FULL LENGTH PAPER

Series A

Convergence rates for sums-of-squares hierarchies with correlative sparsity

Milan Korda[1,2] · Victor Magron[1] · Rodolfo Ríos-Zertuche[3]

Abstract

This work derives upper bounds on the convergence rate of the moment-sum-of-squares hierarchy with correlative sparsity for global minimization of polynomials on compact basic semialgebraic sets. The main conclusion is that both sparse hierarchies based on the Schmüdgen and Putinar Positivstellensätze enjoy a polynomial rate of convergence that depends on the size of the largest clique in the sparsity graph but not on the ambient dimension. Interestingly, the sparse bounds outperform the best currently available bounds for the dense hierarchy when the maximum clique size is sufficiently small compared to the ambient dimension and the performance is measured by the running time of an interior point method required to obtain a bound on the global minimum of a given accuracy.

Contents

✉ Rodolfo Ríos-Zertuche
rodolforiosz@gmail.com

  Milan Korda
  korda@laas.fr

  Victor Magron
  vmagron@laas.fr

[1] LAAS, CNRS, 7 avenue du colonel Roche, 31400 Toulouse, France

[2] Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 16626 Prague, Czech Republic

[3] Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, Norway

1 Introduction

This work provides rates of convergence for the sums-of-squares hierarchy with correlative sparsity. For a positive $n \in \mathbb{N}$, consider the polynomial optimization problem

$$f_{\min} := y \min_{x \in S(\mathbf{g})} f(x)$$

where f is an element of the ring $\mathbb{R}[x]$ of polynomials in $x = (x_1, \ldots, x_n)$, and $S(\mathbf{g})$ is a basic compact semialgebraic set determined by a finite collection of polynomials $\mathbf{g} = \{g_1, \ldots, g_{\bar{k}}\}$ by $S(\mathbf{g}) = \{x \in \mathbb{R}^n : g_i(x) \geq 0, \ i = 1, \ldots, \bar{k}\}$. An approach to attack this problem, first proposed by Lasserre [10] and Parrilo [20], is as follows: Imagine we knew that $f(x) - \lambda$ could be written as

$$f(x) - \lambda = \sum_{j=0}^{\bar{k}} \sigma_j g_j(x) \quad \text{or} \quad f(x) - \lambda = \sum_{J \subseteq \{1,\ldots,\bar{k}\}} \sigma_J \prod_{j \in J} g_j(x),$$

with $g_0(x) = 1$ and σ_j and σ_J being sum-of-squares (SOS) polynomials. Then the right-hand sides of each of these equations would be clearly nonnegative on $S(\mathbf{g})$, so we would know that $f_{\min} \geq \lambda$. By bounding the degree of the SOS polynomials, we obtain the following two hierarchies of lower bounds:

$$\text{lb}_q(f, r) = \max \left\{ \lambda \in \mathbb{R} : f - \lambda = \sum_{j=0}^{\bar{k}} \sigma_j g_j, \deg(\sigma_j g_j) \leq 2r, \ \sigma_j \in \Sigma[x] \right\},$$

$$\text{lb}_p(f, r) = \max \left\{ \lambda \in \mathbb{R} : f - \lambda = \sum_{J \subseteq \{1,\ldots,\bar{k}\}} \sigma_J \prod_{j \in J} g_j, \deg\left(\sigma_J \prod_{j \in J} g_j\right) \leq 2r, \ \sigma_J \in \Sigma[x] \right\},$$

where $\Sigma[x]$ is the convex cone of all sum-of-squares polynomials. These satisfy $\text{lb}_q(f, r) \leq \text{lb}_p(f, r) \leq f_{\min}$. The lower bound $\text{lb}_q(f, r)$ is associated to a so-called *quadratic module certificate*, while $\text{lb}_p(f, r)$ corresponds to a *preordering certificate*; this terminology is justified by the definitions in Sect. 1.2. The well-known Putinar and Schmüdgen Positivstellensätze [21, 23], respectively, guarantee that these bounds converge to $f_{\min}$ as $r \to +\infty$, the former with the additional assumption that the associated quadratic module be Archimedian.[1] Here we prove sparse quantitative versions of these results.

---

[1] This means that there are $R > 0$ and $\sigma_j \in \Sigma[x]$ such that $R - \|x\|^2 = \sum_j \sigma_j g_j(x)$.

Table 1 Known results on the asymptotic error of Lasserre's hierarchies of lower bounds; based in part on [25, Table 1]

Domain $S(\mathbf{g})$	Error	Certificate	Ref
Archimedean	$O(1/\log(r)^c)$	Quadratic module	[18]
Archimedean	$O(1/r^c)$	Quadratic module	[2]
General	$O(1/r^c)$	Preordering	[24]
General	$O(1/r^c)$	Quadratic module & uniform denominators	[16]
S^{n-1}	$O(1/r^2)$	quadratic module/preordering	[4]
$\{0, 1\}^n$	See [26]	Quadratic module/preordering	[26]
B^n	$O(1/r^2)$	Quadratic module/preordering	[25]
$[-1, 1]^n$	$O(1/r^2)$	Preordering	[14]
Δ^n	$O(1/r)$	Preordering	[9]
Δ^n	$O(1/r^2)$	Preordering	[25]

The domain $S(\mathbf{g})$ is assumed to be compact in all cases, $\epsilon \in [0, 1/2)$, $c > 0$, S^{n-1} is the unit sphere, B^n is the unit ball, Δ^n is the standard simplex

Polynomial optimization schemes have generated substantial interest due to their abundant fields of application; see for example [12, 13]. The first proof of convergence, without a convergence rate, was given by Lasserre [10] using the Archimedian Positivstellensatz due to Putinar [21]. Eventually, rates of convergence were obtained; initially in [18] these were logarithmic in the degree of the polynomials involved, and later on they were improved [2, 4, 14, 25] (using ideas of [3, 22]) to polynomial rates; refer to Table 1. The crux of the argument used to obtain those rates is a bound of the deformation incurred by a polynomial strictly-positive on the domain of interest, as it passes through an integral operator that closely approximates the identity and is associated to a strictly-positive polynomial kernel that is itself composed of sums of squares and similar to the Christoffel–Darboux and Jackson kernels (see Definition 10). More recently, results showing an exponential bound on the convergence rate was obtained in [1] with the additional assumption on the positive definiteness of the Hessian at all global minimizers.

The techniques used to obtain these results generally involve linear operators on the space of polynomials (mostly Christoffel–Darboux kernel operators; see [25]) that are close to the identity and that, for positive polynomials, are easily (usually, by construction) proved to output polynomials that are sums of squares and/or of their products with the functions in $\mathbf{g}$. All of these results deal, however, with the dense case.

In this work, we treat the case where the problem possesses the so-called correlative sparsity, where each function g_i depends only on a certain subset of variables and the function f decomposes as a sum of functions depending only on these subsets of variables. This structure can be exploited in order to define sparse lower bounds that are cheaper to compute but possibly weaker. Nevertheless, these sparse lower bounds allow one to tackle large-scale polynomial optimization problems arising from various applications including roundoff error bounds in computer arithmetic, quantum

correlations and robustness certification of deep networks; see the recent survey [15]. In [11] Lasserre proved that these sparse lower bounds converge as the degree of the SOS multipliers tends to infinity provided the variable groups satisfy the so-called running intersection property (see Definition 1). A shorter and more direct proof was provided in [6], and was adapted in [17] to obtain a sparse variant of Reznick's Positivstellensatz. In this work, we show polynomial rates of convergence for sparse hierarchies based on both Schmüdgen and Putinar Positivstellensätze. Importantly, we obtain rates that depend only on the size of the largest clique in the sparsity graph rather than the overall ambient dimension. This allows the perhaps surprising conclusion that, asymptotically, the sparse hierarchy is more accurate than the dense hierarchy for a given computation time of an optimization method, provided that the size of the largest clique is no more than the square root of the ambient dimension. This assumes that the running time of the optimization method is governed by the size of the largest PSD block and the number of such blocks in the semidefinite programming reformulations of the dense and sparse SOS problems which is the case for the interior point method as well as the most commonly used first-order methods.

To the best of our knowledge, these are the first quantitative results of this kind. Our proof techniques rely on an adaption of [6] and utilize heavily the recent results from [2, 14], and can thus be seen as a generalization of these works to the sparse setting.

Since our results are very technical and their full statement necessitates the introduction of a lot of notation, we have prepared a rough summary to help the reader get a glance at what they are, presented next. We urge the reader to mind the fact that we have not fully spelled-out all the details and definitions, for which we refer to the next section.

Definition 1 A collection $\{J_1, \ldots, J_\ell\}$ of subsets of $\{1, \ldots, n\} \supset J_j$ satisfies the *running intersection property* if for all $1 \leq k \leq \ell - 1$ we have

$$J_{k+1} \cap \bigcup_{j=1}^{k} J_j \subset J_s \quad \text{for some } s \leq k. \tag{RIP}$$

Theorem 2 (Rough summary of results) *Let:*

- *$n > 0$, $\ell > 0$, and $\bar{k}$ be positive integers,*
- *$\mathbf{r}_1, \ldots, \mathbf{r}_\ell \in \mathbb{N}^n$, $\mathbf{r}_j = (r_{j,1}, \ldots, r_{j,n})$ be some multi-indices with $r_{j,i} \geq 1$,*
- *$J_1, \ldots, J_\ell \subset \{1, \ldots, n\}$ be sets of indices satisfying (RIP),*
- *$p_1, \ldots, p_\ell$ be polynomials such that p_j depends only on the variables x_i with $i \in J_j$, and its degree in the variable x_i is $\leq r_{j,i}$,*
- *$p = p_1 + p_2 + \cdots + p_\ell$ be a polynomial that this a sum of the polynomials $p_1, \ldots, p_\ell$,*

We will denote by $|J_j|$ the cardinality of the set J_j.

i. (*Schmüdgen-type, Theorem* 6) *Assume that, for large-enough $c_1 > 0$ (determined explicitly in the statement of Theorem* 6 *and depending only on n and $J_1, \ldots, J_\ell$),*

$$p(x) \geq c_1 \frac{\|p\|_{L^\infty([-1,1]^n)}}{r_{j,i}^{\frac{2}{3+\max_j |J_j|}}} \tag{1}$$

for all $1 \leq i \leq n$, $1 \leq j \leq \ell$, and $x \in [-1, 1]^n$. Then p can be written as a sum $p = h_1 + \cdots + h_\ell$ of polynomials h_j that belong to the respective preorderings generated by the polynomials $1 - x_i^2$ with $i \in J_j$ and only depend on the variables x_i with $i \in J_j$. This means that there are polynomials $\sigma_{j,K}$ that are sums of squares, that depend only on x_i for $i \in J_j$, and such that

$$h_j = \sum_{K \subset J_j} \sigma_{j,K} \prod_{m \in K} (1 - x_m^2).$$

The sum is taken over all (possibly empty) subsets K of J_j; the product on the right is understood to be 1 when $K = \emptyset$. Moreover, the degree of each term $\sigma_{j,K} \prod_{m \in J_j} (1 - x_m^2)$ in the variable x_i is bounded by $r_{j,i}$.

ii. (*Putinar-type, Theorem* 8) *Additionally we let*

- $K_1, \ldots, K_\ell \subset \{1, \ldots, \bar{k}\}$ *be sets of indices,*
- $g_1, \ldots, g_{\bar{k}}$ *be polynomials such that, if $m \in K_j$ then g_m depends only on the variables x_i with $i \in J_j$, and satisfying some additional technical assumptions.*[2]

Now, instead of assuming (1), *we assume that, for large-enough $c_2, c_3 > 0$ (determined more-or-less explicitly in the statement of Theorem* 8 *and depending only on n, $J_1, \ldots, J_\ell$, $g_1, \ldots, g_{\bar{k}}$),*

$$p(x) \geq c_2 \frac{(\|p\|_{L^\infty(S(\mathbf{g}))} \deg p_j \sum_i \mathrm{Lip} \, p_i)^{c_3}}{r_{j,i}^{\frac{1}{13+3|J_j|}}},$$

for x in the set $S(\mathbf{g}) = \{x \in \mathbb{R}^n : g_j(x) \geq 0, \ j = 1, \ldots, \bar{k}\}$, and for all $1 \leq i \leq n$, $1 \leq j \leq \ell$. Then p can be written as a sum $p = h_1 + \cdots + h_\ell$ of polynomials h_j that belong to the respective quadratic modules generated by the polynomials $(g_i)_{i \in K_j}$ and only depend on the variables $(x_i)_{i \in J_j}$. This means that there are polynomials $\sigma_{j,0}$ and $\sigma_{j,k}$ that are sums of squares, that depend only on x_i for $i \in J_j$, and such that

$$h_j = \sigma_{j,0} + \sum_{k \in K_j} \sigma_{j,k} g_k.$$

---

[2] Apart from assumptions of scale (3) and an Archimedean condition (4), for this simplified version of the results we have for simplicity taken $\mathsf{L}_j = 1$ and $\mathsf{c}_j = 1$ in the statement of Theorem 8. With these values, the exponent of ε in (5) is $(238 + 55|J_j|)/9$, which we estimate from below with the simpler expression $26 + 6|J_j|$ to derive the statement of Theorem 2.

Moreover, the degree of each term $\sigma_{j,0}$ and $\sigma_{j,k}g_k$ in the variable x_i is bounded by $r_{j,i} + 2$.

Although the exponents of $r_{j,i}$ in the above statement are often much smaller than 2, hence making the rates slower than those that had been obtained (see Table 1) for the dense case, we have also analyzed the complexity involved in solving the corresponding optimization problems, showing that the sparse hierarchies may outperform the dense ones in certain situations.

We proceed to summarize this analysis. When using the sums-of-squares hierarchies, we use, as a proxy for the complexity necessary to obtain a certificate that a lower bound of the minimum of a polynomial p has been found to $\varepsilon > 0$ accuracy, the size and number of the positive-semidefinite (PSD) blocks in the corresponding semidefinite program (SDP).

We denote:

- $B_{\text{dense}}(\varepsilon)$ the size of PSD block in the dense case, which is equal to the number $\binom{n+r}{r}$ of monomials of degree r in n variables, and in our argument we take r to be of the order of $O(\varepsilon^{-1/2})$ as in the best results listed in Table 1,
- $B_{\text{sparseSchm}}(\varepsilon)$ the size of the largest PSD block in the sparse case of Theorem 2i. (i.e., optimizing over $[-1, 1]^n$ using a Schmüdgen-style scheme), multiplied by the number ℓ of blocks,
- $B_{\text{sparsePut}}(\varepsilon)$ the size of the largest PSD block in the sparse case of Theorem 2ii. (i.e., optimizing over $S(\mathbf{g})$ using a Putinar-style scheme), multiplied by the number ℓ of blocks.

The quotients

$$\frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} \quad \text{and} \quad \frac{B_{\text{sparsePut}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)}$$

being < 1 is thus indicative of our rates being better than the dense ones.

Proposition 3 (Summary of SDP size results) *With some technical assumptions, we have:*

1. *(Proposition 7) If $n > |J_j|^2 + 3|J_j|$ for all $1 \leq j \leq \ell$, then, for minimizing p over $[-1, 1]^n$, we have*

$$\lim_{\varepsilon \searrow 0} \frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = 0.$$

2. *(Proposition 9) If $n > 6|J_j|^2 + 26|J_j|$ for all $1 \leq j \leq \ell$, then, for minimizing p over $S(\mathbf{g})$, we have*

$$\lim_{\varepsilon \searrow 0} \frac{B_{\text{sparsePut}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = 0.$$

In other words, our rates improve on those already found for the dense case as long as n is sufficiently large with respect to the sizes $|J_j|$ of the blocks of variables indexed by the sets $J_1, \ldots, J_\ell$, and $\varepsilon > 0$ is sufficiently small.

Let us turn to some concrete examples. The functions in our examples were considered before in [27, §6.1], where sums-of-squares algorithms leveraging sparsity were benchmarked on them.

Example 4 Consider the chain singular function for $x \in [-1, 1]^n$:

$$f_{\text{cs}}(x) = \sum_{i \in P} \left((x_i + 10x_{i+1})^2 + 5(x_{i+2} - x_{i+3})^2 + (x_{i+1} - 2x_{i+2})^4 + 10(x_i - 10x_{i+3})^4 \right),$$

where $P = \{1, 3, 5, \ldots, n - 3\}$ and n is a multiple of 4. Then we can take $J_j = \{j, j + 1, j + 2, j + 3\}$ for $j = 1, \ldots, n - 3$, so that these sets satisfy (RIP) and $|J_j| = 4$. The proofs of Propositions 7 and 9 show that in this case we have, for large-enough n and as $\varepsilon \searrow 0$,

$$\frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = O(\varepsilon^{\frac{n}{2}-14}) \quad \text{and} \quad \frac{B_{\text{sparsePut}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = O(\varepsilon^{\frac{n}{2}-62}).$$

Example 5 Consider the Broyden banded function, defined for each $n \in \mathbb{N}$ by

$$f_{\text{Bb}}(x) = \sum_{i=1}^{n} \left(x_i(2 + 5x_i^2) + 1 - \sum_{j \in P_i}(1 + x_j)x_j \right)^2,$$

where $P_i = \{j : j \neq i, \ \max(1, i-5) \leq j \leq \min(n, i+1)\}$, on the box $[-1, 1]^n$. Then $J_i = P_i \cup \{i\}$, and $|J_i| \leq 7$, and these sets satisfy (RIP). The proofs of Propositions 7 and 9 show that in this case we have, for large-enough n and as $\varepsilon \searrow 0$,

$$\frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = O(\varepsilon^{\frac{n}{2}-35}) \quad \text{and} \quad \frac{B_{\text{sparsePut}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = O(\varepsilon^{\frac{n}{2}-143}).$$

The paper is organized as follows. The results are detailed below in Sect. 1.2 and further discussed in Sect. 1.2.1, after a brief interlude to establish some notations in Sect. 1.1. Some machinery is developed in Sects. 2 and 3, regarding variants of the Jackson kernel and some approximation theory, respectively, and the proofs of the main theorems are presented in Sect. 4.

1.1 Notation

Denote by $\mathbb{R}$ the set of real numbers, by $\mathbb{N}$ the set of positive integers, and by $\mathbb{N}_0 = \{0, 1, \ldots\}$ the set of nonnegative integers. Denote by $e_1, \ldots, e_n$ the vectors of the standard basis of Euclidean space $\mathbb{R}^n$.

For a Lipschitz continuous function $f \colon [-1, 1]^n \to \mathbb{R}$, we set

$$\text{Lip } f = \max \left(1, \sup_{x, y \in [-1, 1]^n} \frac{|f(x) - f(y)|}{\|x - y\|} \right).$$

We take this to be at least 1 to simplify estimates below.

A multi-index $I = (i_1, \ldots, i_n) \in \mathbb{N}_0^n$ is an n-tuple of nonnegative integers i_k, and its weight is denoted by

$$|I| = \sum_{k=1}^{n} i_k.$$

For a multi-index $I = (i_1, \ldots, i_n) \in \mathbb{N}_0^n$ and $J \subset \{1, \ldots, n\}$, we write $I \subseteq J$ to indicate that the index k of every nonzero entry $i_k > 0$ is contained in J, that is, $k \in J$ for all $1 \le k \le n$ such that $i_k > 0$.

Similarly, given a multi-index $I = (i_1, \ldots, i_n) \in \mathbb{N}_0^n$ and a subset $J \subseteq \mathbb{N}_0$, we let I_J be the multi-index whose k-th entry is either i_k if $k \in J$ or 0 if $k \notin J$.

For two multi-indices I and I', we write $I \le I'$ if the entrywise inequalities $i_k \le i_k'$ hold for all $1 \le k \le n$.

We distinguish two special multi-indices:

$$\mathbf{1} = (1, 1, \ldots, 1) \quad \text{and} \quad \mathbf{2} = (2, 2, \ldots, 2).$$

We denote $x^I = x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}$. Also, we denote the *Hamming weight* of $I \in \mathbb{N}_0^n$ by

$$w(I) = \#\{k : i_k > 0, \ 1 \le k \le n\}.$$

In other words, $w(I)$ is the number of nonzero entries in I.

We denote the space of polynomials in n variables by $\mathbb{R}[x]$, and within this set we distinguish the subspace $\mathbb{R}[x]_d$ of polynomials of total degree at most d. We denote, for a polynomial $p(x) = \sum_I c_I x^I$, by $\overline{\deg} \, p$ the vector whose i-th entry is the degree of p in x_i,

$$\overline{\deg} \, p = \left(\max_{c_I \ne 0} i_1, \max_{c_I \ne 0} i_2, \ldots, \max_{c_I \ne 0} i_n \right).$$

Observe that $\deg p \le \left| \overline{\deg} \, p \right| = \sum_{k=1}^{n} \max_{c_I \ne 0} i_k$.

Set also

$$\mathcal{I}_p = \{I \in \mathbb{N}_0^n : c_I \ne 0\}.$$

Given a subset $J \subset \{1, \ldots, n\}$, we let $\mathbb{R}[x_J]$ denote the set of polynomials in the variables $\{x_j\}_{j \in J}$. For a multi-index $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{N}_0^n$, we let $\mathbb{R}[x]_{\mathbf{r}}$ denote the set of polynomials p such that, if $p(x) = \sum_I c_I x^I$ for some real numbers $c_I \in \mathbb{R}$, then for each $I = (i_1, \ldots, i_n)$ with $c_I \ne 0$ we also have $i_k \le r_k$ for $1 \le k \le n$. Finally,

we let $\mathbb{R}[x_J]_\mathbf{r} = \mathbb{R}[x_J] \cap \mathbb{R}[x]_\mathbf{r}$; in other words, $\mathbb{R}[x_J]_\mathbf{r}$ is the set of polynomials p with $\overline{\deg} \, p \leq \mathbf{r}$ in the variables $\{x_j : j \in J\} \subseteq \{x_1, \ldots, x_n\}$.

Given a set X, we write X^n to denote the product

$$X^n = \underbrace{X \times X \times \cdots \times X}_{n}.$$

We denote by $\| \cdot \|_\infty$ the supremum norm on $[-1, 1]^n$.

The notation $\lceil s \rceil$ stands for the least integer $\geq s$.

1.2 Results

Let $\Sigma[x_J]$ denote the set of polynomials p that are sums of squares of polynomials in $\mathbb{R}[x_J]$, that is, of the form $p = p_1^2 + \cdots + p_\ell^2$ for $p_1, \ldots, p_\ell \in \mathbb{R}[x_J]$.

Let $\bar{k} \in \mathbb{N}$ and let $\mathbf{g} = \{g_1, \ldots, g_{\bar{k}}\}$ be a collection of polynomials $g_i \in \mathbb{R}[x]$ defining a set

$$S(\mathbf{g}) = \{x \in \mathbb{R}^n : g_i(x) \geq 0, \ i = 1, \ldots, \bar{k}\}.$$

For convenience, denote also $g_0 = 1$. To the collection $\mathbf{g}$ and a multi-index $\mathbf{r}$, we associate the (*variable- and degree-wise truncated*) *quadratic module* associated to the collection $\mathbf{g}$ and a multi-index $\mathbf{r}$ be

$$Q_{\mathbf{r},J}(\mathbf{g}) = \left\{ \sum_{i=0}^{\bar{k}} \sigma_i g_i : \sigma_i \in \Sigma[x_J], \overline{\deg}(\sigma_i g_i) \leq \mathbf{r} \right\}.$$

Similarly, we have a (*variable- and degree-wise truncated*) *preordering*

$$\mathcal{P}_{\mathbf{r},J}(\mathbf{g}) = Q_{\mathbf{r},J}(\{g_K : K \subseteq \{1, \ldots, \bar{k}\}\})$$

$$= \left\{ \sum_{K \subseteq \{1, \ldots, \bar{k}\}} \sigma_K g_K : \sigma_K \in \Sigma[x_J], \overline{\deg}(\sigma_K g_K) \leq \mathbf{r} \right\},$$

where

$$g_K = \prod_{i \in K} g_i.$$

Denote, for $j = 2, 3, \ldots, \ell$,

$$\mathcal{J}_j = J_j \cap \bigcup_{k < j} J_k. \tag{2}$$

Then (RIP) is the condition that, for all $1 \leq k \leq \ell - 1$, there is some $s \leq k$ such that $\mathcal{J}_{k+1} \subset J_s$.

1.2.1 Sparse Schmüdgen-type representation on $[-1, 1]^n$

Let $\overline{L} := \sum_{k=1}^{\ell} \text{Lip } p_k$, $M := \max_{\substack{1 \le k \le \ell \\ 1 \le m \le n}} (\overline{\text{deg}} \, p_k)_m$, and $\overline{J} := \max_{1 \le k \le \ell} |J_k|$.

Theorem 6 *Let $n > 0$ and $\ell \ge 2$, and let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_\ell \in \mathbb{N}^n$, $\mathbf{r}_j = (r_{j,1}, \ldots, r_{j,n})$, be nowhere-vanishing multi-indices.*

Let also $J_1, \ldots, J_\ell$ be subsets of $\{1, \ldots, n\}$ satisfying (RIP).

Let $p = p_1 + p_2 + \cdots + p_\ell$ be a polynomial that is the sum of finitely many polynomials $p_j \in \mathbb{R}[x_{J_j}]_{\mathbf{r}_j}$. Then if $p \ge \varepsilon$ on $[-1, 1]^n$, we have

$$p \in \mathcal{P}_{\mathbf{r}_1, J_1}(\{1 - x_i^2\}_{i \in J_1}) + \cdots + \mathcal{P}_{\mathbf{r}_\ell, J_\ell}(\{1 - x_i^2\}_{i \in J_\ell})$$

as long as, for all $1 \le j \le \ell$ and all $1 \le i \le n$,

$$r_{j,i}^2 \ge \frac{2^{\overline{J}+3}(\ell+2)n\pi^2\|p\|_\infty}{\varepsilon} \left(\max\left(M, 4C_{\text{Jac}}(\ell+2)\frac{\overline{J}\,\overline{L}}{\varepsilon}\right) + 2\right)^{\overline{J}+2}.$$

For small enough $0 < \varepsilon < 4C_{\text{Jac}}(\ell+2)\overline{J}\,\overline{L}/M$, this boils down to

$$r_{j,i} \ge \sqrt{\frac{A\,\|p\|_\infty}{\varepsilon^{\overline{J}+3}}} = O(\varepsilon^{-\frac{\overline{J}+3}{2}}),$$

with

$$A = n\pi^2(4C_{\text{Jac}}\overline{J}\,\overline{L} + 2)^{\overline{J}+2}(2(\ell+2))^{\overline{J}+3}.$$

The proof is presented in Sect. 4.1.

Discussion

Solving the dense problem considered by [14] using the sum-of-squares hierarchy reduces to a semidefinite program with the largest PSD block of size $\binom{n+r}{r}$ that typical optimization methods (e.g., interior point or first order) can solve in an amount of time proportional to a power of

$$\binom{n+r}{r} \approx \binom{n + C\varepsilon^{-1/2}}{C\varepsilon^{-1/2}} =: B_{\text{dense}}(\varepsilon),$$

at least when certain non-degeneracy conditions are satisfied [5]. The bounds we find in Theorem 6 —in the case in which J_j is the largest of the sets $J_1, \ldots, J_\ell$— give a bound for the complexity of the leading term as (the same power of)

$$\ell\binom{|J_j| + |\mathbf{r}_j|}{|\mathbf{r}_j|} \le \ell\binom{|J_j|(1 + C'\varepsilon^{-\frac{|J_j|+3}{2}})}{|J_j|C'\varepsilon^{-\frac{|J_j|+3}{2}}} =: B_{\text{sparseSchm}}(\varepsilon).$$

The reason we have $|\mathbf{r}_j| \leq |J_j| C' \varepsilon^{-|J_j|-3}$ is that $r_{j,i} \leq O(\varepsilon^{-\frac{|J_j|+3}{2}})$ and there are at most $|J_j|$ values of i with $r_{j,i} \neq 0$.

Proposition 7 *If $n > |J_j|(|J_j| + 3)$ for all $j = 1, \ldots, \ell$, then we have*

$$\lim_{\varepsilon \searrow 0} \frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = 0.$$

Thus, if the size of the largest clique is of the order of square root of the ambient dimension n or smaller, the sparse bound outperforms the best available dense bound if the performance is measured by the amount of time required by an optimization method to find a bound of a given accuracy ε.

Proof of Proposition 7 by Lemma 20 we have, as $\varepsilon \searrow 0$,

$$\frac{B_{\text{sparseSchm}}(\varepsilon)}{B_{\text{dense}}(\varepsilon)} = \frac{\binom{|J_j|(1 + C'\varepsilon^{-\frac{|J_j|+3}{2}})}{|J_j| C' \varepsilon^{-\frac{|J_j|+3}{2}}}}{\binom{n+C\varepsilon^{-1/2}}{C\varepsilon^{-1/2}}} = O(\varepsilon^{\frac{1}{2}(n-|J_j|(|J_j|+3))}),$$

and this tends to 0 if the sparsity of the polynomial p is such that $n > |J_j|(|J_j| + 3)$. □

1.2.2 Sparse Putinar-type representation on arbitrary domains

For a set of polynomials $\mathbf{g} = \{g_1, \ldots, g_{\bar{k}}\}$, denote

$$S(\mathbf{g}) = \{x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for all } i = 1, \ldots, \bar{k}\}$$

and, for a subset $K \subset \{1, \ldots, \bar{k}\}$, denote

$$\mathbf{g}_K = \{g_i : i \in K\}.$$

Theorem 8 *Let $n > 0$, $\bar{k} > 0$, $\ell \geq 2$, $J_1, \ldots, J_\ell \subset \{1, \ldots, n\}$, $\mathbf{r}_1, \ldots, \mathbf{r}_\ell \in \mathbb{N}^n$, and $p = p_1 + \cdots + p_\ell$ with $p_j \in \mathbb{R}[x_{J_j}]_{\mathbf{r}_j}$.*
Assume that the sets $J_1, \ldots, J_\ell$ satisfy (RIP).
Let $K_1, \ldots, K_\ell \subset \{1, \ldots, \bar{k}\}$ and let $\mathbf{g} = \{g_1, \ldots, g_{\bar{k}}\} \subset \mathbb{R}[x]$ be a collection of $\bar{k}$ polynomials such that, if $i \in K_j$ for some $1 \leq j \leq \ell$, then $g_i \in \mathbb{R}[x_{J_j}]$. Assume that

$$\|g_j\|_\infty \leq \frac{1}{2}, \quad j = 1, 2, \ldots, \bar{k}. \tag{3}$$

Assume that $S(\mathbf{g}) \subset [-1, 1]^n$ and that there exist polynomials $s_{j,i} \in \mathbb{R}[x_{J_j}]_{\mathbf{r}_j}$, $j = 1, \ldots, \ell$, $i \in \{0\} \cup K_j$, such that the Archimedean conditions

$$1 - \sum_{i \in J_j} x_i^2 = s_{j,0}(x_{J_j})^2 + \sum_{i \in K_j} s_{j,i}(x_{J_j})^2 g_i(x_{J_j}), \quad j = 1, \ldots, \ell, \tag{4}$$

hold; that is to say, we assume that $1 - \sum_{i \in J_j} x_i^2 \in \mathcal{Q}_{\mathbf{r}_j, J_j}(\mathbf{g}_{K_j})$. Let $\mathsf{c}_1, \ldots, \mathsf{c}_\ell \geq 1$ and $\mathsf{L}_1, \ldots, \mathsf{L}_\ell \geq 1$ be constants such that[3]

$$\text{dist}(x, S(\mathbf{g}_{K_j}))^{\mathsf{L}_j} \leq -\mathsf{c}_j \min \left\{ \{0\} \cup \{ g_i(x) : i \in K_j \} \right\} \quad \textit{for all} \quad x \in [-1, 1]^n.$$

Then there are constants $\mathbf{C}_j > 0$, depending only on $\mathbf{g}, J_1, \ldots, J_\ell$, such that, if $p \geq \varepsilon > 0$ on $S(\mathbf{g})$, we have

$$p \in \mathcal{Q}_{\mathbf{r}_1+\mathbf{2}, J_j}(\mathbf{g}_{K_j}) + \cdots + \mathcal{Q}_{\mathbf{r}_\ell+\mathbf{2}, J_\ell}(\mathbf{g}_{K_\ell})$$

as long as, for all $1 \leq j \leq \ell$ and $1 \leq k \leq n$,

$$(r_{j,k} + 2)^2 \geq \mathbf{C}_j \frac{4(\ell + 2) \left(\sum_i \|p_i\|_\infty \right)^{\mathsf{L}_j+1} (\deg p_j \sum_{i=1}^\ell \text{Lip } p_i)^{(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})}}{\varepsilon^{1 + \mathsf{L}_j + \frac{4\mathsf{L}_j+1}{3}(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})}}, \quad (5)$$

$$(r_{j,k} + 2)^2 \geq \mathbf{C}_j \left(\frac{\left(\sum_{i=1}^\ell \|p_i\|_\infty \right)^{\frac{4\mathsf{L}_j+1}{3}} (\deg p_j \sum_{i=1}^\ell \text{Lip } p_i)^{\frac{8\mathsf{L}_j}{3}}}{\varepsilon^{\frac{12\mathsf{L}_j+1}{3}}} \right)^2. \quad (6)$$

The proof of the theorem can be found in Sect. 4.2.

Discussion By the same arguments we used in the discussion at the end of the previous section, if we assume $\mathsf{L}_1 = \cdots = \mathsf{L}_{\bar{k}} = 1$, the bounds we find in Theorem 8 give a bound for the complexity of the leading term as (a power of)[4]

$$\ell \binom{|J_j| + |\mathbf{r}_j|}{|\mathbf{r}_j|} \approx \binom{|J_j|(1 + C''\varepsilon^{-13-3|J_j|})}{|J_j|C''\varepsilon^{-13-3|J_j|}} =: B_{\text{sparsePut}},$$

at least when certain non-degeneracy conditions are satisfied [5]. The assumption $\mathsf{L}_1 = \cdots = \mathsf{L}_{\bar{k}} = 1$ is realized for example when the so-called *constraint qualification condition* that, at each point $x \in S(\mathbf{g})$ all the active constraints $g_{i_1}, \ldots, g_{i_l}$ (i.e., those satisfying $g_{i_j}(x) = 0$) have linearly independent gradients $\nabla g_{i_1}(x), \ldots, \nabla g_{i_l}(x)$, holds; this latter statement is proved in [2, Thm 2.11].

In this case, we have:

Proposition 9 *If $\frac{n}{\geq} |J_j|(26 + 6|J_j|)$ for all $j = 1, \ldots, \ell$ and if $\mathsf{L}_1 = \cdots = \mathsf{L}_{\bar{k}} = 1$, then we have*

$$\lim_{\varepsilon \searrow 0} \frac{B_{\text{sparsePut}}}{B_{\text{dense}}} = 0.$$

---

[3] This is a version of the *Łojasiewicz inequality*, and its validity (with appropriate constants $\mathsf{c}_j, \mathsf{L}_j$) for semialgebraic functions is justified in [2, Thm. 2.3] and the papers cited therein.

[4] With $\mathsf{L}_1 = \cdots = \mathsf{L}_{\bar{k}} = 1$, the exponent of ε in (5) is $(238 + 55|J_j|)/9$, which we estimate from below with the simpler expression $26 + 6|J_j|$ to derive the exponent of ε here.

Again the implication is that the sparse bound asymptotically outperforms the dense bound provided that the largest clique is sufficiently small.

Proof of Proposition 9 Lemma 20 gives

$$\frac{B_{\text{sparsePut}}}{B_{\text{dense}}} = \frac{\binom{|J_j|(1 + C''\varepsilon^{-13-3|J_j|})}{|J_j|C''\varepsilon^{-13-3|J_j|}}}{\binom{n+C\varepsilon^{-1/2}}{C\varepsilon^{-1/2}}} = O(\varepsilon^{\frac{n}{2} - |J_j|(13+3|J_j|)}),$$

which tends to 0 if $\frac{n}{2} > |J_j|(13 + 3|J_j|)$. □

Organization of the paper The proof of Theorem 6 can be seen as a variable-separated version of the proof in [14], which relies on the Jackson kernel. Therefore in Sect. 2 we derive the suitable ingredients for sparse Jackson kernels while carefully taking into account each variable separately.

A strategy is also required to write a positive polynomial p that is known to be a sum $p = p_1 + \cdots + p_\ell$ with $p_i \in \mathbb{R}[x_{J_i}]$ as a similar sum $p = h_1 + \cdots + h_\ell$ but now with $h_j \in \mathbb{R}[x_{J_j}]$ and $h_j \geq 0$ on $[-1, 1]^{|J_j|}$; this is done in Sect. 3.

Section 4 gives the proofs of Theorems 6 and 8, together with the statement and proof of Lemma 20, which was used in the proofs of Propositions 7 and 9 above.

2 The sparse Jackson kernel

In this section, we derive Corollary 13, one of the main ingredients of the proof of Theorem 6. The corollary guarantees that polynomials bounded from below by $\varepsilon > 0$ on $[-1, 1]^n$ are in the preordering $\mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J})$ (defined in Sect. 1.2) for ε large enough. The corollary follows from Theorem 14, which gives a refined estimate of the distance between a polynomial p and its preimage under a Jackson-style operator that treats each variable separately. We begin with some preliminaries and a useful lemma.

The measure μ_n on the box $[-1, 1]^n$ defined by

$$d\mu_n(x) := \frac{dx_1}{\pi\sqrt{1 - x_1^2}} \cdots \frac{dx_n}{\pi\sqrt{1 - x_n^2}}, \quad x = (x_1, \ldots, x_n) \in [-1, 1]^n,$$

is known as the (*normalized*) *Chebyshev measure*; it is a probability measure on $[-1, 1]^n$. It induces the inner product

$$\langle f, g \rangle_{\mu_n} := \int_{[-1,1]^n} f(x)g(x) \, d\mu_n(x)$$

and the norm $\|f\|_{\mu_n} = \sqrt{\langle f, f \rangle_{\mu_n}}$.

For $k = 0, 1, \ldots$, we let $T_k \in \mathbb{R}[x]$ be the univariate *Chebyshev polynomial* of degree k, defined by

$$T_k(\cos \theta) := \cos(k\theta), \quad \theta \in \mathbb{R}.$$

The Chebyshev polynomials satisfy $|T_k(x)| \leq 1$ for all $x \in [-1, 1]$, and

$$\langle T_a, T_b \rangle_{\mu_1} = \int_{-1}^{1} \frac{T_a(x)T_b(x)}{\pi\sqrt{1-x^2}} dx = \begin{cases} 0, & a \neq b, \\ 1, & a = b = 0, \\ \frac{1}{2}, & a = b \neq 0. \end{cases}$$

For a multi-index $I = (i_1, \ldots, i_n)$, we let

$$T_I(x_1, \ldots, x_n) := y T_{i_1}(x_1) T_{i_2}(x_2) \cdots T_{i_n}(x_n)$$

be the multivariate Chebyshev polynomials, which then satisfy (see for example [28, §II.A.1]), for multi-indices I and I',

$$\deg T_I = |I| \quad \text{and} \quad \langle T_I, T_{I'} \rangle_{\mu_n} = \begin{cases} 0, & I \neq I', \\ 2^{-w(I)}, & I = I'. \end{cases} \tag{7}$$

Thus $p \in \mathbb{R}[x]_d$ can be expanded as $p = \sum_{|I| \leq d} 2^{w(I)} \langle p, T_I \rangle_{\mu_n} T_I$.

If we let, for a finite collection $\Lambda \subseteq \mathbb{R} \times \mathbb{N}_0^n$ of pairs (λ, I) of a real number λ and a multi-index I,

$$K^{\Lambda}(x, y) = \sum_{(\lambda, I) \in \Lambda} 2^{w(I)} \lambda \, T_I(x) T_I(y), \quad x, y \in \mathbb{R}^n,$$

then, for any $p \in \mathbb{R}[x]$, we have

$$\mathbf{K}^{\Lambda}(p)(x) := \int_{[-1,1]^n} K^{\Lambda}(x, y) p(y) \, d\mu_n(y) = \sum_{\substack{(\lambda, I) \in \Lambda \\ |I| \leq d}} 2^{w(I)} \lambda \langle p, T_I \rangle_{\mu_n} T_I(x).$$

This means that, if we set all the nonzero numbers λ equal to 1, then $\mathbf{K}^{\Lambda}$ is the identity operator in the linear span of $\{T_I : \exists \lambda \neq 0 \text{ s.t. } (\lambda, I) \in \Lambda\} \subseteq \mathbb{R}[x]_d$.

We let, for $r, k \in \mathbb{N}$,

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

and

$$\lambda_0^r = 1.$$

We set, for $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{N}_0^n$,

$$\lambda_I^{\mathbf{r}} = \prod_{j=1}^{n} \lambda_{i_j}^{r_j}$$

and

$$\Lambda_{\mathbf{r}} = \{(\lambda_I^{\mathbf{r}}, I) : I \leq \mathbf{r}\}.$$

Then $K_{\mathbf{r}}^{\mathrm{Jac}} = K^{\Lambda_{\mathbf{r}}}$ is the *(**r**-adapted) Jackson kernel*, and its associated linear operator $\mathbf{K}^{\Lambda_{\mathbf{r}}}$ is denoted $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}$.

Lemma 10 *Let $\mathbf{r} \in \mathbb{N}_0^n$ be a multi-index. The operator $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}$ defined above has the following properties:*

i. $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(\mathbb{R}[x_J]_{\mathbf{r}}) \subseteq \mathbb{R}[x_J]_{\mathbf{r}}$.
ii. *We have*

$$\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(T_I) = \begin{cases} \lambda_I^{\mathbf{r}} T_I, & I \leq \mathbf{r}, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(1) = 1$.

iii. $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}$ *is invertible in $\mathbb{R}[x_J]_{\mathbf{r}}$ with $J = \{i : 1 \leq i \leq n, \ r_i > 0\}$.*
iv. $0 < \lambda_I^{\mathbf{r}} \leq 1$ *for all $0 \leq I \leq \mathbf{r}$.*
v. *For $I = (i_1, \ldots, i_n)$ and $\mathbf{r} = (r_1, \ldots, r_n)$ in $\mathbb{N}_0^n$,*

$$|1 - \lambda_I^{\mathbf{r}}| = 1 - \lambda_I^{\mathbf{r}} \leq n\pi^2 \max_j \frac{i_j^2}{(r_j + 2)^2}.$$

vi. *For $I = (i_1, \ldots, i_n)$ and $\mathbf{r} = (r_1, \ldots, r_n)$ in $\mathbb{N}_0^n$ that verify (9), we have*

$$\left| 1 - \frac{1}{\lambda_I^{\mathbf{r}}} \right| \leq 2n\pi^2 \max_j \frac{i_j^2}{(r_j + 2)^2}.$$

vii. *Let $p \in \mathbb{R}[x]_{J,\mathbf{r}}$ with $p(x) \geq 0$ for all $x \in [-1, 1]^n$ and $\|p\|_\infty \leq 1$. Let $\mathbf{r} = (r_1, \ldots, r_n)$ be a multi-index such that $I \leq \mathbf{r}$ for all $I \in \mathcal{I}_p$, and assume that, for all $I = (i_1, \ldots, i_n) \in \mathcal{I}_p$, condition (9) is verified. Then we have*

$$\left\| (\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}})^{-1}(p) - p \right\|_\infty \leq 2n\pi^2 \left(\prod_{1 \leq k \leq n} ((\overline{\deg}\, p)_k + 1) \right) \max_{\substack{I \in \mathcal{I}_p \\ 1 \leq j \leq n}} \left[2^{w(I)/2} \frac{i_j^2}{(r_j + 2)^2} \right].$$

viii. $K_{\mathbf{r}}^{\mathrm{Jac}}(x, y) \geq 0$ *for all $x, y \in [-1, 1]^n$.*
ix. *If $p \in \mathbb{R}[x]$ is such that $p(x) \geq 0$ for $x \in [-1, 1]^n$, then $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(p)(x) \geq 0$ for all $x \in [-1, 1]^n$.*

Proof Throughout, we follow [14].

Item (ii) is immediate from the definitions and (7). Item (i) follows from item (ii) and the fact that $\{T_I : I \leq \mathbf{r}, \ I \subseteq J\}$ is a basis for $\mathbb{R}[x_J]_{\mathbf{r}}$.

Observe that item (ii) means that $\mathbf{K}_\mathbf{r}^{\mathrm{Jac}}$ is diagonal in $\mathbb{R}[x_J]_\mathbf{r}$, so in order to prove item (iii) it suffices to show that $\lambda_I^\mathbf{r} > 0$ for all $I \leq \mathbf{r}, I \subseteq J$. This follows immediately from item (iv), which in turn follows from the definition of $\lambda_I^\mathbf{r}$ and [14, Proposition 6(ii)], which shows that $0 < \lambda_k^r \leq 1$ for all $0 \leq k \leq r$.

Similarly, by [14, Proposition 6(iii)] we have that, if $k \leq r$, then

$$\left|1 - \lambda_k^r\right| = 1 - \lambda_k^r \leq \frac{\pi^2 k^2}{(r+2)^2}.$$

Thus, if $\gamma_j = 1 - \lambda_{i_j}^{r_j} \leq \pi^2 i_j^2/(r_j + 2)^2$ and $\gamma = \max_j \gamma_j$, we also have, using Bernoulli's inequality [14, Lemma 11],

$$
\begin{aligned}
1 - \lambda_I^\mathbf{r} = 1 - \prod_{j=1}^n \lambda_{i_j}^{r_j} \\
= 1 - \prod_{j=1}^n (1 - \gamma_j) \\
\leq 1 - (1 - \gamma)^n \\
\leq n\gamma \\
\leq n\pi^2 \max_j \frac{i_j^2}{(r_j + 2)^2}.
\end{aligned}
$$

This shows item (v). Using it, we can prove item (vi) as follows: condition (9) implies, by item (v), that $|1 - \lambda_I^\mathbf{r}| \leq 1/2$, and hence $|\lambda_I^\mathbf{r}| \geq 1/2$, so

$$\left|1 - \frac{1}{\lambda_I^\mathbf{r}}\right| = \frac{|1 - \lambda_I^\mathbf{r}|}{|\lambda_I^\mathbf{r}|} \leq 2n\pi^2 \max_j \frac{i_j^2}{(r_j + 2)^2},$$

leveraging item (v) again.

Let us show item (vii). From items (ii) and (iii), we have

$$\left\|(\mathbf{K}_\mathbf{r}^{\mathrm{Jac}})^{-1}(p) - p\right\|_\infty = \left\|\sum_I \left[\frac{1}{\lambda_I^\mathbf{r}} 2^{w(I)} \langle p, T_I \rangle_{\mu_n} T_I - 2^{w(I)} \langle p, T_I \rangle_{\mu_n} T_I\right]\right\|_\infty$$

$$\leq \sum_I 2^{w(I)} |\langle p, T_I \rangle_{\mu_n}| \left|1 - \frac{1}{\lambda_I^\mathbf{r}}\right|,$$

because $|T_I(x)| \leq 1$ for all $x \in [-1, 1]^n$. Plugging in the estimate from item (vi), we get

$$\left\|(\mathbf{K}_\mathbf{r}^{\mathrm{Jac}})^{-1}(p) - p\right\|_\infty \leq \sum_I 2^{w(I)/2+1} n\pi^2 \max_j \frac{i_j^2}{(r_j + 2)^2}$$

$$\leq 2n\pi^2 \left(\prod_{1 \leq k \leq n} ((\overline{\deg} \, p)_k + 1) \right) \max_{I \in \mathcal{I}_p} \left[2^{w(I)/2} \max_j \frac{i_j^2}{(r_j + 2)^2} \right],$$

where we have also used

$$|\langle p, T_I \rangle_{\mu_n}| \leq \|p\|_{\mu_n} \|T_I\|_{\mu_n} \leq \|T_I\|_{\mu_n} = 2^{-w(I)/2},$$

which follows from (7).

To prove item (viii), let, for a fixed multi-index $\mathbf{r}$,

$$\begin{aligned} \Lambda_k &= \{(\lambda_I^{\mathbf{r}}, I) : I \leq (0, \ldots, 0, r_k, 0, \ldots, 0)\} \\ &= \{(\lambda_{i_k}^{r_k}, (0, \ldots, 0, i_k, 0, \ldots, 0)) : i_k \leq r_k\}, \quad 1 \leq k \leq n, \end{aligned}$$

and observe that

$$\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}} = \mathbf{K}^{\Lambda_{\mathbf{r}}} = \mathbf{K}_{x_1}^{\Lambda_1} \circ \mathbf{K}_{x_2}^{\Lambda_2} \circ \cdots \circ \mathbf{K}_{x_n}^{\Lambda_n} \tag{8}$$

where $\mathbf{K}_{x_k}^{\Lambda_k}$ is the operator $\mathbf{K}^{\Lambda_k}$ acting in the variable x_k, i.e.,

$$\mathbf{K}_{x_k}^{\Lambda_k}(p)(x) = \int_{-1}^{1} K^{\Lambda_k}(x_k, y) \, p(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_n) \, d\mu_1(y).$$

Equation (8) follows from the identity

$$\begin{aligned} K^{\Lambda_{\mathbf{r}}}(x, y) &= K^{\Lambda_1}(x_1, y_1) K^{\Lambda_2}(x_2, y_2) \cdots K^{\Lambda_n}(x_n, y_n) \\ &= K_{(r_1)}^{\mathrm{Jac}}(x_1, y_1) K_{(r_2)}^{\mathrm{Jac}}(x_2, y_2) \cdots K_{(r_n)}^{\mathrm{Jac}}(x_n, y_n), \end{aligned}$$

that can be checked from the definitions. Item (viii) then follows from the well-known fact that $K_{(r)}^{\mathrm{Jac}}(x, y) \geq 0$ for all $r \in \mathbb{N}_0$ and all $x, y \in [-1, 1]$; see for example [28, §II.C.2–3].

Item (ix) follows immediately from item (viii). $\qquad \square$

Theorem 11 *We have* $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(\mathbb{R}[x_J]_{\mathbf{r}}) \subseteq \mathbb{R}[x_J]_{\mathbf{r}}$, *and if* $p(x) \geq 0$ *on* $[-1, 1]^n$ *then* $\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(p) \geq 0$ *on* $[-1, 1]^n$.

Also, we have:

P1. If $p \in \mathbb{R}[x_J]_{\mathbf{r}}$ *satisfies* $p(x) \geq 0$ *for all* $x \in [-1, 1]^n$,

$$\mathbf{K}_{\mathbf{r}}^{\mathrm{Jac}}(p) \in \mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J}).$$

P2. Let $p \in \mathbb{R}[x_J]_{\mathbf{r}}$ *be a polynomial that satisfies* $0 \leq p(x) \leq 1$ *for all* $x \in [-1, 1]^n$, *and for all* $I = (i_1, \ldots, i_n) \in \mathcal{I}_p$, *assume that* $\mathbf{r} = (r_1, \ldots, r_n)$ *verifies*

$$\frac{i_j^2}{(r_j + 2)^2} \leq \frac{1}{2\pi^2 n}, \quad 1 \leq j \leq n. \tag{9}$$

Assume that

$$\varepsilon \geq 2n n \pi^2 \left(\prod_{1 \leq k \leq n} ((\overline{\deg} \, p)_k + 1) \right) \max_{I \in \mathcal{I}_p} \left[2^{w(I)/2} \max_j \frac{i_j^2}{(r_j + 2)^2} \right] > 0.$$

Then

$$\left\| (\mathbf{K_r^{Jac}})^{-1}(p + \varepsilon) - (p + \varepsilon) \right\|_\infty \leq \varepsilon.$$

Proof The first statement of the theorem corresponds to Lemma 10(i) and 10(ix). Property P2. follows from Lemma 10(vii).

Let us prove property P1.. Take a finite subset $\{z_i\}_i$ of $[-1, 1]^n$ and a corresponding set of positive weights $\{w_i\}_i \subset \mathbb{R}$ giving a quadrature rule for integration of polynomials $q \in \mathbb{R}[x_J]_{\mathbf{r}}$, so that

$$\int_{[-1,1]^n} q(x) \, d\mu_n(x) = \sum_i w_i q(z_i) \quad \text{for all} \quad q \in \mathbb{R}[x_J]_{\mathbf{r}}.$$

Then we have, for p as in the statement of P1.,

$$\mathbf{K_r^{Jac}}(p)(x) = \sum_i w_i \, p(z_i) K_{\mathbf{r}}^{Jac}(z_i, x),$$

with $w_i \, p(z_i) \geq 0$. Since, by Lemma 10(viii) and Theorem 12 below, $K_{\mathbf{r}}^{Jac}(z_i, x)$ is in $\mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J})$, so is $\mathbf{K_r^{Jac}}(p)$. □

Theorem 12 ([8, Th. 10.3]) *If $p \in \mathbb{R}[y]$ is a univariate polynomial of degree d non-negative on the interval $[a, b] \subset \mathbb{R}$, then*

$$\begin{cases} p = \sigma_0 + \sigma_1(b - y)(y - a), & \sigma_0 \in \Sigma_d[y], \quad \sigma_1 \in \Sigma_{d-2}[y] \quad d \text{ even}, \\ p = \sigma_0(y - a) + \sigma_1(b - y), & \sigma_0 \in \Sigma_{d-1}[y], \ \sigma_1 \in \Sigma_{d-1}[y] \quad d \text{ odd}, \end{cases}$$

where Σ_d is the cone of sum-of-squares of polynomials of degree at most d.

Corollary 13 *If $p \in \mathbb{R}[x_J]_{\mathbf{r}}$ satisfies $0 \leq p(x) \leq 1$ for all $x \in [-1, 1]^n$, then*

$$p + \varepsilon \in \mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J})$$

for all multi-indices $\mathbf{r}$ satisfying (9) and

$$\varepsilon \geq 2n \pi^2 \left(\prod_{1 \leq k \leq n} ((\overline{\deg} \, p)_k + 1) \right) \max_{I \in \mathcal{I}_p} \left(2^{w(I)/2} \max_{1 \leq j \leq n} \frac{i_j^2}{(r_j + 2)^2} \right).$$

Here, $\mathbf{r} = (r_1, \ldots, r_n)$ and $I = (i_1, \ldots, i_n)$.

Proof By property P2. in Theorem 11,

$$\left\| (\mathbf{K_r^{Jac}})^{-1}(p + \varepsilon) - (p + \varepsilon) \right\|_\infty \leq \varepsilon.$$

Thus, $(\mathbf{K_r^{Jac}})^{-1}(p + \varepsilon) \geq 0$ on $[-1, 1]^n$. By property P1. and Lemma 10(i),

$$p + \varepsilon = \mathbf{K_r^{Jac}} \circ (\mathbf{K_r^{Jac}})^{-1}(p + \varepsilon) \in \mathcal{P}_{r,J}(\{1 - x_i^2\}_{i \in J}).$$

$\square$

The rest of this section is devoted to results used in the proof of Theorem 11.

3 Sparse approximation theory

In this section, we prove a useful result, Lemma 15, that is crucial to our proof of Theorem 6. Given a positive polynomial $f \geq \varepsilon > 0$ for which we know there is an expression of the type $f = f_1 + \cdots + f_\ell$, with each f_j depending only on the variables indexed by a subset $J_j \subset \{1, \ldots, n\}$ (but with f_j not necessarily positive), the lemma gives us *positive* polynomials $h_1, \ldots, h_\ell$ such that $f = h_1 + \cdots + h_\ell$ and h_j depends only on the variables indexed by the subset J_j. To prove the lemma we need some preliminaries.

For $1 \leq i \leq n$ and a function $f : [-1, 1]^n \to \mathbb{R}$, let

$$\mathrm{Lip}_i f = \sup_{\substack{x \in [-1,1]^n \\ y \in [-1,1]}} \frac{|f(x) - f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)|}{|x_i - y|}.$$

Theorem 14 *Let $f \in C^0([-1, 1]^n)$ be a Lipschitz function with variable-wise Lipschitz constants $\mathrm{Lip}_1 f, \ldots, \mathrm{Lip}_n f$. Then there is a constant $C_{\mathrm{Jac}} > 0$ such that, for each multi-index $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{N}^n$, there is a polynomial $p \in \mathbb{R}[x]_\mathbf{m}$ such that*

$$\sup_{x \in [-1,1]^n} |f(x) - p(x)| \leq C_{\mathrm{Jac}} \sum_{i=1}^n \frac{\mathrm{Lip}_i f}{m_i}$$

and

$$\mathrm{Lip}_i p \leq 2 \, \mathrm{Lip}_i f.$$

The constant C_{Jac} does not depend on n, f, or $\mathbf{m}$.

Proof Jackson [7, p. 2–6] proved that there is a constant $C > 0$ (independent of n, f, $\mathbf{m}$) such that, if $g\colon \mathbb{R} \to \mathbb{R}$ is Lipschitz and π-periodic, $g(0) = g(\pi)$, then

$$\left| g(\theta) - \frac{\int_{-\pi/2}^{\pi/2} g(\theta - \vartheta) \left(\frac{\sin m\vartheta}{m \sin \vartheta} \right)^4 d\vartheta}{\int_{-\pi/2}^{\pi/2} \left(\frac{\sin m\vartheta}{m \sin \vartheta} \right)^4 d\vartheta} \right| \le \frac{C \operatorname{Lip} g}{m}, \quad m \in \mathbb{N}, x \in \mathbb{R}. \tag{10}$$

For a multivariate Lipschitz function $g\colon \mathbb{R}^n \to \mathbb{R}$ and a multi-index $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{N}^n$, let

$$L_i(g)(\theta)$$
$$= \frac{\int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} g(\theta_1 - \vartheta_1, \ldots, \theta_i - \vartheta_i, \theta_{i+1}, \ldots, \theta_n) \prod_{j=1}^{i} \left(\frac{\sin m_j \vartheta_j}{m_j \sin \vartheta_j} \right)^4 d\vartheta_1 \ldots d\vartheta_i}{\prod_{j=1}^{i} \int_{-\pi/2}^{\pi/2} \left(\frac{\sin m_j \vartheta_j}{m_j \sin \vartheta_j} \right)^4 d\vartheta_j}.$$

Then we have, using the triangle inequality and the single-variable inequality (10) at each step,

$$|g(\theta) - L_n(g)(\theta)|$$
$$\le |g(\theta) - L_1(g)(\theta)| + |L_1(g)(\theta) - L_2(g)(\theta)| + \cdots + |L_{n-1}(g)(\theta) - L_n(g)(\theta)|$$
$$\le C \left(\frac{\operatorname{Lip}_1 g}{m_1} + \cdots + \frac{\operatorname{Lip}_n g}{m_n} \right).$$

The function $\prod_j (\sin m_j \theta_j / m_j \sin \theta_j)^4$ is a polynomial of degree m_j in $\cos \theta_j$ (cf. [7, p. 3]). If we replace f with its Lipschitz extension to $[-2, 2]^n$ and apply the results above to $g(\theta) = f(2 \cos \theta_1, \ldots, 2 \cos \theta_n)$ we get a polynomial $L_n(g)(\theta)$ in $\cos \theta_1, \ldots, \cos \theta_n$ satisfying the above inequality. Thus

$$p(x) = L_n(g)(\arccos(x_1/2), \ldots, \arccos(x_n/2)), \quad x \in [-2, 2],$$

is a polynomial with $\overline{\deg} p \le \mathbf{m}$ that satisfies (cf. [7, p. 13–14])

$$|f(x) - p(x)| \le C \left(\frac{\operatorname{Lip}_1 g}{m_1} + \cdots + \frac{\operatorname{Lip}_n g}{m_n} \right) \le 2C \left(\frac{\operatorname{Lip}_1 f}{m_1} + \cdots + \frac{\operatorname{Lip}_n f}{m_n} \right),$$

since $\operatorname{Lip}_i g \le 2 \operatorname{Lip}_i f$. This proves the first statement, setting $C_{\text{Jac}} = 2C$. We also have

$$\left| \frac{d}{dx} \arccos(x/2) \right| = \frac{1}{2\sqrt{1 - (x/2)^2}} \le \frac{1}{\sqrt{3}} \quad \text{for} \quad x \in [-1, 1],$$

and, by linearity and monotonicity of L_n,

$$|L_n g(\theta) - L_n g(\theta_1, \ldots, \theta_{i-1}, \theta_i + t, \theta_{i+1}, \ldots, \theta_n)|$$
$$\le \left| L_n(|t| \operatorname{Lip}_i g)(\theta) \right| = |t| \operatorname{Lip}_i g \le 2|t| \operatorname{Lip}_i f,$$

whence

$$\text{Lip}_i \, p = \text{Lip}_i \, L_n(g)(\arccos(x_1/2), \dots, \arccos(x_n/2))$$

$$\leq \text{Lip}_i \, L_n(g) \left| \frac{d}{dx_i} \arccos \frac{x_i}{2} \right| \leq \text{Lip}_i \, L_n(g) \leq 2 \, \text{Lip}_i \, f. \qquad \square$$

Lemma 15 (a version of [6, Lemma 3]) *Let $J_1, \dots, J_\ell$ be subsets of $\{1, \dots, n\}$ satisfying* (RIP). *Suppose $f = f_1 + \cdots + f_\ell$ with $\ell \geq 2$, $f_j \in \mathbb{R}[x_{J_j}]$. Let $\varepsilon > 0$ be such that $f \geq \varepsilon$ on $S(\mathbf{g}) \subseteq [-1, 1]^n$. Pick numbers $\epsilon, \eta > 0$ so that*

$$\varepsilon = (\ell - 1)\epsilon - (\ell - 2)\eta \quad \text{and} \quad \epsilon > 2\eta.$$

Set, for $2 \leq l \leq \ell$,

$$D_{l,m} = \left\lceil \frac{2 \, C_{\text{Jac}} \, |\mathcal{J}_l| \sum_{k=l}^m \text{Lip} \, f_k}{\epsilon - 2\eta} \right\rceil, \tag{11}$$

with $\mathcal{J}_l$ as in (2), *and $D_{1,m} = D_{2,m}$.*

Then $f = h_1 + \cdots + h_\ell$ for some $h_j \in \mathbb{R}[x_{J_j}]$ with $h_j \geq \eta$ on $S(\mathbf{g}) \subseteq [-1, 1]^n$ and

$$\overline{\deg} \, h_j \leq \max(\overline{\deg} \, f_j, \bar{D}_{j,\ell}, \bar{D}_{j+1,\ell}, \dots, \bar{D}_{\ell,\ell})_{J_j} \tag{12}$$

where $\bar{D}_{j,m}$ is the multi-index whose k-th entry equals $D_{j,m}$ if $k \in \mathcal{J}_j = J_j \cap \bigcup_{k<j} J_i$ and 0 otherwise, and the maximum is taken entry-wise.

Additionally, if $\text{Lip} \, f$ denotes the Lipschitz constant of f on $[-1, 1]^n$, then

$$\text{Lip} \, h_j \leq 3 \sum_{k=j}^\ell \text{Lip} \, f_k.$$

Finally, we have

$$\|h_j\|_\infty \leq 3 \times 2^{\ell-1} \sum_{j=1}^\ell \|f_j\|_\infty.$$

Remark 16 If $S(\mathbf{g}) = [-1, 1]^n$, we also have the obvious estimate $\|h_j\|_\infty \leq \|f\|_\infty$, that follows from $0 \leq h_j \leq f$.

Proof In order to prove the result by induction, let us first consider the case $\ell = 2$. In this case, $\varepsilon = \epsilon$ and $\epsilon > 2\eta$. Assume that $J_1 \cap J_2 \neq \emptyset$. For a subset $J \subset \{1, \dots, n\}$, let π_J denote the projection onto the variables with indices in J, that is, $\pi_J(x) = (x_i)_{i \in J} \in [-1, 1]^J$ for $x \in [-1, 1]^n$.

Define $g \colon [-1, 1]^{J_1 \cap J_2} \to \mathbb{R}$ by

$$g(x) := \min_{y \in \pi_{J_1 \setminus J_2}(S(\mathbf{g})) \subseteq [-1,1]^{J_1 \setminus J_2}} f_2(x, y) - \frac{\varepsilon}{2}, \qquad x \in [-1, 1]^{J_1 \cap J_2}.$$

The function g is Lipschitz continuous on $[-1, 1]^{J_1 \cap J_2}$. To see why, let $x, x' \in [-1, 1]^{J_1 \cap J_2}$ and pick $y, y' \in \pi_{J_1 \setminus J_2}(S(\mathbf{g})) \subseteq [-1, 1]^{J_1 \setminus J_2}$ minimizing $f_2(x, y)$ and $f_2(x', y')$, respectively. Then

$$|g(x) - g(x')| = |f_2(x, y) - f_2(x', y')|$$
$$\leq \max(|f_2(x, y) - f_2(x', y)|, |f_2(x, y') - f_2(x', y')|) \leq \operatorname{Lip}(f_2)|x - x'|,$$

where $\operatorname{Lip}(f_2)$ denotes the Lipschitz constant of f_2 on $[-1, 1]^n$.

The function g also satisfies

$$f_1 + g \geq \frac{\varepsilon}{2} \quad \text{and} \quad f_2 - g \geq \frac{\varepsilon}{2}$$

on $S(\mathbf{g})$. The second inequality follows from the definition of g, and the first one can be shown taking $(x, y, z) \in S(\mathbf{g})$ with $x \in [-1, 1]^{J_1 \cap J_2}$, $y \in [-1, 1]^{J_1 \setminus J_2}$, and $z \in [-1, 1]^{J_2 \setminus J_1}$, taking care to pick y only after x has been chosen, in such a way that the minimum is in the definition of g is realized there, that is, $g(x) = f_2(x, y) - \varepsilon/2$ holds (this is possible by compactness of $S(\mathbf{g})$ and continuity of f); then we have

$$f_1(x, z) + g(x) = f_1(x, z) + f_2(x, y) - \frac{\varepsilon}{2} = f(x, y, z) - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2}.$$

For $j \in J_1 \cap J_2$, let

$$m_j = D_{2,2} = \left\lceil \frac{2 C_{\text{Jac}} |J_1 \cap J_2| \operatorname{Lip}(f_2)}{\varepsilon - 2\eta} \right\rceil.$$

Set $m_j = 0$ for all other $0 \leq j \leq n$, and $\mathbf{m} = (m_1, \ldots, m_n) = \bar{D}_{2,2}$. Then Theorem 14 gives a polynomial p_2 such that

$$\|g - p_2\|_\infty = C_{\text{Jac}} \sum_{j \in J_1 \cap J_2} \frac{\operatorname{Lip}_j g}{m_j} \leq C_{\text{Jac}} |J_1 \cap J_2| \operatorname{Lip}(f_2) \frac{2}{D_{2,2}} \leq \frac{\varepsilon}{2} - \eta.$$

Also,

$$\overline{\deg} p_2 \leq \mathbf{m} = \bar{D}_{2,2}. \tag{13}$$

Let

$$h_1 := f_1 + p_2 \quad \text{and} \quad h_2 := f_2 - p_2,$$

so that $f = h_1 + h_2$, $h_1 \geq \eta$ and $h_2 \geq \eta$ on $S(\mathbf{g})$, and $h_j \in \mathbb{R}[x_{J_j}]$.

The bound (12) follows from the definition of h_j and (13). Observe also that, by the last part of Theorem 14,

$$\operatorname{Lip} p_2 \leq 2 \operatorname{Lip} g \leq 2 \operatorname{Lip}(f_2).$$

Finally, we have

$$\|p_2\|_\infty \le \|g\|_\infty + \frac{\varepsilon}{2} - \eta \le \|f_2\|_\infty + \varepsilon - \eta \le 2\|f_2\|_\infty, \tag{14}$$

so

$$\|h_j\|_\infty \le \|f_j\|_\infty + \|p_2\|_\infty \le \|f_j\|_\infty + 2\|f_2\|_\infty \le 3(\|f_1\|_\infty + \|f_2\|_\infty).$$

For the induction step, let $\ell \ge 3$ and set $\tilde{f} = f_1 + \cdots + f_{\ell-1} - (\ell - 2)(\epsilon - \eta)$, so that we have $f - (\ell - 2)(\epsilon - \eta) = \tilde{f} + f_\ell \ge \epsilon$ since $f \ge \varepsilon = (\ell - 1)\epsilon - (\ell - 2)\eta$. The proof for the case $\ell = 2$ with $\varepsilon = \varepsilon_{\ell-1}$ gives a polynomial $p_\ell \in \mathbb{R}[x_{\mathcal{J}_\ell}]$ such that

$$\tilde{f} - p_\ell \ge \eta \quad \text{and} \quad f_\ell + p_\ell \ge \eta$$

on $S(\mathbf{g})$, and with $\overline{\deg}\, p_\ell = \bar{D}_{\ell,\ell}$, $\mathrm{Lip}\, p_\ell \le 2\,\mathrm{Lip}\, f_\ell$.and, analogously to (14),

$$\|p_\ell\|_\infty \le 2\|f_\ell\|_\infty. \tag{15}$$

Write

$$f'_1 + \cdots + f'_{\ell-1} = f_1 + \cdots + f_{\ell-1} - p_\ell,$$

where $f'_j = f_j - p_\ell$ for the largest j with $\mathcal{J}_\ell \subset J_j$ (which must happen for some j, by (RIP)) and $f'_k = f_k$ for all other $k \ne j$. Thus $f'_j \in \mathbb{R}[x_{J_j}]$,

$$\overline{\deg}\, f'_j \le \max(\overline{\deg}\, f_j, \overline{\deg}\, p_\ell) = \max(\overline{\deg}\, f_j, \bar{D}_{\ell,\ell}), \qquad \mathrm{Lip}\, f'_j \le \mathrm{Lip}\, f_j + \mathrm{Lip}\, p_\ell, \tag{16}$$

The induction hypothesis applies to the polynomial

$$f'_1 + \cdots + f'_{\ell-1} + (\ell - 2)(\epsilon - \eta) = \tilde{f} + (\ell - 2)(\epsilon - \eta) - p_\ell \ge (\ell - 2)\epsilon - (\ell - 3)\eta.$$

This means that there are polynomials $h_1, \ldots, h_{\ell-1}$ such that

- $f'_1 + \cdots + f'_{\ell-1} = f_1 + \cdots + f_{\ell-1} - p_\ell = h_1 + \cdots + h_{\ell-1}$,
- $h_j \in \mathbb{R}[x_{J_j}]$ for all $1 \le j \le \ell - 1$,
- $h_j \ge \eta$ for all $1 \le j \le \ell - 1$,
- We have, for all $1 \le j \le \ell - 1$,

$$\begin{aligned}
\overline{\deg}\, h_j &\le \max(\overline{\deg}\, f'_j, \bar{D}_{j,\ell} \ldots, \bar{D}_{\ell-1,\ell}) \\
&\le \max(\overline{\deg}\, f_j, \overline{\deg}\, p_\ell, \bar{D}_{j,\ell}, \ldots, \bar{D}_{\ell-1,\ell}) \\
&= \max(\overline{\deg}\, f_j, \bar{D}_{j,\ell}, \ldots, \bar{D}_{\ell,\ell}).
\end{aligned}$$

Observe that the second index in each $\bar{D}_{k,\ell}$ is ℓ because of the accumulation of Lipschitz constants resulting from the estimate (16).

- We have, for all $1 \le j \le \ell - 1$, again because of (16),

$$\operatorname{Lip} h_j \le 3 \sum_{k=j}^{\ell} \operatorname{Lip} f_k.$$

- We have, for all $1 \le j \le \ell - 1$, using (15),

$$\|h_j\|_\infty \le 3 \times 2^{\ell-2} \sum_{k=1}^{\ell-1} \|f_k'\|_\infty \le 3 \times 2^{\ell-2} \left(\sum_{k=1}^{\ell-1} \|f_k\|_\infty + \|p_\ell\|_\infty \right)$$

$$\le 3 \times 2^{\ell-1} \sum_{k=1}^{\ell} \|f_k\|_\infty.$$

Let

$$h_\ell = f_\ell + p_\ell.$$

Then again $f_1 + \cdots + f_\ell = h_1 + \cdots + h_\ell$, $h_\ell \in \mathbb{R}[x_{J_\ell}]$, $h_\ell \ge \eta$ on $S(\mathbf{g})$, $\overline{\deg} h_j \le \max(\overline{\deg} f_\ell, \bar{D}_{\ell,\ell})$, $\operatorname{Lip} h_\ell \le \operatorname{Lip} f_\ell + \operatorname{Lip} p_\ell \le 3 \operatorname{Lip} f_\ell$, $\|h_\ell\|_\infty \le \|f_\ell\|_\infty + \|p_\ell\| \le 3\|f_\ell\| \le 3 \times 2^{\ell-1} \sum_{j=1}^{\ell} \|f_j\|_\infty$, so the lemma is proven. $\qquad\square$

4 Proofs

4.1 Proof of Theorem 6

Overview Theorem 6 follows from Theorem 17, which presents a more detailed bound, together with the definitions of $\overline{L}, M, \overline{J}$. To prove the latter theorem, we first use the sparse approximation theory developed in Sect. 3 to represent the sparse polynomial p as a sum of positive polynomials $h_1 + \cdots + h_\ell$, each of them depending on a clique of variables J_j. We then use Corollary 13 to see that each h_j belongs to the preordering.

Theorem 17 *Let $n > 0$ and $\ell \ge 2$, and let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_\ell \in \mathbb{N}^n$, $\mathbf{r}_j = (r_{j,1}, \ldots, r_{j,n})$, be nowhere-vanishing multi-indices. Let also $J_1, \ldots, J_\ell$ be subsets of $\{1, \ldots, n\}$ satisfying (RIP). Let $p = p_1 + p_2 + \cdots + p_\ell$ be a polynomial that is the sum of finitely many polynomials $p_j \in \mathbb{R}[x_{J_j}]_{\mathbf{r}_j}$. Then if $p \ge \varepsilon$ on $[-1, 1]^n$, we have*

$$p \in \mathcal{P}_{\mathbf{r}_1, J_1}(\{1 - x_i^2\}_{i \in J_1}) + \cdots + \mathcal{P}_{\mathbf{r}_\ell, J_\ell}(\{1 - x_i^2\}_{i \in J_\ell})$$

as long as, for all $1 \le j \le \ell$ and all $1 \le i \le n$,

$$(r_{j,k} + 2)^2 \ge \frac{2^{\frac{|J_j|}{2}+2}(\ell + 2)\|p\|_\infty n\pi^2}{\varepsilon}$$

$$\cdot \prod_{1 \le m \le n} \left(\max \left[(\overline{\deg} \, p_j)_m, \max_{\substack{j \le l \le \ell \\ m \in \mathcal{J}_l}} \frac{4C_{\mathrm{Jac}}(\ell+2)|\mathcal{J}_l| \sum_{t=l}^{\ell} \mathrm{Lip} \, p_t}{\varepsilon} \right] + 2 \right)$$

$$\cdot \max_{l \in J_j} \left[(\overline{\deg} \, p_j)_l, \max_{\substack{j \le q \le \ell \\ l \in \mathcal{J}_q}} \frac{4C_{\mathrm{Jac}}(\ell+2)|\mathcal{J}_q| \sum_{t=q}^{\ell} \mathrm{Lip} \, p_t}{\varepsilon} \right]^2, \tag{17}$$

and

$$(r_{j,i}+2)^2 \ge 2\pi^2 n \max \left[\max_{1 \le m \le n} (\overline{\deg} \, p_j)_m, \max_{j \le k \le \ell} \frac{4C_{\mathrm{Jac}}(\ell+2)|\mathcal{J}_k| \sum_{t=k}^{\ell} \mathrm{Lip} \, p_t}{\varepsilon} \right]^2. \tag{18}$$

Proof of Theorem 17 Let

$$\epsilon = \frac{\varepsilon + (\ell-2)\eta}{\ell} \quad \text{and} \quad \eta = \frac{\varepsilon}{2(\ell+2)}. \tag{19}$$

Apply Lemma 15 with $\mathbf{g} = 0$, so that $S(\mathbf{g}) = [-1,1]^n$. From the lemma, we get polynomials $h_1, \ldots, h_\ell$ with

$$h_j \in \mathbb{R}[x_{J_j}],$$
$$p = h_1 + \cdots + h_\ell,$$
$$h_j \ge \eta \text{ on } [-1,1] \text{ for } 1 \le j \le \ell,$$
$$\overline{\deg} \, h_j \le \max(\overline{\deg} \, p_j, \bar{D}_j, \bar{D}_{j+1}, \ldots, \bar{D}_\ell).$$

Here,

$$\bar{D}_l := (\delta_{j \in \mathcal{J}_l} D_l)_{j=1}^\ell,$$
$$\delta_{j \in \mathcal{J}_l} = \begin{cases} 1, & j \in \mathcal{J}_l \\ 0, & \text{otherwise}, \end{cases}$$
$$D_l := \left\lceil \frac{2C_{\mathrm{Jac}}|\mathcal{J}_l| \sum_{t=l}^{\ell} \mathrm{Lip} \, p_t}{\varepsilon_{l-1} - 2\eta} \right\rceil = \left\lceil \frac{4C_{\mathrm{Jac}} \ell |\mathcal{J}_l| \sum_{t=l}^{\ell} \mathrm{Lip} \, p_t}{\eta(\ell+2)} \right\rceil, \quad l = 2, 3, \ldots, \ell,$$

since $\varepsilon_j - 2\eta = \eta(\ell+2)/2\ell$. We also set

$$D_1 := D_2.$$

Thus

$$|\bar{D}_l| = |\mathcal{J}_l| D_l \quad \text{for} \quad 2 \le l \le \ell.$$

Apply Corollary 13 to each of the polynomials

$$H_j = \frac{h_j - \min_{[-1,1]^n} h_j}{\max_{[-1,1]^n} h_j - \min_{[-1,1]^n} h_j}$$

to see that, for

$$\varepsilon_j \geq 2n\pi^2 \left(\prod_{1 \leq k \leq n} ((\overline{\deg} H_j)_k + 1) \right) \max_{I \in \mathcal{I}_{H_j}} \left(2^{w(I)/2} \max_{1 \leq k \leq n} \frac{i_k^2}{(r_{j,k} + 2)^2} \right), \quad (20)$$

(recall that $\mathcal{I}_{H_j}$ is the set of multiindices $I = (i_1, \ldots, i_n)$ corresponding to exponents of $x_1, \ldots, x_n$ in the terms appearing in H_j and $w(I)$ is the number of nonzero entries in I) we have

$$H_j + \epsilon_j \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i \in J_j}); \quad (21)$$

when applying the corollary, note that (18) implies (9) in this case because, if $I = (i_1, \ldots, i_n) \in \mathcal{I}_{H_j}$, then

$$i_k \leq (\overline{\deg} H_j)_k \leq (\overline{\deg} h_j)_k \leq \max(\overline{\deg} p_j, \bar{D}_j, \ldots, \bar{D}_\ell)$$

$$\leq \max \left[\max_{1 \leq m \leq n} (\overline{\deg} p_j)_m, \max_{j \leq k \leq \ell} \frac{4C_{\text{Jac}}(\ell + 2)|\mathcal{J}_k| \sum_{t=k}^{\ell} \text{Lip } p_t}{\varepsilon} \right],$$

by the definition of $\bar{D}_l$. Observe that (21) means also that

$$h_j - \min_{[-1,1]^n} h_j + \epsilon_j \left(\max_{[-1,1]^n} h_j - \min_{[-1,1]^n} h_j \right) \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i \in J_j}). \quad (22)$$

Note that we have $\overline{\deg} H_j = \overline{\deg} h_j$, $\mathcal{I}_{H_j} \backslash \mathcal{I}_{h_j} = \emptyset$, and $\mathcal{I}_{h_j} \backslash \mathcal{I}_{H_j} \subseteq \{(0, \ldots, 0)\}$ since the powers of all terms in h_j and in H_j are the same, with the only possible exception of the constant term, which may appear in one of these and vanish in the other. Now, going back to our choice (19) of η and using (17), we have

$$\eta = \frac{\varepsilon}{2(\ell + 2)}$$

$$\geq \|p\|_\infty n\pi^2 \left(\max \left[\max_{1 \leq m \leq n} (\overline{\deg} p_j)_m, \max_{j \leq k \leq \ell} \frac{2C_{\text{Jac}} |\mathcal{J}_k| \sum_{t=k}^{\ell} \text{Lip } p_t}{\eta} \right] + 2 \right)^{|J_j|+2}$$

$$\cdot \frac{2^{\frac{|J_j|}{2}+1}}{\min_{1 \leq k \leq n}(r_{j,k} + 2)^2}$$

$$\geq \|p\|_\infty n\pi^2 \left(\max \left[\max_{1 \leq m \leq n} (\overline{\deg} p_j)_m, \max_{j \leq k \leq \ell} \left\lceil \frac{2C_{\text{Jac}} \ell |\mathcal{J}_k| \sum_{t=k}^{\ell} \text{Lip } p_t}{\eta(\ell + 2)} \right\rceil \right] + 1 \right)^{|J_j|+2}$$

$$\cdot \frac{2^{\frac{|J_j|}{2}+1}}{\min_{1 \leq k \leq n}(r_{j,k} + 2)^2}.$$

for all $j \in \{1, \ldots, \ell\}$. Notice that after separating two of the $|J_j| + 2$ terms in the product and removing the $+1$ factor from them, we obtain

$$
\eta \geq \|p\|_\infty n\pi^2 \left(\max\left[\max_{1 \leq m \leq n} (\overline{\deg} \, p_j)_m, \max_{j \leq k \leq \ell} \left\lceil \frac{2C_{\text{Jac}} \, \ell \, |\mathcal{J}_k| \sum_{t=k}^\ell \text{Lip} \, p_t}{\eta(\ell+2)} \right\rceil \right] + 1 \right)^{|J_j|} 2^{\frac{|J_j|}{2}+1}
$$

$$
\cdot \frac{\max\left[\max_{1 \leq m \leq n} (\overline{\deg} \, p_j)_m, \max_{j \leq k \leq \ell} \left\lceil \frac{2C_{\text{Jac}} \, \ell \, |\mathcal{J}_k| \sum_{t=k}^\ell \text{Lip} \, p_t}{\eta(\ell+2)} \right\rceil \right]^2}{\min_{1 \leq k \leq n} (r_{j,k} + 2)^2}
$$

$$
\geq \|p\|_\infty n\pi^2 \left(\prod_{k \in J_j} \left([\max(\overline{\deg} \, p_j, \bar{D}_j, \ldots, \bar{D}_\ell)]_k + 1 \right) \right) 2^{\frac{|J_j|}{2}+1}
$$

$$
\cdot \frac{\max_{1 \leq k \leq n} [\max(\overline{\deg} \, p_j, \bar{D}_j, \ldots, \bar{D}_\ell)]_k^2}{\min_{1 \leq k \leq n} (r_{j,k} + 2)^2},
$$

where we have used the definition of $\bar{D}_l$, as well as the fact that each factor has been replaced by one that is smaller or equal, the original expression containing the maximum of them on each factor. Next, use $\overline{\deg} \, H_j \leq \max(\overline{\deg} \, p_j, \bar{D}_j, \ldots, \bar{D}_\ell)$ as well as $w(I) \leq |J_j|$ for every multi-index I in $\mathcal{I}_{H_j}$, which is true because $H_j \in \mathbb{R}[x_{J_j}]$, yielding

$$
\eta \geq \|p\|_\infty \left(2n\pi^2 \left(\prod_{1 \leq k \leq n} \left((\overline{\deg} \, H_j)_k + 1 \right) \right) \max_{I \in \mathcal{I}_{H_j}} \left(2^{w(I)/2} \max_{1 \leq k \leq n} \frac{i_k^2}{(r_{j,k}+2)^2} \right) \right)
$$

$$
\geq \left(\max_{[-1,1]^n} h_j - \min_{[-1,1]^n} h_j \right)
$$

$$
\cdot \left(2n\pi^2 \left(\prod_{1 \leq k \leq n} \left((\overline{\deg} \, H_j)_k + 1 \right) \right) \max_{I \in \mathcal{I}_{H_j}} \left(2^{w(I)/2} \max_{1 \leq k \leq n} \frac{i_k^2}{(r_{j,k}+2)^2} \right) \right),
$$

since we have $\max_{[-1,1]^n} h_j - \min_{[-1,1]^n} h_j \leq \|p\|_\infty$. With this bound for η, together with the fact that $\min_{[-1,1]^n} h_j \geq \eta$, we get

$$
h_j \geq h_j - \min_{[-1,1]^n} h_j + \eta
$$

$$
\geq h_j - \min_{[-1,1]^n} h_j + \left(\max_{[-1,1]^n} h_j - \min_{[-1,1]^n} h_j \right)
$$

$$
\cdot \left(2n\pi^2 \left(\prod_{1 \leq k \leq n} \left((\overline{\deg} \, H_j)_k + 1 \right) \right) \max_{I \in \mathcal{I}_{H_j}} \left(2^{w(I)/2} \max_{1 \leq k \leq n} \frac{i_k^2}{(r_{j,k}+2)^2} \right) \right),
$$

so that, by (20) and (22), $h_j \in \mathcal{P}_{\mathbf{r}, J_j}(\{1 - x_i^2\}_{i \in J_j})$ and hence

$$p = h_1 + \cdots + h_\ell \in \mathcal{P}_{\mathbf{r}, J_1}(\{1 - x_i^2\}_{i \in J_1}) + \cdots + \mathcal{P}_{\mathbf{r}, J_\ell}(\{1 - x_i^2\}_{i \in J_\ell}).$$

□

4.2 Proof of Theorem 8

Overview For this proof, we first use the sparse approximation theory developed in Sect. 3 to represent the sparse polynomial p as a sum of positive polynomials $h_1 + \cdots + h_\ell$, each of them depending on a clique of variables J_j. We then work with each of these polynomials h_j using the tools developed by Baldi–Mourrain [2] to write $h_j = \hat{f}_j + \hat{q}_j$, where $\hat{q}_j$ is by construction obviously an element of the corresponding quadratic module, and $\hat{f}_j$ is strictly positive on $[-1, 1]^n$. Thus Corollary 13 can be applied to $\hat{f}_j$, which shows that it belongs to the preordering, and then one argues (also following the ideas of [2]) that the preordering is contained in the quadratic module, hence giving that $\hat{f}_j$ is contained in the latter as well. In sum, this shows that h_j is in the quadratic module, which is what want. Most of the heavy lifting goes to estimating the minimum of $\hat{f}_j$ to justify the application of Corollary 13.

Proof of Theorem 8 For each $j = 1, \ldots, \ell$, pick $\mathbf{C}_j > 0$ such that the following two bounds are satisfied:

$$\mathbf{C}_j \geq 2\pi^2 |J_j|^{1 + \frac{16 \mathsf{L}_j}{3}} C_d^2 C_{\mathrm{Jac}}^{\frac{16 \mathsf{L}_j}{3}} 2^{1 + 2(4 + 3\frac{8}{3})\mathsf{L}_j} 3^{\frac{(16 + 8\ell)\mathsf{L}_j + 2}{3}} \bar{k}^{-\frac{2}{3}} \mathsf{c}_j^{\frac{8}{3}} (\max_{i \in K_j} \deg g_i)^2 (2(\ell + 2))^{8\mathsf{L}_j}, \tag{23}$$

$$\begin{aligned}
\mathbf{C}_j \geq{} & C_f (C_{\mathrm{Jac}} C_m)^{(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} |J_j| \pi^2 2^{4\mathsf{L}_j + \frac{|J_j|}{2} + 1 + (1 + \frac{4\mathsf{L}_j + 1}{3})(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} \\
& \times 3^{\ell(\mathsf{L}_j + 1) + (2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} (\ell + 2)^{1 + \mathsf{L}_j + \frac{4\mathsf{L}_j + 1}{3}(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} \bar{k} \\
& \times \mathsf{c}_j^{1 + \frac{3}{4}(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} \left(\sum_{i = j}^{\ell} |\mathcal{J}_i|^{2(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})} \right) \\
& \cdot (\max_{k \in K_j} \deg g_k + 1)^{(2\mathsf{L}_j + |J_j| + 2)(1 + \frac{8\mathsf{L}_j}{3})}. \tag{24}
\end{aligned}$$

Note that these only depend on $\mathbf{g}$ and $J_1, \ldots, J_\ell$.

Apply Lemma 15 to $f = p$, $f_i = p_i$, $\epsilon = 3\varepsilon/2\ell$, $\eta = \varepsilon/2(\ell + 2)$ to get polynomials $h_1, \ldots, h_\ell$ such that

$$p = h_1 + \cdots + h_\ell, \quad h_i \in \mathbb{R}[x_{J_i}], \quad h_i(x) \geq \eta = \frac{\varepsilon}{2(\ell + 2)} \text{ for } x \in S(\mathbf{g}), \tag{25}$$

and

$$\overline{\deg} h_i \leq \max(\overline{\deg} p_i, \bar{D}_{i,\ell}, \ldots, \bar{D}_{\ell,\ell})_{J_i}. \tag{26}$$

In the dense setting, Baldi–Mourrain [2] construct a family of single-variable polynomials

$$(\mathsf{h}_{t,m})_{(t,m)\in\mathbb{N}\times\mathbb{N}}$$

providing useful approximation properties that we have adapted to the (separated-variables) sparse setting and collected in Lemma 18. To state this, we set, for all $j = 1, \ldots, \ell$ and for $(t_j, m_j) \in \mathbb{N} \times \mathbb{N}$ as well as for $s_j > 0$,

$$q_{j,t_j,m_j}(x) := \sum_{i\in K_j} \mathsf{h}_{t_i,m_i}(g_i(x))^2 \, g_i(x), \tag{27}$$

$$f_{j,s_j,t_j,m_j}(x) := h_j(x) - s_j \, q_{j,t_j,m_j}(x). \tag{28}$$

Let us give an idea of what these functions do. The single-variable polynomial h_{t_j,m_j} is of degree m_j and roughly speaking approximates the function that equals 1 on $(-\infty, 0)$ and $1/t_j$ elsewhere. Thus q_{j,t_j,m_j} almost vanishes (for large t_j) on $S(\mathbf{g}_{K_j})$, and outside of this domain it is roughly a sum of multiples of the negative parts of $\mathbf{g}_{K_j}$'s entries. The definition of f_{j,s_j,t_j,m_j} is engineered to obtain a polynomial that is almost equal to h_j in $S(\mathbf{g}_{K_j})$ yet remains positive throughout $[-1, 1]^n$. Instead of going into the details of the construction, we record the properties we need in the following lemma. □

Lemma 18 (a version of [2, Props. 2.13, 3.1, and 3.2, Lem. 3.5]) *Assume* (3) *and the Archimedean conditions* (4) *are satisfied. Then for each $j = 1, \ldots, \ell$ there are values s_j, t_j, m_j of the parameters involved in Definition* (27) *and Definition* (28)*, such that the following holds with the shorthands*

$$\hat{f}_j = f_{j,s_j,t_j,m_j} \quad \text{and} \quad \hat{q}_j = s_j q_{j,t_j,m_j} : \tag{29}$$

i. [2, *Prop. 3.1*] *gives*

$$\hat{f}_j(x) \geq \frac{1}{2} \min_{y\in S(\mathbf{g}_{K_j})} h_j(y) \geq \frac{\eta}{2} = \frac{\varepsilon}{4(\ell+2)} \quad \text{for all} \quad x \in [1, 1]^n.$$

ii. We have $\hat{q}_j \in \mathcal{Q}_{\mathbf{r}, J_j}(\mathbf{g})$ for all multi-indices $\mathbf{r} = (r_1, \ldots, r_n)$ with

$$r_i \geq (2m_j + 1) \max_{k\in K_j} (\overline{\deg} \, g_k)_i.$$

iii. [2, *eq.* (20)] *gives the existence of a constant $C_m > 0$ such that*

$$m_j \leq C_m \mathsf{c}_j^{\frac{4}{3}} \bar{k}^{\frac{1}{3}} 2^{4\mathsf{L}_j} (\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{\min_{x\in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\frac{4\mathsf{L}_j+1}{3}}.$$

iv. [2, eq. (16)] *gives the existence of a constant $C_f > 0$ such that*

$$\|\hat{f}_j\|_\infty \le C_f \|h_j\|_\infty 2^{3\mathsf{L}_j} \bar{k} \mathsf{c}_j (\deg h_j)^{2\mathsf{L}_j} \left(\frac{\min_{x \in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\mathsf{L}_j}.$$

v. [2, eq. (17)] *gives the existence of a constant $C_d > 0$ such that*

$$\deg \hat{f}_j \le C_d 2^{4\mathsf{L}_j} \bar{k}^{\frac{1}{3}} \mathsf{c}_j^{\frac{4}{3}} (\max_{i \in K_j} \deg g_i)(\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{\min_{x \in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\frac{4\mathsf{L}_j+1}{3}}.$$

Item (ii.) follows[5] from $\deg \mathsf{h}_{t_j,m_j} = m_j$ and the definition of q_{j,t_j,m_j}. The proofs of the other items can be found in the indicated sources.

Take $s_j, t_j, m_j, \hat{f}_j, \hat{q}_j$ for $j = 1, \ldots, \ell$ satisfying the properties (i)–(v) collected in Lemma 18.

Continuing with the proof of Theorem 8, denote

$$F_j := \frac{\hat{f}_j - \min_{[-1,1]^n} \hat{f}_j}{\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j}.$$

Since $\hat{f}_j \ge \varepsilon/4(\ell+2)$ on $[-1,1]^n$, we may apply Corollary 13 with $p = F_j$ to get that

$$F_j + \epsilon_j \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i \in J_j}) \tag{30}$$

as long as

$$\epsilon_j \ge 2|J_j|\pi^2 \left(\prod_{i \in J_j}((\overline{\deg}\, F_j)_i + 1) \right) \max_{I=(i_1,\ldots,i_n) \in \mathcal{I}_{F_j}} \left(2^{w(I)/2} \max_{1 \le k \le n} \frac{i_k^2}{(r_{j,k}+2)^2} \right), \tag{31}$$

and (9) are verified.

In this context, the condition (9) required in Corollary 13 is equivalent to the theorem's assumption (6); let us show how this works: First, using $i_k \le (\overline{\deg}\, F_j)_k$, $\overline{\deg}\, F_j \le \overline{\deg}\, \hat{f}_j \le (\deg \hat{f}_j)\mathbf{1}$, Lemma 18(v.), we get

$$2\pi^2 |J_j| i_k^2 \le 2\pi^2 |J_j| (\overline{\deg}\, F_j)_k^2 \le 2\pi^2 |J_j| (\deg \hat{f}_j)^2$$

$$\le 2\pi^2 |J_j| \left(C_d 2^{4\mathsf{L}_j} \bar{k}^{\frac{1}{3}} \mathsf{c}_j^{\frac{4}{3}} (\max_{i \in K_j} \deg g_i)(\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{\min_{x \in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\frac{4\mathsf{L}_j+1}{3}} \right)^2.$$

---

[5] This calculation is slightly different to the one in [2, Lem. 3.5] because the definition of q_{j,t_j,m_j} (or in their notations, $f - p$) differs from the one given there in that the functions h_j are squared here, an idea we take from the exposition of the results of [2] in the dissertation of L. Baldi and that is advantageous because then $q_{j,t_j,m_j} \in \mathcal{Q}_{\mathbf{r},J_j}(\mathbf{g})$ automatically. This requires taking m_j twice as large, and we absorb this difference into the constant C_m.

Now use Eq. (26) to get that this is

$$\leq 2\pi^2 |J_j|$$

$$\cdot \left(C_d 2^{4L_j} \bar{k}^{\frac{1}{3}} \mathsf{c}_j^{\frac{4}{3}} (\max_{i \in K_j} \deg g_i) \max \left(\overline{\deg}\, p_j, \bar{D}_{j,\ell}, \dots, \bar{D}_{\ell,\ell} \right)^{\frac{8L_j}{3}} \left(\frac{\min_{x \in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\frac{4L_j+1}{3}} \right)^2$$

$$\leq 2\pi^2 |J_j| \left(C_d 2^{4L_j} \bar{k}^{\frac{1}{3}} \mathsf{c}_j^{\frac{4}{3}} (\max_{i \in K_j} \deg g_i) \max \left(\overline{\deg}\, p_j, \bar{D}_{j,\ell}, \dots, \bar{D}_{\ell,\ell} \right)^{\frac{8L_j}{3}} \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)} \right)^{\frac{4L_j+1}{3}} \right)^2,$$

where we have also used the fact that

$$\|h_j\|_\infty \leq 3 \times 2^{\ell-1} \sum_{i=1}^\ell \|p_i\|_\infty \leq 3^\ell \sum_{i=1}^\ell \|p_i\|_\infty, \tag{32}$$

and the last estimate from (25). Next, use (11), $|\mathcal{J}_j| \leq |J_j|$, $\varepsilon_i - 2\eta_i = \frac{\ell+6}{2\ell(\ell+2)}\varepsilon$ to get

$$2\pi^2 |J_j| i_k^2$$

$$\leq 2\pi |J_j| \left(C_d 2^{4L_j} \bar{k}^{-\frac{1}{3}} \mathsf{c}_j^{\frac{4}{3}} (\max_{i \in K_j} \deg g_i)(\deg p_j)^{\frac{8L_j}{3}} \left(\frac{2C_{\mathrm{Jac}} |J_j|(3\sum_{i=1}^\ell \mathrm{Lip}\, p_i)}{\frac{\ell+6}{2\ell(\ell+2)}\varepsilon} \right)^{\frac{8L_j}{3}} \right.$$

$$\left. \cdot \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)} \right)^{\frac{4L_j+1}{3}} \right)^2$$

$$\leq \mathbf{C}_j \left(\left(\sum_{i=1}^\ell \|p_i\|_\infty \right)^{\frac{4L_j+1}{3}} (\deg p_j)^{\frac{8L_j}{3}} \frac{\left(\sum_{i=1}^\ell \mathrm{Lip}\, p_i \right)^{\frac{8L_j}{3}}}{\varepsilon^{\frac{(8+4)L_j+1}{3}}} \right)^2 \leq (r_{j,k}+2)^2,$$

where we have additionally used Eq. (23) and our assumption (6); this is precisely (9).

We would next like to show that

$$\hat{f}_j \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i \in J_j}). \tag{33}$$

Let us first explain why this will be enough to prove the theorem. Once we have (33), by Lemma 19 $\hat{f}_j$ is also contained in $\mathcal{Q}_{\mathbf{r}_j+2, J_j}(1 - \sum_{i \in J_j} x_i^2)$, and it is our assumption (4) that $\mathcal{Q}_{\mathbf{r}_j+2, J_j}(1 - \sum_{i \in J_j} x_i^2) \subseteq \mathcal{Q}_{\mathbf{r}_j+2, J_j}(\mathbf{g}_{K_j})$. In other words, we have

$$\hat{f}_j \in \mathcal{Q}_{\mathbf{r}_j+2, J_j}(\mathbf{g}_{K_j}).$$

By Lemma 18(ii.), $\hat{q}_j$ also belongs to $\mathcal{Q}_{\mathbf{r}_j+2, J_j}(\mathbf{g}_{K_j})$, so we can conclude that

$$h_j \in \mathcal{Q}_{\mathbf{r}_j+2, J_j}(\mathbf{g}_{K_j}),$$

which is equivalent to the conclusion of the theorem.

Thus we need to prove (33). Let us first show that

$$
\frac{\eta}{2} = \frac{\varepsilon}{4(\ell+2)} \geq (\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j) \, 2|J_j|\pi^2 \left(\prod_{i \in J_j}((\overline{\deg} F_j)_i + 1) \right)
$$
$$
\times \max_{I=(i_1,\dots,i_n)\in\mathcal{I}_{F_j}} \left(2^{w(I)/2} \max_{1\leq k \leq n} \frac{i_k^2}{(r_{j,k}+2)^2} \right) \tag{34}
$$

implies (33) Observe that (30) is equivalent to

$$
\hat{f}_j - \min_{[-1,1]^n} \hat{f}_j + \epsilon_j(\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j) \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i\in J_j}). \tag{35}
$$

If (34) were true, we would then have

$$
\hat{f}_j \geq \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j + \frac{\varepsilon}{4(\ell+2)}
$$
$$
\geq \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j + (\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j) \, 2|J_j|\pi^2 \left(\prod_{i \in J_j}((\overline{\deg} F_j)_i + 1) \right)
$$
$$
\times \max_{I=(i_1,\dots,i_n)\in\mathcal{I}_{F_j}} \left(2^{w(I)/2} \max_{1\leq k \leq n} \frac{i_k^2}{(r_{j,k}+2)^2} \right).
$$

So in view of (31) and (35), we would indeed have $\hat{f}_j \in \mathcal{P}_{\mathbf{r}_j, J_j}(\{1 - x_i^2\}_{i\in J_j})$, which is (33).

Let us now collect some preliminary estimates that will help us to prove (34). For $I \in \mathcal{I}_{F_j}$ we have $w(I) \leq |J_j|$ so we estimate

$$
2^{w(I)/2} \leq 2^{|J_j|/2}. \tag{36}
$$

We also estimate

$$
\frac{i_k^2}{(r_{j,k}+2)^2} \leq \frac{\left(\overline{\deg} \hat{f}_j\right)_k^2}{\min_{1\leq l \leq n}(r_{j,l}+2)^2}. \tag{37}
$$

Now we will estimate $\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j$ from above. Using Lemma 18(i.) and (iv.), we get

$$
\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j \leq C_f \|h_j\|_\infty 2^{3\mathsf{L}_j} \bar{k} \mathsf{c}_j (\deg h_j)^{2\mathsf{L}_j} \left(\frac{\min_{x\in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\mathsf{L}_j}
$$

$$- \frac{\varepsilon}{4(\ell + 2)}.$$

Use (25), (26) and (32) to see that this is

$$\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j$$

$$\leq C_f \left(3^\ell \sum_i \|p_i\|_\infty \right) 2^{3\mathsf{L}_j} \bar{k} \mathsf{c}_j \max(\deg p_j, |\bar{D}_{j,\ell}|, \ldots, |\bar{D}_{\ell,\ell}|)^{2\mathsf{L}_j}$$

$$\cdot \left(\frac{\varepsilon}{2(\ell + 2)3^\ell \sum_i \|p_i\|_\infty} \right)^{-\mathsf{L}_j}$$

$$= C_f \left(3^\ell \sum_i \|p_i\|_\infty \right) 2^{3\mathsf{L}_j} \bar{k} \mathsf{c}_j \max(\deg p_j, |\mathcal{J}_j||D_{j,\ell}|, \ldots, |\mathcal{J}_\ell||D_{\ell,\ell}|)^{2\mathsf{L}_j}$$

$$\cdot \left(\frac{\varepsilon}{2(\ell + 2)3^\ell \sum_i \|p_i\|_\infty} \right)^{-\mathsf{L}_j}. \tag{38}$$

For the last line, we have used the definition of $\bar{D}_{l,m}$ as in Lemma 15.

Additionally, we obtain the following estimate

$$\overline{\deg}\, \hat{f}_j \leq \max\left(\overline{\deg}\, h_j, \overline{\deg}\, \hat{q}_j \right)$$

$$\leq \max\left(\overline{\deg}\, p_j, \bar{D}_{j,\ell}, \ldots, \bar{D}_{\ell,\ell}, (2m_j + 1) \max_{k \in K_j} \overline{\deg}\, g_k \right)$$

$$\leq \max\left(\overline{\deg}\, p_j, \bar{D}_{j,\ell}, \ldots, \bar{D}_{\ell,\ell}, \right.$$
$$\left. \left(2 \left(C_m \mathsf{c}_j^{\frac{3}{4}} \bar{k}^{-\frac{1}{3}} 2^{4\mathsf{L}_j} (\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{\min_{x \in S(\mathbf{g})} h_j(x)}{\|h_j\|_\infty} \right)^{-\frac{4\mathsf{L}_j+1}{3}} \right) + 1 \right) \max_{k \in K_j} \overline{\deg}\, g_k \right)$$

$$\leq \max\left(\overline{\deg}\, p_j, \bar{D}_{j,\ell}, \ldots, \bar{D}_{\ell,\ell}, \right.$$
$$\left. \left(C_m 2^{4\mathsf{L}_j+1} \mathsf{c}^{\frac{3}{4}} \bar{k}^{-\frac{1}{3}} (\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)} \right)^{\frac{4\mathsf{L}_j+1}{3}} + 1 \right) \max_{k \in K_j} \overline{\deg}\, g_k \right). \tag{39}$$

The first inequality comes from (28), the second one from (26) and Lemma 18(ii.), the third one from Lemma 18(iii.), and the last one from (25), (32), and (26). Compare with Lemma 18(v.).

With those estimates under our belt, we now turn to showing that (34) is true. Using (38), (36), (37), as well as $\overline{\deg}\, F_j \leq \overline{\deg}\, \hat{f}_j$, we can start to estimate the right-hand side of (34) by

$$(\max_{[-1,1]^n} \hat{f}_j - \min_{[-1,1]^n} \hat{f}_j)\, 2|J_j|\pi^2 \left(\prod_{i \in J_j} ((\overline{\deg}\, F_j)_i + 1) \right)$$

$$\times \max_{I=(i_1,\ldots,i_n)\in\mathcal{I}_{F_j}} \left(2^{w(I)/2} \max_{1\leq k\leq n} \frac{i_k^2}{(r_{j,k}+2)^2}\right)$$

$$\leq C_f \left(3^\ell \sum_i \|p_i\|_\infty\right) 2^{3\mathsf{L}_j} \bar{k} \mathsf{c}_j \max(\deg p_j, |\mathcal{J}_j|D_{j,\ell},\ldots,|\mathcal{J}_\ell|D_{\ell,\ell})^{2\mathsf{L}_j}$$

$$\times \left(\frac{\varepsilon}{2(\ell+2)3^\ell \sum_i \|p_i\|_\infty}\right)^{-\mathsf{L}_j}$$

$$\times 2|J_j|\pi^2 \left(\max_{1\leq i\leq n} (\overline{\deg}\,\hat{f}_j)_i + 1\right)^{|J_j|} \frac{2^{|J_j|/2} \max_{1\leq i\leq n} \left(\overline{\deg}\,\hat{f}_j\right)_i^2}{\min_{1\leq k\leq n}(r_{j,k}+2)^2}.$$

Next, denote

$$m_{j,\ell} = \max(\deg p_j, |\mathcal{J}_j|D_{j,\ell},\ldots,|\mathcal{J}_\ell|D_{\ell,\ell}),$$

This will help us to reorganize and consolidate the terms. Use (39) to see that this is

$$\leq \varepsilon^{-\mathsf{L}_j} C_f |J_j|\pi^2 \left(3^\ell \sum_i \|p_i\|_\infty\right)^{\mathsf{L}_j+1} 2^{4\mathsf{L}_j+\frac{|J_j|}{2}+1}(\ell+2)^{\mathsf{L}_j}\bar{k}\mathsf{c}_j m_{j,\ell}^{2\mathsf{L}_j}$$

$$\times \left(\max_{1\leq i\leq n} (\overline{\deg} f_j)_i + 1\right)^{|J_j|+2} \frac{1}{\min_{1\leq k\leq n}(r_{j,k}+2)^2}$$

$$\leq \varepsilon^{-\mathsf{L}_j} C_f |J_j|\pi^2 \left(3^\ell \sum_i \|p_i\|_\infty\right)^{\mathsf{L}_j+1} 2^{4\mathsf{L}_j+\frac{|J_j|}{2}+1}(\ell+2)^{\mathsf{L}_j}\bar{k}\mathsf{c}_j m_{j,\ell}^{2\mathsf{L}_j}$$

$$\times \left(\max\left(\overline{\deg}\,p_j, \bar{D}_{j,\ell},\ldots,\bar{D}_{\ell,\ell},\right.\right.$$

$$\left(C_m 2^{4\mathsf{L}_j+1}\mathsf{c}^{\frac{3}{4}}\bar{k}^{-\frac{1}{3}}(\deg h_j)^{\frac{8\mathsf{L}_j}{3}} \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)}\right)^{\frac{4\mathsf{L}_j+1}{3}} + 1\right)$$

$$\left.\left.\times \max_{k\in K_j} \overline{\deg}\, g_k\right) + 1\right)^{|J_j|+2} \frac{1}{\min_{1\leq k\leq n}(r_{j,k}+2)^2}$$

$$\leq \varepsilon^{-\mathsf{L}_j} C_f |J_j|\pi^2 \left(3^\ell \sum_i \|p_i\|_\infty\right)^{\mathsf{L}_j+1} 2^{4\mathsf{L}_j+\frac{|J_j|}{2}+1}(\ell+2)^{\mathsf{L}_j}\bar{k}\mathsf{c}_j m_{j,\ell}^{2\mathsf{L}_j}$$

$$\times \left(\max\left(\overline{\deg}\,p_j, \bar{D}_{j,\ell},\ldots,\bar{D}_{\ell,\ell},\right.\right.$$

$$\left(C_m 2^{4\mathsf{L}_j+1} \mathsf{c}^{\frac{3}{4}} \bar{k}^{-\frac{1}{3}} m_{j,\ell}^{\frac{8\mathsf{L}_j}{3}} \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)} \right)^{\frac{4\mathsf{L}_j+1}{3}} + 1 \right)$$

$$\left. \times \max_{k \in K_j} \overline{\deg} \, g_k + 1 \right)^{|J_j|+2} \frac{1}{\min_{1 \le k \le n}(r_{j,k}+2)^2}.$$

Now use (11) as well as $\varepsilon_i - 2\eta_i = \frac{\ell+6}{2\ell(\ell+2)}\varepsilon$ to see that the above is bounded by

$$\varepsilon^{-\mathsf{L}_j} C_f |J_j| \pi^2 \left(3^\ell \sum_i \|p_i\|_\infty \right)^{\mathsf{L}_j+1} 2^{4\mathsf{L}_j + \frac{|J_j|}{2}+1} (\ell+2)^{\mathsf{L}_j} \bar{k} \mathsf{c}_j$$

$$\times \left(\max\left[\deg p_j, |\mathcal{J}_j|^2 \frac{2\ell(\ell+2)}{\ell+6} \frac{2C_{\mathrm{Jac}} \sum_{k=j}^\ell \mathrm{Lip}\,\hat{f}_k}{\varepsilon} + 1, \ldots, \right. \right.$$

$$|\mathcal{J}_\ell|^2 \frac{2\ell(\ell+2)}{\ell+6} \frac{2C_{\mathrm{Jac}} \mathrm{Lip}\,\hat{f}_\ell}{\varepsilon} + 1,$$

$$\left. \left(C_m 2^{4\mathsf{L}_j+1} \mathsf{c}^{\frac{3}{4}} \bar{k}^{-\frac{1}{3}} \left(\frac{3^\ell \sum_i \|p_i\|_\infty}{\varepsilon/2(\ell+2)} \right)^{\frac{4\mathsf{L}_j+1}{3}} + 1 \right) \max_{k \in K_j} \overline{\deg}\, g_k \right]$$

$$\left. + 1 \right)^{(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \frac{1}{\min_{1 \le k \le n}(r_{j,k}+2)^2}$$

$$\le \varepsilon^{-\mathsf{L}_j - \frac{4\mathsf{L}_j+1}{3}(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} C_f (C_{\mathrm{Jac}} C_m)^{(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} |J_j| \pi^2 \left(3^\ell \sum_i \|p_i\|_\infty \right)^{\mathsf{L}_j+1}$$

$$\times 2^{4\mathsf{L}_j + \frac{|J_j|}{2}+1+(1+\frac{4\mathsf{L}_j+1}{3})(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} (\ell+2)^{1+\mathsf{L}_j+\frac{4\mathsf{L}_j+1}{3}(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \bar{k}$$

$$\times \mathsf{c}_j^{1+\frac{3}{4}(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \left(\sum_{i=j}^\ell |\mathcal{J}_i|^{2(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \right)$$

$$\times (3 \deg p_j \sum_{i=1}^\ell \mathrm{Lip}\, p_i)^{(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})}$$

$$\times (\max_{k \in K_j} \deg g_k + 1)^{(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \frac{1}{\min_{1 \le k \le n}(r_{j,k}+2)^2}.$$

Finally use (24) and then (5) to get that the above is less than

$$\mathsf{C}_j \frac{\varepsilon^{-\mathsf{L}_j - \frac{4\mathsf{L}_j+1}{3}(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})} \left(\sum_i \|p_i\|_\infty \right)^{\mathsf{L}_j+1} (\deg p_j \sum_{i=1}^\ell \mathrm{Lip}\, p_i)^{(2\mathsf{L}_j+|J_j|+2)(1+\frac{8\mathsf{L}_j}{3})}}{\min_{1 \le k \le n}(r_{j,k}+2)^2}$$

$$\le \frac{\varepsilon}{4(\ell+2)}.$$

This shows that (34) holds, and hence also (33), which proves the theorem. □

Lemma 19 ([2, Lemma 3.8]) *Let $J \subset \{1, \ldots, n\}$, and let $\mathbf{r} = (r_1, \ldots, r_n)$ be a multi-index such that $r_i > 0$ only if $i \in J$.*

The quadratic module $\mathcal{Q}_{\mathbf{r}+2, J}(1 - \sum_{i \in J} x_i^2)$ contains the preordering $\mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J})$,

$$\mathcal{P}_{\mathbf{r}, J}(\{1 - x_i^2\}_{i \in J}) \subseteq \mathcal{Q}_{\mathbf{r}+2, J}(1 - \sum_{i \in J} x_i^2).$$

Proof This follows from

$$1 \pm x_i = \frac{1}{2}(1 - x_i^2 + (1 \pm x_i)^2) = \frac{1}{2}((1 - \|x\|^2) + \sum_{\substack{j \in J \\ j \neq i}} x_j^2 + (1 \pm x_i)^2)$$

and

$$1 - x_i^2 = (1 - x_i)(1 + x_i)$$

$$= \frac{1}{4}(1 - \|x\|_J^2)\left(2\sum_{j \neq i} x_j^2 + (1 - x_i)^2 + (1 + x_i)^2\right)$$

$$+ \frac{1}{4}(1 - \|x\|_J^2)^2 + \left(\sum_{j \neq i} x_j^2\right)^2 + \sum_{j \neq i} x_j^2((1 - x_i)^2 + (1 + x_i)^2).$$

The increase of **2** in **r** stems from the fact that $\deg(1 - x_i^2) = 2$ while the degree of the right-hand side above is 4. $\square$

4.3 An asymptotic lemma

Lemma 20 *For $a, b, c, d, p, q > 0$, with $cq - ap \neq 0$, we have*

$$\lim_{\varepsilon \searrow 0} \frac{\log\left[\left(\dfrac{a + b\varepsilon^{-p}}{b\varepsilon^{-p}}\right) \Big/ \left(\dfrac{c + d\varepsilon^{-q}}{d\varepsilon^{-q}}\right)\right]}{(cq - ap)\log \varepsilon} = 1.$$

In other words, as $\varepsilon \searrow 0$,

$$\frac{\left(\dfrac{a + b\varepsilon^{-p}}{b\varepsilon^{-p}}\right)}{\left(\dfrac{c + d\varepsilon^{-q}}{d\varepsilon^{-q}}\right)} \approx \varepsilon^{cq - ap} \to 0.$$

Proof Recall the Stirling series can be used [19, p. 293–294] to see that

$$\log(n!) = n \log n + n + \frac{1}{2}\log(2\pi n) + O(1/n)$$

for all large $n > 0$. We use it to get that,

$$\frac{\log\left[\binom{a+b\varepsilon^{-p}}{b\varepsilon^{-p}} \Big/ \binom{c+d\varepsilon^{-q}}{d\varepsilon^{-q}}\right]}{(cq-ap)\log\varepsilon}$$

$$= \frac{\log\left[\frac{(a+b\varepsilon^{-p})!}{a!(b\varepsilon^{-p})!} \Big/ \frac{(c+d\varepsilon^{-q})!}{c!(d\varepsilon^{-q})!}\right]}{(cq-ap)\log\varepsilon}$$

$$= \frac{\log(a+b\varepsilon^{-p})! - \log a! - \log(b\varepsilon^{-p})! - \log(c+d\varepsilon^{-q})! + \log c! + \log(d\varepsilon^{-q})!}{(cq-ap)\log\varepsilon}$$

$$= \frac{1}{(cq-ap)\log\varepsilon}\Big[(a+b\varepsilon^{-p})\log(a+b\varepsilon^{-p}) - b\varepsilon^{-p}\log(b\varepsilon^{-p}) \tag{40}$$

$$- (c+d\varepsilon^{-q})\log(c+d\varepsilon^{-q}) + d\varepsilon^{-q}\log(d\varepsilon^{-q}) \tag{41}$$

$$+ (a+b\varepsilon^{-p}) - (b\varepsilon^{-p}) - (c+d\varepsilon^{-q}) + (d\varepsilon^{-q}) \tag{42}$$

$$+ \frac{1}{2}\Big(\log(2\pi(a+b\varepsilon^{-p})) - \log(2\pi b\varepsilon^{-p}) - \log(2\pi(c+d\varepsilon^{-q})) + \log(2\pi d\varepsilon^{-q})\Big) \tag{43}$$

$$+ \log c! - \log a! + O(\varepsilon^p + \varepsilon^q)\Big]. \tag{44}$$

Notice that the terms in both (42)—where the non-constant terms cancel out—and in (44) are asymptotically much smaller than the absolute value of the denominator, which tends to $+\infty$. Line (43) is

$$\frac{1}{2}\frac{\log(2\pi(a+b\varepsilon^{-p})) - \log(2\pi b\varepsilon^{-p}) - \log(2\pi(c+d\varepsilon^{-q})) + \log(2\pi d\varepsilon^{-q})}{(cq-ap)\log\varepsilon}$$

$$= \frac{\log(\frac{a+b\varepsilon^{-p}}{b\varepsilon^{-p}}\frac{d\varepsilon^{-q}}{c+d\varepsilon^{-q}})}{2(cq-ap)\log\varepsilon} = \frac{\log(\frac{a\varepsilon^p+b}{b}\frac{d}{c\varepsilon^q+d})}{2(cq-ap)\log\varepsilon}$$

As $\varepsilon \searrow 0$, the factors inside the logarithm tend to 1, so the numerator tends to 0, while the denominator tends to $\pm\infty$, and the quotient tends to 0. Let us now show that the remaining two lines (40)–(41) together tend to 1 in the limit. Now,

$$\lim_{\varepsilon\searrow0} b\varepsilon^{-p}\left(\log(a+b\varepsilon^{-p}) - \log(b\varepsilon^{-p})\right) = a \quad \text{and}$$

$$\lim_{\varepsilon\searrow0} d\varepsilon^{-p}\left(\log(c+d\varepsilon^{-p}) - \log(d\varepsilon^{-p})\right) = c,$$

so we get

$$\lim_{\varepsilon\searrow0} \frac{1}{(cq-ap)\log\varepsilon}\Big((a+b\varepsilon^{-p})\log(a+b\varepsilon^{-p}) - b\varepsilon^{-p}\log(b\varepsilon^{-p})$$

$$- (c+d\varepsilon^{-q})\log(c+d\varepsilon^{-q}) + d\varepsilon^{-q}\log(d\varepsilon^{-q})\Big)$$

$$= \lim_{\varepsilon\searrow0} \frac{a - c + a\log(a+b\varepsilon^{-p}) - c\log(c+d\varepsilon^{-q})}{(cq-ap)\log\varepsilon}$$

$$= \lim_{\varepsilon \searrow 0} \frac{a \log(a + b\varepsilon^{-p}) - c \log(c + d\varepsilon^{-q})}{(cq - ap) \log \varepsilon}.$$

In this quotient, both the numerator and the denominator tend to $\pm\infty$, so we can apply a version of the l'Hôpital rule, which states that, if the limit of the quotient of their derivatives exists, then the original limit above equals that limit. Taking the limit of the quotient of the derivatives gives

$$\lim_{\varepsilon \searrow 0} \frac{\frac{cdq\varepsilon^{-1-q}}{c+d\varepsilon^{-q}} - \frac{abp\varepsilon^{-1-p}}{a+b\varepsilon^{-p}}}{(cq - ap)/\varepsilon} = \lim_{\varepsilon \searrow 0} \frac{cdq\varepsilon^{-q}(a + b\varepsilon^{-p}) - abp\varepsilon^{-p}(c + d\varepsilon^{-q})}{(cq - ap)(a + b\varepsilon^{-p})(c + d\varepsilon^{-q})}$$

$$= \lim_{\varepsilon \searrow 0} \frac{cdq(a\varepsilon^p + b) - abp(c\varepsilon^q + d)}{(cq - ap)(a\varepsilon^q + b)(c\varepsilon^p + d)}$$

$$= \lim_{\varepsilon \searrow 0} \frac{cdqb - abpd}{(cq - ap)bd} = 1. \qquad \square$$

Declarations

Conflict of interest The authors have no interests to delcare.

References

1. Bach, F., Rudi, A.: Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials (2022). arXiv:2211.04889
2. Baldi, L., Mourrain, B.: On the effective Putinar's Positivstellensatz and moment approximation. Math. Program. **200**, 71–103 (2023)
3. Doherty, A.C, Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere (2012). arXiv:1210.5048
4. Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere and applications in quantum information theory. Math. Program. **190**(1), 331–360 (2021)

5. Gribling, S., Polak, S., Slot, L.: A note on the computational complexity of the moment-SOS hierarchy for polynomial optimization. In Proceedings of the 2023 International Symposium on Symbolic and Algebraic Computation, ISSAC '23, pp. 280–288. ACM, New York, NY, USA (2023)

6. Grimm, D., Netzer, T., Schweighofer, M.: A note on the representation of positive polynomials with structured sparsity. Arch. Math. **89**(5), 399–403 (2007)

7. Jackson, D.: The Theory of Approximation, volume 11 of Colloquium Publications. American Mathematical Society, Providence (1930)

8. Karlin, S., Shapley, L.S.: Geometry of Moment Spaces, volume 12. American Mathematical Society, Providence (1953)

9. Kirschner, F., De Klerk, E.: Convergence rates of RLT and Lasserre-type hierarchies for the generalized moment problem over the simplex and the sphere. Optim. Lett. **16**, 2191–2208 (2022)

10. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)

11. Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM J. Optim. **17**(3), 822–843 (2006)

12. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications, vol. 1. World Scientific, Singapore (2009)

13. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry, pp. 157–270. Springer, Berlin (2009)

14. Laurent, M., Slot, L.: An effective version of Schmüdgen's Positivstellensatz for the hypercube. Optim. Lett. **17**, 515–530 (2023)

15. Magron, V., Wang, J.: Sparse Polynomial Optimization: Theory and Practice. Series on Optimization and Its Applications. World Scientific Press, Singapore (2023)

16. Mai, N.H.A., Magron, V.: On the complexity of Putinar–Vasilescu's Positivstellensatz. J. Complex. **72**, 101663 (2022)

17. Mai, N.H.A., Magron, V., Lasserre, J.: A sparse version of Reznick's Positivstellensatz. Math. Oper. Res. **48**(2), 812–833 (2022)

18. Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. J. Complex. **23**(1), 135–150 (2007)

19. Olver, F.: Asymptotics and Special Functions. CRC Press, Boca Raton (1997)

20. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Math. Program. **96**(2), 293–320 (2003)

21. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**(3), 969–984 (1993)

22. Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Math. Z. **220**(1), 75–97 (1995)

23. Schmüdgen, K.: The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**(1), 203–206 (1991)

24. Schweighofer, M.: On the complexity of Schmüdgen's Positivstellensatz. J. Complex. **20**(4), 529–543 (2004)

25. Slot, L.: Sum-of-squares hierarchies for polynomial optimization and the Christoffel–Darboux kernel. SIAM J. Optim. **32**(4), 2612–2635 (2022)

26. Slot, L., Laurent, M.: Sum-of-squares hierarchies for binary polynomial optimization. Math. Program. **197**, 621–660 (2022)

27. Waki, H., Kim, S., Kojima, M., Muramatsu, M.: Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. SIAM J. Optim. **17**(1), 218–242 (2006)

28. Weiße, A., Wellein, G., Alvermann, A., Fehske, H.: The kernel polynomial method. Rev. Mod. Phys. **78**(1), 275 (2006)