

---

# THE LINK BETWEEN 1-NORM APPROXIMATION AND EFFECTIVE POSITIVSTELLENSATZE FOR THE HYPERCUBE

---

Etienne de Klerk \*

Juan C. Vera †

December 19, 2024

**Abstract**

The Schmüdgen Positivstellensatz gives a certificate to verify positivity of a strictly positive polynomial  $f$  on a compact, basic, semi-algebraic set  $\mathbf{K} \subset \mathbb{R}^n$ . A Positivstellensatz of this type is called *effective* if one may bound the degrees of the polynomials appearing in the certificate in terms of properties of  $f$ . If  $\mathbf{K} = [-1, 1]^n$  and  $0 < f_{\min} := \min_{x \in \mathbf{K}} f(x)$ , then the degrees of the polynomials appearing in the certificate may be bounded by  $O\left(\sqrt{\frac{f_{\max} - f_{\min}}{f_{\min}}}\right)$ , where  $f_{\max} := \max_{x \in \mathbf{K}} f(x)$ , as was recently shown by Laurent and Slot [*Optimization Letters* 17:515–530, 2023]. The big-O notation suppresses dependence on  $n$  and the degree  $d$  of  $f$ . In this paper we show a similar result, but with a better dependence on  $n$  and  $d$ . In particular, our bounds depend on the 1-norm of the coefficients of  $f$ , that may readily be calculated.

**Keywords** Polynomial kernel method · semidefinite programming · Positivstellensatz

**1 Introduction**

We first recall a *Positivstellensatz* of Schmüdgen [26], which gives representations for positive polynomials on a basic, closed, semi-algebraic set; see also, e.g., [21],[16, Theorem 3.16], [18].

**Theorem 1** (Schmüdgen [26]). *Consider the set  $\mathbf{K} = \{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) \geq 0, \dots, g_m(\mathbf{x}) \geq 0\}$  with  $g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$ , where  $\mathbb{R}[\mathbf{x}]$  denotes the ring of real polynomials with variables  $\mathbf{x} = (x_1, \dots, x_n)$ . Assume that  $\mathbf{K}$  is compact. If  $p \in \mathbb{R}[x]$  is positive on  $\mathbf{K}$ , then  $p$  can be written as*

$$p = \sum_{I \subseteq [m]} \sigma_I \prod_{i \in I} g_i, \quad (1)$$

where  $[m] := \{1, \dots, m\}$ , and each  $\sigma_I$  ( $I \subseteq [m]$ ) belongs to the cone of sums-of-squares of polynomials, denoted by  $\Sigma[\mathbf{x}]$ .

A polynomial that allows a representation of the type (1) is said to lie in the pre-ordering associated with the (generating) polynomials  $g_1, \dots, g_m$ , denoted by  $\mathcal{T}(g_1, \dots, g_m)$ .

We denote the real polynomials of degree at most  $d$  by  $\mathbb{R}[\mathbf{x}]_d$ , and define the truncated pre-ordering of degree  $r \in \mathbb{N}$  by

$$\mathcal{T}(g_1, \dots, g_m)_r := \left\{ p \in \mathbb{R}[\mathbf{x}] : p = \sum_{I \subseteq [m]} \sigma_I \prod_{i \in I} g_i, \deg\left(\sigma_I \prod_{i \in I} g_i\right) \leq r, \sigma_I \in \Sigma[x] \forall I \in [m] \right\}. \quad (2)$$

For the hypercube  $\mathbf{K} = [-1, 1]^n$ , described by the polynomial inequalities  $1 - x_1^2 \geq 0, \dots, 1 - x_n^2 \geq 0$ , we consider polynomials that allow a Schmüdgen-type representation of bounded degree  $r$ :

$$p(\mathbf{x}) = \sum_{I \subseteq [n]} \sigma_I(\mathbf{x}) \prod_{i \in I} (1 - x_i^2) \in \mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_r, \quad (3)$$

---

\*Tilburg University, e.deklerk@tilburguniversity.edu

†Tilburg University, j.c.veralizcano@tilburguniversity.edu

where the polynomials  $\sigma_I$  are sum-of-squares polynomials with degree at most  $r - 2|I|$  (to ensure that the degree of each term in the sum in (2) is at most  $r$ ).

We call a *Positivstellensatz* effective if one may bound the degree  $r$  in terms of properties of  $p$ . The best bound of this type is due to Slot and Laurent [17] (with an alternative proof by Bach and Rudi [3]).

**Theorem 2** (Laurent and Slot [17]). *Assume  $f \in \mathbb{R}[x]_d$  and  $f_{\min} := \min_{\mathbf{x} \in [-1,1]^n} f(\mathbf{x}) > 0$ . Then*

$$f \in \mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_{(r+1)n}$$

provided that

$$r \geq \max \left\{ \pi d \sqrt{2n}, \frac{1}{\sqrt{f_{\min}}} \sqrt{C(n, d)(f_{\max} - f_{\min})} \right\},$$

where  $C(n, d) > 0$  depends only on  $n$  and  $d$ , and may be chosen to be polynomial in  $n$  for fixed  $d$ , or polynomial in  $d$  for fixed  $n$ . In particular

$$C(n, d) \leq 2\pi^2 d^2 n 2^{n/2} (d+1)^n \text{ and } C(n, d) \leq 2\pi^2 d^2 n 2^{d/2} (n+1)^d.$$

One may formulate this result in a different way after introducing the following semidefinite programming (SDP) approximation of  $f_{\min} := \min_{\mathbf{x} \in [-1,1]^n} f(\mathbf{x})$ :

$$f_{(r)} := \sup \{ \lambda \mid f - \lambda \in \mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_r \} \leq f_{\min}. \quad (4)$$

In particular, one has the following result.

**Theorem 3** (Laurent and Slot [17]). *One has*

$$f_{\min} - f_{((r+1)n)} \leq \frac{1}{r^2} C(n, d)(f_{\max} - f_{\min}) \quad \forall r \geq \pi d \sqrt{2n},$$

where  $C(n, d)$  is as in Theorem 2.

## Contributions

In this paper we prove similar results to Theorems 2 and 3. We will consider a stronger SDP relaxation than (4), by considering the truncated pre-ordering

$$\mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r := \mathcal{T}(1 + x_1, 1 - x_1, \dots, 1 + x_n, 1 - x_n)_r$$

instead of  $\mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_r$ . As noted by, e.g. Powers and Reznick [20, p. 4682] and Baldi and Slot [4, §6], these two truncated pre-orderings are closely related through the identities:

$$\begin{aligned} 1 \pm x_i &= \frac{1}{2} \left( (1 \pm x_i)^2 + 1 - x_i^2 \right) \\ 1 - x_i^2 &= (1 - x_i)(1 + x_i). \end{aligned}$$

In particular, the identities imply

$$\mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_r \subset \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r \subset \mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)_{2r}. \quad (5)$$

Of course, without truncation, the two pre-orderings coincide, i.e.  $\mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n) = \mathcal{T}(1 - x_1^2, \dots, 1 - x_n^2)$ .

Thus we consider the SDP lower bound on  $f_{\min}$  given by

$$\hat{f}_{(r)} := \sup \{ \lambda \mid f - \lambda \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r \} \geq f_{(r)}. \quad (6)$$

Due to the relation (5) between the two truncated pre-orderings, one has  $f_{(r)} \leq \hat{f}_{(r)} \leq f_{(2r)}$ .

We will show (see Theorem 9 and Proposition 2) that

$$f_{\min} - \hat{f}_{(rn)} \leq \|f\|_{1,T} \frac{\pi^2 d^2}{(r+2)^2} \quad \forall r \geq \pi d, \quad (7)$$

where  $\|f\|_{1,T}$  is the sum of the absolute values of the coefficients of  $f$  in the Chebyshev basis, i.e. the 1-norm of the coefficients in this basis. This result is clearly of the same flavor as Theorem 3, but the right-hand-side may be computed in an elementary way, and does not explicitly depend on  $n$ . Moreover, the norm  $\|f\|_{1,T}$  can be of the same order as  $(f_{\max} - f_{\min})$  for some classes of polynomials. In particular, as discussed in Section 2.3, this is the case for the standard polynomial optimization reformulation of the maximum cut problem in graphs. Furthermore, by comparing different norms of polynomials of bounded degrees, one may use our result to derive a result very similar to Theorem 3.

We will also show an upper bound of the form  $f_{\min} - \hat{f}_{(r)} \leq \Theta_{n,d}^r \|f\|_{1,T}$ , that is valid for all  $r$  (not only multiples of  $n$ ), where the quantity  $\Theta_{n,d}^r$  depends on  $(r, n, d)$  only, and may be upper bounded by solving an SDP problem; see Theorem 9. Thus an upper bound on  $\Theta_{n,d}^r \|f\|_{1,T}$  may readily be computed if tables of SDP upper bounds on  $\Theta_{n,d}^r$  are given. We give such (partial) tables in an appendix to this paper for small values of  $n$ .

---

## Outline

Our paper is structured as follows. Section 2 contains notation and background material, including a review of some properties of Chebyshev polynomials (§2.2), and an overview of relations between some (equivalent) norms on spaces of polynomials of fixed degree (§2.3). We derive a key technical lemma in Section 3 that relates the SDP hierarchy errors  $f_{\min} - \hat{f}_{(r)}$  to the 1-norm distance of the objective function to the truncated pre-ordering  $\mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r$ . After we established the link with 1-norm approximation, we explore the implications from the literature on projections of polynomials onto sum-of-squares polynomials in the 1-norm in Section 4. In particular, we show there how results by Lasserre and Netzer [14], and Lasserre [12], can be combined with our approach to yield insights into the convergence rate of the SDP hierarchies (4) and (6). Our main results are presented in Sections 5 and 6 where we show how to obtain convergence rates by using suitable kernels of the Jackson type. In particular, we prove the error bound (7) there in Theorem 9 and Proposition 2. In the penultimate Section 7, we show how to bound the aforementioned quantity  $\Theta_{n,d}^r$  numerically via the solution of an SDP problem. We conclude with some discussion on future research directions and related problems in Section 8.

## 2 Notation and preliminaries

### 2.1 Further notation

We will denote  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ , and

$$\mathbb{N}_d^n := \left\{ \alpha \in (\mathbb{N}_0)^n \mid |\alpha| := \sum_{i=1}^n \alpha_i \leq d \right\},$$

and  $\mathbf{x}^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ . We will be considering polynomials of different types of variables, and will need to keep track of the degree of these polynomials with respect to only some of the variables. To do this, given a multivariate polynomial  $p$  and a subset of the variables  $S$ , we define  $\deg_S p$  as the *degree of  $p$  with respect to the variables in  $S$* . That is, when all the variables not in  $S$  are considered to be constants (or more correctly parameters). For instance let  $p(\mathbf{u}, \mathbf{x}, \mathbf{y}) = \mathbf{u}^\alpha \mathbf{x}^\beta \mathbf{y}^\gamma$ , then  $\deg p = |\alpha| + |\beta| + |\gamma|$  while  $\deg_{\mathbf{x}} p = |\beta|$  and  $\deg_{\mathbf{ux}} p = |\alpha| + |\beta|$ .

If we consider functions of two distinct types of variables on the hypercube, say  $(\mathbf{x}, \mathbf{y}) \in [-1, 1]^n \times [-1, 1]^n$ , then we define an associated truncated pre-ordering as

$$\mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r_1, r_2} := \left\{ \sum_{I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+ \subseteq [n]} \sigma_{I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+}(\mathbf{x}, \mathbf{y}) \prod_{i \in I_{\mathbf{x}}^-} (1 - x_i) \prod_{i \in I_{\mathbf{x}}^+} (1 + x_i) \prod_{i \in I_{\mathbf{y}}^-} (1 - y_i) \prod_{i \in I_{\mathbf{y}}^+} (1 + y_i) \right\},$$

where, for all  $I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+ \subseteq [n]$ , one has  $\sigma_{I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+} \in \Sigma[\mathbf{x}, \mathbf{y}]$ , as well as  $|I_{\mathbf{x}}^-| + |I_{\mathbf{x}}^+| + \deg_{\mathbf{x}} \sigma_{I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+} \leq r_1$ , and  $|I_{\mathbf{y}}^-| + |I_{\mathbf{y}}^+| + \deg_{\mathbf{y}} \sigma_{I_{\mathbf{x}}^-, I_{\mathbf{x}}^+, I_{\mathbf{y}}^-, I_{\mathbf{y}}^+} \leq r_2$ . In other words, in this special truncated pre-ordering, the degree of each term is at most  $r_1$  in the  $\mathbf{x}$  variables, and at most  $r_2$  in the  $\mathbf{y}$  variables.

### 2.2 Chebyshev polynomials

In this section we review known results on Chebyshev polynomials for later use. The interested reader may find more information in the classic text by Rivlin [25].

Let  $\mathbf{K} = [-1, 1]$  and fix the measure  $\mu$  on  $\mathbf{K}$  defined by  $d\mu(x) = (\pi\sqrt{1-x^2})^{-1}dx$ ,  $x \in \mathbf{K}$ . The Chebyshev polynomials of the first kind form a system of orthogonal polynomials for  $\mu$ . We will refer to the  $k$ -th Chebyshev polynomial of first kind as  $T_k(x)$ . We have, for  $k \in \mathbb{N}_0$ ,

$$T_k(x) = \cos(k \arccos x). \tag{8}$$

Define for  $f, g : [-1, 1] \rightarrow \mathbb{R}$

$$\langle f, g \rangle_\mu = \int_{-1}^1 \frac{f(x)g(x)}{\pi\sqrt{1-x^2}} dx$$

to obtain the following orthogonality relations for the Chebyshev polynomials of the first kind

$$\langle T_k, T_m \rangle_\mu = \frac{1 + \delta_{k,0}}{2} \delta_{k,m}. \tag{9}$$

We will denote the normalized Chebyshev polynomials by

$$\hat{T}_0 = T_0, \quad \hat{T}_k = \sqrt{2}T_k \quad (k \in \mathbb{N}).$$

It is straightforward to generalize the Chebyshev polynomials to the multivariate case by taking products of univariate Chebyshev polynomials. Let  $\mathbf{K} = [-1, 1]^n$  and define

$$d\mu(\mathbf{x}) := \prod_{i=1}^n \frac{1}{\pi \sqrt{1-x_i^2}} d\mathbf{x}. \quad (10)$$

Then, for  $\alpha \in (\mathbb{N}_0)^n$  the corresponding (normalized) multivariate Chebyshev polynomial of the first kind is defined as

$$\hat{T}_\alpha(\mathbf{x}) = \prod_{i=1}^n \hat{T}_{\alpha_i}(x_i).$$

The orthogonality relations extend in the following way

$$\langle \hat{T}_\alpha, \hat{T}_\beta \rangle_\mu = \int_{\mathbf{K}} \hat{T}_\alpha(\mathbf{x}) \hat{T}_\beta(\mathbf{x}) d\mu(\mathbf{x}) = \prod_{i=1}^n \int_{-1}^1 \frac{\hat{T}_{\alpha_i}(x_i) \hat{T}_{\beta_i}(x_i)}{\pi \sqrt{1-x_i^2}} dx_i = \prod_{i=1}^n \delta_{\alpha_i, \beta_i}.$$

One has the following formula to convert from the univariate monomial basis to the Chebyshev basis:

$$x^k = 2^{1-k} \sum_{\substack{j=0 \\ k-j \text{ even}}}^k \binom{k}{(k-j)/2} T_j(x), \quad (k \in \mathbb{N}), \quad (11)$$

where the prime at the sum symbol means the first term (at  $j = 0$  for even  $n$ ) should be halved.

### 2.3 Norms of polynomials

The  $n$ -variate polynomials of degree at most  $d$ , denoted by  $\mathbb{R}[\mathbf{x}]_d$ , may be associated with  $\mathbb{R}^{\binom{n+d}{d}}$  via the polynomial coefficients in a given basis.

Using the standard monomial basis, any  $f \in \mathbb{R}[\mathbf{x}]_d$  may be written as  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \mathbf{x}^\alpha$ , with associated 1-norm:

$$\|f\|_1 := \sum_{\alpha \in \mathbb{N}_d^n} |f_\alpha|.$$

Since the 1-norm depends on the basis, we will use the notation  $\|\cdot\|_{1,T}$  (resp.  $\|\cdot\|_{1,\hat{T}}$ ) for the Chebyshev (resp. normalized Chebyshev) basis. Note that  $\|f\|_{1,\hat{T}} \leq \|f\|_{1,T} \leq 2^{d/2} \|f\|_{1,\hat{T}}$  if  $f \in \mathbb{R}[x]_d$  and  $\|f\|_{1,\hat{T}} = \frac{1}{\sqrt{2}^d} \|f\|_{1,T}$  if  $f$  is homogeneous of degree  $d$ .

**Lemma 1.** *One has*

$$\|f\|_{1,T} \leq \|f\|_1 \quad \forall f \in \mathbb{R}[\mathbf{x}]. \quad (12)$$

*Proof.* Since  $T_j(1) = 1$  for all  $j \in \mathbb{N}_0$ , all the coefficients on the right-hand-side of (11) are positive and sum to one. This implies that, for any (multivariate) monomial  $f(\mathbf{x}) = \mathbf{x}^\alpha$  with  $\alpha \in \mathbb{N}^n$ , one has  $\|f\|_{1,T} = \|f\|_1 = 1$ . Thus, if  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \mathbf{x}^\alpha$  is a general degree  $d$  polynomial, one has (with slight abuse of notation):

$$\|f\|_1 = \sum_{\alpha \in \mathbb{N}_d^n} |f_\alpha| = \sum_{\alpha \in \mathbb{N}_d^n} \|f_\alpha \mathbf{x}^\alpha\|_{1,T} \geq \left\| \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \mathbf{x}^\alpha \right\|_{1,T} \equiv \|f\|_{1,T},$$

where the inequality is the triangle inequality.  $\square$

We will only be concerned with the 1-norm for different bases, but in general one may define the  $p$ -norm for any  $p > 0$ :

$$\|f\|_p := \left( \sum_{\alpha \in \mathbb{N}_d^n} |f_\alpha|^p \right)^{1/p}.$$

To avoid confusion with the usual  $\ell_p$  norm, we will use the notation, for any  $p \in \mathbb{N}$ ,

$$\|f\|_{\ell_p} := \left( \int_{\mathbf{K}} |f(\mathbf{x})|^p d\mathbf{x} \right)^{1/p},$$

where we only consider the case  $\mathbf{K} = [-1, 1]^n$ , and the sup-norm:

$$\|f\|_{\ell_\infty} := \sup_{\mathbf{K}} |f(\mathbf{x})|.$$

The relations between the (equivalent) norms  $\|\cdot\|_p$  and  $\|\cdot\|_{\ell_p}$  on  $\mathbb{R}[x]_d$  is a classical problem that is still a topic of active research; see, e.g. [2] for a recent overview. Perhaps the best-known result along these lines is the Bohnenblust-Hille inequality:

$$\|f\|_p \leq C_d \|f\|_{\ell_\infty} \quad \deg(f) = d, \quad p = \frac{2d}{d+1}, \quad (13)$$

where the constant  $C_d$  depends on  $d$  but not on  $n$ . For values of  $p < \frac{2d}{d+1}$ , the corresponding constant does depend on  $n$ . One may obtain a bound for the 1-norm instead by using the known inequalities between  $p$ -norms: if  $0 < q < p$  one has

$$\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q \leq k^{1/q-1/p} \|\mathbf{x}\|_p \quad \forall \mathbf{x} \in \mathbb{R}^k.$$

Setting  $q = 1$ ,  $p = \frac{2d}{d+1}$ , and  $k = \binom{n+d}{d}$ , (13) implies

$$\|f\|_1 \leq \binom{n+d}{d}^{1-\frac{d+1}{2d}} C_d \|f\|_{\ell_\infty} = \binom{n+d}{d}^{\frac{d-1}{2d}} C_d \|f\|_{\ell_\infty} < \binom{n+d}{d}^{\frac{1}{2}} C_d \|f\|_{\ell_\infty}, \quad (14)$$

where  $C_d$  is the same constant as in (13).

Some stronger results are known in special cases, e.g, when  $n = 2$  and  $d = 2$ , one has

$$\|f\|_{\ell_\infty} \leq \|f\|_1 \leq (1 + \sqrt{2}) \|f\|_{\ell_\infty}.$$

As another example, if  $f(x) = \frac{1}{4} x^T L x$  where  $L$  is the Laplacian matrix of a graph, then  $\|f\|_{\ell_\infty}$  is the maximum cut size in the graph, while  $\|f\|_1$  equals the number of edges in the graph. Thus, in this case one has  $\|f\|_{\ell_\infty} \leq \|f\|_1 \leq 2\|f\|_{\ell_\infty}$ , i.e. the constant in the upper bound (2 in this case) need not depend on  $n$ .

### 3 A key lemma

As before, let  $T_k$  denote the univariate Chebyshev polynomials of the first kind. Then one has the well-known recursive relation  $T_0(t) = 1$ ,  $T_1(t) = t$  and  $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$ . Recall that we define the multivariate version by  $T_\alpha(\mathbf{x}) = \prod_{i=1}^n T_{\alpha_i}(x_i)$ . Recall that  $1 \pm T_\alpha$  is nonnegative on  $[-1, 1]^n$ . Our first goal is to prove the stronger result that  $1 \pm T_\alpha$  belongs to the truncated pre-ordering of order  $|\alpha|$ , namely  $1 \pm T_\alpha(\mathbf{x}) \in \mathcal{T}(1 - x_1, 1 + x_1, \dots, 1 - x_n, 1 + x_n)_{|\alpha|}$ .

In the proof, we will need the a classical result of this type for the special case of univariate polynomials.

**Theorem 4** (Fekete, Markov-Lukács (see, e.g., [20])). *Let  $p$  be a univariate polynomial of degree  $2m$ . Then  $p$  is nonnegative on the interval  $[a, b]$  if and only if*

$$p(t) = q_1^2(t) + (t-a)(t-x)q_2^2(t),$$

for some polynomials  $q_1$  of degree  $m$  and  $q_2$  of degree  $m-1$ . If the degree of  $p$  is  $2m+1$  then  $p$  is nonnegative on  $[a, b]$  if and only if

$$p(t) = (t-a)q_1^2(t) + (b-t)q_2^2(t),$$

for some polynomials  $q_1$  and  $q_2$  of degree  $m$ .

We now give a multivariate generalization of this result.

**Lemma 2.** *For any  $\alpha \in (\mathbb{N}_0)^n$  we have  $1 \pm T_\alpha(\mathbf{x}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_{|\alpha|}$ .*

*Proof.* By applying the identities,

$$1 + ab = [(1-a)(1+b) + (1+a)(1-b)]/2 \quad \text{and} \quad 1 - ab = [(1+a)(1+b) + (1-a)(1-b)]/2$$

repeatedly as necessary, we obtain that  $1 \pm T_\alpha(\mathbf{x}) = 1 \pm \prod_{i=1}^n T_{\alpha_i}(x_i)$  can be expressed as sums of products of terms of the form  $1 \pm T_{\alpha_i}(x_i)$ .

Thus, it is enough to show that  $1 \pm T_k(t) \in \mathcal{T}(1-t, 1+t)_k$ . But this follows from the fact that every  $p \in \mathbb{R}[t]_k$  non-negative on  $[-1, 1]$  is in  $\mathcal{T}(1-t, 1+t)_k$ , by Theorem 4, and the fact that  $1 \pm T_k$  is nonnegative on  $[-1, 1]$ .  $\square$

The main result of this section now follows as an immediate consequence of the last lemma.

**Theorem 5.** For all  $p \in \mathbb{R}[\mathbf{x}]$ , one has  $\|p\|_{1,T} - p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_{\deg p}$ .

*Proof.* Letting  $p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha T_\alpha(\mathbf{x})$ , we get  $\|p\|_{1,T} - p = \sum_{\alpha \in \mathbb{N}_d^n} |p_\alpha| (1 - \text{sign}(p_\alpha) T_\alpha(\mathbf{x}))$  and thus the statement follows by the last lemma.  $\square$

**Corollary 1.** If  $f, p \in \mathbb{R}[\mathbf{x}]_d$ , and  $p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_d$ , then

$$f + \|p - f\|_{1,T} \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_d.$$

In particular, replacing  $f$  by  $f - f_{\min}$ , one has

$$\hat{f}_{(r)} \geq f_{\min} - \min_{p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r} \|p - (f - f_{\min})\|_{1,T}, \quad (15)$$

for  $r \geq \deg f$ , where  $\hat{f}_{(r)}$  is the SDP bound defined in (6).

**Remark:** these results remain true if the Chebyshev basis is replaced by the standard monomial basis, with our prior convention that, if  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \mathbf{x}^\alpha$ , then  $f$  has 1-norm  $\|f\|_1 = \sum_{\alpha \in \mathbb{N}_d^n} |f_\alpha|$ .

## 4 A quantitative Schmüdgen-type theorem for the hypercube via sums-of-squares approximations

We can use Corollary 1 to derive a quantitative version of Schmüdgen-type theorem for the  $[-1, 1]^n$  hypercube. We will need the following result from Lasserre and Netzer [14].

**Theorem 6** (Lasserre and Netzer [14]). *If  $f \in \mathbb{R}[\mathbf{x}]$  is nonnegative on  $[-1, 1]^n$ , then, for any  $\epsilon > 0$ , there exists an  $r_\epsilon \in \mathbb{N}$  such that*

$$f_\epsilon(\mathbf{x}) := f(\mathbf{x}) + \epsilon \left( 1 + \sum_{i=1}^n x_i^{2r_\epsilon} \right) \in \Sigma[\mathbf{x}]_{2r_\epsilon}. \quad (16)$$

We now have the following Schmüdgen-type theorem.

**Theorem 7.** *If  $f \in \mathbb{R}[\mathbf{x}]$  is nonnegative on  $[-1, 1]^n$ , then, for any  $\epsilon > 0$ , there exists an  $r \in \mathbb{N}$  such that*

$$f + (n+1)\epsilon \in \mathcal{T}(1 - x_1, 1 + x_1, \dots, 1 - x_n, 1 + x_n)_{2r}.$$

The value  $r$  is upper bounded by the smallest integer  $r_\epsilon$  such that (16) holds.

*Proof.* Note that  $\|f - f_\epsilon\|_1 = (n+1)\epsilon$  in the standard monomial basis. The result now follows from Corollary 1.  $\square$

Lasserre [12] characterized the best sums-of-squares approximation in 1-norm.

**Theorem 8** (Lasserre [12]). *Let  $f \in \mathbb{R}[\mathbf{x}]$ , and  $d \in \mathbb{N}$  such that  $\deg f \leq 2d$ . Then there exist nonnegative  $\lambda_0^*, \dots, \lambda_n^*$ , such that*

$$\mathbf{x} \mapsto f(\mathbf{x}) + \lambda_0^* + \sum_{i=1}^n \lambda_i^* x_i^{2d} \in \arg \min_{p \in \Sigma[\mathbf{x}]_{2d}} \|f - p\|_1.$$

Moreover, the values  $\lambda_0^*, \dots, \lambda_n^*$  are given by the optimal solution of the SDP

$$\rho_d := \min_{\lambda \geq 0} \left\{ \sum_{i=0}^n \lambda_i : \mathbf{x} \mapsto f(\mathbf{x}) + \lambda_0 + \sum_{i=1}^n \lambda_i x_i^{2d} \in \Sigma[\mathbf{x}]_{2d} \right\},$$

i.e.

$$\rho_d = \sum_{i=0}^n \lambda_i^* = \min_{p \in \Sigma[\mathbf{x}]_{2d}} \|f - p\|_1. \quad (17)$$

By Corollary 1, we have the following immediate implication of the last theorem.

**Corollary 2.** *If  $f$  has  $\deg f < 2d$ , then*

$$f + \rho_d \in \mathcal{T}(1 - x_1, 1 + x_1, \dots, 1 - x_n, 1 + x_n)_{2d},$$

where  $\rho_d$  is defined in (17).

The corollary implies that, if  $\deg f < 2d$  and  $r \geq d$ , one has  $\hat{f}_{(2r)} \geq -\rho_r$ .

**Example 1** (Lasserre [12]). Consider the Motzkin-type polynomial  $\mathbf{x} \mapsto f(\mathbf{x}) = x_1^2 x_2^2 (x_1^2 + x_2^2 - 1) + 1/27$  of degree 6, which is nonnegative but not a sum of squares, and with a global minimum  $f_{\min} = 0$  on  $[-1, 1]^2$ , attained at four global minimizers  $x^* = (\pm(1/3)^{1/2}, \pm(1/3)^{1/2})$ . The values  $\rho_d$  are displayed in Table 1 for  $d = 3, 4, 5$ . For this

$d$	$\lambda^*$	$\rho_d$
3	$\approx 10^{-3} (5.445, 5.367, 5.367)$	$\approx 1.6 \cdot 10^{-2}$
4	$\approx 10^{-4} (2.4, 9.36, 9.36)$	$\approx 2 \cdot 10^{-3}$
5	$\approx 10^{-5} (0.04, 4.34, 4.34)$	$\approx 8 \cdot 10^{-5}$

Table 1: Best 1-norm approximation for the Motzkin polynomial.

example one has  $f \in \mathcal{T}(1 - x_1, 1 + x_1, 1 - x_2, 1 + x_2)_6$  so that the lower bounds  $-\rho_d$  on  $f_{\min} = 0$  via Corollary 2 are not tight for  $d = 3, 4, 5$ .

## 5 1-norm approximation by kernels

Corollary 1 showed that one may bound the errors in the Schmüdgen SDP hierarchy (6) in terms of suitable 1-norm approximations of the objective function. In this section we therefore discuss a systematic approach to 1-norm approximation of polynomials via positive kernels. This is a classical approach in approximation theory, see e.g. [1], and the survey [29].

We now consider a kernel  $K_r : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  given by

$$K_r(\mathbf{x}, \mathbf{y}) := \sum_{\alpha \in \mathbb{N}_r^n} \lambda_\alpha \hat{T}_\alpha(\mathbf{x}) \hat{T}_\alpha(\mathbf{y}),$$

for given constants  $\lambda_\alpha \in [0, 1]$  for  $\alpha \in \mathbb{N}_r^n$ . The convolution operator:

$$\mathcal{K}^{(r)}(f)(\mathbf{x}) := \int_{\mathbf{K}} f(\mathbf{y}) K_r(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{y}), \quad (18)$$

maps any integrable function  $f$  to a polynomial of degree at most  $r$ . More precisely,

$$\mathcal{K}^{(r)}(f)(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_r^n} b_\alpha \hat{T}_\alpha(\mathbf{x}), \text{ where } b_\alpha = \langle \hat{T}_\alpha, f \rangle_\mu \lambda_\alpha.$$

The coefficients  $\lambda_\alpha$  of the kernel  $K_r$  determine the approximation. In particular, if  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha T_\alpha(\mathbf{x})$  is a polynomial of degree  $d \leq r$ , then

$$\mathcal{K}^{(r)}(f)(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} \lambda_\alpha f_\alpha T_\alpha(\mathbf{x}),$$

so that the operator  $\mathcal{K}^{(r)}$  is a projection if all  $\lambda_\alpha = 1$  (the so-called Dirichlet or Christoffel-Darboux kernel, see e.g. [15]). If  $\lambda_0 = 1$ , then the convolution operator preserves constant functions. We will assume this property throughout.

Also note that, if  $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha T_\alpha(\mathbf{x})$ , and  $r \geq d$ ,

$$\|\mathcal{K}^{(r)}(f) - f\|_{1, T} = \sum_{\alpha \in \mathbb{N}_d^n} |1 - \lambda_\alpha| |f_\alpha| \leq \max_{\alpha \in \mathbb{N}_d^n} |1 - \lambda_\alpha| \|f\|_{1, T}.$$

We are interested in positive kernels that map nonnegative functions on  $\mathbf{K} = [-1, 1]^n$  to nonnegative functions, i.e. positive linear approximation operators in the sense of Korovkin; see e.g. [1]. More precisely, for given degree  $d \in \mathbb{N}$ , we are interested in kernels where the coefficients  $\lambda_\alpha$  are such that

$$\mathcal{K}^{(r)}(f)(\mathbf{x}) := \int_{\mathbf{K}} f(\mathbf{y}) K_r(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r \quad \forall f \in \mathbb{R}[\mathbf{x}]_d \text{ nonnegative on } \mathbf{K}. \quad (19)$$

Note that the Dirichlet kernel is not positive; see e.g. [29]. However, we would like the coefficients  $\lambda_\alpha$  as close to 1 as possible, to approximate the Dirichlet kernel, since this kernel gives a spectral rate of convergence for analytic functions; see e.g. [28].

This leads us naturally to the following abstract optimization problem, that aims to construct the best kernel that meets our requirements.

$$\Theta_{n,d}^r := \inf \left\{ \max_{\alpha \in \mathbb{N}_d^n} |1 - \lambda_\alpha| : K_r(\mathbf{x}, \mathbf{y}) = 1 + \sum_{\alpha \in \mathbb{N}_d^n \setminus \{\mathbf{0}\}} \lambda_\alpha \hat{T}_\alpha(\mathbf{x}) \hat{T}_\alpha(\mathbf{y}) \text{ satisfies (19)} \right\}. \quad (20)$$

One may bound the optimal value  $\hat{f}_{(r)}$  of the problem (6) in terms of  $\Theta_{n,d}^r$  as follows.

**Theorem 9.** *Let  $f \in \mathbb{R}[\mathbf{x}]_d$  and  $f_{\min} = \min_{\mathbf{x} \in \mathbf{K}} f(\mathbf{x})$ . One has, for each  $r \in \mathbb{N}$  with  $r \geq d$ ,*

$$f - (f_{\min} - \Theta_{n,d}^r \|f\|_{1,T}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r$$

*which implies  $f_{\min} - \hat{f}_{(r)} \leq \Theta_{n,d}^r \|f\|_{1,T}$ , where  $\hat{f}_{(r)}$  is the SDP bound from (6).*

*Proof.* If  $\lambda_\alpha$  ( $\alpha \in \mathbb{N}_d^n$ ) denotes a feasible solution of problem (20), one has

$$\|\mathcal{K}^{(r)}(f) - f\|_{1,T} = \sum_{\alpha \in \mathbb{N}_d^n} |1 - \lambda_\alpha| |f_\alpha| \leq \max_{\alpha \in \mathbb{N}_d^n} |1 - \lambda_\alpha| \|f\|_{1,T}.$$

Also, with reference to (15), one has:

$$\mathcal{K}^{(r)}(f - f_{\min}) - (f - f_{\min}) = \mathcal{K}^{(r)}(f) - f_{\min} - (f - f_{\min}) = \mathcal{K}^{(r)}(f) - f,$$

since  $\mathcal{K}^{(r)}$  preserves constant functions. Finally, by assumption one has  $\mathcal{K}^{(r)}(f - f_{\min}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r$ , so that the required result follows from (15).  $\square$

Some comments on the theorem:

- The values  $\Theta_{n,d}^r$  do not depend on  $f$ , and may be upper-bounded a priori via SDP, as we will show later. Thus the right-hand-side of the inequality  $f_{\min} - \hat{f}_{(r)} \leq \Theta_{n,d}^r \|f\|_{1,T}$  may readily be upper bounded, given a table of SDP upper bounds on the values  $\Theta_{n,d}^r$ .
- We will show in the next section that  $\Theta_{n,d}^r = O(1/r^2)$  by considering a specific kernel that is feasible for problem (20). This is essentially the same argument used by Laurent and Slot [17]. In particular, it suffices to use products of so-called *univariate Jackson kernels*.

## 6 Kernels of the Jackson type

For  $r \in \mathbb{N}$ , consider the univariate kernel  $K_r^{\text{ja}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  given by:

$$K_r^{\text{ja}}(x, y) := 1 + \sum_{k=1}^r \lambda_k^r \hat{T}_k(x) \hat{T}_k(y), \quad (21)$$

where the coefficients  $\lambda_k^r$  are given by

$$\lambda_k^r = \frac{1}{r+2} ((r+2-k) \cos(k\theta_r) + \frac{\sin(k\theta_r)}{\sin(\theta_r)} \cos(\theta_r)) \quad (1 \leq k \leq r), \quad (22)$$

with  $\theta_r = \frac{\pi}{r+2}$ .

This kernel is called the *univariate Jackson kernel* in [29] and [17]. In the approximation theory literature there is another object which is known by that name, that corresponds to a kernel in the original analysis of Jackson [8, 9], where he showed that the famous approximation bounds by Bernstein [6] were tight. In [10], the kernel (21) is called the *minimum resolution kernel*, since it minimizes the resolution over all positive univariate kernels of a given degree, where the resolution is defined by

$$\sigma_r := \left( \int_{[-1,1] \times [-1,1]} (x-y)^2 K_r(x, y) d\mu(x) d\mu(y) \right)^{1/2}.$$

In this paper we will also speak about the univariate Jackson kernel, with the caveat that the term is ambiguous. The interested reader may consult [10] for more details on the history and terminology.

The following properties of the univariate Jackson kernel were shown in [17]; see also [29] and the references therein.

**Proposition 1** (Proposition 6 in Laurent and Slot [17]). *For every  $r \in \mathbb{N}$  we have:*

- (i)  $K_r^{\text{ja}}(x, y) \geq 0$  for all  $x, y \in [-1, 1]$ ,
- (ii)  $1 \geq \lambda_k^r > 0$  for all  $1 \leq k \leq r$ , and
- (iii)  $|1 - \lambda_k^r| = 1 - \lambda_k^r \leq \frac{\pi^2 k^2}{(r+2)^2}$  for all  $1 \leq k \leq r$ .

One may obtain a multivariate kernel by multiplying univariate Jackson kernels in the obvious way:

$$K_{nr}(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^n K_r^{\text{ja}}(x_i, y_i), \quad (23)$$

where  $K_r^{\text{ja}}$  is the univariate Jackson kernel from (21). The coefficients of this kernel are given by

$$\lambda_\alpha^r := \prod_{i=1}^n \lambda_{\alpha_i}^r \quad (\alpha \in \mathbb{N}_r^n).$$

The following is a refinement of a lemma by Laurent and Slot [17, Lemma 12].

**Lemma 3.** *Let  $\pi d < r + 2$ . For any  $\alpha \in \mathbb{N}_d^n$  we have  $|1 - \lambda_\alpha^r| \leq \frac{\pi^2 d^2}{(r+2)^2}$ .*

*Proof.* From Proposition 1 part (iii), we know that  $1 \geq \lambda_k^r \geq 1 - \frac{\pi^2 k^2}{(r+2)^2} \geq 0$ . Then,

$$1 \geq \lambda_\alpha^r \geq \prod_{i=1}^n \left(1 - \frac{\pi^2 (\alpha_i)^2}{(r+2)^2}\right) \geq 1 - \frac{\pi^2}{(r+2)^2} \sum_{i=1}^n (\alpha_i)^2 \geq 1 - \frac{\pi^2 d^2}{(r+2)^2},$$

where the third inequality is the generalised Bernoulli inequality:

$$\prod_{i=1}^n (1 - y_i) \geq 1 - \sum_{i=1}^n y_i \quad \text{if } y_i < 1 \ \forall i \in [n].$$

Note that the condition  $y_i < 1$  for all  $i$  in the Bernoulli inequality corresponds to the condition  $\pi d < r + 2$  in the statement of the lemma, since  $\alpha \in \mathbb{N}_d^n$ .  $\square$

We obtain the following upper bound on  $\Theta_{n,d}^{rn}$ .

**Proposition 2.**  $\Theta_{n,d}^{rn} \leq \frac{\pi^2 d^2}{(r+2)^2}$  for all  $r$  and  $d$  such that  $\pi d < r + 2$ .

*Proof.* The result follows from the definition of  $\Theta_{n,d}^r$  in (20) and Lemma 3, provided that the multivariate Jackson kernel (23) satisfies the property (19). This indeed holds, as is shown in [17, Lemma 9].  $\square$

We will need the argument from [17, Lemma 9] once more in the next section in a more general context, and there we will give a self-contained proof for the sake of completeness; see Proposition 3.

## 7 SDP bounds on $\Theta_{n,d}^r$

In this section we will prove that the following inequality holds.

**Theorem 10.** *The value  $\Theta_{n,d}^r$  as defined in (20) is upper bounded as follows:*

$$\Theta_{n,d}^r \leq \inf \left\{ \max\{|1 - p_\alpha| : |\alpha| \leq d\} : \mathbf{x} \mapsto 1 + \sum_{\alpha \in \mathbb{N}_d^n \setminus \{\mathbf{0}\}} 2^{\omega(\alpha)} p_\alpha T_\alpha(\mathbf{x}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r \right\}, \quad (24)$$

where  $\omega(\alpha)$  denotes the number of nonzero elements of  $\alpha$ .

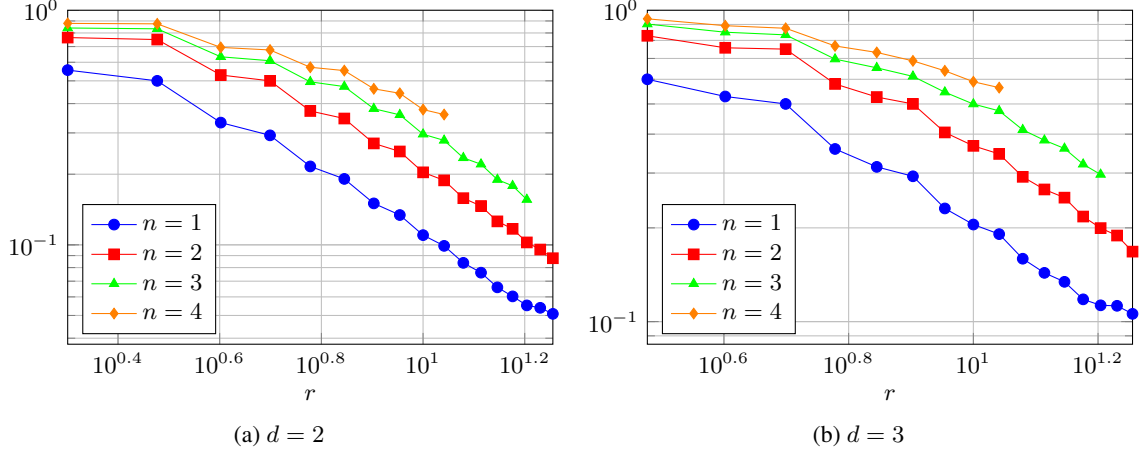


Figure 1: The SDP upper bound (24) on  $\Theta_{n,d}^r$  for  $d = 2, 3$ .

The right-hand-side of the inequality (24) may readily be computed via SDP for given (small) values of  $n, r, d$ ; we list the resulting numerical bounds for various values of  $n, r, d$  in an appendix to this paper. In Figure 1 we plot some of the values of the SDP bound, for illustrative purposes.

To prove Theorem 10, we will show that the following two statements are equivalent:

1.  $p(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}_d^n} 2^{\omega(\alpha)} p_\alpha T_\alpha(\mathbf{x}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r$
2.  $K_r(\mathbf{x}, \mathbf{y}) := \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha \hat{T}_\alpha(\mathbf{x}) \hat{T}_\alpha(\mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r,r}$ .

Note that item 1 appears in the statement of problem (24). Also, item 2 is a sufficient condition for (19) to hold, again by a similar argument as in [17, Lemma 9]; we give a proof of this fact in Proposition 3 below.

Also note that proving the equivalence of the two conditions is the same as proving the equivalence of the following two statements (after re-scaling):

1.  $p(\mathbf{x}) := \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha T_\alpha(\mathbf{x}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r$
2.  $K_r(\mathbf{x}, \mathbf{y}) := \sum_{\alpha \in \mathbb{N}_d^n} p_\alpha T_\alpha(\mathbf{x}) T_\alpha(\mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r,r}$ .

For the sake of completeness, we give a proof in Proposition 3 below that item 2 is a sufficient condition for (19) to hold, using a similar argument as in [17, Lemma 9]. The proof relies on a celebrated theorem<sup>3</sup> by Tchakaloff [27] concerning the existence of cubature rules for numerical integration. (Modern expositions and generalizations of Tchakaloff's theorem are given in [23] and [7].) Recall that a positive cubature rule of degree  $d$  for numerical integration with respect to a measure  $\mu$  over a compact set  $\mathbf{K}$  is defined by a set of  $N$  nodes  $\mathbf{y}^{(\ell)} \in \mathbf{K}$  and weights  $w_\ell \geq 0$  ( $\ell \in [N]$ ) with the property that

$$\int_{\mathbf{K}} f(\mathbf{y}) d\mu(\mathbf{y}) = \sum_{\ell=1}^N w_\ell f(\mathbf{y}^{(\ell)}), \quad \forall f \in \mathbb{R}[\mathbf{y}]_d.$$

**Theorem 11** (Tchakaloff [27]). *For any compact set  $\mathbf{K} \subset \mathbb{R}^n$ , and finite, positive Borel measure supported on  $\mathbf{K}$ , there exists a positive cubature rule of degree  $d$  with  $N = \binom{n+d}{d}$  nodes.*

One may use the Tchakaloff theorem to show the following.

**Proposition 3.** *Assume  $K : [-1, 1]^n \times [-1, 1]^n \rightarrow \mathbb{R}$  and*

$$(\mathbf{x}, \mathbf{y}) \mapsto K(\mathbf{x}, \mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r,r}.$$

<sup>3</sup>The same result was obtained independently at the same time by Richter [24].

Then, for any  $f \in \mathbb{R}[\mathbf{x}]_d$  that is nonnegative on  $[-1, 1]^n$ , one has

$$\mathbf{x} \mapsto \int_{[-1, 1]^n} f(\mathbf{y})K(\mathbf{x}, \mathbf{y})d\mu(\mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r,$$

where  $\mu$  is defined in (10).

*Proof.* By Tchakaloff's theorem there exist  $\mathbf{y}^{(\ell)} \in [-1, 1]^n$  and the weights  $w_\ell \geq 0$  ( $\ell \in [N]$ ) with  $N = \binom{n+d}{d}$  such that

$$\int_{[-1, 1]^n} f(\mathbf{y})K(\mathbf{x}, \mathbf{y})d\mu(\mathbf{y}) = \sum_{\ell=1}^N w_\ell K(\mathbf{x}, \mathbf{y}^{(\ell)})f(\mathbf{y}^{(\ell)}).$$

Using  $f(\mathbf{y}^{(\ell)}) \geq 0$  for all  $\ell \in [N]$ , as well as

$$\mathbf{x} \mapsto K(\mathbf{x}, \mathbf{y}^{(\ell)}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r \quad \forall \ell \in [N],$$

yields the required result.  $\square$

For clarity of exposition, we will first prove the equivalence of the two statements in the univariate case, before proceeding to the multivariate case.

### 7.1 Univariate case

First, we work out the univariate case.

**Lemma 4.** For any  $k \geq 0$ , and  $x, y \in [-1, 1]$ , one has

$$T_k(x)T_k(y) = \frac{1}{2} \left( T_k(xy + \sqrt{(1-x^2)(1-y^2)}) + T_k(xy - \sqrt{(1-x^2)(1-y^2)}) \right).$$

*Proof.* Let  $x_1, x_2 \in [-1, 1]$ . Then there are  $\theta_1, \theta_2 \in [0, \pi]$  such that  $x_i = \cos(\theta_i)$ . Then  $\sin(\theta_i) \geq 0$  and thus  $\sin(\theta_i) = \sqrt{1 - \cos^2(\theta_i)} = \sqrt{1 - x_i^2}$ . Thus

$$x_1x_2 \pm \sqrt{(1-x_1^2)(1-x_2^2)} = \cos\theta_1 \cos\theta_2 \pm \sin\theta_1 \sin\theta_2 = \cos(\theta_1 \mp \theta_2).$$

Using  $T_k(\cos(\theta)) = \cos(k\theta)$  we have

$$\begin{aligned} T_k \left( x_1x_2 + \sqrt{(1-x_1^2)(1-x_2^2)} \right) + T_k \left( x_1x_2 - \sqrt{(1-x_1^2)(1-x_2^2)} \right) &= T_k(\cos(\theta_1 - \theta_2)) + T_k(\cos(\theta_1 + \theta_2)) \\ &= \cos(k\theta_1 - k\theta_2) + \cos(k\theta_1 + k\theta_2) \\ &= \cos(k\theta_1) \cos(k\theta_2) \\ &= T_k(x_1)T_k(x_2). \end{aligned}$$

$\square$

Now given  $p \in \mathbb{R}[t]$  define

$$K_p(x, y) := \frac{1}{2} \left( p(xy + \sqrt{(1-x^2)(1-y^2)}) + p(xy - \sqrt{(1-x^2)(1-y^2)}) \right).$$

From Lemma 4 we obtain that if  $p(t) = \sum_{k \geq 0} p_k T_k(t)$  then  $K_p(x, y) = \sum_{k \geq 0} p_k T_k(x)T_k(y)$ . Moreover, every kernel is of this form as given a kernel  $K(x, y) = \sum_{k \geq 0} c_k T_k(x)T_k(y)$ , defining  $p_K(t) = K(t, 1)$  we have that  $p_K(t) = \sum_{k \geq 0} c_k T_k(t)$ , since  $T_k(1) = 1$  for all  $k$ .

We have that  $|xy \pm \sqrt{(1-x^2)(1-y^2)}| \leq 1$  when  $|x|, |y| \leq 1$ . Thus,  $p$  is nonnegative on  $[-1, 1]$  if and only if  $K_p$  is non-negative on  $[-1, 1]^2$ . Now,  $p$  is non-negative on  $[-1, 1]$  if, and only if,  $p \in \mathcal{T}(1 \pm t)_{\deg p}$ , by Theorem 4. Next, we show that we also have  $K_p \in \mathcal{T}(1 \pm x; 1 \pm y)_{\deg p, \deg p}$  when  $p$  nonnegative. To this end, we first prove the following lemma.

**Lemma 5.** Let  $d \geq 0$ . Let  $q \in \mathbb{R}[t]_d$ .

1.  $K_{q(t)^2} \in \mathcal{T}(1 \pm x; 1 \pm y)_{2d, 2d}$ .
2.  $K_{(1 \pm t)q(t)^2} \in \mathcal{T}(1 \pm x; 1 \pm y)_{2d+1, 2d+1}$ .
3.  $K_{(1-t^2)q(t)^2} \in \mathcal{T}(1 \pm x; 1 \pm y)_{2d+2, 2d+2}$ .

*Proof.* First notice that by expanding and collecting terms, we can write

$$q(xy + \sqrt{(1-x^2)(1-y^2)}) = q_0(x, y) + \sqrt{(1-x^2)(1-y^2)}q_1(x, y)$$

and

$$q(xy - \sqrt{(1-x^2)(1-y^2)}) = q_0(x, y) - \sqrt{(1-x^2)(1-y^2)}q_1(x, y)$$

with  $q_0 \in \mathbb{R}[x, y]_{d, d}$  and  $q_1 \in \mathbb{R}[x, y]_{d-1, d-1}$ .

We then have

$$q(xy \pm \sqrt{(1-x^2)(1-y^2)})^2 = q_0^2 + (1-x^2)(1-y^2)q_1^2 \pm 2\sqrt{(1-x^2)(1-y^2)}q_0q_1. \quad (25)$$

From (25)  $K_{q^2} = q_0^2 + (1-x^2)(1-y^2)q_1^2$  follows, proving 1. Now, to prove 2, we write

$$\begin{aligned} 2K_{(1-t)q(t)^2} &:= (1-xy - \sqrt{(1-x^2)(1-y^2)})q(xy + \sqrt{(1-x^2)(1-y^2)})^2 \\ &\quad + (1-xy + \sqrt{(1-x^2)(1-y^2)})q(xy - \sqrt{(1-x^2)(1-y^2)})^2 \\ &= (1-xy - \sqrt{(1-x^2)(1-y^2)})(q_0^2 + (1-x^2)(1-y^2)q_1^2 + 2\sqrt{(1-x^2)(1-y^2)}q_0q_1) \\ &\quad + (1-xy + \sqrt{(1-x^2)(1-y^2)})(q_0^2 + (1-x^2)(1-y^2)q_1^2 - 2\sqrt{(1-x^2)(1-y^2)}q_0q_1), \end{aligned}$$

where the second equality follows from (25).

Using  $(a-b)(c+d) + (a+b)(c-d) = 2ac - 2bd$  we obtain,

$$2K_{(1-t)q(t)^2} = 2(1-xy)(q_0^2 + (1-x^2)(1-y^2)q_1^2) - 4(1-x^2)(1-y^2)q_0q_1.$$

Let  $A = (1-x)(1+y)$  and  $B = (1+x)(1-y)$ . We have then  $A+B = 2(1-xy)$  and  $AB = (1-x^2)(1-y^2)$ . Therefore,

$$\begin{aligned} 2K_{(1-t)q(t)^2} &= (A+B)(q_0^2 + ABq_1^2) - 4ABq_0q_1 = A(q_0 - Bq_1)^2 + B(q_0 - Aq_1)^2 \\ &= (1-x)(1+y)(q_0 - (1+x)(1-y)q_1)^2 + (1+x)(1-y)(q_0 - (1-x)(1+y)q_1)^2. \end{aligned}$$

Similarly,

$$2K_{(1+t)q(t)^2} = (1+x)(1+y)(q_0 + (1-x)(1-y)q_1)^2 + (1-x)(1-y)(q_0 + (1+x)(1+y)q_1)^2.$$

To prove 3, we write,

$$\begin{aligned} 2K_{(1-t^2)q(t)^2} &:= \left(1 - (xy + \sqrt{(1-x^2)(1-y^2)})^2\right) q(xy + \sqrt{(1-x^2)(1-y^2)})^2 \\ &\quad + \left(1 - (xy - \sqrt{(1-x^2)(1-y^2)})^2\right) q(xy - \sqrt{(1-x^2)(1-y^2)})^2 \\ &= \left(1 - (xy)^2 - (1-x^2)(1-y^2) - 2xy\sqrt{(1-x^2)(1-y^2)}\right) \\ &\quad \cdot \left(q_0^2 + (1-x^2)(1-y^2)q_1^2 + 2\sqrt{(1-x^2)(1-y^2)}q_0q_1\right) \\ &\quad + \left(1 - (xy)^2 - (1-x^2)(1-y^2) + 2xy\sqrt{(1-x^2)(1-y^2)}\right) \\ &\quad \cdot \left(q_0^2 + (1-x^2)(1-y^2)q_1^2 - 2\sqrt{(1-x^2)(1-y^2)}q_0q_1\right), \end{aligned}$$

where we have used (25) to obtain the last equality.

Using  $(a-b)(c+d) + (a+b)(c-d) = 2ac - 2bd$  again, we obtain,

$$2K_{(1-t^2)q(t)^2} = 2\left(1 - (xy)^2 - (1-x^2)(1-y^2)\right)(q_0^2 + (1-x^2)(1-y^2)q_1^2) - 8xy(1-x^2)(1-y^2)q_0q_1.$$

Now, let  $A = (1 - x^2)$  and  $B = (1 - y^2)$  we have then  $(1 - x^2)(1 - y^2) = AB$  and  $1 - (xy)^2 = AB + x^2B + y^2A$ . Therefore,

$$\begin{aligned} K_{(1-t^2)q(t)^2} &= (x^2B + y^2A)(q_0^2 + ABq_1^2) - 4xyABq_0q_1 \\ &= B(xq_0 - yAq_1)^2 + A(yq_0 - xBq_1)^2 \\ &= (1 - y^2)(xq_0 - y(1 - x^2)q_1)^2 + (1 - x^2)(yq_0 - x(1 - y^2)q_1)^2. \end{aligned}$$

□

Now we are ready to prove our main result in this section.

**Theorem 12.** *Let  $p \in \mathcal{T}(1 \pm t)_r$ . Then  $K_p \in \mathcal{T}(1 \pm x; 1 \pm y)_{r,r}$ .*

*Proof.* We have  $p = \sigma_0 + (1 - t)\sigma_1 + (1 + t)\sigma_2 + (1 - t^2)\sigma_3$  where  $\sigma_i \in \Sigma[t]$  and  $\deg(\sigma_0), \deg(\sigma_1) + 1, \deg(\sigma_2) + 1, \deg(\sigma_3) + 2 \leq r$ . By the linearity of the operator  $p \mapsto K_p$  and Lemma 5, we obtain  $K_p \in \mathcal{T}(1 \pm x, 1 \pm y)_{r,r}$ . □

**Corollary 3.** *Let  $K(x, y) = \sum_{k=0}^r c_k T_k(x)T_k(y)$  be a nonnegative kernel. Then  $K \in \mathcal{T}(1 \pm x, 1 \pm y)_{r,r}$ .*

*Proof.* Let  $p(t) = K(t, 1)$ . We have then  $K = K_p$  and  $p$  is nonnegative on  $[-1, 1]$ . By Theorem 4, we have  $p \in \mathcal{T}(1 \pm t)_r$ , and by Theorem 12 the statement follows. □

## 7.2 Multivariate case

The natural question is whether the results from the previous section translate to the general case. Given an  $n$ -variate polynomial  $p \in \mathbb{R}[\mathbf{u}]$ , it is natural to define

$$K_p(\mathbf{x}, \mathbf{y}) = \frac{1}{2^n} \sum p(x_1y_1 \pm \sqrt{(1 - x_1^2)(1 - y_1^2)}, \dots, x_ny_n \pm \sqrt{(1 - x_n^2)(1 - y_n^2)}),$$

where the summation is over all the possible  $2^n$  sign combinations.

As in the univariate case, from Lemma 4 we obtain that if  $p(\mathbf{u}) = \sum_{\alpha} p_{\alpha} T_{\alpha}(\mathbf{u})$  then  $K_p(\mathbf{x}, \mathbf{y}) = \sum_{\alpha} p_{\alpha} T_{\alpha}(\mathbf{x})T_{\alpha}(\mathbf{y})$ . Moreover, every kernel is of this form, as, given a kernel  $K(\mathbf{x}, \mathbf{y}) = \sum_{\alpha} c_{\alpha} T_{\alpha}(\mathbf{x})T_{\alpha}(\mathbf{y})$ , we have  $K = K_p$  for  $p(\mathbf{u}) = K(\mathbf{u}, \mathbf{1})$ , where  $\mathbf{1}$  denotes the all-ones vector.

We have then,  $p$  nonnegative on  $[-1, 1]^n$  if and only if  $K_p$  non-negative on  $[-1, 1]^{2n}$ . Next we prove  $p \in \mathcal{T}(1 \pm u_1, \dots, 1 \pm u_n)_r$  implies  $K_p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r,r}$ . The proof is basically the same as in the univariate case but requires induction. To do this, for each  $i = 1, \dots, n$  we define the operator  $\kappa^i : \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}] \rightarrow \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  by

$$\begin{aligned} p(u_1, \dots, u_i, \dots, u_n, \mathbf{x}, \mathbf{y}) &\mapsto \kappa^i(p)(\mathbf{u}, \mathbf{x}, \mathbf{y}) \\ &:= \frac{1}{2} \left( p(u_1, \dots, u_{i-1}, x_iy_i + \sqrt{(1 - x_i^2)(1 - y_i^2)}, u_{i+1}, \dots, u_n, \mathbf{x}, \mathbf{y}) \right. \\ &\quad \left. + p(u_1, \dots, u_{i-1}, x_iy_i - \sqrt{(1 - x_i^2)(1 - y_i^2)}, u_{i+1}, \dots, u_n, \mathbf{x}, \mathbf{y}) \right). \end{aligned}$$

Now, given  $p \in \mathbb{R}[\mathbf{u}]$  define  $P^i \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  for  $i = 0, \dots, n$  by  $P^0 = p$ , and  $P^i = \kappa^i(P^{i-1})$  for  $i = 1, \dots, n$ . Notice that in the  $i$ -th step the variable  $u_i$  is eliminated and the variables  $x_i$  and  $y_i$  introduced. Moreover, we obtain that  $K_p(\mathbf{x}, \mathbf{y}) = P^n$ . Now we rewrite the statement of Lemma 5 in a bit more general form.

**Lemma 6.** *Let  $q \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$ . Then, there are  $q_0 \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  and  $q_1 \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  such that*

$$q(u_1, \dots, u_{i-1}, x_iy_i \pm \sqrt{(1 - x_i^2)(1 - y_i^2)}, u_{i+1}, \dots, u_n, \mathbf{x}, \mathbf{y}) = q_0(\mathbf{u}, \mathbf{x}, \mathbf{y}) \pm \sqrt{(1 - x_i^2)(1 - y_i^2)}q_1(\mathbf{u}, \mathbf{x}, \mathbf{y}), \quad (26)$$

with  $\deg_{\mathbf{u}\mathbf{x}}(q_0) \leq \deg_{\mathbf{u}\mathbf{x}}(q)$ ,  $\deg_{\mathbf{u}\mathbf{y}}(q_0) \leq \deg_{\mathbf{u}\mathbf{y}}(q)$ ,  $\deg_{\mathbf{u}\mathbf{x}}(q_1) \leq \deg_{\mathbf{u}\mathbf{x}}(q) - 1$ , and  $\deg_{\mathbf{u}\mathbf{y}}(q_1) \leq \deg_{\mathbf{u}\mathbf{y}}(q) - 1$ .

Moreover, we have

$$\begin{aligned} \kappa^i(q^2) &= q_0^2 + (1 - x_i^2)(1 - y_i^2)q_1^2, \\ 2\kappa^i((1 - u_i)q^2) &= (1 - x_i)(1 + y_i)(q_0 - (1 + x_i)(1 - y_i)q_1)^2 + (1 + x_i)(1 - y_i)(q_0 - (1 - x_i)(1 + y_i)q_1)^2, \\ 2\kappa^i((1 + u_i)q^2) &= (1 + x_i)(1 + y_i)(q_0 + (1 - x_i)(1 - y_i)q_1)^2 + (1 - x_i)(1 - y_i)(q_0 + (1 + x_i)(1 + y_i)q_1)^2, \\ \kappa^i((1 - u_i^2)q^2) &= (1 - y_i^2)(x_iq_0 - y_i(1 - x_i^2)q_1)^2 + (1 - x_i^2)(y_iq_0 - x_i(1 - y_i^2)q_1)^2. \end{aligned}$$

*Proof.* First notice that by expanding and collecting terms, we obtain  $q_0$  and  $q_1$  satisfying all the degree requirements. The rest of the statement is proven following the proof of Lemma 5 line by line after equation (25), but only using equation (26) instead of (25).  $\square$

Now we can prove the main result of this section.

**Proposition 4.** *Let  $p \in \mathbb{R}[\mathbf{u}]$  be given. Then,  $p \in \mathcal{T}(1 \pm u_1, \dots, 1 \pm u_n)_r$  if and only if  $K_p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, 1 \pm y_n)_{r,r}$ .*

*Proof.* Assume  $K_p \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, 1 \pm y_n)_{r,r}$ , then as  $p(\mathbf{u}) = K(\mathbf{u}, \mathbf{1})$  we obtain  $p \in \mathcal{T}(1 \pm u_1, \dots, 1 \pm u_n)_r$ . Now, we prove the other direction. Assume  $p \in \mathcal{T}(1 \pm u_1, \dots, 1 \pm u_n)_r$  is given. Define  $P^i \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  for  $i = 0, \dots, n$  by  $P^0 = p$ , and  $P^i = \kappa^i(P^{i-1})$  for  $i = 1, \dots, n$ . We obtain then that  $K_p(\mathbf{x}, \mathbf{y}) = P^n$ . Write

$$P^0(\mathbf{u}, \mathbf{x}, \mathbf{y}) = p(\mathbf{u}) = \sum_{J^-, J^+ \subseteq [n]} \sigma_{J^-, J^+}(\mathbf{u}) \prod_{j \in J^-} (1 - u_j) \prod_{j \in J^+} (1 + u_j),$$

where, for all  $J^-, J^+ \subseteq [n]$ ,  $\sigma_{J^-, J^+} \in \Sigma[\mathbf{u}]$ , and  $|J^-| + |J^+| + \deg \sigma_{J^-, J^+} \leq r$ .

Notice that for all  $i \in [n]$  and  $q_1, q_2 \in \mathbb{R}[\mathbf{u}, \mathbf{x}, \mathbf{y}]$  such that  $u_i$  does not appear in  $q_2$  (that is  $\deg_{u_i}(q_2) = 0$ ), we have from the definition of  $\kappa^i$ , that  $\kappa^i(q_1 q_2) = q_2 \kappa^i(q_1)$ . Thus by linearity of  $\kappa^i$  and Lemma 6 we obtain by induction,

$$P^i = \sum_{\substack{J^-, J^+ \subseteq [n] \setminus [i] \\ I_x^-, I_x^+ \subseteq [i] \\ I_y^-, I_y^+ \subseteq [i]}} \sigma_{J^-, J^+, I_x^-, I_x^+, I_y^-, I_y^+} \prod_{j \in J^-} (1 - u_j) \prod_{j \in J^+} (1 + u_j), \prod_{j \in I_x^-} (1 - x_j) \prod_{j \in I_x^+} (1 + x_j) \prod_{j \in I_y^-} (1 - y_j) \prod_{j \in I_y^+} (1 + y_j),$$

where, for all  $J^-, J^+ \subseteq [n] \setminus [i]$ , all  $I_x^-, I_x^+ \subseteq [i]$  and all  $I_y^-, I_y^+ \subseteq [i]$  we have  $\sigma_{J^-, J^+, I_x^-, I_x^+, I_y^-, I_y^+} \in \Sigma[\mathbf{u}, \mathbf{x}, \mathbf{y}]$ , and  $|J^-| + |J^+| + |I_x^-| + |I_x^+| + \deg_{\mathbf{u}, \mathbf{x}} \sigma_{J^-, J^+, I_x^-, I_x^+, I_y^-, I_y^+} \leq r$  and  $|J^-| + |J^+| + |I_y^-| + |I_y^+| + \deg_{\mathbf{u}, \mathbf{y}} \sigma_{J^-, J^+, I_x^-, I_x^+, I_y^-, I_y^+} \leq r$ .

In particular,  $K_p = P^n \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, 1 \pm y_n)_{r,r}$ .  $\square$

We now conclude the section with a proof of Theorem 10.

### Proof of Theorem 10

*Proof.* Assume that we have a feasible solution of problem (24), i.e. we have a polynomial in the truncated pre-ordering, say

$$\mathbf{x} \mapsto 1 + \sum_{\alpha \in \mathbb{N}_d^n \setminus \{\mathbf{0}\}} 2^{\omega(\alpha)} p_\alpha T_\alpha(\mathbf{x}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n)_r.$$

By Proposition 4, we then have that

$$(\mathbf{x}, \mathbf{y}) \mapsto 1 + \sum_{\alpha \in \mathbb{N}_d^n \setminus \{\mathbf{0}\}} p_\alpha \hat{T}_\alpha(\mathbf{x}) \hat{T}_\alpha(\mathbf{y}) \in \mathcal{T}(1 \pm x_1, \dots, 1 \pm x_n; 1 \pm y_1, \dots, 1 \pm y_n)_{r,r}.$$

By Proposition 3, we have that this kernel satisfies property (19). Thus this kernel is feasible for the problem (20). The required result follows.  $\square$

## 8 Concluding remarks

Our main results may best be framed in terms of error bounds for the SDP hierarchy (6). In particular, we have shown the following sequence of inequalities:

$$\begin{aligned}
f_{\min} - \hat{f}_{(r)} &\leq \Theta_{n,d}^r \|f\|_{1,T} \quad (\text{by Theorem 9}) \\
&\leq \frac{\pi^2 d^2 n^2}{r^2} \|f\|_{1,T} \quad (\text{by Proposition 2 if } n|r \text{ and } r \geq \pi dn) \\
&\leq \frac{\pi^2 d^2 n^2}{r^2} \|f\|_1 \quad (\text{by (12)}) \\
&\leq \frac{\pi^2 d^2 n^2}{r^2} \binom{n+d}{d}^{\frac{1}{2}} C_d \|f\|_{\ell_\infty} \quad (\text{by Bohnenblust-Hille ineq. (13)}) \\
&\leq \frac{\pi^2 d^2 n^2}{r^2} \binom{n+d}{d}^{\frac{1}{2}} C_d (f_{\max} - f_{\min}) \quad (\text{w.l.o.g. if } f(0) = 0).
\end{aligned}$$

We may readily compare the outermost inequality to the result by Laurent and Slot [17], as reproduced in Theorem 3, to see that we obtain a very similar result. The added value of our analysis is therefore in the first inequality, since, for special classes of polynomials,  $\|f\|_{1,T}$  may relate more favorably to  $\|f\|_{\ell_\infty}$  than in the general case. We should mention that this is not the first time where changing or comparing norms gives new insights in polynomial optimization. For example, in [5] the authors improve the known dependence of the general Putinar's theorem [22] on the number of variables, by switching from the supremum norm to the 1-norm in the Bernstein basis.

The object  $\Theta_{n,d}^r$  is of independent interest, as it corresponds to families of positive approximation operators, and may be upper bounded by an SDP as in Theorem 10. A key question is if one could find an analytical expression for (an upper bound on)  $\Theta_{n,d}^r$  in terms of  $r, n, d$  in a way that does not involve the product of univariate (Jackson) kernels.

Our research links two, seemingly different, topics:

1. Deriving error bounds for SDP hierarchies for polynomial optimization, in the spirit of [17] and [3].
2. Constructing optimal multivariate approximation kernels using SDP, in the spirit of [10].

In particular, we showed how kernels that are optimized for 1-norm polynomial approximation on the hypercube yield error bounds for SDP hierarchies of the Schmüdgen type.

A natural question is if our analysis may be extended to SDP hierarchies that use the Putinar Positivstellensatz [22] instead of the one by Schmüdgen. This is of theoretical as well as computational interest, since the famous Lasserre SDP hierarchy [11] for polynomial optimization is based on the Putinar Positivstellensatz. In particular, Baldi and Slot [4] have recently extended the results by Laurent and Slot [17] to obtain the best-known effective version of the Putinar Positivstellensatz. It would be interesting to investigate if our approach allows a similar extension. The main obstacle is that our key technical result, namely Corollary 1, only applies to pre-orderings, and not to quadratic modules.

### Acknowledgments

The authors would like to thank Monique Laurent for valuable feedback on the first version of this manuscript, and two anonymous referees for their helpful comments that led to an improved presentation of the results.

### Appendix: numerical SDP bounds

In this appendix we present the SDP upper bound (24) on  $\Theta_{n,d}^r$  for different values of  $(n, r, d)$ , as indicated in the tables. We may mention that, when  $r$  is a multiple of  $n$  in the tables, the SDP upper bound on  $\Theta_{n,d}^r$  is strictly lower than the value obtained from products of univariate Jackson kernels. In particular, for the appropriate values the tables where  $r$  is a multiple of  $n$ , the inequality in Proposition 2 is strict. The values in the tables were computed using the SDP solver MOSEK [19] under Matlab.

	d=1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r=1	0.5000														
2	0.2929	0.5556													
3	0.1910	0.5001	0.6001												
4	0.1340	0.3320	0.5284	0.6214											
5	0.0991	0.2929	0.5001	0.5556	0.6405										
6	0.0762	0.2159	0.3585	0.5177	0.5783	0.6527									
7	0.0604	0.1910	0.3140	0.5001	0.5374	0.6001	0.6641								
8	0.0490	0.1502	0.2929	0.3723	0.5123	0.5556	0.6108	0.6723							
9	0.0406	0.1340	0.2309	0.3320	0.5001	0.5284	0.5707	0.6214	0.6802						
10	0.0341	0.1101	0.2050	0.3066	0.3834	0.5090	0.5418	0.5848	0.6310	0.6862					
11	0.0291	0.0991	0.1910	0.2929	0.3461	0.5001	0.5217	0.5556	0.6001	0.6405	0.6921				
12	0.0251	0.0839	0.1592	0.2389	0.3202	0.3909	0.5070	0.5336	0.5668	0.6070	0.6467	0.6968			
13	0.0219	0.0762	0.1434	0.2160	0.3027	0.3587	0.5001	0.5177	0.5447	0.5793	0.6152	0.6542	0.7015		
14	0.0193	0.0660	0.1342	0.2012	0.2930	0.3341	0.4169	0.5056	0.5311	0.5606	0.6041	0.6271	0.6960	0.7603	
15	0.0171	0.0604	0.1179	0.1925	0.2570	0.3188	0.3933	0.5001	0.5159	0.5409	0.5772	0.6120	0.6311	0.6904	0.9302
16	0.0153	0.0552	0.1128	0.1830	0.2544	0.3114	0.3764	0.4940	0.5147	0.5434	0.5781	0.6136	0.6471	0.8190	0.9657
17	0.0137	0.0539	0.1124	0.1804	0.2549	0.3061	0.3706	0.4850	0.5072	0.5323	0.5644	0.5944	0.6284	0.7332	0.8572
18	0.0137	0.0508	0.1059	0.1712	0.2273	0.2978	0.3590	0.4652	0.5030	0.5282	0.5776	0.5872	0.6259	0.6649	0.8315

Table 2: The SDP upper bound (24) on  $\Theta_{n,d}^r$  for  $n = 1$

	d=1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r=1	0.7500														
2	0.5001	0.7648													
3	0.3455	0.7501	0.8278												
4	0.2501	0.5295	0.7575	0.8338											
5	0.1883	0.5001	0.7501	0.8025	0.8620										
6	0.1465	0.3720	0.5798	0.7566	0.8212	0.8640									
7	0.1170	0.3455	0.5263	0.7501	0.7859	0.8401	0.8805								
8	0.0955	0.2706	0.5001	0.5920	0.7557	0.7968	0.8459	0.8800							
9	0.0794	0.2501	0.4051	0.5537	0.7501	0.7776	0.8142	0.8566	0.8918						
10	0.0670	0.2038	0.3668	0.5167	0.6166	0.7549	0.7882	0.8219	0.8633	0.8912					
11	0.0573	0.1883	0.3455	0.5001	0.5722	0.7501	0.7712	0.8025	0.8359	0.8708	0.8999				
12	0.0496	0.1581	0.2917	0.4140	0.5378	0.6191	0.7542	0.7801	0.8123	0.8413	0.8747	0.8989			
13	0.0433	0.1465	0.2657	0.3852	0.5132	0.5884	0.7501	0.7674	0.7925	0.8222	0.8503	0.8795	0.9063		
14	0.0381	0.1259	0.2501	0.3592	0.5001	0.5501	0.6340	0.7547	0.7761	0.8018	0.8304	0.8581	0.8833	0.9058	
15	0.0338	0.1170	0.2175	0.3455	0.4288	0.5294	0.5986	0.7501	0.7645	0.7865	0.8104	0.8403	0.8606	0.8902	0.9283
16	0.0302	0.1025	0.1997	0.3039	0.4001	0.5114	0.5678	0.6644	0.7544	0.7720	0.7983	0.8213	0.8535	0.8728	0.9043
17	0.0271	0.0955	0.1890	0.2810	0.3759	0.5007	0.5440	0.6154	0.7502	0.7646	0.7917	0.8089	0.8388	0.8604	0.8881
18	0.0245	0.0877	0.1678	0.2659	0.3600	0.4742	0.5336	0.6081	0.6705	0.7560	0.7771	0.8029	0.8307	0.8489	0.8624

Table 3: The SDP upper bound (24) on  $\Theta_{n,d}^r$  for  $n = 2$

	d=1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r=1	0.8333														
2	0.5918	0.8401													
3	0.4727	0.8334	0.9026												
4	0.3546	0.6331	0.8500	0.9048											
5	0.2787	0.6092	0.8334	0.8889	0.9319										
6	0.2197	0.4952	0.6967	0.8408	0.9016	0.9322									
7	0.1785	0.4727	0.6532	0.8334	0.8795	0.9201	0.9464								
8	0.1467	0.3804	0.6132	0.7150	0.8410	0.8847	0.9246	0.9455							
9	0.1230	0.3582	0.5456	0.6925	0.8334	0.8732	0.9039	0.9343	0.9547						
10	0.1042	0.2961	0.4996	0.6374	0.7380	0.8402	0.8797	0.9073	0.9378	0.9537					
11	0.0895	0.2788	0.4751	0.6189	0.7097	0.8334	0.8661	0.8959	0.9197	0.9439	0.9602				
12	0.0776	0.2348	0.4129	0.5523	0.6682	0.7461	0.8410	0.8716	0.9027	0.9221	0.9462	0.9591			
13	0.0679	0.2207	0.3823	0.5289	0.6421	0.7260	0.8335	0.8621	0.8887	0.9118	0.9305	0.9503	0.9642		
14	0.0599	0.1897	0.3598	0.4921	0.6181	0.6826	0.7598	0.8402	0.8686	0.8927	0.9161	0.9316	0.9536	0.9634	
15	0.0533	0.1786	0.3201	0.4759	0.5752	0.6643	0.7374	0.8340	0.8585	0.8838	0.9029	0.9228	0.9385	0.9555	0.9677
16	0.0477	0.1560	0.2968	0.4278	0.5470	0.6348	0.7039	0.7671	0.8479	0.8647	0.8911	0.9069	0.9282	0.9444	0.9591

Table 4: The SDP upper bound (24) on  $\Theta_{n,d}^r$  for  $n = 3$

## References

- [1] Francesco Altomare and Michele Campiti. *Korovkin-type Approximation Theory and its Applications*. De Gruyter, 2011.
- [2] Gustavo Araújo, P. Jiménez-Rodríguez, Gustavo A. Muñoz-Fernández, Juan B. Seoane-Sepúlveda. Equivalent norms in polynomial spaces and applications, *Journal of Mathematical Analysis and Applications*, 445(2): 1200–1220, 2017.

	d=1	2	3	4	5	6	7	8	9	10	11
r=1	0.8750										
2	0.6465	0.8788									
3	0.5510	0.8751	0.9385								
4	0.4272	0.6943	0.8919	0.9376							
5	0.3575	0.6770	0.8751	0.9286	0.9613						
6	0.2866	0.5713	0.7679	0.8841	0.9380	0.9604					
7	0.2398	0.5530	0.7320	0.8788	0.9226	0.9543	0.9723				
8	0.1985	0.4624	0.6876	0.7828	0.8883	0.9255	0.9575	0.9713			
9	0.1689	0.4425	0.6390	0.7680	0.8779	0.9183	0.9443	0.9662	0.9784		
10	0.1436	0.3773	0.5893	0.7122	0.8087	0.8854	0.9235	0.9453	0.9679	0.9776	
11	0.1244	0.3593	0.5641	0.6976	0.7873	0.8794	0.9130	0.9388	0.9562	0.9731	0.9821

Table 5: The SDP upper bound (24) on  $\Theta_{n,d}^r$  for  $n = 4$

- [3] Francis Bach and Alessandro Rudi. Exponential Convergence of Sum-of-Squares Hierarchies for Trigonometric Polynomials, *SIAM Journal on Optimization*, 33(3):2137–2159, 2023.
- [4] Lorenzo Baldi and Lucas Slot. Degree Bounds for Putinar’s Positivstellensatz on the Hypercube. *SIAM Journal on Applied Algebra and Geometry*, 8(1):1–25, 2024.
- [5] Lorenzo Baldi, Bernard Mourrain, and Adam Parusinski. On Łojasiewicz Inequalities and the Effective Putinar’s Positivstellensatz. *Journal of Algebra*, 662, 741–767, 2025.
- [6] Sergei Bernstein. Sur l’ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné. *Mémoire couronné, Brussels*, 1912.
- [7] Philipp J. di Dio and Konrad Schmüdgen, The multidimensional truncated moment problem: The moment cone, *Journal of Mathematical Analysis and Applications*, 511(1), 126066, 2022.
- [8] Dunham Jackson. *Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung*. Universität Göttingen, June 1911.
- [9] Dunham Jackson. On Approximation by Trigonometric Sums and Polynomials. *Transactions of the American Mathematical Society*, 13(4):491–515, Oct 1912.
- [10] Felix Kirschner and Etienne de Klerk. Construction of Multivariate Polynomial Approximation Kernels via Semidefinite Programming. *SIAM Journal on Optimization*, 33(2):513–537, 2023.
- [11] Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11:796–817, 2001.
- [12] Jean-Bernard Lasserre. Best  $\ell_1$ -approximation of nonnegative polynomials by sums of squares. arXiv preprint: <https://arxiv.org/pdf/1012.3271.pdf>, December 15th, 2010.
- [13] Jean-Bernard Lasserre. A sum of squares approximation of nonnegative polynomials. *SIAM Review* 49:651–669, 2007.
- [14] Jean-Bernard Lasserre and Tim Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. *Math. Zeitschrift* 256:99–112, 2006.
- [15] Jean-Bernard Lasserre, Eduard Pauwels, and Mihai Putinar. *The Christoffel-Darboux Kernel for Data Analysis*, Cambridge University Press, 2021.
- [16] Monique 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.
- [17] Monique Laurent and Lucas Slot. An effective version of Schmüdgen’s Positivstellensatz for the hypercube. *Optimization Letters* 17:515–530, 2023.
- [18] Murray Marshall. *Positive Polynomials and Sums of Squares*, Mathematical Surveys and Monographs 146, American Mathematical Society, 2008.
- [19] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0*, 2019.

- 
- [20] Victoria Powers and Bruce Reznick. Polynomials that are positive on an interval. *Trans. Amer. Math. Soc.*, **352**, 4677–4692, 2000.
- [21] Alexander Prestel, Charles N. Delzell. *Positive Polynomials - From Hilbert's 17th Problem to Real Algebra*, Springer Monographs in Mathematics, Springer, 2001.
- [22] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Ind. Univ. Math. J.* 42:969–984, 1993.
- [23] Mihai Putinar. A note on Tchakaloff's theorem. *Proceedings of the American Mathematical Society*, 125(8), 2409–2414 (1997)
- [24] Hans Richter. Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Blatter der DGVFM*, 3, 147–162, 1957.
- [25] Theodore J. Rivlin. *Chebyshev Polynomials*. Second edition. Courier Dover Publications, 2020.
- [26] Konrad Schmüdgen. The  $K$ -moment problem for compact semi-algebraic sets. *Mathematische Annalen* 289, 203–206, 1991.
- [27] Ljubomir Tchakaloff. Formules de cubature mécanique à coefficients non négatifs, *Bull. Sci. Math.*, 81, 123–134 (1957)
- [28] Lloyd N. Trefethen. Multivariate polynomial approximation in the hypercube. *Proceedings of the American Mathematical Society*, 145(11):4837–4844, 2017.
- [29] Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. The kernel polynomial method. *Rev. Mod. Phys.*, 78:275–306, Mar 2006.