

# ERROR BOUNDS FOR SOME SEMIDEFINITE PROGRAMMING APPROACHES TO POLYNOMIAL MINIMIZATION ON THE HYPERCUBE

ETIENNE DE KLERK\* AND MONIQUE LAURENT†

**Abstract.** We consider the problem of minimizing a polynomial on the hypercube  $[0, 1]^n$  and derive new error bounds for the hierarchy of semidefinite programming approximations to this problem corresponding to the Positivstellensatz of Schmüdgen [26]. The main tool we employ is Bernstein approximations of polynomials, which also gives constructive proofs and degree bounds for positivity certificates on the hypercube.

**Key words.** Positivstellensatz, positive polynomial, sum of squares of polynomials, bound constrained optimization of polynomials, multivariate Bernstein approximation, semidefinite programming.

**AMS subject classifications.** 90C60, 90C56, 90C26

**1. Introduction.** In this paper we study the problem:

$$p_{\min, Q} := \min \{p(x) \mid x \in Q\}, \quad (1.1)$$

of minimizing a polynomial  $p$  over the unit hypercube  $Q = [0, 1]^n$ . When  $p$  is quadratic, this problem includes e.g. the maximum cut problem in graphs. Indeed, for a graph  $G = (V, E)$ , the size of the maximum cut is given by the quadratic program

$$\max_{x \in [-1, 1]^{|V|}} \frac{1}{4} x^T L x = \max_{x \in [0, 1]^{|V|}} \frac{1}{4} (2x - e)^T L (2x - e), \quad (1.2)$$

where  $e \in \mathbb{R}^V$  denotes the all-ones vector and  $L \in \mathbb{R}^{V \times V}$  is the Laplacian matrix of  $G$ , with  $L_{ii}$  being the number of nodes adjacent to  $i \in V$  and  $L_{ij} = -1$  if  $ij \in E$ ,  $L_{ij} = 0$  otherwise, for  $i \neq j \in V$ . For the maximum cut problem there is a celebrated 0.878-approximation result due to Goemans and Williamson [7], and related approximation results for quadratic optimization over the hypercube were given by Nesterov *et al.* [20]. On the negative side, the maximum cut problem cannot be approximated within  $16/17 \approx 0.941$  [10].

Another example is the maximum stable set problem in graphs. Recall that a stable set in a graph is a subset of vertices such that no two vertices in this subset are adjacent. The cardinality of the largest stable set in a graph  $G = (V, E)$  is called the stable set number of  $G$ , and is usually denoted by  $\alpha(G)$ . One may show that

$$\alpha(G) = \max_{x \in [0, 1]^{|V|}} \left( \frac{1}{2} x^T L x - \sum_{i=1}^{|V|} \left( \frac{1}{2} d_i - 1 \right) x_i \right),$$

where  $d_i$  denotes the degree of vertex  $i$ , and  $L$  the Laplacian matrix of  $G$ , as before. It is known that there does not exist a fixed  $\epsilon > 0$  such that one can always approximate  $\alpha(G)$  to within a factor  $|V|^{1-\epsilon}$  in polynomial time, unless  $P=NP$  [9, 29].

\*Tilburg University, Department of Econometrics and Operations Research, 5000 LE Tilburg, Netherlands. (E.deKlerk@uvt.nl).

†Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, and Tilburg University, Department of Econometrics and Operations Research, 5000 LE Tilburg, Netherlands. (M.Laurent@cwi.nl).

A recent approach to approximate a polynomial optimization problem like (1.1) is to use sums of squares representations for polynomials positive on the feasible region, which can then be computed efficiently using semidefinite programming. Some error bounds for the approximations obtained from the Positivstellensätze of Schmüdgen [26] and of Putinar [24] have been derived by Schweighofer [27] and by Nie and Schweighofer [21]. These bounds involve some constants that depend on the given data and are generally hard to compute. Practical error bounds (not involving unknown constants) are known only in sporadic cases, most notably for polynomial optimization over the standard simplex [1, 3]. In this paper we derive explicit (stronger) error bounds for the hierarchy of semidefinite programming relaxations obtained from Schmüdgen's Positivstellensatz for optimization over the hypercube. Our approach is elementary and relies on using Bernstein approximations. Moreover it provides explicit positivity certificates for positive polynomials on the hypercube.

The paper is organized as follows. In the introduction, we introduce the Positivstellensätze of Handelman and of Schmüdgen, we present the bounds of Schweighofer [27], and we summarize our main results; in the last section, we recall some basic facts on Bernstein approximations which will play a central role in our approach. Sections 2 and 3 contain our new degree and error bounds for optimization on the hypercube. In our analysis, we will distinguish between quadratic polynomials (Section 2) and higher degree polynomials (Section 3) — the quadratic case is of independent interest due to its applications, its treatment is simpler, and sharper bounds may be obtained than for the general case.

**Some notation.** For an integer  $n \geq 1$ , we set  $[n] := \{1, \dots, n\}$ ,  $[n]_0 := \{0, 1, \dots, n\}$ , and  $e_1, \dots, e_n$  denote the standard unit vectors in  $\mathbb{R}^n$ .  $\mathbb{R}[x] = \mathbb{R}[x_1, \dots, x_n]$  denotes the ring of multivariate polynomials in  $n$  variables, and  $\mathbb{R}[x]_m$  the subspace of polynomials with degree at most  $m$ . Monomials in  $\mathbb{R}[x]$  are denoted as  $x^k = x_1^{k_1} \dots x_n^{k_n}$  for  $k \in \mathbb{N}^n$ , with degree  $|k| := \sum_{i=1}^n k_i$ , and we set  $k! := k_1! \dots k_n!$ . For a polynomial  $p = \sum_k p_k x^k \in \mathbb{R}[x]$ ,  $\deg(p) = \max_{k|p_k \neq 0} |k|$  (set to  $-\infty$  if  $p = 0$ ) and we set

$$L(p) := \max_k |p_k| \frac{k!}{|k|!}. \quad (1.3)$$

Given a subset  $S \subseteq \mathbb{R}^n$ , we say that  $p$  is positive on  $S$  when  $p(x) > 0$  for all  $x \in S$ . Given  $g_1, \dots, g_s \in \mathbb{R}[x]$  and  $k \in \mathbb{N}^s$ , we often use the notation  $g^k = g_1^{k_1} \dots g_s^{k_s}$ , with  $g^0 := 1$ .

**1.1. Positivity certificates and relaxations for polynomial optimization problems.** Problem (1.1) falls in the general class of polynomial optimization problems, where the hypercube  $Q$  is replaced by an arbitrary compact basic closed semi-algebraic set

$$S := \{x \in \mathbb{R}^n : g_j(x) \geq 0, j = 1, \dots, s\}, \quad (1.4)$$

and  $g_j \in \mathbb{R}[x]$  are given polynomials. We find the hypercube  $S = Q$  when considering the  $s = 2n$  (linear) polynomials

$$g_1 = x_1, g_2 = 1 - x_1, \dots, g_{2n-1} = x_n, g_{2n} = 1 - x_n. \quad (1.5)$$

The general problem is to minimize a polynomial  $p$  over  $S$ , i.e., to compute

$$p_{\min, S} := \min\{p(x) : x \in S\}. \quad (1.6)$$

In the last few years there has been significant interest in designing tractable approximations for this problem, following the seminal works [13, 22] (see e.g. [18] for an overview). The starting point is to reformulate  $p_{\min,S}$  as

$$p_{\min,S} = \min_{x \in S} p(x) = \sup \{t \mid p(x) - t > 0 \forall x \in S\}.$$

Then the strategy is to use a *Positivstellensatz* (i.e. some result describing positive polynomials on  $S$ ) to replace the positivity condition, which is hard to test, by some easier condition. When the set  $S$  is a polytope, we may use the following result of Handelman.

**THEOREM 1.1** (Handelman [8]). *Let  $p \in \mathbb{R}[x]$  and let  $S$  be a polytope, as defined in (1.4) where all  $g_j$  are linear polynomials. If  $p$  is positive on  $S$ , then  $p$  admits a representation*

$$p = \sum_{k \in \mathbb{N}^s} \lambda_k g^k \quad \text{for some scalars } \lambda_k \geq 0. \quad (1.7)$$

A constructive proof is given by Powers and Reznick [23], who use a reduction to optimization over the simplex together with a result by Pólya. When  $S$  is a general compact basic semi-algebraic set, we can use the following Positivstellensatz of Schmüdgen.

**THEOREM 1.2** (Schmüdgen [26]). *Let  $p \in \mathbb{R}[x]$  and let  $S$  be as in (1.4) assumed to be compact. If  $p$  is positive on  $S$ , then  $p$  admits a representation*

$$p = \sum_{k \in \{0,1\}^s} \sigma_k g^k \quad \text{where } \sigma_k \text{ are sums of squares of polynomials.} \quad (1.8)$$

Schweighofer [27] gives a proof which moreover provides explicit bounds on the degree of the representation as well as an error analysis; both are recalled in Theorem 1.3 below. The maximum degree:  $\max_{k \mid \lambda_k \neq 0} \deg(g^k)$  in the representation (1.7), or  $\max_{k \mid \sigma_k \neq 0} \deg(\sigma_k g^k)$  in (1.8), is called the *maximum degree of the representation*. For an integer  $r \geq 1$ , it is convenient to introduce the following sets<sup>1</sup>:

$$H_r(g) := \left\{ \sum_{k \in \mathbb{N}^s} \lambda_k g^k \mid \deg(\lambda_k g^k) \leq r, \lambda_k \in \mathbb{R}_+ \right\}, \quad (1.9)$$

that corresponds to the Handelman representation (1.7), and

$$T_r(g) := \left\{ \sum_{k \in \{0,1\}^s} \sigma_k g^k \mid \deg(\sigma_k g^k) \leq r, \sigma_k \text{ sums of squares of polynomials} \right\}, \quad (1.10)$$

that corresponds to the Schmüdgen representation (1.8). Obviously,  $H_r(g) \subseteq T_r(g)$ . Indeed, if  $k \in \mathbb{N}^s$  and  $\lambda_k > 0$  then, by pulling out all squares in  $g^k$ , we can write  $\lambda_k g^k = \sigma g^{k'}$ , where  $k' \in \{0,1\}^s$ ,  $k'_i = 1$  precisely when  $k_i$  is an odd integer, and  $\sigma$  is a sum of squares of polynomials (in fact, a single square). We can define the following lower bounds on  $p_{\min,S}$ :

$$p_{\text{han},g}^{(r)} := \sup \left\{ t \mid p - t \in H_r(g) \right\}, \quad (1.11)$$

<sup>1</sup> $H_r(g)$  is known as the truncated preprime and  $T_r(g)$  as the truncated preordering generated by  $g_1, \dots, g_s$ .

and

$$p_{\text{sch},g}^{(r)} := \sup \left\{ t \mid p - t \in T_r(g) \right\}, \quad (1.12)$$

which satisfy

$$p_{\text{han},g}^{(r)} \leq p_{\text{sch},g}^{(r)} \leq p_{\min,S}. \quad (1.13)$$

Getting explicit tight error bounds for the parameters  $p_{\text{han},g}^{(r)}$  and  $p_{\text{sch},g}^{(r)}$  is the main motivation of this paper.

Theorem 1.1 implies directly that the bounds (1.11) converge asymptotically to  $p_{\min,S}$  (as  $r$  goes to  $\infty$ ) when  $S$  is a polytope, and the asymptotic convergence of the bounds (1.12) to  $p_{\min,S}$  follows directly from Theorem 1.2. For fixed  $r$ , the bound (1.11) can be computed via a linear program in the variables  $\lambda_k$ , obtained by equating the coefficients of the polynomials on both sides of the equality  $p - t = \sum_k \lambda_k g^k$ . For fixed  $r$ , the bound (1.12) can be computed via a semidefinite program. Indeed, as is well known, testing whether a polynomial  $\sigma$  of degree  $2d$  is a sum of squares of polynomials amounts to testing whether there exists a positive semidefinite matrix  $M$  of order  $\binom{n+d}{d}$  satisfying  $\sigma(x) = [x]_d^T M [x]_d$ , where  $[x]_d = (x^k)_{k \in \mathbb{N}^n, |k| \leq d}$ . Similarly, testing for membership in the set  $T_r(g)$  can be reformulated as a semidefinite program involving  $2^s$  positive semidefinite matrices of order  $\binom{n+r/2}{r/2}$ . Although the bounds (1.11) of Handelman type might appear to be simpler than the sums of squares bounds (1.12) as they can be computed via LP instead of SDP, they have several drawbacks as pointed out by Lasserre [15]. In particular, no decomposition of the form  $p - p_{\min,S} = \sum_k \lambda_k g^k$  with  $\lambda_k \geq 0$ , exists when  $p$  attains its minimum at an interior point of the polytope  $S$ ; in other words, in that case, there cannot be *finite* convergence of the bounds (1.11) to  $p_{\min,S}$ . However, if we allow sums of squares as multipliers instead of nonnegative scalars, then finite convergence can be proved for several problem instances (e.g. in the finite variety case [14, 17], for optimization over the gradient variety [6], or in the convex case [16]).

Given an integer  $d \geq 1$ , by minimizing the polynomial  $p$  over the set

$$Q(d) := \{x \in Q \mid dx \in \mathbb{N}^n\} = \left\{ \frac{k}{d} \mid k \in [d]_0^n \right\}$$

of rational points with denominator  $d$  in the hypercube  $Q$ , we obtain the following upper bound

$$p_{\min,Q(d)} := \min_{x \in Q(d)} p(x) \geq p_{\min,Q} \quad (1.14)$$

for the minimum of  $p$  over  $Q$ .

It turns out that our approach – based on Bernstein approximations – will produce representations of Handelman type for positive polynomials on the hypercube and error bounds for the lower approximations of  $p_{\min,Q}$  by the parameters  $p_{\text{han},g}^{(r)}$  and  $p_{\text{sch},g}^{(r)}$ . Moreover, it will also give error bounds for the upper approximations  $p_{\min,Q(d)}$ .

**1.2. Error analysis for sums of squares representations.** We recall the result of Schweighofer [27] which gives quantitative information about the sums of squares representation of Schmüdgen for positive polynomials on  $S$ , namely degree bounds for the representation (1.8) and an error analysis for the approximation of  $p_{\min,S}$  by the parameters  $p_{\text{sch},g}^{(r)}$ .

**THEOREM 1.3** (Schweighofer [27]). *Let  $S$  be as in (1.4) and assume that  $S \subseteq (-1, 1)^n$ . Then there exist integer constants  $c, c' > 0$  satisfying the following properties.*

- (i) *Every polynomial  $p$  of degree  $m$  which is positive on  $S$  belongs to  $T_r(g)$  for some integer  $r$  satisfying*

$$r \leq cm^2 \left( 1 + \left( m^2 n^m \frac{L(p)}{p_{\min, S}} \right)^c \right).$$

- (ii) *For every polynomial  $p$  of degree  $m$  and for all integers  $r \geq c' m^{c'} n^{c' m}$ , we have*

$$p - p_{\min, S} + \frac{c' m^4 n^{2m}}{c' \sqrt[r]{r}} L(p) \in T_r(g), \quad \text{and thus,} \quad p_{\min, S} - p_{\text{sch}, g}^{(r)} \leq \frac{c' m^4 n^{2m}}{c' \sqrt[r]{r}} L(p). \quad (1.15)$$

The bounds in Theorem 1.3 depend on three parameters: the constants  $c$  and  $c'$  (which depend only on the description of  $S$  by the polynomials  $g_j$ ), the degree  $m$  of  $p$ , and the quantity  $L(p)/p_{\min, S}$  (which measures how close  $p$  is to having a zero on  $S$ ). Schweighofer [27] shows that  $c' = (4c)^c$  is a valid choice and notes that the constant  $c$  could in principle be deduced from his proof, although the analysis would probably be too tedious (cf. [27, Remark 10]). It thus remains a nontrivial task to compute the constants explicitly for concrete sets  $S$ . In this note we show – using simple direct arguments – that  $c = c' = 1$  are (roughly speaking) suitable choices in the case of the hypercube  $S = [0, 1]^n$ .

More precisely, we show the following results.

**THEOREM 1.4.** *Let  $S = Q = [0, 1]^n$  be described by the polynomials  $g_j$  from (1.5), and let  $p$  be a polynomial of degree  $m$ .*

- (i) *If  $p$  is positive on  $Q$ , then  $p \in H_r(g) \subseteq T_r(g)$  for some integer  $r \leq n \left\lceil \frac{L(p)}{p_{\min, Q}} \binom{m+1}{3} n^m \right\rceil$ .*  
 (ii) *For any integer  $d \geq 1$ , we have*

$$p - p_{\min, Q} + \frac{L(p)}{d} \binom{m+1}{3} n^m \in H_r(g) \subseteq T_r(g) \quad \text{for some integer } r \leq \max(dn, m).$$

- (iii) *For any integer  $d \geq m$ , we have*

$$\max \left\{ p_{\min, Q} - p_{\text{sch}, g}^{(dn)}, p_{\min, Q} - p_{\text{han}, g}^{(dn)}, p_{\min, Q(d)} - p_{\min, Q} \right\} \leq \frac{L(p)}{d} \binom{m+1}{3} n^m.$$

This is our main result, which will follow from Theorem 3.4; sharper bounds are given in Theorem 2.1 in the quadratic case  $m = 2$ . This result thus adds to the small number of explicit error bounds that are known for semidefinite programming approximations of polynomial optimization problems.

We conclude with a brief comparison with the results of Theorem 1.3 applied to the hypercube  $S = Q$ . We show in Table 1.1 the order of magnitude for the degree bounds (of positivity certificates on the hypercube) and for the error bounds obtained for the approximations based on Schmüdgen type representations (Theorem 1.3), compared to our results from Theorem 1.4 for the general case  $m = \deg(p) \geq 1$  and from Theorem 2.1 for the quadratic case  $m = \deg(p) = 2$ . We see that, in the case of optimization over the hypercube, we can choose the constants  $c = c' = 1$  in Theorems 1.3. Our bounds improve the bounds in this theorem (except we lose a factor  $n$  with respect to the degree bound of Theorem 1.3 for general degree  $m \geq 1$ ).

	Theorem 2.1 $m = 2$	Theorem 1.4 $m \geq 1$	Theorem 1.3 with $c = c' = 1$
degree bound	$n^2 \frac{L(p)}{p_{\min, Q}}$	$\frac{m^3 n^{m+1}}{6} \frac{L(p)}{p_{\min, Q}}$	$m^4 n^m \frac{L(p)}{p_{\min, Q}}$
error bound at order $r = nd$	$n \frac{L(p)}{d}$ for $d \geq 2$	$\frac{m^3 n^m}{6} \frac{L(p)}{d}$ for $d \geq m$	$m^4 n^{2m-1} \frac{L(p)}{d}$ for $d \geq mn^{m-1}$

TABLE 1.1

Comparing degree and error bounds for optimization over the hypercube

**1.3. Bernstein operators on the hypercube.** We recall here some basic results on Bernstein approximations of functions on  $[0, 1]^n$ . We will indeed use the Bernstein operator as a crucial ingredient for constructing positivity certificates and approximating polynomials on the hypercube.

In the univariate case, the Bernstein basis for the space of univariate polynomials of degree at most  $d$  consists of the polynomials

$$p_{d,k} := \binom{d}{k} x^k (1-x)^{d-k} \quad (k = 0, \dots, d). \quad (1.16)$$

Then the Bernstein approximation of a function  $f \in C[0, 1]$  is the polynomial  $B_d(f) \in \mathbb{R}[x]_d$  defined by

$$B_d(f) := \sum_{k=0}^d f\left(\frac{k}{d}\right) p_{d,k}.$$

It is well known that  $B_d(f)$  converges uniformly to  $f$  as  $d \rightarrow \infty$  (see e.g. [19]). Clearly,  $B_d$  is a linear operator and  $B_d$  preserves positivity, i.e.,  $f > 0$  on  $[0, 1]$  implies  $B_d(f) > 0$  on  $[0, 1]$ . We will use the Bernstein approximations  $B_d(x^m)$  of monomials  $x^m$  ( $m \in \mathbb{N}$ ). Clearly,  $B_d(1) = 1$  and  $B_d(x) = x$ , and thus  $B_d$  preserves linear polynomials. Moreover,

$$B_d(x^2) = x^2 + \frac{1}{d}x(1-x). \quad (1.17)$$

Closed form expressions for  $B_d(x^m)$  ( $m > 2$ ) are given in [12, Theorem 4.1]<sup>2</sup>; in particular,

$$B_d(x^m) = \frac{1}{d^m} \sum_{i=0}^m b_{m,i} d^i x^i \quad (1.18)$$

where  $d^i = d(d-1)\cdots(d-i+1)$ , and  $b_{m,i}$  is the number of ways to partition a  $m$ -elements set into  $i$  nonempty subsets; thus  $b_{m,1} = b_{m,m} = 1$  and  $b_{m,i} > 0$ . (The numbers  $b_{m,i}$  are known as Sterling numbers of the second kind.)

In the multivariate case, the  $n$ -variate Bernstein polynomials are defined by

$$P_{d,k} := \prod_{i=1}^n p_{d,k_i}(x_i) = \prod_{i=1}^n \binom{d}{k_i} x_i^{k_i} (1-x_i)^{d-k_i} \quad (k \in [d]_0^n = \{0, 1, \dots, d\}^n), \quad (1.19)$$

<sup>2</sup>The paper [12] deals with higher moments of the binomial distribution, and the polynomial  $d^m B_d(x^m)$  evaluated at  $x = a$  is precisely the  $m$ th moment of a binomial distribution with parameters  $a$  and  $d$ .

where  $p_{d,k_i}$  are the univariate Bernstein polynomials as in (1.16). Then,  $\mathcal{B}_d := \{P_{d,k} \mid k \in [d]_0^n\}$  forms the so-called Bernstein basis of the space  $\mathbb{R}[x]_{d,\dots,d}$ , consisting of the  $n$ -variate polynomials having degree at most  $d$  in each variable  $x_i$ . A simple useful fact is that the Bernstein polynomials sum up to 1; namely,

$$\sum_{k=0}^d p_{d,k} = 1 \quad (\text{in the univariate case}), \quad \sum_{k \in [d]_0^n} P_{d,k} = 1 \quad (\text{in the multivariate case}). \quad (1.20)$$

The Bernstein approximation of order  $d$  of a function  $f \in C([0,1]^n)$  is the polynomial (of degree  $dn$ )

$$B_d(f) := \sum_{k_1=0}^d \cdots \sum_{k_n=0}^d f\left(\frac{k_1}{d}, \dots, \frac{k_n}{d}\right) P_{d,k}.$$

It follows from the definition that the Bernstein operator is multiplicative when applying it to functions that are products of functions in disjoint sets of variables. In particular,

$$B_d(x_1^{k_1} \cdots x_n^{k_n}) = \prod_{i=1}^n B_d(x_i^{k_i}) \quad \text{for } k \in \mathbb{N}^n, \quad \text{thus } B_d(x^k) = x^k \quad \text{for } k \in \{0,1\}^n. \quad (1.21)$$

Any polynomial  $p \in \mathbb{R}[x]_{d,\dots,d}$  can be written in the basis  $\mathcal{B}_d$  as  $p = \sum_{k \in [d]_0^n} b_k^{(d)} P_{d,k}$  for some scalars  $b_k^{(d)}$ , called the Bernstein coefficients of  $p$ . Let

$$p_{\text{Ber}}^{(d)} := \min_{k \in [d]_0^n} b_k^{(d)} \quad (1.22)$$

denote the smallest Bernstein coefficient of  $p$  in the basis  $\mathcal{B}_d$ . Using (1.20), we see that the polynomial

$$p - p_{\text{Ber}}^{(d)} = \sum_{k \in [d]_0^n} (b_k^{(d)} - p_{\text{Ber}}^{(d)}) P_{d,k} \quad (1.23)$$

has nonnegative coefficients in the basis  $\mathcal{B}_d$ . Therefore, this polynomial is nonnegative on  $Q$ , which implies  $p_{\text{Ber}}^{(d)} \leq p_{\min,Q}$ . Moreover, as each  $P_{k,d}$  belongs to the set  $H_{dn}(g)$  (introduced in (1.9), where  $g$  stands for the polynomials  $x_i$  and  $1-x_i$  for  $i = 1, \dots, n$ ), (1.23) shows that  $p - p_{\text{Ber}}^{(d)} \in H_{dn}(g)$ . This implies

$$p_{\text{Ber}}^{(d)} \leq p_{\text{han},g}^{(dn)}$$

and thus, combined with (1.13) and (1.14),

$$p_{\text{Ber}}^{(d)} \leq p_{\text{han},g}^{(dn)} \leq p_{\text{sch},g}^{(dn)} \leq p_{\min,Q} \leq p_{\min,Q(d)}. \quad (1.24)$$

Our approach will produce bounds on the quantity  $p_{\min,Q(d)} - p_{\text{Ber}}^{(d)}$ , which thus also implies bounds for the approximation of  $p_{\min,Q}$  by  $p_{\text{han},g}^{(dn)}$  or  $p_{\text{sch},g}^{(dn)}$ .

**2. Results for quadratic polynomials.** Here we consider a quadratic polynomial of the form  $p = x^T Ax + b^T x + c$ , where  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$  and  $c \in \mathbb{R}$ . Set

$$I_+ := \{i \in [n] \mid A_{ii} > 0\}, \quad I_- := \{i \in [n] \mid A_{ii} < 0\}. \quad (2.1)$$

Fix an integer  $d \geq 1$ . Then the Bernstein approximation of order  $d$  of  $p$  takes the form

$$B_d(p) = p + \frac{1}{d} \sum_{i=1}^n A_{ii} x_i (1 - x_i) \quad (2.2)$$

(which follows directly from the linearity of the Bernstein operator and (1.17), (1.21)). As we now see, it can be used to derive various information about the minimum of  $p$  over the hypercube  $Q$ . Indeed (2.2) implies:

$$\begin{aligned} p &= B_d(p) + p - B_d(p) = B_d(p) - \frac{1}{d} \sum_{i=1}^n A_{ii} x_i (1 - x_i) \\ &= B_d(p) - \frac{1}{d} \sum_{i \in I_-} A_{ii} x_i (1 - x_i) + \frac{1}{d} \sum_{i \in I_+} A_{ii} ((x_i - 1)^2 + x_i) - \frac{1}{d} \sum_{i \in I_+} A_{ii}, \end{aligned}$$

where we used the identity  $-x_i(1 - x_i) = (x_i - 1)^2 + x_i - 1$  for the last equality. This gives the identity:

$$p - p_{\min, Q} = B_d(p - p_{\min, Q}) + \underbrace{\frac{1}{d} \sum_{i \in I_-} |A_{ii}| x_i (1 - x_i)}_{:=q_1} + \underbrace{\frac{1}{d} \sum_{i \in I_+} A_{ii} ((x_i - 1)^2 + x_i)}_{:=q_2} - \underbrace{\frac{1}{d} \sum_{i \in I_+} A_{ii}}_{:=C(d, p)}. \quad (2.3)$$

Here,  $B_d(p - p_{\min, Q}) = \sum_{k \in [d]_0^n} \left( p \binom{k}{d} - p_{\min, Q} \right) P_{d, k} \in H_{dn}(g)$ ,  $q_1, q_2 \in H_2(g)$ , and the constant  $C(d, p) = \frac{1}{d} \sum_{i \in I_+} A_{ii}$  depends only on  $p$  and on the order  $d$  of the Bernstein operator.

**THEOREM 2.1.** *Let  $p = x^T Ax + b^T x + c$  be a quadratic polynomial, where  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$  and  $c \in \mathbb{R}$ , let  $I_+$  be as in (2.1), and let  $g$  be as in (1.5).*

- (i) *For any integer  $d \geq 1$ ,  $p - p_{\min, Q} + \frac{1}{d} \sum_{i \in I_+} A_{ii} \in H_r(g)$  for some integer  $r \leq \max(dn, 2)$ .*
- (ii) *If  $p$  is positive on the hypercube  $Q$ , then  $p \in H_r(g)$  for some integer  $r \leq \max(nd_p, 2)$ , where  $d_p := \left\lceil \frac{\sum_{i \in I_+} A_{ii}}{p_{\min, Q}} \right\rceil$ .*
- (iii) *For any integer  $d \geq 2$ ,*

$$p_{\min, Q(d)} - p_{\text{Ber}}^{(d)} \leq \frac{1}{d} \sum_{i \in I_+} A_{ii} \quad (2.4)$$

$$\text{and thus, } p_{\min, Q} - p_{\text{han}, g}^{(dn)} \leq p_{\min, Q} - p_{\text{Ber}}^{(d)} \leq \frac{1}{d} \sum_{i \in I_+} A_{ii}.$$

- (iv) *For any integer  $d \geq 2$ ,*

$$p_{\min, Q(d)} - p_{\min, Q} \leq \frac{1}{4d} \sum_{i \in I_+} A_{ii}. \quad (2.5)$$



**Proof.** (i) follows directly from (2.3); the same for (ii) after choosing for  $d$  the smallest integer making the constant  $p_{\min, Q} - \frac{1}{d} \sum_{i \in I_+} A_{ii}$  nonnegative.

(iii) We verify that  $q_1 + q_2$  has nonnegative Bernstein coefficients in the basis  $\mathcal{B}_d$ . Indeed each of  $x_i^2 - x_i + 1$  and  $x_i(1 - x_i)$  has nonnegative coefficients in  $\mathcal{B}_d$ , which follows directly from the fact that

$$x_i^2 + 1 - x_i = ((1 - x_i)^2 + x_i(1 - x_i) + x_i^2)(x_i + (1 - x_i))^{d-2} \prod_{j \in [n], j \neq i} (x_j + (1 - x_j))^d,$$

$$x_i(1 - x_i) = x_i(1 - x_i)(x_i + (1 - x_i))^{d-2} \prod_{j \in [n], j \neq i} (x_j + (1 - x_j))^d,$$

and that, when expanding in the basis  $\mathcal{B}_d$ , all coefficients remain nonnegative. Hence  $q_1 + q_2 = \sum_{k \in [d]_0^n} a_k P_{d,k}$  for some  $a_k \geq 0$  and thus

$$p = B_d(p) + q_1 + q_2 - C(d, p) = \sum_{k \in [d]_0^n} \left( p \binom{k}{d} + a_k - C(d, p) \right) P_{d,k}.$$

This implies  $p_{\text{Ber}}^{(d)} \geq p_{\min, Q(d)} - C(d, p)$  which, combined with (1.24), shows (iii).

To show (iv), use again the identity  $p = B_d(p) + q_1 + q_2 - C(d, p)$ , together with  $B_d(p) \geq p_{\min, Q(d)}$ ,  $q_1 \geq 0$  and  $q_2 \geq \frac{3}{4}C(d, p)$  on  $Q$ , which follows from the fact that  $(x_i - 1)^2 + x_i \geq \frac{3}{4}$  on  $[0, 1]$ .  $\square$

Compared to the results from Theorem 1.4 for the general case  $m \geq 1$ , note that we gain a factor  $n$  in the quadratic case  $m = 2$ . To see this, use the fact that  $A_{ii} \leq L(p)$  for all  $i$ , thus implying  $\sum_{i \in I_+} A_{ii} \leq nL(p)$ . We refer to Table 1.1 for a comparison with the results of Theorem 1.3.

We conclude this section with comparing our bound (2.5) with another bound that can be derived from the following known result for the minimization of quadratic forms (i.e. homogeneous polynomials) over the simplex.

**PROPOSITION 2.2** (cf. the proof of Theorem 3.2 in [1]; see also Theorem 3.2 in [3]). *Let  $q \in \mathbb{R}[x]$  be a form of degree 2,  $d \geq 1$  an integer,  $q_{\min, \Delta} := \min_{x \in \Delta} q(x)$ , and  $q_{\min, \Delta(d)} := \min_{x \in \Delta(d)} q(x)$ , where*

$$\Delta := \left\{ x \in \mathbb{R}^n, \sum_{i=1}^n x_i = 1, x_i \geq 0 \quad (i = 1, \dots, n) \right\}, \quad \Delta(d) := \{x \in \Delta \mid dx \in \mathbb{N}^n\}.$$

*are, respectively, the standard simplex and the set of rational points with denominator  $d$  in  $\Delta$ . Then,*

$$q_{\min, \Delta(d)} - q_{\min, \Delta} \leq \frac{1}{d} \left( \max_{i=1, \dots, n} q(e_i) - q_{\min, \Delta} \right).$$

One may apply Proposition 2.2 to the hypercube  $Q$ , by viewing  $Q$  as the convex hull of its vertices, i.e. by mapping  $Q$  to a simplex  $\Delta_N$  in  $\mathbb{R}^N$  ( $N := 2^n$ ).

**COROLLARY 2.3.** *Let  $p = x^T A x + b^T x + c$  be a polynomial of degree 2. Then,*

$$p_{\min, Q(d)} - p_{\min, Q} \leq \frac{1}{d} \left( \max_{x \in \{0,1\}^n} p(x) - p_{\min, Q} \right).$$

**Proof.** Consider the linear mapping  $\phi$  that maps  $y = (y_I)_{I \subseteq [n]} \in \mathbb{R}^N$  to  $\phi(y) = \sum_{I \subseteq [n]} y_I \chi^I \in \mathbb{R}^n$ , where  $\chi^I \in \{0, 1\}^n$  denotes the incidence vector of the subset  $I \subseteq [n]$ . Define the polynomial  $q$  in the variables  $y = (y_I)_{I \subseteq [n]}$  by  $q(y) = \phi(y)^T A \phi(y) + (\sum_I y_I) b^T \phi(y) + c$ . Thus  $q(y) = p(\phi(y))$  for  $y \in \Delta_N$ . Moreover,  $\phi(\Delta_N) = Q$ . Indeed,  $\phi(\Delta_N) \subseteq Q$  is obvious. Conversely, if  $x \in Q$  with say  $0 \leq x_1 \leq \dots \leq x_n \leq 1$ , then

$$x = x_1 \chi^{[n]} + (x_2 - x_1) \chi^{[n] \setminus \{1\}} + \dots + (x_n - x_{n-1}) \chi^{[n] \setminus \{1, \dots, n-1\}} + (1 - x_n) \chi^\emptyset,$$

showing  $x \in \phi(\Delta_N)$ . This also shows  $\phi(\Delta_N(d)) = Q(d)$ . Therefore,  $p_{\min, Q} = q_{\min, \Delta_N}$  and  $p_{\min, Q(d)} = q_{\min, \Delta_N(d)}$ . Thus the corollary follows from Proposition 2.2, as  $\max_{I \subseteq [n]} q(e_I) = \max_{x \in \{0, 1\}^n} p(x)$ .  $\square \square$

We now show that our new bound (2.5) dominates the bound from Corollary 2.3.

PROPOSITION 2.4. *Let  $p = x^T A x + b^T x + c$  be a polynomial of degree 2. Then,*

$$\frac{1}{4} \sum_{i \in I_+} A_{ii} \leq \max_{x \in \{0, 1\}^n} p(x) - p_{\min, Q}.$$

**Proof.** Assume first that  $I_+ = [n]$ . Let  $x \in \{0, 1\}^n$  be a global maximizer of  $p$  over  $\{0, 1\}^n$  and set  $S := \{i \in [n] \mid x_i = 1\}$ ,  $T = [n] \setminus S$ . Then,  $p(x - e_i) \leq p(x)$  for all  $i \in S$  and  $p(x + e_i) \leq p(x)$  for all  $i \in T$ . This implies  $A_{ii} - b_i - 2x^T A e_i \leq 0$  for  $i \in S$  and  $A_{ii} + b_i + 2x^T A e_i \leq 0$  for  $i \in T$ . Summing over  $S$  and over  $T$  we obtain:

$$\sum_{i \in S} A_{ii} - b^T x - 2x^T A x \leq 0, \quad \sum_{i \in T} A_{ii} + b^T (e - x) + 2x^T A (e - x) \leq 0.$$

Summing these two relations implies  $\sum_{i \in [n]} A_{ii} \leq 4x^T A x + 2b^T x - b^T e - 2x^T A e$  and thus

$$\frac{1}{4} \sum_{i \in [n]} A_{ii} \leq \underbrace{x^T A x + \frac{1}{2} b^T x - \frac{1}{4} b^T e - \frac{1}{2} x^T A e}_{:= \text{RHS}}.$$

One can verify that  $\text{RHS} = \frac{1}{4} p(e - x) + \frac{3}{4} p(x) - p(e/2)$ , which implies  $\text{RHS} \leq p(x) - p_{\min, Q}$  and concludes the proof in the case when  $I_+ = [n]$ . Suppose now  $I_+ \subseteq [n]$ . Define the polynomial  $q$  in the variables  $y = (x_i \mid i \in I_+)$  by  $q(y) := p(y, 0, \dots, 0) = y^T A_+ y + b_+^T y$ , where  $A_+$  is the principal submatrix of  $A$  indexed by  $I_+$  and  $b_+ := (b_i \mid i \in I_+)$ . The above argument shows that  $\frac{1}{4} \sum_{i \in I_+} A_{ii} \leq \max_{y \in \{0, 1\}^{I_+}} q(y) - q_{\min, [0, 1]^{I_+}}$ . The result now follows after observing that  $\max_{y \in \{0, 1\}^{I_+}} q(y) \leq \max_{x \in \{0, 1\}^n} p(x)$  and  $\min_{y \in [0, 1]^{I_+}} q(y) \geq \min_{x \in [0, 1]^n} p(x)$ .  $\square \square$

**3. Results for polynomials of higher degree .** We now show how the results from the preceding section for quadratic polynomials extend to polynomials with an arbitrary degree  $m \geq 1$ . The basic idea is the same; namely we will establish an identity analogous to (2.3) using Bernstein approximations. The technical details are however a bit more involved and we will only work with bounds on the constant  $C(d, p)$  appearing in (2.3), not an explicit expression as in the quadratic case.

We start with an easy, but useful result, which shows how to express any term  $-\lambda x^h (1 - x)^k$  with  $\lambda > 0$  and of degree  $t$ , as  $-\lambda + q$  for some  $q \in H_t(g)$ .

LEMMA 3.1. *For any  $h, k \in \mathbb{N}^n$ ,  $-\lambda x^h (1 - x)^k + 1 \in H_t(g)$ , where  $t := |h + k| = \deg(x^h (1 - x)^k)$ .* **Proof.** The proof is by induction on the number  $n$  of variables.

First we show the result in the univariate case  $n = 1$ ; say  $x = x_1$  is a univariate variable. We use induction on  $h+k$ . If  $h+k = 0$  there is nothing to prove. Let  $h+k \geq 1$ . If  $h \geq 1$ , then  $-x^h(1-x)^k = -x \cdot x^{h-1}(1-x)^k = (1-x) \cdot x^{h-1}(1-x)^k - x^{h-1}(1-x)^k = x^{h-1}(1-x)^{k+1} - x^{h-1}(1-x)^k$  and we can conclude using the induction assumption applied to the term  $-x^{h-1}(1-x)^k$ . If  $h = 0$ , then  $-(1-x)^k = -(1-x) \cdot (1-x)^{k-1} = x(1-x)^{k-1} - (1-x)^{k-1}$  and again conclude using induction applied to the term  $-(1-x)^{k-1}$ . We now consider the multivariate case  $n \geq 2$ . We have just proved that  $-x_1^{h_1}(1-x_1)^{k_1} = -1 + \sum_{r_1, s_1 \in \mathbb{N} | r_1 + s_1 \leq h_1 + k_1} \lambda_{r_1, s_1} x_1^{r_1} (1-x_1)^{s_1}$  with  $\lambda_{r_1, s_1} \geq 0$ . Hence  $-x^h(1-x)^k = -\prod_{i=2}^n x_i^{h_i} (1-x_i)^{k_i} + \sum_{r_1, s_1 | r_1 + s_1 \leq h_1 + k_1} \lambda_{r_1, s_1} x_1^{r_1} (1-x_1)^{s_1} \prod_{i=2}^n x_i^{h_i} (1-x_i)^{k_i}$  and the result follows using the induction assumption for the case  $n-1$ .  $\square$

As in the quadratic case, our strategy is now to write  $p$  as

$$p = B_d(p - p_{\min}) + q + p_{\min, Q} - C(d, p), \quad (3.1)$$

where  $q \in H_t(g)$  (for some suitable  $t$ ), and  $C(d, p)$  is a constant which depends only on  $p$  and  $d$ . It is useful to consider first the univariate case, whose treatment will be used afterwards in the multivariate case.

**3.1. The univariate case.** We first establish some facts on the Bernstein approximation of monomials. Recall the Bernstein approximation of a monomial:  $B_d(x^k) = \frac{1}{d^k} \sum_{i=0}^k b_{k,i} d^i x^i$ , from property (1.18), where  $d^i = d(d-1)\cdots(d-i+1)$ ,  $b_{k,i} > 0$  and  $b_{k,k} = 1$ . Moreover,

$$\frac{1}{d^k} \sum_{i=0}^k b_{k,i} d^i = 1, \quad (3.2)$$

which follows from the fact that  $B_d(f)(1) = f(1) = 1$  for  $f(x) = x^k$ .

LEMMA 3.2.

- (i)  $d^k = d^k - \binom{k}{2} d^{k-1} + \sum_{i=0}^{k-2} (-1)^{k-i} \gamma_i^{(k)} d^i$ , for some scalars  $\gamma_i^{(k)}$  satisfying  $\gamma_i^{(k)} \geq 0$  and  $1 - \binom{k}{2} + \sum_{i=0}^{k-2} (-1)^{k-i} \gamma_i^{(k)} = 0$ .
- (ii)  $B_d(x^k) = x^k + \sum_{i=0}^k a_i^{(k)} x^i$ , for some scalars  $a_i^{(k)}$  satisfying  $-\frac{1}{d} \binom{k}{2} \leq a_k^{(k)} \leq 0$ ,  $0 \leq a_i^{(k)}$  for  $i \leq k-1$ , and  $\sum_{i=0}^{k-1} a_i^{(k)} \leq \frac{1}{d} \binom{k}{2}$ .

**Proof.** (i) follows by expanding the univariate polynomial  $x(x-1)\cdots(x-k+1)$ . (ii) By construction, we have  $a_i^{(k)} = \frac{d^i}{d^k} b_{k,i} > 0$  for  $i \leq k-1$ , and  $a_k^{(k)} = \frac{d^k}{d^k} b_{k,k} - 1 = \frac{d^k}{d^k} - 1$ . Using (3.2) combined with (i), we obtain:  $1 = \sum_{i=0}^k b_{k,i} \frac{d^i}{d^k} = \sum_{i=0}^{k-1} a_i^{(k)} + \frac{d^k}{d^k}$ , and thus  $\sum_{i=0}^{k-1} a_i^{(k)} = -a_k^{(k)} = 1 - \frac{d^k}{d^k} = \frac{1}{d} \binom{k}{2} - \frac{1}{d^k} \left( \sum_{i=0}^{k-2} (-1)^{k-i} \gamma_i^{(k)} d^i \right)$ . Hence it suffices now to verify that  $\sum_{i=0}^{k-2} (-1)^{k-i} \gamma_i^{(k)} d^i \geq 0$ . Using again (i), we find that  $\sum_{i=0}^{k-2} (-1)^{k-i} \gamma_i^{(k)} d^i = d^k - d^k + \binom{k}{2} d^{k-1}$ , which can be easily verified to be nonnegative using induction on  $k$ .  $\square$

Let  $p = \sum_{k=0}^m p_k x^k$  be a univariate polynomial of degree  $m$ . Then, we have

$$p - B_d(p) = \sum_{k=0}^m p_k (x^k - B_d(x^k)) = - \sum_{k=0}^m \sum_{i=0}^k p_k a_i^{(k)} x^i.$$

Now we split the sum depending on the signs of  $p_k$  and of  $a_i^{(k)}$  (negative for  $i = k$  and positive for  $i \leq k-1$ ) and we use Lemma 3.1 to write

$$p - B_d(p) = q - \underbrace{\left( \sum_{k \geq 1 |p_k \geq 0} p_k \sum_{i=0}^{k-1} a_i^{(k)} + \sum_{k \geq 0 |p_k < 0} |p_k| |a_k^{(k)}| \right)}_{:=C(d,p)},$$

where  $q \in H_m(g)$ . Next, since  $|p_k| \leq L(p)$  and, by Lemma 3.2 (ii),  $|a_k^{(k)}|, \sum_{i=0}^{k-1} a_i^{(k)} \leq \frac{1}{d} \binom{k}{2}$ , we can bound  $C(d, p)$  as follows:

$$C(d, p) \leq \frac{1}{d} L(p) \sum_{k=0}^m \binom{k}{2} = \frac{1}{d} L(p) \binom{m+1}{3}. \quad (3.3)$$

Thus we have shown:

**PROPOSITION 3.3.** *Let  $p$  be a univariate polynomial of degree  $m$ ,  $Q = [0, 1]$ , and  $g$  stand for the polynomials  $x, 1 - x$ . For any integer  $d \geq 1$ , we have*

$$p = B_d(p - p_{\min, Q}) + q + p_{\min, Q} - C(d, p), \quad (3.4)$$

where  $q \in H_m(g)$ ,  $B_d(p - p_{\min, Q}) \in H_d(g)$ , and  $C(d, p)$  is a scalar satisfying  $C(d, p) \leq \frac{1}{d} L(p) \binom{m+1}{3}$ . If  $p$  is positive on  $[0, 1]$  and if  $d \geq \left\lceil \frac{L(p)}{p_{\min, Q}} \binom{m+1}{3} \right\rceil$ , then  $p_{\min, Q} - C(d, p) \geq 0$  and thus (3.4) is a decomposition of Handelman type with degree  $d$ .

**3.2. The multivariate case.** In the multivariate case we have to deal with the terms  $x^k - B_d(x^k)$ , where  $x^k = \prod_{i=1}^n x_i^{k_i}$  is a multivariate monomial and  $B_d(x^k) = \prod_{i=1}^n B_d(x_i^{k_i})$ . For this we use the identity:

$$a_1 \cdots a_n - b_1 \cdots b_n = \sum_{j=1}^n b_1 \cdots b_{j-1} a_{j+1} \cdots a_n (a_j - b_j)$$

and write:

$$\prod_{j=1}^n x_j^{k_j} - \prod_{j=1}^n B_d(x_j^{k_j}) = \sum_{j=1}^n B_d(x_1^{k_1}) \cdots B_d(x_{j-1}^{k_{j-1}}) x_{j+1}^{k_{j+1}} \cdots x_n^{k_n} (x_j^{k_j} - B_d(x_j^{k_j})).$$

Let  $p = \sum_{k \in \mathbb{N}^n, |k| \leq m} p_k x^k$  be a polynomial of degree  $m$ . Fix an integer  $d \geq 1$ . We have:

$$p - B_d(p) = \sum_{k, |k| \leq m} p_k (x^k - B_d(x^k)) = \sum_{k, |k| \leq m} p_k \left( \sum_{j=1}^n \underbrace{B_d(x_1^{k_1}) \cdots B_d(x_{j-1}^{k_{j-1}}) x_{j+1}^{k_{j+1}} \cdots x_n^{k_n}}_{:=q_{k,j}} (x_j^{k_j} - B_d(x_j^{k_j})) \right).$$

Here, each polynomial  $q_{k,j}$  has degree  $|k| - k_j$  and belongs to the set  $H_{|k|-k_j}(g)$ . Moreover, in view of (3.2), each  $q_{k,j}$  can be written as  $\sum_{h,k} \lambda_{h,k} x^h (1-x)^k$  for some  $\lambda_{h,k} \geq 0$  satisfying  $\sum_{h,k} \lambda_{h,k} = 1$ . Next, applying Lemma 3.2 (ii), we can reformulate each  $x_j^{k_j} - B_d(x_j^{k_j})$  and write:

$$p - B_d(p) = - \sum_{k, |k| \leq m} p_k \left( \sum_{j=1}^n q_{k,j} \left( \sum_{i_j=0}^{k_j} a_{i_j}^{(k_j)} x_j^{i_j} \right) \right).$$

As in the univariate case, we split the sum depending on the signs of  $p_k$  and of  $a_{i_j}^{(k_j)}$ . Then, using Lemma 3.1 and the fact that each  $q_{k,j}$  can be written in  $H_{|k|-k_j}(g)$  with coefficients summing up to 1, we obtain:

$$p - B_d(p) = q - \underbrace{\sum_{k|p_k < 0} |p_k| \left( \sum_{j=1}^n |a_{k_j}^{(k_j)}| \right) + \sum_{k|p_k > 0} p_k \left( \sum_{j=1}^n \sum_{i_j=0}^{k_j-1} a_{i_j}^{(k_j)} \right)}_{:=C(d,p)}, \quad (3.5)$$

where  $q \in H_m(g)$ . We now bound the constant  $C(d,p)$ . As in the univariate case, we use the fact that  $|a_{k_j}^{(k_j)}|, \sum_{i_j=0}^{k_j-1} a_{i_j}^{(k_j)} \leq \frac{1}{d} \binom{k_j}{2}$ , combined with  $|p_k| \leq L(p) \frac{|k|!}{k!}$ , to derive:

$$C(d,p) \leq \frac{L(p)}{d} \left( \sum_{k||k| \leq m} \frac{|k|!}{k!} \left( \sum_{j=1}^n \binom{k_j}{2} \right) \right).$$

Next, using the fact that  $\sum_{j=1}^n \binom{k_j}{2} \leq \binom{|k|}{2}$  and  $\sum_{k \in \mathbb{N}^n, |k|=l} \frac{|k|!}{k!} = n^l$ , we obtain:

$$\sum_{k||k| \leq m} \frac{|k|!}{k!} \left( \sum_{j=1}^n \binom{k_j}{2} \right) \leq \sum_{l=0}^m n^l \binom{l}{2} \leq n^m \sum_{l=0}^m \binom{l}{2} = n^m \binom{m+1}{3}.$$

Therefore,

$$C(d,p) \leq \frac{L(p)}{d} \binom{m+1}{3} n^m. \quad (3.6)$$

Hence we have obtained a decomposition of the form (3.1), where  $q \in H_m(g)$  and  $C(d,p)$  is a constant which is bounded as in (3.6). We can now formulate our main result, the analogue of Theorem 2.1 for the quadratic case.

**THEOREM 3.4.** *Let  $p$  be a polynomial of degree  $m$ ,  $Q = [0, 1]^n$ , and  $g$  as in (1.5).*

- (i) *For any integer  $d \geq 1$ ,  $p - p_{\min,Q} + \frac{L(p)}{d} \binom{m+1}{3} n^m \in H_r(g)$  for some integer  $r \leq \max(dn, m)$ .*
- (ii) *If  $p$  is positive on  $Q$ , then  $p \in H_r(g)$  for some integer  $r \leq n \left\lceil \frac{L(p)}{p_{\min,Q}} \binom{m+1}{3} n^m \right\rceil$ .*
- (iii) *For any integer  $d \geq m$ ,  $\max(p_{\min,Q} - p_{Ber}^{(d)}, p_{\min,Q(d)} - p_{\min,Q}) \leq p_{\min,Q(d)} - p_{Ber}^{(d)} \leq \frac{L(p)}{d} \binom{m+1}{3} n^m$ .*

**Proof.** (i) follows directly from (3.5), (3.6), and (ii) follows by choosing for  $d$  the smallest integer for which  $p_{\min,Q} - \frac{L(p)}{d} n^m \binom{m+1}{3} \geq 0$ . The proof for (iii) is similar to that of Theorem 2.1 (iii) and is thus omitted.  $\square \square$

**4. Concluding remarks.** We have seen how Bernstein approximations can be used for obtaining stronger degree and error bounds for the semidefinite programming approximations obtained by applying Schmüdgen's Positivstellensatz to the hypercube. This approach via Bernstein approximations is also suited for optimization over the simplex and, although not explicitly mentioned there, it underlies some of the results of de Klerk *et al.* [3]. One of the main results of [3] is that the problem of computing the minimum of an  $n$ -variate form  $q$  over the simplex  $\Delta$  admits a polynomial time approximation scheme when fixing the degree  $m$  of  $q$ . This PTAS is provided simply by computing the minimum  $q_{\min,\Delta(d)}$  of  $q$  over the set of rational

points in  $\Delta$  with given denominator  $d$ . Indeed, the following inequality is shown in [3]: For any integer  $d \geq 1$ ,

$$q_{\min, \Delta(d)} - q_{\min, \Delta} \leq \frac{1}{d} c_m (q_{\max, \Delta} - q_{\min, \Delta}), \quad (4.1)$$

where  $c_m$  is a constant depending only on the degree  $m$  of  $q$ , and  $q_{\max, \Delta} := \max_{x \in \Delta} q(x)$ ; moreover, one can compute  $q_{\min, \Delta(d)}$  in polynomial time, simply by enumeration since  $|\Delta(d)| = O(n^d)$ . Note that  $q_{\min, \Delta(d)}$  is the analogue of the parameter  $p_{\min, Q(d)}$  introduced in (1.14); however, the cardinality of  $Q(d)$  is exponential in the number of variables and computing  $p_{\min, Q(d)}$  is a hard problem. The inapproximability results for the maximum cut and stable set problems described in the introduction show that no polynomial time approximation scheme can exist for polynomial optimization over the hypercube if  $P \neq NP$ . (See the recent survey [2] for the relevant definitions and detailed complexity results.) The fact that our error bounds apply only to the SDP relaxations of order  $r = \Omega(n)$  (namely,  $r = dn$ , for  $d$  integer), should be seen in this light.

Using the correspondence between the hypercube  $Q$  and the simplex in the space  $\mathbb{R}^{2^n}$  introduced in the proof of Corollary 2.3, one can derive directly from (4.1) the following analogous error estimate for the parameter  $p_{\min, Q(d)}$ :

$$p_{\min, Q(d)} - p_{\min, Q} \leq \frac{1}{d} c_m (p_{\max, Q} - p_{\min, Q}), \quad (4.2)$$

where  $c_m$  is the constant from (4.1) depending only on the degree  $m$  of  $p$ . It would be interesting to see whether there is an analogous error estimate for the parameters  $p_{\text{Ber}}^{(d)}$  or  $p_{\text{han}, g}^{(r)}$ .

As mentioned in the introduction, the semidefinite bound  $p_{\text{sch}, g}^{(r)}$  based on Schmüdgen's Positivstellensatz can be computed by a semidefinite program involving  $2^s$  matrices constrained to be positive semidefinite, thus it can be efficiently computed only when fixing the order  $r$  of the relaxations and the number  $s$  of polynomial inequalities describing  $S$ . For this reason Lasserre [13] considers semidefinite approximations based on the Positivstellensatz of Putinar [24], which claims that (under a mild assumption on the polynomials  $g_j$  describing  $S$ ) every positive polynomial on  $S$  belongs to

$$M(g) := \left\{ \sigma_0 + \sum_{j=1}^m \sigma_j g_j \mid \sigma_0, \sigma_j \text{ are sums of squares of polynomials} \right\},$$

the *quadratic module* generated by the  $g_j$ 's. When bounding the degrees, i.e. when considering the truncated quadratic module  $M_r(g)$  where  $\deg(\sigma_0), \deg(\sigma_j g_j) \leq r$ , the corresponding bound

$$p_{\text{put}, g}^{(r)} := \sup \left\{ t \mid p - t \in M_r(g) \right\}$$

can be computed with a semidefinite program involving only  $s + 1$  semidefinite matrices, thus efficiently when fixing the order  $r$  of the relaxation. Nie and Schweighofer [21] give degree and error bounds for these approximations which, analogously to those in Theorem 1.3, also depend on some unknown constant (although there is now an additional exponential dependency). It would be very interesting to give explicit degree and error bounds for Putinar type representations in the case of the hypercube. In some recent work, Nie [5] gives error bounds of the form  $c(p_{\max, S} - p_{\min, S})$  for the

Putinar type approximation  $p_{\text{put},g}^{(r)}$  and  $S$  as in (1.4); for the hypercube, the constant  $c$  is of the form  $\Omega(n^r)$  for  $r \geq \deg(p)$  and thus the error bounds do not tend to zero as  $r \rightarrow \infty$ .

One possible way for giving explicit degree and error bounds for Putinar type representations on the hypercube would be to relate the quadratic module  $M_r(g)$  and the preordering  $T_r(g)$ , when considering the polynomials  $g_i := x_i - x_i^2$  ( $i = 1, \dots, n$ ) describing the hypercube. Obviously  $M_r(g) \subseteq T_r(g)$ . The reverse inclusion is not true since the monomial  $\prod_{i=1}^s x_i$  does not belong to  $M(g)$  (for  $1 < s \leq n$ ). However, this monomial belongs to  $M(g)$  after adding a suitable constant. Namely, we can show the following: For  $n$  even,

$$\prod_{i=1}^n x_i + C_n \in M_n(g) \quad (4.3)$$

for some constant  $C_n \leq 1$ .

Here are two arguments why this is true. The first one relies on showing that the polynomial

$$\prod_{i=1}^n x_i + 1 + \sum_{i=1}^n (x_i^2 - x_i)x_i^{n-2}$$

is a sum of squares of polynomials (which follows directly applying [4, Corollary 2.2]). The second argument (communicated to us by V. Powers and B. Reznick) relies on the following identity:

$$\prod_{i=1}^n x_i + 1 = \prod_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n x_i^n + \frac{1}{n} \sum_{i=1}^n (1 - x_i^n).$$

Then note that the polynomial  $\prod_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n x_i^n$  is a sum of squares of polynomials (shown by Hurwitz, see [25], or use [4, Theorem 2.1]) and note that each polynomial  $1 - x_i^n$  lies in  $M_n(x_i - x_i^2) \subseteq M_n(g)$ , since it can be written as  $1 - x_i^n = (1 - x_i)^2 + (x_i - x_i^2)(2 + x_i + x_i^2 + \dots + x_i^{n-2})$  where the univariate polynomial  $2 + x_i + x_i^2 + \dots + x_i^{n-2} = 1 + \frac{x_i^{n-1} - 1}{x_i - 1}$  is nonnegative and thus a sum of squares.

We conjecture that for  $n$  even the smallest constant  $C_n$  for which (4.3) holds is  $C_n = 1/n(n+2)$ . We verified that this is true for  $n = 2, 4, 6$  (using computer for  $n = 4, 6$ ). For  $n = 2$  the identity

$$x_1 x_2 + \frac{1}{8} = \frac{1}{2} \left( x_1 + x_2 - \frac{1}{2} \right)^2 + \frac{1}{2} (x_1 - x_1^2) + \frac{1}{2} (x_2 - x_2^2)$$

shows  $C_2 \leq 1/8$ .

The idea is that, using (4.3), one may easily show

$$B_d(p - p_{\min,Q}) + (p_{\max,Q} - p_{\min,Q})C_r 2^{nd} \in M_r(g) \quad \text{for even } r \geq nd. \quad (4.4)$$

This in turn allows us to derive error bounds for the Lasserre hierarchy of approximations. For example, if  $p$  is a quadratic polynomial, we may use (2.3) and (4.4) to obtain:

$$p - \left( p_{\min,Q} - \frac{1}{d} \sum_{i \in I_+} A_{ii} - (p_{\max,Q} - p_{\min,Q})C_r 2^{nd} \right) \in M_r(g) \quad \text{for even } r \geq nd.$$

If the conjecture  $C_r = \frac{1}{r(r+2)}$  holds, we may set  $r = 2^{nd}$  for any given integer  $d \geq 1$  to obtain

$$p - \left( p_{\min, Q} - \frac{n}{\log_2 r} \sum_{i \in I_+} A_{ii} - \frac{p_{\max, Q} - p_{\min, Q}}{r + 2} \right) \in M_r(g) \quad \text{for } r = 2^{nd}, d \geq 1.$$

Thus we obtain the following error bound for the Lasserre hierarchy:

$$p_{\min, Q} - p_{\text{put}, g}^{(r)} \leq \frac{n}{\log_2 r} \sum_{i \in I_+} A_{ii} + \frac{p_{\max, Q} - p_{\min, Q}}{r + 2} \quad \text{for } r = 2^{nd}, d \geq 1.$$

**Acknowledgements.** The authors are indebted to Bruce Reznick for pointing out a mistake in an earlier version of this paper, and would also like to thank him, Marianna Nagy, Vicki Powers, and Markus Schweighofer for interesting discussions on the conjecture  $C_n = 1/n(n+2)$  in (4.3).

#### REFERENCES

- [1] 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.
- [2] E. de Klerk. The complexity of optimizing over a simplex, hypercube or sphere: A short survey. *Central European Journal of Operations Research*, **16(2)**, 111–125, 2008.
- [3] E. de Klerk, M. Laurent, P. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, **361(2-3)**:210–225, 2006.
- [4] M. Ghasemi and M. Marshall. Lower bounds for a polynomial in terms of its coefficients. Preprint, 2010.
- [5] J. Nie. An approximation bound analysis for Lasserre’s relaxation in multivariate polynomial optimization. Preprint, 2009.
- [6] J. Nie, J. Demmel, and B. Sturmfels. Minimizing polynomials via sums of squares over the gradient ideal. *Mathematical Programming Series A*, **106**:587–606, 2006.
- [7] 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.
- [8] D. Handelman. Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific Journal of Mathematics*, **132**:35–62, 1988.
- [9] J. Håstad. Clique is hard to approximate within  $|V|^{1-\epsilon}$ . *Acta Mathematica*, 182, 105–142, 1999.
- [10] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, **48**:798–859, 2001.
- [11] D. Henrion and J.-B. Lasserre. GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi. *ACM Transactions Math. Soft.*, **29**:165194, 2003.
- [12] A. Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM Journal on Applied Mathematics*, **69(1)**:197–204, 2008.
- [13] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, **11**:296–817, 2001.
- [14] J.B. Lasserre. An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM Journal on Optimization*, **12**:756–769, 2002.
- [15] J. B. Lasserre. Semidefinite programming vs. LP relaxations for polynomial programming. *Mathematics of Operations Research*, **27**:347–360, 2002.
- [16] J.B. Lasserre. Convexity in Semialgebraic geometry and polynomial optimization. *SIAM Journal on Optimization*, **19**:1995–2014, 2009.
- [17] M. Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, **109**:1–26, 2007.
- [18] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging Applications of Algebraic Geometry*, Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 157–270, 2009.
- [19] G.G. Lorentz. *Bernstein Polynomials*, 2nd ed., Chelsea, 1986.



- [20] Yu. Nesterov, H. Wolkowicz, and Y. Ye. Semidefinite programming relaxations of nonconvex quadratic optimization. In *Handbook of semidefinite programming*, H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.), pages 361–419. Kluwer Academic Publishers, Norwell, MA, 2000.
- [21] J. Nie and M. Schweighofer. On the complexity of Putinar’s Positivstellensatz. *Journal of Complexity*, **23**(1):135–150, 2007.
- [22] P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology, 2000.
- [23] V. Powers and B. Reznick. A new bound for Pólya’s theorem with applications to polynomials positive on polyhedra. *Journal of pure and Applied Algebra*, **164**:221–229, 2001.
- [24] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, **42**:969–984, 1993.
- [25] B. Reznick. A quantitative version of Hurwitz’ theorem on the arithmetic-geometric inequality. *J. Reine Angew. Math.*, **377**:108–112, 1987.
- [26] K. Schmüdgen. The  $K$ -moment problem for compact semi-algebraic sets. *Mathematische Annalen*, **289** (2):203–206, 1991.
- [27] M. Schweighofer. On the complexity of Schmüdgen’s Positivstellensatz. *Journal of Complexity*, **20**:529–543, 2004.
- [28] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization*, **15**(3):805–825, 2005.
- [29] D. Zuckerman: Linear degree extractors and the inapproximability of max clique and chromatic number, *Proc. 38th ACM Symp. Theory of Computing*, 681-690, 2006.