.

Corollary 3 *For any square-free polynomial $f \in \mathcal{H}_{n,d}$ and $r \geq 1$, one has*

$$f_{\Delta(n,r)} - \underline{f} \leq \left(1 - \frac{r^{\underline{d}}}{r^d}\right) (\overline{f} - \underline{f}).$$

This result was first shown by Nesterov [16, Theorem 2] (see also De Klerk, Laurent, and Parrilo [11, Remark 3.4]). In fact, our proof is again closely related to the one by Nesterov; see Section 6 for the details.

Example 4 Consider the square-free polynomial $f = -x_1 x_2$. Then, $B_r(f)(x) = -\frac{r-1}{r} x_1 x_2$ and one can check that $\overline{f} = 0$, $\underline{f} = -\frac{1}{4}$, and $\min_{x \in \Delta_2} B_r(f)(x) = -\frac{1}{4}\frac{r-1}{r}$. Moreover,

$$f_{\Delta(2,r)} = \begin{cases} -\frac{1}{4} & \text{if } r \text{ is even,} \\ -\frac{1}{4} + \frac{1}{4r^2} & \text{if } r \text{ is odd.} \end{cases}$$

Hence, for any integer $r \geq 2$, one has strict inequality: $\min_{x \in \Delta_2} B_r(f)(x) > f_{\Delta(n,r)}$. Moreover, as $\min_{x \in \Delta_2} B_r(f)(x) - \underline{f} = \max_{x \in \Delta_2} \{B_r(f)(x) - f(x)\} = \frac{1}{4r} = \frac{1}{r}(\overline{f} - \underline{f})$, the upper bound from Theorem 6 is tight on this example. On the other hand, $f_{\Delta(2,r)} - \underline{f} = \frac{1}{4r^2} = \frac{1}{r^2}(\overline{f} - \underline{f})$ for odd r, and thus the range $f_{\Delta(2,r)} - \underline{f}$ grows proportionally to $\frac{1}{r^2}$.

5 PTAS for general polynomial optimization over the simplex

In this section we deal with the minimization of an arbitrary polynomial $f \in \mathcal{H}_{n,d}$. In order to be able to bound the minimum of $B_r(f)$ over Δ_n we need an explicit description of the Bernstein approximation of f.

5.1 Bernstein approximation over the simplex of an arbitrary monomial

Here we work out an explicit description of the Bernstein approximation of arbitrary monomials $\phi_\beta(x) = x^\beta$ ($\beta \in I(n,d)$). The key ingredient is to express it in terms of the moments of the multinomial distribution.

Fix $x = (x_1, \ldots, x_n) \in \Delta_n$ and consider the multinomial distribution with n categories and r independent trials, where the probability for the i-th category is given by x_i. Then, given $\alpha \in I(n,r)$, the probability of drawing α_i times the i-th category for each $i \in [n]$ is equal to $\frac{r!}{\alpha!} x^\alpha$. Therefore, for $\beta \in \mathbb{N}^n$, the β-th moment of this multinomial distribution is given by

$$m^\beta_{(n,r)} := \sum_{\alpha \in I(n,r)} \alpha^\beta \frac{r!}{\alpha!} x^\alpha.$$

Comparing with the definition of the Bernstein approximation of $\phi_\beta(x) = x^\beta$ we find the identity

$$B_r(\phi_\beta)(x) = \sum_{\alpha \in I(n,r)} (\frac{\alpha}{r})^\beta \frac{r!}{\alpha!} x^\beta = \frac{1}{r^{|\beta|}} m^\beta_{(n,r)}.$$

Combining [9, relation (34.18)] and [9, relation (35.5)], we can obtain an explicit formula for the moments $m^\beta_{(n,r)}$ of the multinomial distribution in terms of the Stirling numbers of the second kind. Recall that, for integers $n, k \in \mathbb{N}$, the *Stirling number of the second kind* $S(n,k)$ counts the number of ways of partitioning a set of n objects into k nonempty subsets. Thus $S(n,k) = 0$ if $k > n$, $S(n,0) = 0$ if $n \geq 1$, and $S(0,0) = 1$ by convention.

Theorem 7 *For $\beta \in \mathbb{N}^n$, one has*

$$m^\beta_{(n,r)} = \sum_{\alpha \in \mathbb{N}^n : \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^n S(\beta_i, \alpha_i),$$

where $S(\beta_i, \alpha_i)$ are Stirling numbers of the second kind.

Therefore, we can deduce the explicit formula of the Bernstein approximation for any monomial.

Corollary 4 *For any monomial $\phi_\beta(x) = x^\beta$, one has*

$$B_r(\phi_\beta)(x) = \frac{1}{r^{|\beta|}} \sum_{\alpha \in \mathbb{N}^n : \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^n S(\beta_i, \alpha_i) \quad \forall x \in \Delta_n.$$

For completeness, we will give a self-contained proof for Corollary 4 in the Appendix.

5.2 Error bound analysis

We show the following error bound for the Bernstein approximation of order r of an arbitrary polynomial on the simplex.

Theorem 8 *For any polynomial $f \in \mathcal{H}_{n,d}$ and $r \geq 1$, one has*

$$
\begin{aligned}
\max_{x \in \Delta_n}\{B_r(f)(x) - f(x)\} &\leq \left(1 - \tfrac{r^{\underline{d}}}{r^d}\right)\left(\max_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} - \min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!}\right) \\
&\leq \left(1 - \tfrac{r^{\underline{d}}}{r^d}\right)\binom{2d-1}{d}d^d(\overline{f} - \underline{f}).
\end{aligned}
$$

For the proof we will need two auxiliary results about the Stirling numbers of the second kind. The first result is implied by [13, relation (3.2)], and we therefore only sketch the proof.

Lemma 3 *For positive integers d and $r \geq 1$, one has*

$$
\sum_{k=1}^{d-1} r^{\underline{k}} S(d,k) = r^d - r^{\underline{d}}.
$$

Proof The proof is by induction on d, and using relation (20) (in the appendix to this paper) for the induction step. □

The second result gives an alternative expression for Stirling numbers of the second kind. We provide a full proof, since we could not find this result in the literature.

Lemma 4 *Given $\alpha \in I(n,k)$ and $d > k$, one has*

$$
S(d,k) = \frac{\alpha!}{k!}\sum_{\beta \in I(n,d)} \frac{d!}{\beta!}\prod_{i=1}^{n} S(\beta_i, \alpha_i). \tag{12}
$$

Proof For integers $d, k \geq 0$, let $S_{d,k}$ denote the number of surjective maps from a d-elements set to a k-elements set. It is not difficult to see the following relation between $S_{d,k}$ and $S(d,k)$:

$$
S_{d,k} = k! S(d,k).
$$

Indeed, let $B = [d]$ and $A = [k]$. In order to choose a surjective map f from B to A one needs to select the pre-image $B_i = f^{-1}(i) \subseteq B$ for each element $i \in [k]$. So to define a surjective map f, one first selects a partition of B into k non-empty subsets $B_1, \ldots, B_k$, which can be done in $S(d,k)$ ways. As any permutation of the $B_i's$ gives rise to a distinct surjective map, there are $k!S(d,k)$ surjective maps from $[d]$ to $[k]$.

Now, the identity (12) about the Stirling numbers $S(d,k)$ can be equivalently reformulated as the following identity about the numbers $S_{d,k}$: For any $\alpha \in I(n,k)$,

$$
S_{d,k} = \sum_{\beta \in I(n,d)} \frac{d!}{\beta!}\prod_{i=1}^{n} S_{\beta_i, \alpha_i}.
$$

Again set $B = [d]$ and $A = [k]$. Say, α has p non-zero coordinates, i.e., $\alpha_1, \ldots, \alpha_p \geq 1$ and $\alpha_1 + \ldots + \alpha_p = k$. Fix a partition of $A = [k]$ into p subsets $A_1, \ldots, A_p$ where $|A_i| = \alpha_i$ for $i \in [p]$.

Then, a surjection f from B to A defines a surjection from $B_i = f^{-1}(A_i)$ to A_i for each $i \in [p]$. Setting $\beta_i = |B_i|$, we have $\beta_1 + \ldots + \beta_p = d$ since the $B_i's$ partition B. Hence one can count the number of surjections from B to A as follows.

First, select $\beta_1, \ldots, \beta_p \geq 1$ such that $\beta_1 + \ldots + \beta_p = d$. Then split the d elements of B into an ordered sequence of p disjoint subsets $B_1, \ldots, B_p$ where $|B_i| = \beta_i$ for $i \in [p]$; there are $\frac{d!}{\beta!}$ ways of doing so. Once $B_1, \ldots, B_p$ are selected, there are S_{β_i, α_i} possible surjections from B_i to A_i for each $i \in [p]$ and thus a total of $\prod_{i=1}^{p} S_{\beta_i, \alpha_i}$ possibilities. Therefore, we get that the total number of surjections from B to A is equal to $\sum_{\beta \in I(p,d)} \frac{d!}{\beta!} \prod_{i=1}^{p} S_{\beta_i, \alpha_i}$, which shows the result. □

We are now ready to prove Theorem 8.

Proof (of Theorem 8) Consider a polynomial $f = \sum_{\beta \in I(n,d)} f_\beta x^\beta \in \mathcal{H}_{n,d}$ and $x \in \Delta_n$. Applying Corollary 4, we can write the Bernstein approximation of f at $x \in \Delta_n$ as follows:

$$B_r(f)(x) = \frac{1}{r^d} \sum_{\beta \in I(n,d)} f_\beta \sum_{\alpha : 0 \leq \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^{n} S(\beta_i, \alpha_i).$$

Therefore,

$$r^d B_r(f)(x) = r^{\underline{d}} f(x) + \sum_{\beta \in I(n,d)} f_\beta \sum_{\alpha : 0 \leq \alpha \leq \beta, \alpha \neq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^{n} S(\beta_i, \alpha_i),$$

and thus

$$r^d (B_r(f)(x) - f(x)) = -(r^d - r^{\underline{d}}) f(x) + \sum_{\beta \in I(n,d)} f_\beta \sum_{\alpha : 0 \leq \alpha \leq \beta, \alpha \neq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^{n} S(\beta_i, \alpha_i).$$

Using (4), we have $f(x) \geq \min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!}$ and $f_\beta \frac{\beta!}{d!} \leq \max_{\beta' \in I(n,d)} f_{\beta'} \frac{\beta'!}{d!}$, which permits to derive the following inequality:

$$r^d (B_r(f)(x) - f(x)) \leq$$

$$-(r^d - r^{\underline{d}}) \min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} + \max_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} \left(\underbrace{\sum_{\beta \in I(n,d)} \frac{d!}{\beta!} \sum_{\alpha : 0 \leq \alpha \leq \beta, \alpha \neq \beta} x^\alpha r^{\underline{|\alpha|}} \prod_{i=1}^{n} S(\beta_i, \alpha_i)}_{\sigma} \right). \tag{13}$$

It now suffices to upper bound the right handside of the inequality (13) and to show that the inner summation σ is equal to $r^d - r^{\underline{d}}$. Indeed,

$$
\begin{aligned}
\sigma &= \sum_{\beta \in I(n,d)} \frac{d!}{\beta!} \sum_{\alpha:0 \leq \alpha \leq \beta, \alpha \neq \beta} x^\alpha r^{\underline{|\alpha|}} \prod_{i=1}^n S(\beta_i, \alpha_i) \\
&= \sum_{\alpha \in \mathbb{N}^n} x^\alpha r^{\underline{|\alpha|}} \sum_{\beta \in I(n,d):\beta \geq \alpha, \beta \neq \alpha} \frac{d!}{\beta!} \prod_{i=1}^n S(\beta_i, \alpha_i) \\
&= \sum_{k=1}^{d-1} \sum_{\alpha \in I(n,k)} x^\alpha r^{\underline{|\alpha|}} \left(\sum_{\beta \in I(n,d):\beta \geq \alpha} \frac{d!}{\beta!} \prod_{i=1}^n S(\beta_i, \alpha_i) \right) \\
&= \sum_{k=1}^{d-1} r^{\underline{k}} \sum_{\alpha \in I(n,k)} x^\alpha \left(\frac{k!}{\alpha!} S(d,k) \right) \qquad \text{[using Lemma 4]} \\
&= \sum_{k=1}^{d-1} r^{\underline{k}} S(d,k) (\sum_i x_i)^k = \sum_{k=1}^{d-1} r^{\underline{k}} S(d,k) = r^d - r^{\underline{d}}. \qquad \text{[using Lemma 3]}
\end{aligned}
$$

Using this identity for the summation σ in the inequality (13) we obtain

$$
r^d \left(\min_{x \in \Delta_n} B_r(f)(x) - \underline{f} \right) \leq r^d \max_{x \in \Delta_n} \{ B_r(f)(x) - f(x) \} \leq (r^d - r^{\underline{d}}) \left(\max_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} - \min_{\beta \in I(n,d)} f_\beta \frac{\beta!}{d!} \right).
$$

By combining with Theorem 3 we obtain the claimed inequalities of Theorem 8 and this concludes the proof. $\square$

Combining Theorem 8 with Lemma 1, we obtain the following error bound, which was first shown in [11, Theorem 1.3].

Corollary 5 *For any polynomial* $f \in \mathcal{H}_{n,d}$ *and* $r \geq 1$, *one has*

$$
f_{\Delta(n,r)} - \underline{f} \leq \left(1 - \frac{r^{\underline{d}}}{r^d} \right) \binom{2d-1}{d} d^d (\overline{f} - \underline{f}).
$$

Example 5 We consider here the problem of minimizing the polynomial $f = \sum_{i=1}^n x_i^d$ $(n \geq 2)$ over the simplex for any degree $d \geq 2$, thus extending the case $d = 2$ considered in Example 2 and the case $d = 3, n = 2$ considered in Example 3. As f is convex on $\mathbb{R}_+^n$ it follows that $\overline{f} = 1$ and $\underline{f} = \frac{1}{n^{d-1}}$.

We now compute the minimum over the regular grid $\Delta(n,r)$. As in Example 2 set $r = kn + s$ where $k, s \in \mathbb{N}$ with $s \leq n-1$. We show that $f_{\Delta(n,r)}$ is attained at any point x having s components equal to $\frac{k+1}{r}$ and $n - s$ components equal to $\frac{k}{r}$, so that

$$
f_{\Delta(n,r)} = s \left(\frac{k+1}{r} \right)^d + (n-s) \left(\frac{k}{r} \right)^d.
$$

For this pick a minimizer x of f over $\Delta(n,r)$ and it suffices to show that $x_i - x_j \leq \frac{1}{r}$ for all $i, j \in [n]$. If (say) $x_2 - x_1 > \frac{1}{r}$ then we claim that $f(x_1 + \frac{1}{r}, x_2 - \frac{1}{r}, x_3, \ldots, x_n) < f(x_1, x_2, x_3, \ldots, x_n)$,

which contradicts the minimality assumption on x. One can see the above claim as follows: set $\sigma = 1 - \sum_{i=3}^{n} x_i$, consider the function $\phi(t) = t^d + (\sigma - t)^d$ for t satisfying $0 \leq t < \frac{1}{2}(\sigma - \frac{1}{r})$, and verify (using elementary calculus) that $\phi(t + \frac{1}{r}) < \phi(t)$ for any such t. Therefore, we have

$$
\begin{aligned}
f_{\Delta(n,r)} - \underline{f} &= (n - s)\left(\frac{k}{r}\right)^d + s\left(\frac{k+1}{r}\right)^d - \frac{n}{n^d} \\
&= (n - s)\left(\left(\frac{k}{r}\right)^d - \frac{1}{n^d}\right) + s\left(\left(\frac{k+1}{r}\right)^d - \frac{1}{n^d}\right) \\
&= \frac{n - s}{n^d}\left(\left(1 - \frac{s}{r}\right)^d - 1\right) + \frac{s}{n^d}\left(\left(1 - \frac{s-n}{r}\right)^d - 1\right) \\
&= \frac{s(n-s)}{n^d}\sum_{i=2}^{d}\binom{d}{i}\frac{(n-s)^{i-1} + (-1)^i s^{i-1}}{r^i}.
\end{aligned}
$$

Using the fact that $s, n - s \leq n$, for any $r \geq n$ we can bound the above summation as follows:

$$
\sum_{i=2}^{d}\binom{d}{i}\frac{(n-s)^{i-1} + (-1)^i s^{i-1}}{r^i} \leq \frac{2}{n}\sum_{i=2}^{d}\binom{d}{i}\left(\frac{n}{r}\right)^i \leq \frac{2}{n}\left(\frac{n}{r}\right)^2\sum_{i=2}^{d}\binom{d}{i} \leq \frac{2}{n}\left(\frac{n}{r}\right)^2 2^d = \frac{2^{d+1}n}{r^2}.
$$

Combining with the bound $s(n-s) \leq \frac{n^2}{4}$, we deduce that

$$
f_{\Delta(n,r)} - \underline{f} \leq \frac{n^2}{4}\frac{1}{n^d}\frac{2^{d+1}n}{r^2} = \frac{2^{d-1}}{n^{d-3}r^2}.
$$

Therefore,

$$
\frac{f_{\Delta(n,r)} - \underline{f}}{\overline{f} - \underline{f}} \leq \frac{2^{d-1}}{n^{d-3}r^2}\frac{n^{d-1}}{n^{d-1} - 1} = \frac{2^{d-1}}{r^2}\frac{n^2}{n^{d-1} - 1} \leq \frac{2^d}{r^2} \quad \text{for any } r \geq n \geq 2 \text{ and } d \geq 3. \tag{14}
$$

Hence we see that for any degree $d \geq 3$ the ratio is in the order $\frac{1}{r^2}$. Recall that for degree $d = 2$ it was observed in Example 2 that it can be in the order $\frac{1}{r}$ for certain values of r (e.g., for $r = \frac{3n}{2}$).

6 Concluding remarks

Nesterov [16] proposed an alternative probabilistic argument for estimating the quality of the bounds $f_{\Delta(n,r)}$. He considered a random walk on the simplex Δ_n, which generates a sequence of random points $x^{(r)} \in \Delta(n, r)$ $(r \geq 1)$. Thus the expected value $E(f(x^{(r)}))$ of the evaluation of a polynomial $f \in \mathcal{H}_{n,d}$ at $x^{(r)}$ satisfies:

$$
f_{\Delta(n,r)} \leq E(f(x^{(r)})).
$$

Nesterov's approach goes as follows. Let $x \in \Delta_n$ and let ζ be a discrete random variable taking values in $\{1, \ldots, n\}$ distributed according to the multinomial distribution with n categories and where the probability of the i-th category is given by x_i. That is,

$$
\text{Prob}(\zeta = i) = x_i \ (i = 1, \ldots, n). \tag{15}
$$

Consider the random process:

$$y^{(0)} = 0 \in \mathbb{R}^n, \ y^{(r)} = y^{(r-1)} + e_{\zeta_r} \ (r \geq 1)$$

where ζ_r are independent random variables distributed according to (15). In other words, $y^{(r)}$ equals $y^{(r-1)} + e_i$ with probability x_i. Finally, define

$$x^{(r)} = \frac{1}{r} y^{(r)} \in \Delta(n, r) \quad (r \geq 1).$$

For a given $\alpha \in I(n, r)$, the probability of the event $y^{(r)} = \alpha$ is given by

$$\mathrm{Prob}(y^{(r)} = \alpha) = \frac{r!}{\alpha!} x^\alpha,$$

by the properties of the multinomial distribution. Thus one also has $\mathrm{Prob}(x^{(r)} = \alpha/r) = \frac{r!}{\alpha!} x^\alpha$, and it immediately follows that

$$E(f(x^{(r)})) = \sum_{\alpha \in I(n,r)} \mathrm{Prob}(x^{(r)} = \alpha/r) f(\alpha/r) = \sum_{\alpha \in I(n,r)} \frac{r!}{\alpha!} x^\alpha f\left(\frac{\alpha}{r}\right) = B_r(f)(x).$$

In this sense, the approach of our paper using Bernstein approximation is equivalent to the analysis of Nesterov [16] (although the equivalence is not obvious a priori). On the other hand, in [16] the link with Bernstein approximation is not made, and the author calculated the values $E(f(x^{(r)}))$ from first principles for polynomials up to degree four and for square-free polynomials. Based on this Nesterov [16] gave the error bounds from Theorems 4 and 6 for the quadratic and square-free cases. However he did not consider the general case. Thus the analysis in this paper completes the analysis in [16] by placing it in the well-studied framework of Bernstein approximation and clarifying the link to the multinomial distribution.

We conclude with a general comment regarding a further interpretation of the upper bound $B_r(f)(x)$ (where $x \in \Delta_n$) for the minimum $f_{\Delta(n,r)}$ over the regular grid, within the general framework introduced by Lasserre [14] based on reformulating polynomial optimization problems as optimization problems over measures. A basic, fundamental idea of Lasserre [14] to compute the minimum of a polynomial f over a compact set $K \subseteq \mathbb{R}^n$ is to reformulate the problem as a minimization problem over the set $\mathcal{M}(K)$ of (Borel) probability measures on the set K. (We assume K compact for simplicity but Lasserre's idea works for K closed). Namely,

$$\min_{x \in K} f(x) = \min_{\mu \in \mathcal{M}(K)} E_\mu(f),$$

setting $E_\mu(f) = \int_K f(x)\mu(dx)$. The above identity is simple. As $f(x) \geq \min_{x \in K} f(x)$ for all $x \in K$, one can integrate both sides with respect to any measure $\mu \in \mathcal{M}(K)$, which gives the inequality $\min_{x \in K} f(x) \leq \min_{\mu \in \mathcal{M}(K)} E_\mu(f)$. For the converse inequality, let μ be the Dirac measure at a global minimizer x of f over K, so that $E_\mu(f) = \min_{x \in K} f(x)$.

Applying this idea to polynomial minimization over the regular grid $\Delta(n, r)$, one has

$$f_{\Delta(n,r)} = \min_{\mu \in \mathcal{M}(\Delta(n,r))} E_\mu(f).$$

Thus in order to upper bound $f_{\Delta(n,r)}$ it suffices to choose a suitable probability measure on the regular grid $\Delta(n, r)$ and, according to our discussion above, this is precisely what the bound $B_r(f)(x)$ boils down to.

Indeed, by considering the multinomial distribution with n categories and with probability x_i for the i-th category, after r independent trials we get a probability distribution over $I(n, r)$ where $\alpha \in I(n, r)$ is picked with probability $\frac{r!}{\alpha!} x^\alpha$. This in turn gives a probability distribution μ_r on $\Delta(n, r) = \frac{1}{r} I(n, r)$ where $\frac{\alpha}{r}$ is picked with the same probablity $\frac{r!}{\alpha!} x^\alpha$. Now, as was shown above, $E_{\mu_r}(f) = B_r(f)(x)$ is thus an upper bound on $f_{\Delta(n,r)}$.

A final comment concerns the asymptotic convergence rate of the sequence

$$\rho_r(f) = \frac{f_{\Delta(n,r)} - \underline{f}}{\overline{f} - \underline{f}} \quad r = 1, 2, \dots$$

for a given polynomial $f \in \mathcal{H}_{n,d}$. In all the examples presented in this paper, one has $\rho_r(f) = O(1/r^2)$; see Examples 2, 3, 4 and 5. It remains an open problem to determine the asymptotic convergence rate in general.

References

1. Altomare, F., and Campiti, M. *Korovkin-type approximation theory and its applications.* De Guyter studies in mathematics 17, Walter de Guyter publishers, Berlin, (1994).
2. Bellare, M., Rogaway, P.: The complexity of approximating a nonlinear program. *Math. Program.* 69, 429-441 (1995)
3. Bomze, I.M., Klerk, E. de.: Solving standard quadratic optimization problems via semidefinite and copositive programming. *J Global. Optim.* 24(2), 163-185 (2002)
4. Bomze, I.M., Gollowitzer, S., Yildirim, E.A.: Rounding on the standard simplex: Regular grids for global optimization. Preprint, submitted (2013)
5. Bos, L.P. Bounding the Lebesque function for Lagrange interpolation in a simplex. *Journal of Approximation Theory*, 38, 43–59 (1983)
6. Ditzian, Z. Inverse theorems for multidimensional Bernstein operators, *Pacific Journal of Mathematics* 121 (2), 293–319 (1986)
7. Ditzian, Z. Best polynomial approximation and Bernstein polynomial approximation on a simplex, *Indagationes Mathematicae* 92, 243–256 (1989)
8. Ditzian, Z., Zhou, X. Optimal approximation class for multivariate Bernstein operators, *Pacific Journal of Mathematics* 158, 93–120 (1993)
9. Johnson, N.L., Kotz, S., Balakrishnan, N.: *Discrete Multivariate Distributions.* Wiley, (1997)
10. Klerk, E. de, Hertog, D. den, Elabwabi, G.: On the complexity of optimization over the standard simplex. *Eur J. Oper. Res.* 191(3), 773-785 (2008)
11. Klerk, E. de, Laurent, M., Parrilo, P.: A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theor. Comput. Sci.* 361(2-3), 210-225 (2006)
12. Klerk, E. de, Laurent, M.: Error bounds for some semidefinite programming approaches to polynomial minimization over the hypercube. *SIAM J. Optim.* 20(6), 3104-3120 (2010)
13. Knoblauch, A.: Closed-form expressions for the moments of the binomial probability distribution. *SIAM J. Appl. Math.* 69(1), 197-204, (2008)
14. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11, 796-817 (2001)
15. Motzkin, T.S., Straus, E.G.: Maxima for graphs and a new proof of a theorem of Túran, *Candian. J. Math.* 17, 533-540 (1965)
16. Nesterov, Y.: Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper 2003/71, CORE-UCL, Louvain-La-Neuve (2003)

17. Nesterov, Y., Wolkowicz, H., Ye, Y.: Semidefinite programming relaxations of nonconvex quadratic optimization. In: H. Wolkowicz, R. Saigal, L. Vandenberghe (eds.), Handbook of Semidefinite Programming, Kluwer Academic Publishers, Norwell, MA, 361-419 (2000)
18. Sagol, G., Yildirim, E.A.: Analysis of copositive optimization based bounds on standard quadratic optimization. Technical Report, Department of Industrial Engineering, Koc University, Sariyer, Istanbul, Turkey (2013)
19. Vavasis, S.: Approximation algorithms for concave quadratic programming. In: C.A. Floudas, P. Pardalos (eds.), Recent Advances in Global Optimization, Princeton University Press, Princeton, NJ, 3-18 (1998)
20. Yildirim, E.A.: On the accuracy of uniform polyhedral approximations of the copositive cone. *Optim. Method. Softw.* 27(1), 155-173 (2012)

A Proof of Theorem 7

In this Appendix, we give a self-contained proof for Theorem 7, which provides an explicit description of the moments of the multinomial distribution in terms of the Stirling numbers of the second kind (and as a direct application the explicit formulation for the Bernstein approximation on the simplex from Corollary 4).

Given $x \in \Delta_n$, we consider the multinomial distribution with n categories and r independent trials, where the probability of drawing the i-th category is given by x_i. Hence, given $\alpha \in I(n,r)$, the probability of drawing α_i times the i-th category for each $i \in [n]$ is equal to $\frac{r!}{\alpha!}x^\alpha$ and, for any $\beta \in \mathbb{N}^n$, the β-th moment of the multinomial distribution is given by

$$m^\beta_{(n,r)} := \sum_{\alpha \in I(n,r)} \alpha^\beta \frac{r!}{\alpha!} x^\alpha. \tag{16}$$

Our objective is to show the following reformulaton of the β-th moment in terms of the Stirling numbers of the second kind:

$$m^\beta_{(n,r)} = \sum_{\alpha \in \mathbb{N}^n : \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \prod_{i=1}^n S(\beta_i, \alpha_i). \tag{17}$$

Our proof is elementary in the sense that we will obtain the moments of the multinomial distribution using its moment generating function. One of the ingredients which we will use is the fact that the identity (17) holds for the case $n = 2$ of the binomial distribution when $\beta \in \mathbb{N}^2$ is of the form $\beta = (\beta_1, 0)$. Namely, the following identity is shown in [13] (see Theorem 2.2 and relation (3.1) there).

Lemma 5 *[13] Given $\beta_1 \in \mathbb{N}$ and $x_1 \in \mathbb{R}$ such that $0 \leq x_1 \leq 1$, one has*

$$m^{(\beta_1,0)}_{(2,r)} = \sum_{\alpha_1=0}^r \alpha_1^{\beta_1} \binom{r}{\alpha_1} x_1^{\alpha_1} (1-x_1)^{r-\alpha_1} = \sum_{\alpha_1=0}^{\beta_1} r^{\underline{\alpha_1}} x_1^{\alpha_1} S(\beta_1, \alpha_1).$$

This implies that the identity (17) holds for the moments of the multinomial distribution when the order β has a single non-zero coordinate, i.e., β is of the form $\beta = \beta_i e_i$ with $\beta_i \in \mathbb{N}$.

Corollary 6 *Given $\beta_i \in \mathbb{N}$ and $x \in \Delta_n$, one has*

$$m^{\beta_i e_i}_{(n,r)} = \sum_{\alpha_i=0}^{\beta_i} r^{\underline{\alpha_i}} x_i^{\alpha_i} S(\beta_i, \alpha_i).$$

Proof By (16), we have

$$m^{(\beta_i e_i)}_{(n,r)} = \sum_{\alpha \in I(n,r)} \alpha_i^{\beta_i} \frac{r!}{\alpha!} x^\alpha = \sum_{\alpha_i=0}^r \alpha_i^{\beta_i} \frac{r!}{\alpha_i!(r-\alpha_i)!} x_i^{\alpha_i} \left(\sum_{\overline{\alpha} \in I(n-1,r-\alpha_i)} \frac{(r-\alpha_i)!}{\overline{\alpha}!} x^{\overline{\alpha}} \right)$$

$$= \sum_{\alpha_i=0}^r \alpha_i^{\beta_i} \binom{r}{\alpha_i} x_i^{\alpha_i} \left(\sum_{j \neq i} x_j \right)^{r-\alpha_i} = \sum_{\alpha_i=0}^r \alpha_i^{\beta_i} \binom{r}{\alpha_i} x_i^{\alpha_i} (1-x_i)^{r-\alpha_i},$$

which is equal to $\sum_{\alpha_i=0}^{\beta_i} r^{\underline{\alpha_i}} x_i^{\alpha_i} S(\beta_i, \alpha_i)$ by Lemma 5.   □

In order to determine the moments of the multinomial distribution we use its moment generating function

$$t \in \mathbb{R}^n \mapsto M_x^r(t) := \left(\sum_{i=1}^{n} x_i e^{t_i} \right)^r.$$

Then, for $\beta \in \mathbb{N}^n$, the β-th moment of the multinomial distribution is equal to the β-th derivative of the moment generating function evaluated at $t = 0$. Namely,

$$m_{(n,r)}^{\beta} = \left. \frac{\partial^{|\beta|} M_x^r(t)}{\partial t_1^{\beta_1} \cdots \partial t_n^{\beta_n}} \right|_{t=0}. \tag{18}$$

By Corollary 6 we know that, for any $\beta_i \in \mathbb{N}$,

$$m_{(n,r)}^{(\beta_i e_i)} = \left. \frac{\partial^{\beta_i} M_x^r(t)}{\partial t_i^{\beta_i}} \right|_{t=0} = \sum_{\alpha_i=0}^{\beta_i} S(\beta_i, \alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i}. \tag{19}$$

Next we show an analogue of the above relation (19) for the evaluation of the $\beta_i e_i$-th derivative of the moment generating function at any point $t \in \mathbb{R}^n$.

Lemma 6 *For $x \in \Delta_n$, $\beta_i \in \mathbb{N}$ and $t \in \mathbb{R}^n$, one has*

$$\frac{\partial^{\beta_i} M_x^r(t)}{\partial t_i^{\beta_i}} = \sum_{\alpha_i=0}^{\beta_i} S(\beta_i, \alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M_x^{r-\alpha_i}(t).$$

For the proof we will use the following recursive relation for the Stirling numbers of the second kind.

Lemma 7 *For any integers $\beta \geq 0$ and $\alpha \geq 1$, one has*

$$S(\beta + 1, \alpha) = S(\beta, \alpha - 1) + \alpha S(\beta, \alpha). \tag{20}$$

Proof This well known fact can be easily checked as follows. By definition, $S(\beta + 1, \alpha)$ counts the number of ways of partitioning the set $\{1, \ldots, \beta, \beta + 1\}$ into α nonempty subsets. Considering the last element $\beta + 1$, one can either put it in a singleton subset (so that there are $S(\beta, \alpha - 1)$ such partitions), or partition $\{1, \ldots, \beta\}$ into α nonempty subsets and then assign the last element $\beta + 1$ to one of them (so that there are $\alpha S(\beta, \alpha)$ such partitions). This shows the result. $\qquad \square$

Proof (of Lemma 6) To simplify notation we set $M^r = M_x^r(t)$. We show the result using induction on $\beta_i \geq 0$. The result holds clearly for $\beta_i = 0$ and also for $\beta_i = 1$ in which case we have

$$\frac{\partial M^r}{\partial t_i} = r x_i e^{t_i} M^{r-1}. \tag{21}$$

We now assume that the result holds for β_i and we show that it also holds for $\beta_i + 1$. For this, using the induction assumption, we obtain

$$\frac{\partial^{\beta_i+1} M^r}{\partial t_i^{\beta_i+1}} = \frac{\partial}{\partial t_i} \frac{\partial^{\beta_i} M^r}{\partial t_i^{\beta_i}} = \frac{\partial}{\partial t_i} \left(\sum_{\alpha_i=0}^{\beta_i} S(\beta_i, \alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M^{r-\alpha_i} \right) = \sum_{\alpha_i=0}^{\beta_i} S(\beta_i, \alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} \frac{\partial}{\partial t_i} (e^{\alpha_i t_i} M^{r-\alpha_i}). \tag{22}$$

Now, using (21), we can compute the last term as follows:

$$\frac{\partial}{\partial t_i} (e^{\alpha_i t_i} M^{r-\alpha_i}) = \alpha_i e^{\alpha_i t_i} M^{r-\alpha_i} + (r - \alpha_i) x_i e^{(\alpha_i+1)t_i} M^{r-\alpha_i-1}.$$

Plugging this into relation (22), we deduce

$$
\begin{aligned}
\frac{\partial^{\beta_i+1} M^r}{\partial t_i^{\beta_i+1}} &= \sum_{\alpha_i=0}^{\beta_i} \alpha_i S(\beta_i,\alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M^{r-\alpha_i} + \sum_{\alpha_i=0}^{\beta_i} S(\beta_i,\alpha_i) r^{\underline{\alpha_i+1}} x_i^{\alpha_i+1} e^{(\alpha_i+1)t_i} M^{r-\alpha_i-1} \\
&= \sum_{\alpha_i=0}^{\beta_i} \alpha_i S(\beta_i,\alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M^{r-\alpha_i} + \sum_{\alpha_i'=1}^{\beta_i+1} S(\beta_i,\alpha_i'-1) r^{\underline{\alpha_i'}} x_i^{\alpha_i'} e^{\alpha_i' t_i} M^{r-\alpha_i'} \\
&= \sum_{\alpha_i=0}^{\beta_i} \underbrace{(\alpha_i S(\beta_i,\alpha_i) + S(\beta_i,\alpha_i-1))}_{=S(\beta_i+1,\alpha_i) \text{ by (20)}} r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M^{r-\alpha_i} + r^{\underline{\beta_i+1}} x_i^{\beta_i+1} e^{(\beta_i+1)t_i} M^{r-\beta_i-1} \\
&= \sum_{\alpha_i=0}^{\beta_i+1} S(\beta_i+1,\alpha_i) r^{\underline{\alpha_i}} x_i^{\alpha_i} e^{\alpha_i t_i} M^{r-\alpha_i},
\end{aligned}
$$

which concludes the proof.   □

We now extend the result of Lemma 6 to an arbitrary derivative of the moment generating function.

Theorem 9 *For any $x \in \Delta_n$, $\beta \in \mathbb{N}^n$ and $t \in \mathbb{R}^n$, one has*

$$
\frac{\partial^{|\beta|} M_x^r(t)}{\partial t_1^{\beta_1} \cdots \partial t_n^{\beta_n}} = \sum_{\alpha \in \mathbb{N}^n : \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha M_x^{r-|\alpha|}(t) \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i,\alpha_i) \right).
$$

Proof We show the result using induction on the size k of the support of β, i.e., $k = |\{i \in [n] : \beta_i \neq 0\}|$. The result holds clearly for $k = 0$ and, for $k = 1$, the result holds by Lemma 6. We now assume that the result holds for k and we show that it also holds for $k + 1$. For this, consider the sequences $\beta' = (\beta_1, \ldots, \beta_k, \beta_{k+1}, 0, \ldots, 0)$ and $\beta = (\beta_1, \ldots, \beta_k, 0, 0, \ldots, 0) \in \mathbb{N}^n$, where $\beta_1, \ldots, \beta_{k+1} \geq 1$. By the induction assumption we know that

$$
\frac{\partial^{|\beta|} M^r}{\partial t_1^{\beta_1} \cdots \partial t_k^{\beta_k}} = \sum_{0 \leq \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha M^{r-|\alpha|} \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i,\alpha_i) \right), \tag{23}
$$

setting again $M^r = M_x^r(t)$ for simplicity. Using (23), we obtain

$$
\frac{\partial^{|\beta'|} M^r}{\partial t_1^{\beta_1} \cdots \partial t_{k+1}^{\beta_{k+1}}} = \frac{\partial^{\beta_{k+1}}}{\partial t_{k+1}^{\beta_{k+1}}} \frac{\partial^{|\beta|} M^r}{\partial t_1^{\beta_1} \cdots \partial t_k^{\beta_k}} = \frac{\partial^{\beta_{k+1}}}{\partial t_{k+1}^{\beta_{k+1}}} \left(\sum_{0 \leq \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha M^{r-|\alpha|} \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i,\alpha_i) \right) \right). \tag{24}
$$

Note that $\alpha_{k+1} = 0$ since $\alpha_{k+1} \leq \beta_{k+1}$ and $\beta_{k+1} = 0$. Hence, $M^{r-|\alpha|}$ is the only term containing the variable t_{k+1} and thus (24) implies

$$
\frac{\partial^{|\beta'|} M^r}{\partial t_1^{\beta_1} \cdots \partial t_{k+1}^{\beta_{k+1}}} = \sum_{0 \leq \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i,\alpha_i) \right) \frac{\partial^{\beta_{k+1}} M^{r-|\alpha|}}{\partial t_{k+1}^{\beta_{k+1}}}. \tag{25}
$$

We now use Lemma 6 to compute the last term:

$$
\frac{\partial^{\beta_{k+1}} M^{r-|\alpha|}}{\partial t_{k+1}^{\beta_{k+1}}} = \sum_{\theta_{k+1}=0}^{\beta_{k+1}} S(\beta_{k+1},\theta_{k+1})(r-|\alpha|)^{\underline{\theta_{k+1}}} x_{k+1}^{\theta_{k+1}} e^{\theta_{k+1} t_{k+1}} M^{r-|\alpha|-\theta_{k+1}}. \tag{26}
$$

Plugging (26) into (25) we obtain

$$
\frac{\partial^{|\beta'|} M^r}{\partial t_1^{\beta_1} \cdots \partial t_{k+1}^{\beta_{k+1}}}
$$

$$
= \sum_{0 \leq \alpha \leq \beta} r^{\underline{|\alpha|}} x^\alpha \left(\sum_{\theta_{k+1}=0}^{\beta_{k+1}} S(\beta_{k+1}, \theta_{k+1})(r - |\alpha|)^{\underline{\theta_{k+1}}} x_{k+1}^{\theta_{k+1}} e^{\theta_{k+1} t_{k+1}} M^{r-|\alpha|-\theta_{k+1}} \right) \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i, \alpha_i) \right)
$$

$$
= \sum_{0 \leq \alpha \leq \beta} \sum_{\theta_{k+1}=0}^{\beta_{k+1}} r^{\underline{|\alpha|+\theta_{k+1}}} x^{\alpha + e_{k+1}\theta_{k+1}} S(\beta_{k+1}, \theta_{k+1}) e^{\theta_{k+1} t_{k+1}} M^{r-(|\alpha|+\theta_{k+1})} \left(\prod_{i=1}^n e^{\alpha_i t_i} S(\beta_i, \alpha_i) \right)
$$

$$
= \sum_{0 \leq \alpha' \leq \beta'} r^{\underline{|\alpha'|}} x^{\alpha'} M^{r-|\alpha'|} \left(\prod_{i=1}^n e^{\alpha'_i t_i} S(\beta'_i, \alpha'_i) \right),
$$

after setting $\alpha' = \alpha + e_{k+1}\theta_{k+1}$. This concludes the proof of Theorem 9. $\qquad\square$