FULL LENGTH PAPER

Series B

Sum-of-squares hierarchies for binary polynomial optimization

Lucas Slot[1] · Monique Laurent[1,2]

Abstract

We consider the sum-of-squares hierarchy of approximations for the problem of minimizing a polynomial f over the boolean hypercube $\mathbb{B}^n = \{0, 1\}^n$. This hierarchy provides for each integer $r \in \mathbb{N}$ a lower bound $f_{(r)}$ on the minimum $f_{\min}$ of f, given by the largest scalar λ for which the polynomial $f - \lambda$ is a sum-of-squares on $\mathbb{B}^n$ with degree at most $2r$. We analyze the quality of these bounds by estimating the worst-case error $f_{\min} - f_{(r)}$ in terms of the least roots of the Krawtchouk polynomials. As a consequence, for fixed $t \in [0, 1/2]$, we can show that this worst-case error in the regime $r \approx t \cdot n$ is of the order $1/2 - \sqrt{t(1-t)}$ as n tends to ∞. Our proof combines classical Fourier analysis on $\mathbb{B}^n$ with the polynomial kernel technique and existing results on the extremal roots of Krawtchouk polynomials. This link to roots of orthogonal polynomials relies on a connection between the hierarchy of lower bounds $f_{(r)}$ and another hierarchy of upper bounds $f^{(r)}$, for which we are also able to establish the same error analysis. Our analysis extends to the minimization of a polynomial over the q-ary cube $(\mathbb{Z}/q\mathbb{Z})^n$. Furthermore, our results apply to the setting of matrix-valued polynomials.

Keywords Binary polynomial optimization · Lasserre hierarchy · Sum-of-squares polynomials · Fourier analysis · Krawtchouk polynomials · Polynomial kernels · Semidefinite programming · Polynomial matrices

✉ Lucas Slot
lucas.slot@cwi.nl

Monique Laurent
monique.laurent@cwi.nl

1 Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

2 Tilburg University, Tilburg, The Netherlands

1 Introduction

We consider the problem of minimizing a polynomial $f \in \mathbb{R}[x]$ of degree $d \leq n$ over the n-dimensional boolean hypercube $\mathbb{B}^n = \{0, 1\}^n$, i.e., of computing

$$f_{\min} := \min_{x \in \mathbb{B}^n} f(x). \tag{1}$$

This optimization problem is NP-hard in general. Indeed, as is well-known, one can model an instance of MAX- CUT on the complete graph K_n with edge weights $w = (w_{ij})$ as a problem of the form (1) by setting:

$$f(x) = - \sum_{1 \leq i < j \leq n} w_{ij}(x_i - x_j)^2.$$

As another example one can compute the stability number $\alpha(G)$ of a graph $G = (V, E)$ via the program

$$\alpha(G) = \max_{x \in \mathbb{B}^{|V|}} \sum_{i \in V} x_i - \sum_{\{i,j\} \in E} x_i x_j.$$

One may replace the boolean cube $\mathbb{B}^n = \{0, 1\}^n$ by the discrete cube $\{\pm 1\}^n$, in which case maximizing a quadratic polynomial $x^\top A x$ has many other applications, e.g., to MAX- CUT [13], to the cut norm [1], or to correlation clustering [4]. Approximation algorithms are known depending on the structure of the matrix A (see [1,6,13]), but the problem is known to be NP-hard to approximate within any factor less than 13/11 [2].

Problem (1) also permits to capture polynomial optimization over a general region of the form $\mathbb{B}^n \cap P$ where P is a polyhedron [17] and thus a broad range of combinatorial optimization problems. The general intractability of problem (1) motivates the search for tractable bounds on the minimum value in (1). In particular, several lift-and-project methods have been proposed, based on lifting the problem to higher dimension by introducing new variables modelling higher degree monomials. Such methods also apply to constrained problems on $\mathbb{B}^n$ where the constraints can be linear or polynomial; see, e.g., [3,18,27,35,39,44]. In [21] it is shown that the sums-of-squares hierarchy of Lasserre [18] in fact refines the other proposed hierarchies. As a consequence the sum-of-squares approach for polynomial optimization over $\mathbb{B}^n$ has received a great deal of attention in the recent years and there is a vast literature on this topic. Among many other results, let us just mention its use to show lower bounds on the size of semidefinite programming relaxations for combinatorial problems such as max-cut, maximum stable sets and TSP in [25], and the links to the Unique Game Conjecture in [5]. For background about the sum-of-squares hierarchy applied to polynomial optimization over general semialgebraic sets we refer to [16,19,24,32] and further references therein.

This motivates the interest in gaining a better understanding of the quality of the bounds produced by the sum-of-squares hierarchy. Our objective in this paper is to

investigate such an error analysis for this hierarchy applied to binary polynomial optimization as in (1).

1.1 The sum-of-squares hierarchy on the boolean cube

The *sum-of-squares hierarchy* was introduced by Lasserre [16,18] and Parrilo [32] as a tool to produce tractable lower bounds for polynomial optimization problems. When applied to problem (1) it provides for any integer $r \in \mathbb{N}$ a lower bound $f_{(r)} \leq f_{\min}$ on $f_{\min}$, given by:

$$f_{(r)} := \sup_{\lambda \in \mathbb{R}} \left\{ f(x) - \lambda \text{ is a sum-of-squares of degree at most } 2r \text{ on } \mathbb{B}^n \right\}. \quad (2)$$

The condition '$f(x) - \lambda$ is a sum-of-squares of degree at most $2r$ on $\mathbb{B}^n$' means that there exists a sum-of-squares polynomial $s \in \Sigma_r$ such that $f(x) - \lambda = s(x)$ for all $x \in \mathbb{B}^n$, or, equivalently, the polynomial $f - \lambda - s$ belongs to the ideal generated by the polynomials $x_1 - x_1^2, \ldots, x_n - x_n^2$. Throughout, Σ_r denotes the set of sum-of-squares polynomials with degree at most $2r$, i.e., of the form $\sum_i p_i^2$ with $p_i \in \mathbb{R}[x]_r$.

As sums of squares of polynomials can be modelled using semidefinite programming, problem (2) can be reformulated as a semidefinite program of size polynomial in n for fixed r [16,32], see also [23]. In the case of unconstrained boolean optimization, the resulting semidefinite program is known to have an optimum solution with small coefficients (see [31,33]). For fixed r, the parameter $f_{(r)}$ may therefore be computed efficiently (up to any precision).

The bounds $f_{(r)}$ have finite convergence: $f_{(r)} = f_{\min}$ for $r \geq n$ [18,21]. In fact, it has been shown in [36] that the bound $f_{(r)}$ is exact already for $2r \geq n + d - 1$. That is,

$$f_{(r)} = f_{\min} \text{ for } r \geq \frac{n + d - 1}{2}. \quad (3)$$

In addition, it is shown in [36] that the bound $f_{(r)}$ is exact for $2r \geq n + d - 2$ when the polynomial f has only monomials of even degree. This extends an earlier result of [12] shown for quadratic forms ($d = 2$), which applies in particular to the case of MAX- CUT. Furthermore, this result is tight for MAX- CUT, since one needs to go up to order $2r \geq n$ in order to reach finite convergence (in the cardinality case when all edge weights are 1) [22]. Similarly, the result (3) is tight when d is even and n is odd [15].

The main contribution of this work is an analysis of the quality of the bounds $f_{(r)}$ for parameters $r, n \in \mathbb{N}$ which fall outside of this regime, i.e., $2r < n + d - 1$. The following is our main result, which expresses the error of the bound $f_{(r)}$ in terms of the roots of Krawtchouk polynomials, which are classical univariate orthogonal polynomials with respect to a discrete measure on the set $\{0, 1, \ldots, n\}$ (see Sect. 2 for details).

Theorem 1 *Fix $d \leq n$ and let $f \in \mathbb{R}[x]$ be a polynomial of degree d. For $r, n \in \mathbb{N}$, let ξ_r^n be the least root of the degree r Krawtchouk polynomial (19) with parameter n.*

Then, if $(r + 1)/n \leq 1/2$ and $d(d + 1) \cdot \xi_{r+1}^n/n \leq 1/2$, we have:

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \leq 2C_d \cdot \xi_{r+1}^n/n. \tag{4}$$

Here $C_d > 0$ is an absolute constant depending only on d and we set $\|f\|_\infty :=$ $\max_{x \in \mathbb{B}^n} |f(x)|$.

The extremal roots of Krawtchouk polynomials are well-studied in the literature. The following result of Levenshtein [26] shows their asymptotic behaviour.

Theorem 2 ([26], Section 5) *For $t \in [0, 1/2]$, define the function*

$$\varphi(t) = 1/2 - \sqrt{t(1 - t)}. \tag{5}$$

Then the least root ξ_r^n of the degree r Krawtchouk polynomial with parameter n satisfies

$$\xi_r^n/n \leq \varphi(r/n) + c \cdot (r/n)^{-1/6} \cdot n^{-2/3} \tag{6}$$

for some universal constant $c > 0$.

Applying (6) to (4), we find that the relative error of the bound $f_{(r)}$ in the regime $r \approx t \cdot n$ behaves as the function $\varphi(t) = 1/2 - \sqrt{t(1 - t)}$, up to a term in $O(1/n^{2/3})$, which vanishes as n tends to ∞. As an illustration, Fig. 1 in Appendix A shows the function $\varphi(t)$.

1.2 A second hierarchy of bounds

In addition to the *lower* bound $f_{(r)}$, Lasserre [20] also defines an *upper* bound $f^{(r)} \geq f_{\min}$ on $f_{\min}$ as follows:

$$f^{(r)} := \inf_{s \in \Sigma_r} \left\{ \int_{\mathbb{B}^n} f(x) \cdot s(x) d\mu(x) : \int_{\mathbb{B}^n} s(x) d\mu(x) = 1 \right\}, \tag{7}$$

where μ is the uniform probability measure on $\mathbb{B}^n$. For fixed r, similarly to $f_{(r)}$, one may compute $f^{(r)}$ (up to any precision) efficiently by reformulating problem (7) as a semidefinite program [20]. Furthermore, as shown in [20] the bound is exact for some order r, and it is not difficult to see that the bound $f^{(r)}$ is exact at order $r = n$ and that this is tight (see Sect. 5).

Essentially as a side result in the proof of Theorem 1, we can show the following analog for the upper bounds $f^{(r)}$, which we believe to be of independent interest.

Theorem 3 *Fix $d \leq n$ and let $f \in \mathbb{R}[x]$ be a polynomial of degree d. Then, for any $r, n \in \mathbb{N}$ with $(r + 1)/n \leq 1/2$, we have:*

$$\frac{f^{(r)} - f_{\min}}{\|f\|_\infty} \leq C_d \cdot \xi_{r+1}^n/n,$$

where $C_d > 0$ is the constant mentioned in Theorem 1.

So we have the same estimate of the relative error for the upper bounds $f^{(r)}$ as for the lower bounds $f_{(r)}$ (up to a constant factor 2) and indeed we will see that our proof relies on an intimate connection between both hierarchies. Note that the above analysis of $f^{(r)}$ does not require any condition on the size of ξ_{r+1}^n as was necessary for the analysis of $f_{(r)}$ in Theorem 1. Indeed, as will become clear later, the condition put on ξ_{r+1}^n follows from a technical argument (see Lemma 6), which is not required in the proof of Theorem 3.

1.3 Asymptotic analysis for both hierarchies

The results above show that the relative error of both hierarchies is bounded asymptotically by the function $\varphi(t)$ from (5) in the regime $r \approx t \cdot n$. This is summarized in the following corollary, which can be seen as an asymptotic version of Theorem 1 and Theorem 3.

Corollary 1 *Fix $d \le n$ and for $n, r \in \mathbb{N}$ write*

$$E_{(r)}(n) := \sup_{f \in \mathbb{R}[x]_d} \left\{ f_{\min} - f_{(r)} : \|f\|_\infty = 1 \right\},$$

$$E^{(r)}(n) := \sup_{f \in \mathbb{R}[x]_d} \left\{ f^{(r)} - f_{\min} : \|f\|_\infty = 1 \right\}.$$

Let C_d be the constant of Theorem 1 and let $\varphi(t)$ be the function from (5). Then, for any $t \in [0, 1/2]$, we have:

$$\lim_{r/n \to t} E^{(r)}(n) \le C_d \cdot \varphi(t)$$

and, if $d(d + 1) \cdot \varphi(t) \le 1/2$, we also have:

$$\lim_{r/n \to t} E_{(r)}(n) \le 2 \cdot C_d \cdot \varphi(t).$$

Here, the limit notation $r/n \to t$ means that the claimed convergence holds for all sequences $(n_j)_j$ and $(r_j)_j$ of integers such that $\lim_{j \to \infty} n_j = \infty$ and $\lim_{j \to \infty} r_j/n_j = t$.

We close with some remarks. First, note that $\varphi(1/2) = 0$. Hence Corollary 1 tells us that the relative error of both hierarchies tends to 0 as $r/n \to 1/2$. We thus 'asymptotically' recover the exactness result (3) of [36].

Our results in Theorems 1 and 3 and Corollary 1 extend directly to the case of polynomial optimization over the discrete cube $\{\pm 1\}^n$ instead of the boolean cube $\mathbb{B}^n = \{0, 1\}^n$, as can easily be seen by applying a change of variables $x \in \{0, 1\} \mapsto 2x - 1 \in \{\pm 1\}$. In addition, as we show in Appendix A, our results extend to the case of polynomial optimization over the q-ary cube $\{0, 1, \ldots, q - 1\}^n$ for $q > 2$.

After replacing f by its negation $-f$, one may use the *lower* bound $f_{(r)}$ on $f_{\min}$ in order to obtain an *upper* bound on the *maximum* $f_{\max}$ of f over $\mathbb{B}^n$. Similarly, one may obtain a *lower* bound on $f_{\max}$ using the *upper* bound $f^{(r)}$ on $f_{\min}$. To avoid possible confusion, we will also refer to $f_{(r)}$ as the *outer* Lasserre hierarchy (or simply the sum-of-squares hierarchy), whereas we will refer to $f^{(r)}$ as the *inner* Lasserre hierarchy. This terminology (borrowed from [7]) is motivated by the following observations. As is well-known (and easy to see) the parameter $f_{\min}$ can be reformulated as an optimization problem over the set $\mathcal{M}$ of Borel measures on $\mathbb{B}^n$:

$$f_{\min} = \min \left\{ \int_{\mathbb{B}^n} f(x) d\nu(x) : \nu \in \mathcal{M}, \int_{\mathbb{B}^n} d\nu(x) = 1 \right\}.$$

If we replace the set $\mathcal{M}$ by its *inner* approximation consisting of all measures $\nu(x) = s(x) d\mu(x)$ with polynomial density $s \in \Sigma_r$ with respect to a given fixed measure μ, then we obtain the bound $f^{(r)}$. On the other hand, any $\nu \in \mathcal{M}$ corresponds to a linear functional $L_\nu : p \in \mathbb{R}[x]_{2r} \mapsto \int_{\mathbb{B}^n} p(x) d\nu(x)$ which is nonnegative on sums-of-squares on $\mathbb{B}^n$. These linear functionals thus provide an *outer* approximation for $\mathcal{M}$ and maximizing $L_\nu(p)$ over it gives the bound $f_{(r)}$ (in dual formulation).

1.4 Related work

As mentioned above, the bounds $f_{(r)}$ defined in (2) are known to be exact when $2r \geq n + d - 1$. The case $d = 2$ (which includes MAX- CUT) was treated in [12], positively answering a question posed in [22]. Extending the strategy of [12], the general case was settled in [36]. These exactness results are best possible when d is even and n is odd [15].

In [14], the sum-of-squares hierarchy is considered for approximating instances of KNAPSACK. This can be seen as a variation on the problem (1), restricting to a linear polynomial objective with positive coefficients, but introducing a single, linear constraint, of the form $a_1 x_1 + \cdots + a_n x_n \leq b$ with $a_i > 0$. There, the authors show that the outer hierarchy has relative error at most $1/(r - 1)$ for any integer $r \geq 2$. To the best of our knowledge this is the only known case where one can analyze the quality of the outer bounds for *all* orders $r \leq n$.

For optimization over sets other than the boolean cube, the following results on the quality of the outer hierarchy $f_{(r)}$ are available. When considering general semi-algebraic sets (satisfying a compactness condition), it has been shown in [30] that there exists a constant $c > 0$ (depending on the semi-algebraic set) such that $f_{(r)}$ converges to $f_{\min}$ at a rate in $O(1/\log(r/c)^{1/c})$ as r tends to ∞. This rate can be improved to $O(1/r^{1/c})$ if one considers a variation of the sum-of-squares hierarchy which is stronger (based on the preordering instead of the quadratic module), but much more computationally intensive [38]. Specializing to the hypersphere S^{n-1}, better rates in $O(1/r)$ were shown in [10,34], and recently improved to $O(1/r^2)$ in [11]. Similar improved results exist also for the case of polynomial optimization on the simplex and the continuous hypercube $[-1, 1]^n$; we refer, e.g., to [7] for an overview.

The results for semi-algebraic sets other than $\mathbb{B}^n$ mentioned above all apply in the asymptotic regime where the dimension n is fixed and $r \to \infty$. This makes it difficult

to compare them directly to our new results. Indeed, we have to consider a different regime in the case of the boolean cube $\mathbb{B}^n$, as the hierarchy always converges in at most n steps. The regime where we are able to provide an analysis in this paper is when $r \approx t \cdot n$ with $0 < t \leq 1/2$.

Turning now to the *inner* hierarchy (7), as far as we are aware, nothing is known about the behaviour of the bounds $f^{(r)}$ on $\mathbb{B}^n$. For full-dimensional compact sets, however, results are available. It has been shown that, on the hypersphere [8], the unit ball and the simplex [40], and the unit box [9], the bound $f^{(r)}$ converges at a rate in $O(1/r^2)$. A slightly weaker convergence rate in $O(\log^2 r/r^2)$ is known for general (full-dimensional) semi-algebraic sets [40,42]. Again, these results are all asymptotic in r, and thus hard to compare directly to our analysis on $\mathbb{B}^n$.

1.5 Overview of the proof

Here, we give an overview of the main ideas that we use to show our results. Our broad strategy follows the one employed in [11] to obtain information on the sum-of-squares hierarchy on the hypersphere. The following four ingredients will play a key role in our proof:

1. we use the *polynomial kernel technique* in order to produce low-degree sum-of-squares representations of polynomials that are positive over $\mathbb{B}^n$, thus allowing an analysis of $f_{\min} - f_{(r)}$;
2. using classical *Fourier analysis* on the boolean cube $\mathbb{B}^n$ we are able to exploit symmetry and reduce the search for a *multivariate kernel* to a *univariate sum-of-squares polynomial* on the discrete set $[0 : n] := \{0, 1, \ldots, n\}$;
3. we find this univariate sum-of-squares by applying the *inner* Lasserre hierarchy to an appropriate univariate optimization problem on $[0 : n]$;
4. finally, we exploit a known connection between the inner hierarchy and the *extremal roots of corresponding orthogonal polynomials* (in our case, the Krawtchouk polynomials).

Following these steps we are able to analyze the sum-of-squares hierarchy $f_{(r)}$ as well as the inner hierarchy $f^{(r)}$. We now sketch how our proof articulates along these four main steps.

Let $f \in \mathbb{R}[x]_d$ be the polynomial with degree d for which we wish to analyze the bounds $f_{(r)}$ and $f^{(r)}$. After rescaling, and up to a change of coordinates, we may assume w.l.o.g. that f attains its minimum over $\mathbb{B}^n$ at $0 \in \mathbb{B}^n$ and that $f_{\min} = 0$ and $f_{\max} = 1$. So we have $\|f\|_\infty = 1$. To simplify notation, we will make these assumptions throughout.

The first key idea is to consider a *polynomial kernel K* on $\mathbb{B}^n$ of the form:

$$K(x, y) = u^2(d(x, y)), \tag{8}$$

where $u \in \mathbb{R}[t]_r$ is a univariate polynomial of degree at most r and $d(x, y)$ is the Hamming distance between x and y. Such a kernel K induces an operator $\mathbf{K}$, which acts linearly on the space of polynomials on $\mathbb{B}^n$ by:

$$p \in \mathbb{R}[x] \mapsto \mathbf{K}p(x) := \int_{\mathbb{B}^n} p(y)K(x, y)d\mu(y) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} p(y)K(x, y).$$

Recall that μ is the uniform probability distribution on $\mathbb{B}^n$. An easy but important observation is that, if p is nonnegative on $\mathbb{B}^n$, then $\mathbf{K}p$ is a sum-of-squares (on $\mathbb{B}^n$) of degree at most $2r$. We use this fact as follows.

Given a scalar $\delta \geq 0$, define the polynomial $\tilde{f} := f + \delta$. Assuming that the operator $\mathbf{K}$ is non-singular, we can express $\tilde{f}$ as $\tilde{f} = \mathbf{K}(\mathbf{K}^{-1}\tilde{f})$. Therefore, if $\mathbf{K}^{-1}\tilde{f}$ is nonnegative on $\mathbb{B}^n$, then we find that $\tilde{f}$ is a sum-of-squares on $\mathbb{B}^n$ with degree at most $2r$, which implies $f_{\min} - f_{(r)} \leq \delta$.

One way to guarantee that $\mathbf{K}^{-1}\tilde{f}$ is indeed nonnegative on $\mathbb{B}^n$ is to select the operator $\mathbf{K}$ in such a way that $\mathbf{K}(1) = 1$ and

$$\|\mathbf{K}^{-1} - I\| := \sup_{p \in \mathbb{R}[x]_d} \frac{\|\mathbf{K}^{-1}p - p\|_\infty}{\|p\|_\infty} \leq \delta. \tag{9}$$

We collect this as a lemma for further reference.

Lemma 1 *If the kernel operator $\mathbf{K}$ associated to $u \in \mathbb{R}[t]_r$ via relation (8) satisfies $\mathbf{K}(1) = 1$ and $\|\mathbf{K}^{-1} - I\| \leq \delta$, then we have $f_{\min} - f_{(r)} \leq \delta$.*

Proof With $\tilde{f} = f + \delta$, we have: $\|\mathbf{K}^{-1}\tilde{f} - \tilde{f}\|_\infty = \|\mathbf{K}^{-1}f - f\|_\infty \leq \delta\|f\|_\infty = \delta$, where we use the fact that $\mathbf{K}(1) = 1 = \mathbf{K}^{-1}(1)$ and thus $\mathbf{K}^{-1}\tilde{f} - \tilde{f} = \mathbf{K}^{-1}f - f$ for the first equality and (9) for the inequality. The inequality $\|\mathbf{K}^{-1}\tilde{f} - \tilde{f}\|_\infty \leq \delta$ then implies $\mathbf{K}^{-1}\tilde{f}(x) \geq \tilde{f}(x) - \delta = f(x) \geq f_{\min} = 0$ on $\mathbb{B}^n$. □

In other words, we want the operator $\mathbf{K}^{-1}$ (and thus $\mathbf{K}$) to be 'close to the identity operator' in a certain sense. As kernels of the form (8) are invariant under the symmetries of $\mathbb{B}^n$, we are able to use classical Fourier analysis on the boolean cube to express the eigenvalues λ_i of $\mathbf{K}$ in terms of the polynomial u. More precisely, it turns out that these eigenvalues are given by the coefficients of the expansion of u^2 in the basis of *Krawtchouk polynomials*. As we show later (see (33)), inequality (9) then holds for $\delta \approx \sum_i |\lambda_i^{-1} - 1|$.

It therefore remains to find a univariate polynomial $u \in \mathbb{R}[t]_r$ for which these coefficients, and thus the eigenvalues of $\mathbf{K}$, are sufficiently close to 1. Interestingly, this problem boils down to analyzing the quality of the inner bound $g^{(r)}$ (see 7) for a particular univariate polynomial g (given in (35)).

In order to perform this analysis and conclude the proof of Theorem 1, we make use of a connection between the inner Lasserre hierarchy and the least roots of orthogonal polynomials (in this case the Krawtchouk polynomials).

Finally, we generalize our analysis of the inner hierarchy (for the special case of the selected polynomial g in Step 3 to arbitrary polynomials) to obtain Theorem 3.

1.6 Matrix-valued polynomials

As we explain in this section the results of Theorems 1 and 3 may be extended to the setting of matrix-valued polynomials.

For fixed $k \in \mathbb{N}$, let $\mathcal{S}^k \subseteq \mathbb{R}^{k \times k}$ be the space of $k \times k$ real symmetric matrices. We write $\mathcal{S}^k[x] \subseteq \mathbb{R}^{k \times k}[x]$ for the space of n-variate polynomials whose coefficients lie in $\mathcal{S}^k$, i.e., the set of $k \times k$ symmetric polynomial matrices. The polynomial optimization problem (1) may be generalized to polynomials $F \in \mathcal{S}^k[x]$ as:

$$F_{\min} := \min_{x \in \mathbb{B}^n} \lambda_{\min}(F(x)), \tag{10}$$

where $F \in \mathcal{S}^k[x]$ is a polynomial matrix and $\lambda_{\min}(F(x))$ denotes the smallest eigenvalue of $F(x)$. That is, $F_{\min}$ is the largest value for which $F(x) - F_{\min} I \succeq 0$ for all $x \in \mathbb{B}^n$, where $\succeq$ is the positive semidefinite Löwner order.

The Lasserre hierarchies (2) and (7) may also be defined in this setting. A polynomial matrix $S \in \mathcal{S}^k[x]$ is called a sum-of-squares polynomial of degree $2r$ if it can be written as:

$$S(x) = \sum_i U_i(x) U_i(x)^\top \quad (U_i \in \mathbb{R}^{k \times m}[x], \quad \deg U_i \leq r, \quad m \in \mathbb{N}).$$

We write $\Sigma_r^{k \times k}$ for the set of all such polynomials. See, e.g., [37] for background and results on sum-of-squares polynomial matrices. Note that a sum-of-squares polynomial matrix $S \in \Sigma_r^{k \times k}$ satisfies $S(x) \succeq 0$ for all $x \in \mathbb{B}^n$. For a matrix $A \in \mathcal{S}^k$, $\|A\|$ denotes its spectral (or operator) norm, defined as the largest absolute value of an eigenvalue of A. Then, for $F \in \mathcal{S}^k[x]$, we set

$$\|F\|_\infty = \max_{x \in \mathbb{B}^n} \|F(x)\|.$$

The outer hierarchy $F_{(r)}$ for the polynomial F is then given by:

$$F_{(r)} := \sup_{\lambda \in \mathbb{R}} \left\{ F(x) - \lambda \cdot I = S(x) \text{ on } \mathbb{B}^n \text{ for some } S \in \Sigma_r^{k \times k} \right\}. \tag{11}$$

Similarly, the inner hierarchy $F^{(r)}$ is defined as:

$$F^{(r)} := \inf_{S \in \Sigma_r^{k \times k}} \left\{ \int_{\mathbb{B}^n} \mathrm{Tr}\big(F(x) S(x)\big) d\mu(x) : \int_{\mathbb{B}^n} \mathrm{Tr}\big(S(x)\big) d\mu(x) = 1 \right\}. \tag{12}$$

Indeed, (2) and (7) are the special cases of the above programs when $k = 1$. As before, the parameters $F_{(r)}$ and $F^{(r)}$ may be computed efficiently for fixed r (and fixed k) using semidefinite programming, and they satisfy

$$F_{(r)} \leq F_{(r+1)} \leq F_{\min} \leq F^{(r+1)} \leq F^{(r)}.$$

As was already noted in [11] (in the context of optimization over the unit sphere), the proof strategy outlined in Sect. 1.5 naturally extends to the setting of polynomial matrices. This yields the following generalizations of our main theorems.

Theorem 4 *Fix $d \leq n$ and let $F \in \mathcal{S}^k[x]$ ($k \geq 1$) be a polynomial matrix of degree d. For $r, n \in \mathbb{N}$, let ξ_r^n be the least root of the degree r Krawtchouk polynomial* (19) *with parameter n. Then, if $(r + 1)/n \leq 1/2$ and $d(d + 1) \cdot \xi_{r+1}^n/n \leq 1/2$, we have:*

$$\frac{F_{\min} - F_{(r)}}{\|F\|_\infty} \leq 2C_d \cdot \xi_{r+1}^n/n, \tag{13}$$

where $C_d > 0$ is the absolute constant depending only on d from Theorem 1.

Theorem 5 *Fix $d \leq n$ and let $F \in \mathcal{S}^k[x]$ be a matrix-valued polynomial of degree d. Then, for any $r, n \in \mathbb{N}$ with $(r + 1)/n \leq 1/2$, we have:*

$$\frac{F^{(r)} - F_{\min}}{\|F\|_\infty} \leq C_d \cdot \xi_{r+1}^n/n,$$

where $C_d > 0$ is the constant of Theorem 1.

Organization

The rest of the paper is structured as follows. We review the necessary background on Fourier analysis on the boolean cube in Sect. 2. In Sect. 3, we recall a connection between the inner Lasserre hierarchy and the roots of orthogonal polynomials. Then, in Sect. 4, we give a proof of Theorem 1. In Sect. 5, we discuss how to generalize the proofs of Sect. 4 to obtain Theorem 3. We group the proofs of some technical results needed to prove Lemma 5 in Sect. 6. In Appendix A, we indicate how our arguments extend to the case of polynomial optimization over the q-ary hypercube $\{0, 1, \ldots, q - 1\}^n$ for $q > 2$. Finally, we give the proofs of our results in the matrix-valued setting (Theorems 4 and 5) in Appendix B.

2 Fourier analysis on the Boolean cube

In this section, we cover some standard Fourier analysis on the boolean cube. For a general reference on Fourier analysis on (finite, abelian) groups, see e.g. [43]. See also Section 4 of [45] for the boolean case.

Some notation For $n \in \mathbb{N}$, we write $\mathbb{B}^n = \{0, 1\}^n$ for the boolean hypercube of dimension n. We let μ denote the uniform probability measure on $\mathbb{B}^n$, given by $\mu = \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} \delta_x$, where δ_x is the Dirac measure at x. Further, we write $|x| = \sum_i x_i = |\{i \in [n] : x_i = 1\}|$ for the *Hamming weight* of $x \in \mathbb{B}^n$, and $d(x, y) = |\{i \in [n] : x_i \neq y_i\}|$ for the *Hamming distance* between $x, y \in \mathbb{B}^n$. We let $\mathrm{Sym}(n)$ denote the set of permutations of the set $[n] = \{1, \ldots, n\}$.

We consider polynomials $p : \mathbb{B}^n \to \mathbb{R}$ on $\mathbb{B}^n$. The space $\mathcal{R}[x]$ of such polynomials is given by the quotient ring of $\mathbb{R}[x]$ over the equivalence relation $p \sim q$ if $p(x) = q(x)$ on $\mathbb{B}^n$. In other words, $\mathcal{R}[x] = \mathbb{R}[x]/\mathcal{I}$, where $\mathcal{I}$ is the ideal generated by the polynomials $x_i - x_i^2$ for $i \in [n]$, which can also be seen as the vector space spanned by the (multilinear) polynomials $\prod_{i \in I} x_i$ for $I \subseteq [n]$.

For $a \leq b \in \mathbb{N}$, we let $[a : b]$ denote the set of integers $a, a + 1, \ldots, b$.

The character basis Let $\langle \cdot, \cdot \rangle_\mu$ be the inner product on $\mathcal{R}[x]$ given by:

$$\langle p, q \rangle_\mu = \int_{\mathbb{B}^n} p(x)q(x)d\mu(x) = \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} p(x)q(x).$$

The space $\mathcal{R}[x]$ has an orthonormal basis w.r.t. $\langle \cdot, \cdot \rangle_\mu$ given by the *characters*:

$$\chi_a(x) := (-1)^{x \cdot a} = \prod_{i:a_i=1} (1 - 2x_i) \quad (a \in \mathbb{B}^n). \tag{14}$$

In other words, the set $\{\chi_a : a \in \mathbb{B}^n\}$ of all characters on $\mathbb{B}^n$ forms a basis for $\mathcal{R}[x]$ and

$$\langle \chi_a, \chi_b \rangle_\mu = \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} \chi_a(x)\chi_b(x) = \delta_{a,b} \quad \forall a, b \in \mathbb{B}^n. \tag{15}$$

Then any polynomial $p \in \mathcal{R}[x]$ can be expressed in the basis of characters, known as its *Fourier expansion*:

$$p(x) = \sum_{a \in \mathbb{B}^n} \widehat{p}(a)\chi_a(x) \quad \forall x \in \mathbb{B}^n \tag{16}$$

with *Fourier coefficients* $\widehat{p}(a) := \langle p, \chi_a \rangle_\mu \in \mathbb{R}$.

The group $\mathrm{Aut}(\mathbb{B}^n)$ of automorphisms of $\mathbb{B}^n$ is generated by the coordinate permutations, of the form $x \mapsto \sigma(x) := (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for $\sigma \in \mathrm{Sym}(n)$, and the automorphisms corresponding to bit-flips, of the form $x \in \mathbb{B}^n \mapsto x \oplus a \in \mathbb{B}^n$ for $a \in \mathbb{B}^n$. If we set

$$H_k := \mathrm{span}\{\chi_a : |a| = k\} \quad (0 \leq k \leq n),$$

then each H_k is an irreducible, $\mathrm{Aut}(\mathbb{B}^n)$-invariant subspace of $\mathcal{R}[x]$ of dimension $\binom{n}{k}$. Using (16), we may then decompose $\mathcal{R}[x]$ as the direct sum

$$\mathcal{R}[x] = H_0 \perp H_1 \perp \cdots \perp H_n,$$

where the subspaces H_k are pairwise orthogonal w.r.t. $\langle \cdot, \cdot \rangle_\mu$. In fact, we have that $\mathcal{R}[x]_d = H_0 \perp H_1 \perp \cdots \perp H_d$ for all $d \leq n$, and we may thus write any $p \in \mathcal{R}[x]_d$ (in a unique way) as

$$p = p_0 + p_1 + \cdots + p_d \quad (p_k \in H_k). \tag{17}$$

The polynomials $p_k \in H_k$ $(k = 0, \ldots, d)$ are known as the *harmonic components* of p and the decomposition (17) as the *harmonic decomposition* of p. We will make extensive use of this decomposition throughout.

Let $\mathrm{St}(0) \subseteq \mathrm{Aut}(\mathbb{B}^n)$ be the set of automorphisms fixing $0 \in \mathbb{B}^n$, which consists of the coordinate permutations $x \mapsto \sigma(x) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for $\sigma \in \mathrm{Sym}(n)$. The subspace of functions in H_k that are invariant under $\mathrm{St}(0)$ is one-dimensional and it is spanned by the function

$$X_k(x) := \sum_{|a|=k} \chi_a(x). \tag{18}$$

These functions X_k are known as the *zonal spherical functions* with pole $0 \in \mathbb{B}^n$.

Krawtchouk polynomials For $k \in \mathbb{N}$, the *Krawtchouk polynomial* of degree k (and with parameter n) is the univariate polynomial in t given by:

$$\mathcal{K}_k^n(t) := \sum_{i=0}^{k}(-1)^i \binom{t}{i}\binom{n-t}{k-i} \tag{19}$$

(see, e.g. [41]). The Krawtchouk polynomials form an orthogonal basis for $\mathbb{R}[t]$ with respect to the inner product given by the following discrete probability measure on the set $[0:n] = \{0, 1, \ldots, n\}$:

$$\omega := \frac{1}{2^n}\sum_{t=0}^{n} w(t)\delta_t, \text{ with } w(t) := \binom{n}{t}.$$

Indeed, for all $k, k' \in \mathbb{N}$ we have:

$$\langle \mathcal{K}_k^n, \mathcal{K}_{k'}^n \rangle_\omega := \frac{1}{2^n}\sum_{t=0}^{n} \mathcal{K}_k^n(t)\mathcal{K}_{k'}^n(t)w(t) = \delta_{k,k'}\binom{n}{k}. \tag{20}$$

The following (well-known) lemma explains the connection between the Krawtchouk polynomials and the character basis on $\mathcal{R}[x]$.

Lemma 2 *Let $t \in [0:n]$ and choose $x, y \in \mathbb{B}^n$ so that $d(x, y) = t$. Then for any $0 \le k \le n$ we have:*

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(x)\chi_a(y). \tag{21}$$

In particular, we have:

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(1^t 0^{n-t}) = X_k(1^t 0^{n-t}), \tag{22}$$

where $1^t 0^{n-t} \in \mathbb{B}^n$ is given by $(1^t 0^{n-t})_i = 1$ if $1 \le i \le t$ and $(1^t 0^{n-t})_i = 0$ if $t + 1 \le i \le n$.

Proof Noting that $\chi_a(x)\chi_a(y) = \chi_a(x+y)$ and $|x+y| = d(x,y) = t$, we have:

$$\sum_{|a|=k} \chi_a(x)\chi_a(y) = \sum_{i=0}^{k} (-1)^i \cdot \#\{|a| = k : a \cdot (x+y) = i\}$$

$$= \sum_{i=0}^{k} (-1)^i \binom{t}{i}\binom{n-t}{k-i} = \mathcal{K}_k^n(t).$$

$\square$

From this, we see that any polynomial $p \in \mathbb{R}[x]_d$ that is invariant under the action of St(0) is of the form $\sum_{i=1}^{d} \lambda_i \mathcal{K}_i^n(|x|)$ for some scalars λ_i, and thus $p(x) = u(|x|)$ for some univariate polynomial $u \in \mathbb{R}[t]_d$.

It will sometimes be convenient to work with a different normalization of the Krawtchouk polynomials, given by:

$$\widehat{\mathcal{K}}_k^n(t) := \mathcal{K}_k^n(t)/\mathcal{K}_k^n(0) \quad (k \in \mathbb{N}). \tag{23}$$

So $\widehat{\mathcal{K}}_k^n(0) = 1$. Note that for any $k \in \mathbb{N}$, we have

$$\|\mathcal{K}_k^n\|_\omega^2 := \langle \mathcal{K}_k^n, \mathcal{K}_k^n \rangle_\omega = \binom{n}{k} = \mathcal{K}_k^n(0),$$

meaning that $\widehat{\mathcal{K}}_k^n(t) = \mathcal{K}_k^n(t)/\|\mathcal{K}_k^n\|_\omega^2$.

Finally we give a short proof of two basic facts on Krawtchouk polynomials that will be used below.

Lemma 3 *We have:*

$$\widehat{\mathcal{K}}_k^n(t) \le \widehat{\mathcal{K}}_0^n(t) = 1$$

for all $0 \le k \le n$ and $t \in [0 : n]$.

Proof Given $t \in [0 : n]$ consider an element $x \in \mathbb{B}^n$ with Hamming weight t, for instance the element $1^t 0^{n-t}$ from Lemma 2. By (22) we have

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(x) \le \binom{n}{k} = \mathcal{K}_k^n(0),$$

making use of the fact that $|\chi_a(x)| = 1$ for all $a \in \mathbb{B}^n$. $\square$

Lemma 4 *We have:*

$$|\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t+1)| \le \frac{2k}{n}, \qquad (t = 0, 1, \ldots, n-1)$$

$$|\widehat{\mathcal{K}}_k^n(t) - 1| \le \frac{2k}{n} \cdot t \qquad (t = 0, 1, \ldots, n) \tag{24}$$

for all $0 \leq k \leq n$.

Proof Let $t \in [0 : n-1]$ and $0 < k \leq d$. Consider the elements $1^t 0^{n-t}$ and $1^{t+1} 0^{n-t-1}$ of $\mathbb{B}^n$ from Lemma 2. We have:

$$|\mathcal{K}_k^n(t) - \mathcal{K}_k^n(t+1)| \overset{(22)}{=} |\sum_{|a|=k} \chi_a(1^t 0^{n-t}) - \chi_a(1^{t+1} 0^{n-t-1})|$$

$$\leq 2 \cdot \#\{a \in \mathbb{B}^n : |a| = k, a_{t+1} = 1\} = 2\binom{n-1}{k-1},$$

where for the inequality we note that $\chi_a(1^t 0^{n-t}) = \chi_a(1^{t+1} 0^{n-t-1})$ if $a_{t+1} = 0$. As $\mathcal{K}_k^n(0) = \binom{n}{k}$, this implies that:

$$|\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t+1)| \leq 2\binom{n-1}{k-1} / \binom{n}{k} = \frac{2k}{n}.$$

This shows the first inequality of (24). The second inequality follows using the triangle inequality, a telescope summation argument and the fact that $\widehat{\mathcal{K}}_k^n(0) = 1$. □

Invariant kernels and the Funk–Hecke formula Given a univariate polynomial $u \in \mathbb{R}[t]$ of degree r with $2r \geq d$, consider the kernel $K : \mathbb{B}^n \times \mathbb{B}^n \to \mathbb{R}$ defined by

$$K(x, y) := u^2(d(x, y)). \tag{25}$$

A kernel of the form (25) coincides with a polynomial of degree $2\deg(u)$ in x on the binary cube $\mathbb{B}^n$, as $d(x, y) = \sum_i (x_i + y_i - 2x_i y_i)$ for $x, y \in \mathbb{B}^n$. Furthermore, it is invariant under $\text{Aut}(\mathbb{B}^n)$, in the sense that:

$$K(x, y) = K(\pi(x), \pi(y)) \quad \forall x, y \in \mathbb{B}^n, \pi \in \text{Aut}(\mathbb{B}^n).$$

The kernel K acts as a linear operator $\mathbf{K} : \mathcal{R}[x] \to \mathcal{R}[x]$ by:

$$\mathbf{K}p(x) := \int_{\mathbb{B}^n} p(y) K(x, y) d\mu(y) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} p(y) K(x, y). \tag{26}$$

We may expand the univariate polynomial $u^2 \in \mathbb{R}[t]_{2r}$ in the basis of Krawtchouk polynomials as:

$$u^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t) \quad (\lambda_i \in \mathbb{R}). \tag{27}$$

As we show now, the eigenvalues of the operator $\mathbf{K}$ are given precisely by the coefficients λ_i occurring in this expansion. This relation is analogous to the classical Funk–Hecke formula for spherical harmonics (see, e.g., [11]).

Theorem 6 (Funk–Hecke) *Let $p \in \mathcal{R}[x]_d$ with harmonic decomposition $p = p_0 + p_1 + \cdots + p_d$. Then we have:*

$$\mathbf{K}p = \lambda_0 p_0 + \lambda_1 p_1 + \cdots + \lambda_d p_d. \tag{28}$$

Proof It suffices to show that $\mathbf{K}\chi_z = \lambda_{|z|}\chi_z$ for all $z \in \mathbb{B}^n$. So we compute for $x \in \mathbb{B}^n$:

$$\mathbf{K}\chi_z(x) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} \chi_z(y) u^2(d(x,y)) \overset{(27)}{=} \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} \chi_z(y) \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(d(x,y))$$

$$\overset{(21)}{=} \sum_{i=0}^{2r} \lambda_i \sum_{y \in \mathbb{B}^n} \chi_z(y) \sum_{|a|=i} \chi_a(x)\chi_a(y)$$

$$= \sum_{i=0}^{2r} \lambda_i \sum_{|a|=i} \left(\sum_{y \in \mathbb{B}^n} \chi_z(y)\chi_a(y) \right) \chi_a(x)$$

$$\overset{(15)}{=} \frac{1}{2^n} \sum_{i=0}^{2r} \lambda_i \sum_{|a|=i} 2^n \delta_{z,a} \chi_a(x)$$

$$= \lambda_{|z|}\chi_z(x).$$

$\square$

Finally, we note that since the Krawtchouk polynomials form an *orthogonal* basis for $\mathbb{R}[t]$, we may express the coefficients λ_i in the decomposition (27) of u^2 in the following way:

$$\lambda_i = \langle \mathcal{K}_i^n, u^2 \rangle_\omega / \|\mathcal{K}_i^n\|_\omega^2 = \langle \widehat{\mathcal{K}}_i^n, u^2 \rangle_\omega. \tag{29}$$

In addition, since in view of Lemma 3 we have $\widehat{\mathcal{K}}_i^n(t) \leq \widehat{\mathcal{K}}_0^n(t)$ for all $t \in [0:n]$, it folllows that

$$\lambda_i \leq \lambda_0 \quad \text{for } 0 \leq i \leq 2r. \tag{30}$$

3 The inner Lasserre hierachy and orthogonal polynomials

The inner Lasserre hierachy, which we have defined for the boolean cube in (7), may be defined more generally for the minimization of a polynomial g over a compact set $M \subseteq \mathbb{R}^n$ equipped with a measure ν with support M, by setting:

$$g^{(r)} := g_{M,\nu}^{(r)} = \inf_{s \in \Sigma_r} \left\{ \int_M g \cdot s \, d\nu : \int_M s \, d\nu = 1 \right\} \tag{31}$$

for any integer $r \in \mathbb{N}$. So we have: $g^{(r)} \geq g_{\min} := \min_{x \in M} g(x)$. A crucial ingredient of the proof of our main theorem below is an analysis of the error $g^{(r)} - g_{\min}$ in this hierarchy for a special choice of $M \subseteq \mathbb{R}$, g and ν.

Here, we recall a technique which may be used to perform such an analysis in the univariate case, which was developed in [9] and further employed for this purpose, e.g., in [8,40].

First, we observe that we may always replace g by a suitable *upper estimator* $\widehat{g}$ which satisfies $\widehat{g}_{\min} = g_{\min}$ and $\widehat{g}(x) \geq g(x)$ for all $x \in M$. Indeed, it is clear that for such $\widehat{g}$ we have:

$$g^{(r)} - g_{\min} \leq \widehat{g}^{(r)} - g_{\min} = \widehat{g}^{(r)} - \widehat{g}_{\min}.$$

Next, we consider the special case when $M \subseteq \mathbb{R}$ and $g(t) = t$. Here, the bound $g^{(r)}$ may be expressed in terms of the orthogonal polynomials on M w.r.t. the measure ν, i.e., the polynomials $p_i \in \mathbb{R}[t]_i$ determined by the relation:

$$\int_M p_i p_j d\nu = 0 \text{ if } i \neq j.$$

Theorem 7 ([9]) *Let $M \subseteq \mathbb{R}$ be an interval and let ν be a measure supported on M with corresponding orthogonal polynomials $p_i \in \mathbb{R}[t]_i$ ($i \in \mathbb{N}$). Then the Lasserre inner bound $g^{(r)}$ (from 31) of order r for the polynomial $g(t) = t$ equals*

$$g^{(r)} = \xi_{r+1},$$

where ξ_{r+1} is the smallest root of the polynomial p_{r+1}.

Remark 1 The upshot of the above result is that, for any polynomial $g : \mathbb{R} \to \mathbb{R}$ which is upper bounded on an interval $M \subseteq \mathbb{R}$ by a linear polynomial $\widehat{g}(t) = ct$ for some $c > 0$, we have:

$$g^{(r)} - g_{\min} \leq c \cdot \xi_{r+1}, \tag{32}$$

where ξ_{r+1} is the smallest root of the corresponding orthogonal polynomial of degree $r + 1$.

4 Proof of Theorem 1

Throughout, $d \leq n$ is a fixed integer (the degree of the polynomial f to be minimized over $\mathbb{B}^n$). Recall $u \in \mathbb{R}[t]$ is a univariate polynomial with degree r (that we need to select) with $2r \geq d$. Consider the associated kernel K defined in (25) and the corresponding linear operator $\mathbf{K}$ from (26). Recall from (9) that we are interested in bounding the quantity:

$$\|\mathbf{K}^{-1} - I\| := \sup_{p \in \mathbb{R}[x]_d} \frac{\|\mathbf{K}^{-1}p - p\|_\infty}{\|p\|_\infty}.$$

Our proof consists of two parts. First, we relate the coefficients λ_i, that appear in the decomposition (27) of $u^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t)$ into the basis of Krawtchouk polynomials, to the quantity $\|\mathbf{K}^{-1} - I\|$.

Then, using this relation and the connection between the *inner* Lasserre hierarchy and extremal roots of orthogonal polynomials outlined in Sect. 3, we show that u may be chosen such that $\|\mathbf{K}^{-1} - I\|$ is of the order ξ_{r+1}^n/n, with ξ_{r+1}^n the smallest root of the degree $r + 1$ Krawtchouk polynomial (with parameter n).

Bounding $\|\mathbf{K}^{-1} - I\|$ in terms of the λ_i We need the following technical lemma, which bounds the sup-norm $\|p_k\|_\infty$ of the harmonic components p_k of a polynomial $p \in \mathcal{R}[x]$ in terms of $\|p\|_\infty$, the sup-norm of p itself. The key point is that this bound is independent of the dimension n. We delay the proof which is rather technical to Sect. 6.

Lemma 5 *There exists a constant $\gamma_d > 0$, depending only on d, such that for any $p = p_0 + p_1 + \ldots + p_d \in \mathcal{R}[x]_d$, we have:*

$$\|p_k\|_\infty \leq \gamma_d \|p\|_\infty \text{ for all } 0 \leq k \leq d.$$

Corollary 2 *Assume that $\lambda_0 = 1$ and $\lambda_k \neq 0$ for $1 \leq k \leq d$. Then we have:*

$$\|\mathbf{K}^{-1} - I\| \leq \gamma_d \cdot \Lambda, \text{ where } \Lambda := \sum_{i=1}^{d} |\lambda_i^{-1} - 1|. \tag{33}$$

Proof By assumption, the operator $\mathbf{K}$ is invertible and, in view of Funk–Hecke relation (28), its inverse is given by $\mathbf{K}^{-1}p = \sum_{i=0}^{d} \lambda_i^{-1} p_k$ for any $p = p_0 + p_1 + \ldots + p_d \in \mathcal{R}[x]_d$. Then we have:

$$\begin{aligned}
\|\mathbf{K}^{-1}p - p\|_\infty &= \|\sum_{i=1}^{d}(\lambda_i^{-1} - 1)p_i\|_\infty \\
&\leq \sum_{i=1}^{d} |\lambda_i^{-1} - 1| \|p_i\|_\infty \\
&\leq \sum_{i=1}^{d} |\lambda_i^{-1} - 1| \cdot \gamma_d \|p\|_\infty,
\end{aligned} \tag{34}$$

where we use Lemma 5 for the last inequality. $\qquad\square$

The expression Λ in (33) is difficult to analyze. Therefore, following [11], we consider instead the simpler expression:

$$\tilde{\Lambda} := \sum_{i=1}^{d}(1 - \lambda_i) = d - \sum_{i=1}^{d} \lambda_i,$$

which is linear in the λ_i. Under the assumption that $\lambda_0 = 1$, we have $\lambda_i \leq \lambda_0 = 1$ for all i (recall relation (30)). Thus, Λ and $\tilde{\Lambda}$ are both minimized when the λ_i are close to 1. The following lemma makes this precise.

Lemma 6 *Assume that $\lambda_0 = 1$ and that $\tilde{\Lambda} \leq 1/2$. Then we have $\Lambda \leq 2\tilde{\Lambda}$, and thus*

$$\|\mathbf{K}^{-1} - I\| \leq 2\gamma_d \cdot \tilde{\Lambda}.$$

Proof As we assume $\tilde{\Lambda} \leq 1/2$, we must have $1/2 \leq \lambda_i \leq 1$ for all i. Therefore, we may write:

$$\Lambda = \sum_{i=1}^d |\lambda_i^{-1} - 1| = \sum_{i=1}^d |(1 - \lambda_i)/\lambda_i| = \sum_{i=1}^d (1 - \lambda_i)/\lambda_i \leq 2 \sum_{i=1}^d (1 - \lambda_i) = 2\tilde{\Lambda}.$$

$\square$

Optimizing the choice of the univariate polynomial u In light of Lemma 6, and recalling (29), we wish to find a univariate polynomial $u \in \mathbb{R}[t]_r$ for which

$$\lambda_0 = \langle 1, u^2 \rangle_\omega = 1, \text{ and}$$

$$\tilde{\Lambda} = d - \sum_{i=1}^d \lambda_i = d - \sum_{i=1}^d \langle \widehat{\mathcal{K}}_i^n, u^2 \rangle_\omega \text{ is small.}$$

Unpacking the definition of $\langle \cdot, \cdot \rangle_\omega$, we thus need to solve the following optimization problem:

$$\inf_{u \in \mathbb{R}[t]_r} \left\{ \int g \cdot u^2 d\omega : \int u^2 d\omega = 1 \right\}, \text{ where } g(t) := d - \sum_{i=1}^d \widehat{\mathcal{K}}_i^n(t). \quad (35)$$

(Indeed $\int g u^2 d\omega = \langle g, u^2 \rangle_\omega = \tilde{\Lambda}$ and $\int u^2 d\omega = \langle 1, u^2 \rangle_\omega$.) We recognize this program to be the same as the program (31) defining the inner Lasserre bound[1] of order r for the minimum $g_{\min} = g(0) = 0$ of the polynomial g over $[0 : n]$, computed with respect to the measure $d\omega(t) = 2^{-n} \binom{n}{t}$. Hence the optimal value of (35) is equal to $g^{(r)}$ and, using Lemma 6, we may conclude the following.

Theorem 8 *Let g be as in (35). Assume that $g^{(r)} - g_{\min} \leq 1/2$. Then there exists a polynomial $u \in \mathbb{R}[t]_r$ such that $\lambda_0 = 1$ and*

$$\|\mathbf{K}^{-1} - I\| \leq 2\gamma_d \cdot (g^{(r)} - g_{\min}).$$

Here, $g^{(r)}$ is the inner Lasserre bound on $g_{\min}$ of order r, computed on $[0, n]$ w.r.t. ω, via the program (35), and γ_d is the constant of Lemma 5.

---

[1] Technically, the program (31) allows for the density to be a *sum of squares*, whereas the program (35) requires the density to be an actual square. This is no true restriction, though, since, as a straightforward convexity argument shows, the optimum solution to (31) can in fact always be chosen to be a square [20].

It remains, then, to analyze the range $g^{(r)} - g_{\min}$. For this purpose, we follow the technique outlined in Sect. 3. We first show that g can be upper bounded by its linear approximation at $t = 0$.

Lemma 7 *We have:*

$$g(t) \leq \widehat{g}(t) := d(d+1) \cdot (t/n) \quad \forall t \in [0:n].$$

Furthermore, the minimum $\widehat{g}_{\min}$ of $\widehat{g}$ on $[0:n]$ clearly satisfies $\widehat{g}_{\min} = \widehat{g}(0) = g(0) = g_{\min}$.

Proof Using (24), we find for each $k \leq n$ that:

$$\widehat{\mathcal{K}}_k^n(t) \geq \widehat{\mathcal{K}}_k^n(0) - \frac{2k}{n} \cdot t = 1 - \frac{2k}{n} \cdot t \quad \forall t \in [0:n].$$

Therefore, we have:

$$g(t) := d - \sum_{k=1}^{d} \widehat{\mathcal{K}}_k^n(t) \leq \sum_{k=1}^{d} \frac{2k}{n} \cdot t = d(d+1) \cdot (t/n) \quad \forall t \in [0:n].$$

$\square$

Lemma 8 *We have:*

$$g^{(r)} - g_{\min} \leq d(d+1) \cdot (\xi_{r+1}^n/n), \tag{36}$$

where ξ_{r+1}^n is the smallest root of the Krawtchouk polynomial $\mathcal{K}_{r+1}^n(t)$.

Proof This follows immediately from Lemma 7 and Remark 1 at the end of Sect. 3, noting that the Krawtchouk polynomials are indeed orthogonal w.r.t. the measure ω on $[0:n]$ (cf. 20). $\square$

Putting things together, we may prove our main result, Theorem 1.

Proof of Theorem 1 Assume that r is large enough so that $d(d+1) \cdot (\xi_{r+1}^n/n) \leq 1/2$. By Lemma 8, we then have

$$g^{(r)} - g_{\min} \leq d(d+1) \cdot (\xi_{r+1}^n/n) \leq 1/2.$$

By Theorem 8 we are thus able to choose a polynomial $u \in \mathbb{R}[t]_r$ whose associated operator $\mathbf{K}$ satisfies $\mathbf{K}(1) = 1$ and

$$\|\mathbf{K}^{-1} - I\| \leq 2\gamma_d \cdot d(d+1) \cdot (\xi_{r+1}^n/n).$$

We may then use Lemma 1 to obtain Theorem 1 with constant $C_d := \gamma_d \cdot d(d+1)$.

$\square$

5 Proof of Theorem 3

We turn now to analyzing the inner hierarchy $f^{(r)}$ defined in (7) for a polynomial $f \in \mathbb{R}[x]_d$ on the boolean cube, whose definition is repeated for convenience:

$$f^{(r)} := \min_{s \in \Sigma[x]_r} \left\{ \int_{\mathbb{B}^n} f(x) \cdot s(x) d\mu : \int_{\mathbb{B}^n} s(x) d\mu = 1 \right\} \geq f_{\min}. \qquad (37)$$

As before, we may assume w.l.o.g. that $f_{\min} = f(0) = 0$ and that $f_{\max} = 1$. To facilitate the analysis of the bounds $f^{(r)}$, the idea is to restrict in (37) to polynomials $s(x)$ that are invariant under the action of $\mathrm{St}(0) \subseteq \mathrm{Aut}(\mathbb{B}^n)$, i.e., depending only on the Hamming weight $|x|$. Such polynomials are of the form $s(x) = u(|x|)$ for some univariate polynomial $u \in \mathbb{R}[t]$. Hence this leads to the following, weaker hierarchy, where we now optimize over *univariate* sums-of-squares:

$$f_{\mathrm{sym}}^{(r)} := \min_{u \in \Sigma[t]_r} \left\{ \int_{\mathbb{B}^n} f(x) \cdot u(|x|) d\mu(x) : \int_{\mathbb{B}^n} u(|x|) d\mu(x) = 1 \right\}.$$

By definition, we must have $f_{\mathrm{sym}}^{(r)} \geq f^{(r)} \geq f_{\min}$, and so an analysis of $f_{\mathrm{sym}}^{(r)}$ extends immediately to $f^{(r)}$.

The main advantage of working with the hierarchy $f_{\mathrm{sym}}^{(r)}$ is that we may now assume that f is itself invariant under $\mathrm{St}(0)$, after replacing f by its symmetrization:

$$\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} f(\sigma(x)).$$

Indeed, for any $u \in \Sigma[t]_r$, we have that:

$$\int_{\mathbb{B}^n} f(x) u(|x|) d\mu(x) = \frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \int_{\mathbb{B}^n} f(\sigma(x)) u(|\sigma(x)|) d\mu(\sigma(x))$$

$$= \int_{\mathbb{B}^n} \frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} f(\sigma(x)) u(|x|) d\mu(x).$$

So we now assume that f is $\mathrm{St}(0)$-invariant, and thus we may write:

$$f(x) = F(|x|) \text{ for some polynomial } F(t) \in \mathbb{R}[t]_d.$$

By the definitions of the measures μ on $\mathbb{B}^n$ and ω on $[0:n]$ we have the identities:

$$\int_{\mathbb{B}^n} u(|x|) d\mu(x) = \int_{[0:n]} u(t) d\omega(t),$$

$$\int_{\mathbb{B}^n} F(|x|) u(|x|) d\mu(x) = \int_{[0:n]} F(t) u(t) d\omega(t).$$

Hence we get

$$f_{\text{sym}}^{(r)} = \min_{u \in \Sigma[t]_r} \left\{ \int_{[0:n]} F(t) \cdot u(t) d\omega(t) : \int_{[0:n]} u(t) d\omega(t) = 1 \right\} = F_{[0:n],\omega}^{(r)}. \quad (38)$$

In other words, the behaviour of the *symmetrized* inner hierarchy $f_{\text{sym}}^{(r)}$ over the boolean cube w.r.t. the uniform measure μ is captured by the behaviour of the *univariate* inner hierarchy $F_{[0:n],\omega}^{(r)}$ over $[0:n]$ w.r.t. the discrete measure ω.

Now, we are in a position to make use again of the technique outlined in Sect. 3. First we find a linear upper estimator $\widehat{F}$ for F on $[0:n]$.

Lemma 9 *We have*

$$F(t) \le \widehat{F}(t) := d(d+1) \cdot \gamma_d \cdot t/n \quad \forall t \in [0:n],$$

where γ_d is the same constant as in Lemma 5.

Proof Write $F(t) = \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^n(t)$ for some scalars λ_i. By assumption, $F(0) = 0$ and thus $\sum_{i=0}^{d} \lambda_i = 0$. We now use an analogous argument as for Lemma 7:

$$F(t) = \sum_{i=0}^{d} \lambda_i (\widehat{\mathcal{K}}_i^n(t) - 1) \le \sum_{i=0}^{d} |\lambda_i| |\widehat{\mathcal{K}}_i^n(t) - 1| \overset{(24)}{\le} \max_i |\lambda_i| \cdot t \cdot \sum_{i=0}^{d} \frac{2i}{n}$$

$$\le \max_i |\lambda_i| \cdot t \cdot \frac{d(d+1)}{n}.$$

As $\|f\|_\infty = 1$, using Lemma 5, we can conclude that:

$$|\lambda_i| = \max_{t \in [0:n]} |\lambda_i \widehat{\mathcal{K}}_i^n(t)| \le \gamma_d$$

which gives the desired result. $\qquad\square$

In light of Remark 1 in Sect. 3, we may now conclude that

$$F_{[0:n],\omega}^{(r)} \le d(d+1)\gamma_d \cdot \xi_{r+1}^n/n.$$

As $f^{(r)} \le f_{\text{sym}}^{(r)} = F_{[0:n],\omega}^{(r)}$, we have thus shown Theorem 3 with constant $C_d = d(d+1)\gamma_d$. Note that in comparison to Lemma 8, we only have the additional constant factor γ_d.

Exactness of the inner hierarchy As is the case for the outer hierarchy, the inner hierarchy is exact when r is large enough. Whereas the outer hierarchy, however, is exact for $r \ge (n + d - 1)/2$, the inner hierarchy is exact in general if and only if $r \ge n$. We give a short proof of this fact below, for reference.

Lemma 10 *Let f be a polynomial on $\mathbb{B}^n$. Then $f^{(r)} = f_{\min}$ for all $r \ge n$.*

Proof We may assume w.l.o.g. that $f(0) = f_{\min}$. Consider the interpolation polynomial:

$$s(x) := \sqrt{2^n} \prod_{i=1}^{n} (1 - x_i) \in \mathbb{R}[x]_n,$$

which satisfies $s^2(0) = 2^n$ and $s^2(x) = 0$ for all $0 \neq x \in \mathbb{B}^n$. Clearly, we have:

$$\int f s^2 d\mu = f(0) = f_{\min} \text{ and } \int s^2 d\mu = 1,$$

and so $f^{(n)} = f_{\min}$. □

The next lemma shows that this result is tight, by giving an example of polynomial f for which the bound $f^{(r)}$ is exact only at order $r = n$.

Lemma 11 *Let* $f(x) = |x| = x_1 + \cdots + x_n$. *Then* $f^{(r)} - f_{\min} > 0$ *for all* $r < n$.

Proof Suppose not. That is, $f^{(r)} = f_{\min} = 0$ for some $r \leq n-1$. As $f(x) > 0 = f_{\min}$ for all $0 \neq x \in \mathbb{B}^n$, this implies that there exists a polynomial $s \in \mathbb{R}[x]_r$ such that s^2 is interpolating at 0, i.e. such that $s^2(0) = 1$ and $s^2(x) = 0$ for all $0 \neq x \in \mathbb{B}^n$. But then s is itself interpolating at 0 and has degree $r < n$, a contradiction. □

6 Proof of Lemma 5

In this section we give a proof of Lemma 5, where we bound the sup-norm $\|p_k\|_\infty$ of the harmonic components p_k of a polynomial p by $\gamma_d \|p\|_\infty$ for some constant γ_d depending only on the degree d of p. The following definitions will be convenient.

Definition 1 For $n \geq d \geq k \geq 0$ integers, we write:

$$\rho(n, d, k) := \sup\{\|p_k\|_\infty : p = p_0 + p_1 + \cdots + p_d \in \mathcal{R}[x]_d, \|p\|_\infty \leq 1\}, \text{ and}$$
$$\rho(n, d) := \max_{0 \leq k \leq d} \rho(n, d, k).$$

We are thus interested in finding a bound γ_d depending only on d such that:

$$\gamma_d \geq \rho(n, d) \text{ for all } n \in \mathbb{N}. \tag{39}$$

We will now show that in the computation of the parameter $\rho(n, d, k)$ we may restrict to feasible solutions p having strong structural properties. First, we show that we may assume that the sup-norm of the harmonic component p_k of p is attained at $0 \in \mathbb{B}^n$.

Lemma 12 *We have*

$$\rho(n, d, k) = \sup_{p \in \mathcal{R}[x]_d^n} \{p_k(0) : \|p\|_\infty \leq 1\}. \tag{40}$$

Proof Let p be a feasible solution for $\rho(n, d, k)$ and let $x \in \mathbb{B}^n$ for which $p_k(x) = \|p_k\|_\infty$ (after possibly replacing p by $-p$). Now choose $\sigma \in \text{Aut}(\mathbb{B}^n)$ such that $\sigma(0) = x$ and set $\widehat{p} = p \circ \sigma$. Clearly, $\widehat{p}$ is again a feasible solution for $\rho(n, d, k)$. Moreover, as H_k is invariant under $\text{Aut}(\mathbb{B}^n)$, we have:

$$\|\widehat{p}_k\|_\infty = \widehat{p}_k(0) = (p \circ \sigma)_k(0) = (p_k \circ \sigma)(0) = \|p_k\|_\infty,$$

which shows the lemma. □

Next we show that we may in addition restrict to polynomials that are highly symmetric.

Lemma 13 *In the program* (40) *we may restrict the optimization to polynomials of the form*

$$p(x) = \sum_{i=0}^d \lambda_i \sum_{|a|=i} \chi_a(x) = \sum_{i=0}^d \lambda_i \mathcal{K}_i^n(|x|) \quad \text{where } \lambda_i \in \mathbb{R}.$$

Proof Let p be a feasible solution to (40). Consider the following polynomial $\widehat{p}$ obtained as symmetrization of p under action of $\text{St}(0)$, the set of automorphism of $\mathbb{B}^n$ corresponding to the coordinate permutations:

$$\widehat{p}(x) = \frac{1}{|\text{St}(0)|} \sum_{\sigma \in \text{St}(0)} (p \circ \sigma)(x).$$

Then $\|\widehat{p}\|_\infty \le 1$ and $\widehat{p}_k(0) = p_k(0)$, so $\widehat{p}$ is still feasible for (40) and has the same objective value as p. Furthermore, for each i, $\widehat{p}_i$ is invariant under $\text{St}(0)$, which implies that $\widehat{p}_i(x) = \lambda_i X_i(x) = \lambda_i \sum_{|a|=i} \chi_a(x) = \lambda_i \mathcal{K}_i^n(|x|)$ for some $\lambda_i \in \mathbb{R}$ (see 18). □

A simple rescaling $\lambda_i \leftarrow \lambda_i \cdot \binom{n}{i}$ allows us to switch from $\mathcal{K}_i^n$ to $\widehat{\mathcal{K}}_i^n = \mathcal{K}_i^n / \binom{n}{i}$ and to obtain the following reformulation of $\rho(n, d, k)$ as a linear program.

Lemma 14 *For any $n \ge d \ge k$ we have:*

$$\rho(n, d, k) = \max \quad \lambda_k$$
$$s.t. \quad -1 \le \sum_{i=0}^d \lambda_i \widehat{\mathcal{K}}_i^n(t) \le 1 \quad (t = 0, 1, \dots, n). \tag{41}$$

Limit functions The idea now is to prove a bound on $\rho(n, d, d)$ which holds for fixed d and is independent of n. We will do this by considering 'the limit' of problem (41) as $n \to \infty$. For each $k \in \mathbb{N}$, we define the limit function:

$$\widehat{\mathcal{K}}_k^\infty(t) := \lim_{n \to \infty} \widehat{\mathcal{K}}_k^n(nt),$$

which, as shown in Lemma 15 below, is in fact a polynomial. We first present the polynomial $\widehat{\mathcal{K}}_k^\infty(t)$ for small k as an illustration.

Example 1 We have:

$$\widehat{\mathcal{K}}_0^n(nt) = 1 \implies \widehat{\mathcal{K}}_0^\infty(t) = 1,$$
$$\widehat{\mathcal{K}}_1^n(nt) = -2t + 1 \implies \widehat{\mathcal{K}}_1^\infty(t) = -2t + 1,$$
$$\widehat{\mathcal{K}}_2^n(nt) = \frac{2n^2t^2 - 2n^2t + \binom{n}{2}}{\binom{n}{2}} \implies \widehat{\mathcal{K}}_2^\infty(t) = 4t^2 - 4t + 1 = (1 - 2t)^2.$$

Lemma 15 *We have:* $\widehat{\mathcal{K}}_k^\infty(t) = (1 - 2t)^k$ *for all* $k \in \mathbb{N}$.

Proof The Krawtchouk polynomials satisfy the following three-term recurrence relation (see, e.g., [28]):

$$(k + 1)\mathcal{K}_{k+1}^n(t) = (n - 2t)\mathcal{K}_k^n(t) - (n - k + 1)\mathcal{K}_{k-1}^n(t)$$

for $1 \leq k \leq n - 1$. By evaluating the polynomials at nt we obtain:

$$(k + 1)\mathcal{K}_{k+1}^n(nt) = (n - 2nt)\mathcal{K}_k^n(nt) - (n - k + 1)\mathcal{K}_{k-1}^n(nt),$$
$$\implies (k + 1)\binom{n}{k + 1}\widehat{\mathcal{K}}_{k+1}^n(nt)$$
$$= (n - 2nt)\binom{n}{k}\widehat{\mathcal{K}}_k^n(nt) - (n - k + 1)\binom{n}{k - 1}\widehat{\mathcal{K}}_{k-1}^n(nt),$$
$$\implies \widehat{\mathcal{K}}_{k+1}^n(nt) = \frac{n(1 - 2t)}{(n - k)} \cdot \widehat{\mathcal{K}}_k^n(nt) - \frac{k}{n - k} \cdot \widehat{\mathcal{K}}_{k-1}^n(nt),$$
$$\implies \widehat{\mathcal{K}}_{k+1}^\infty(t) = (1 - 2t)\widehat{\mathcal{K}}_k^\infty(t).$$

As $\widehat{\mathcal{K}}_0^\infty(t) = 1$ and $\widehat{\mathcal{K}}_1^\infty(t) = 1 - 2t$ we can conclude that indeed $\widehat{\mathcal{K}}_k^\infty(t) = (1 - 2t)^k$ for all $k \in \mathbb{N}$. □

Next, we show that solutions to (41) remain feasible after increasing the dimension n.

Lemma 16 *Let* $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_d)$ *be a feasible solution to* (41) *for a certain* $n \in \mathbb{N}$. *Then it is also feasible to* (41) *for* $n + 1$ *(and thus for any* $n' \geq n + 1$*). Therefore,* $\rho(n + 1, d, k) \geq \rho(n, d, k)$ *for all* $n \geq d \geq k$ *and thus* $\rho(n + 1, d) \geq \rho(n, d)$ *for all* $n \geq d$.

Proof We may view $\mathbb{B}^n$ as a subset of $\mathbb{B}^{n+1}$ via the map $a \mapsto (a, 0)$, and analogously $\mathcal{R}[x_1, \ldots, x_n]$ as a subset of $\mathcal{R}[x_1, \ldots, x_n, x_{n+1}]$ via $\chi_a \mapsto \chi_{(a,0)}$. Now for $m, i \in \mathbb{N}$ we consider again the zonal spherical harmonic (18):

$$X_i^m = \sum_{|a|=i, a \in \mathbb{B}^m} \chi_a.$$

Consider the set $\mathrm{St}(0) \subseteq \mathrm{Aut}(\mathbb{B}^{n+1})$ of automorphisms fixing $0 \in \mathbb{B}^{n+1}$, i.e., the coordinate permutations arising from $\sigma \in \mathrm{Sym}(n + 1)$. We will use the following identity:

$$\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \frac{X_i^n}{\binom{n}{i}} \circ \sigma = \frac{X_i^{n+1}}{\binom{n+1}{i}}. \tag{42}$$

To see that (42) holds note that its left hand side is equal to

$$\frac{1}{(n+1)!\binom{n}{i}} \sum_{\sigma \in \mathrm{Sym}(n+1)} \sum_{a \in \mathbb{B}^n, |a|=i} \chi_{(a,0)} \circ \sigma = \frac{1}{(n+1)!\binom{n}{i}} \sum_{b \in \mathbb{B}^{n+1}, |b|=i} N_b \chi_b,$$

where N_b denotes the number of pairs (σ, a) with $\sigma \in \mathrm{Sym}(n+1)$, $a \in \mathbb{B}^n$, $|a| = i$ such that $b = \sigma(a, 0)$. As there are $\binom{n}{i}$ choices for a and $i!(n+1-i)!$ choices for σ we have $N_b = \binom{n}{i}i!(n+1-i)!$ and thus (42) holds.

Assume λ is a feasible solution of (41) for a given value of n. Then, in view of (21), this means

$$\left| \sum_{i=0}^d \lambda_i \cdot \frac{X_i^n(x)}{\binom{n}{i}} \right| \le 1 \quad \text{for all } x \in \mathbb{B}^n, \quad \text{and thus for all } x \in \mathbb{B}^{n+1}.$$

Using (42) we obtain:

$$\left| \sum_{i=0}^d \lambda_i \frac{X_i^{n+1}(x)}{\binom{n+1}{i}} \right| = \left| \sum_{i=0}^d \lambda_i \cdot \frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \frac{X_i^n(\sigma(x))}{\binom{n}{i}} \right|$$

$$= \left| \left(\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \left(\sum_{i=0}^d \lambda_i \frac{X_i^n}{\binom{n}{i}} \right) \circ \sigma \right)(x) \right| \le 1$$

for all $x \in \mathbb{B}^{n+1}$. Using (21) again, this shows that λ is a feasible solution of program (41) for $n + 1$. $\qquad\square$

Example 2 To illustrate the identity (42), we give a small example with $n = i = 2$. Consider:

$$X_2^2 = \sum_{|a|=2, a \in \mathbb{B}^2} \chi_a = \chi_{11}.$$

The automorphisms in $\mathrm{St}(0) \subseteq \mathrm{Aut}(\mathbb{B}^3)$ fixing $0 \in \mathbb{B}^3$ are the permutations of x_1, x_2, x_3. So we get:

$$\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} X_2^2 \circ \sigma = \frac{1}{6}(\chi_{110} + \chi_{101} + \chi_{110} + \chi_{011} + \chi_{101} + \chi_{011})$$

$$= \frac{2}{6}(\chi_{110} + \chi_{101} + \chi_{011}) = \frac{1}{3}X_2^3,$$

and indeed $\binom{2}{2}/\binom{3}{2} = 1/3$.

Lemma 17 *For $d \geq k \in \mathbb{N}$, define the program:*

$$\rho(\infty, d, k) := \max \quad \lambda_k$$
$$s.t. \quad -1 \leq \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(t) \leq 1 \quad (t \in [0, 1]). \tag{43}$$

Then, for any $n \geq d$, we have: $\rho(n, d, k) \leq \rho(\infty, d, k)$.

Proof Let λ be a feasible solution to (41) for (n, d, k). We show that λ is feasible for (43). For this fix $t \in [0, 1] \cap \mathbb{Q}$. Then there exists a sequence of integers $(n_j)_j \to \infty$ such that $n_j \geq n$ and $tn_j \in [0, n_j]$ is integer for each $j \in \mathbb{N}$. As $n_j \geq n$, we know from Lemma 16 that λ is also a feasible solution of program (41) for (n_j, d, k). Hence, since $n_j t \in [0 : n_j]$ we obtain

$$| \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{n_j}(n_j t)| \leq 1 \quad \forall j \in \mathbb{N}.$$

But this immediately gives:

$$| \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(t)| = \lim_{j \to \infty} | \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{n_j}(n_j t)| \leq 1. \tag{44}$$

As $[0, 1] \cap \mathbb{Q}$ lies dense in $[0, 1]$ (and the $\widehat{\mathcal{K}}_i^{\infty}$'s are continuous) we may conclude that (44) holds for all $t \in [0, 1]$. This shows that λ is feasible for (43) and we thus have $\rho(n, d, k) \leq \rho(\infty, d, k)$, as desired. $\qquad \square$

It remains now to compute the optimum solution to the program (43). In light of Lemma 15, and after a change of variables $x = 1 - 2t$, this program may be reformulated as:

$$\max \quad |\lambda_k|$$
$$s.t. \quad -1 \leq \sum_{i=0}^{d} \lambda_i x^i \leq 1 \quad (x \in [-1, 1]). \tag{45}$$

In other words, we are tasked with finding a polynomial $p(x)$ of degree d satisfying $|p(x)| \leq 1$ for all $x \in [-1, 1]$, whose k-th coefficient is as large as possible in absolute value. This is a classical extremal problem solved by V. Markov.

Theorem 9 (see, e.g., Theorem 7, pp. 53 in [29]) *For $m \in \mathbb{N}$, let $T_m(x) = \sum_{i=0}^{m} t_{m,i} x^i$ be the Chebyshev polynomial of degree m. Then the optimum solution λ to (45) is given by:*

$$\sum_{i=0}^{d} \lambda_i x^i = \begin{cases} T_d(x) & \text{if } k \equiv d \mod 2, \\ T_{d-1}(x) & \text{if } k \not\equiv d \mod 2. \end{cases}$$

Table 1 Values of the constant γ_d

d	1	2	3	4	5	6	7	8	9	10
γ_d	1	2	4	8	20	48	112	256	576	1280

In particular, $\rho(\infty, d, k)$ is equal to $|t_{d,k}|$ (resp. $|t_{d-1,k}|$).

As the coefficients of the Chebyshev polynomials are known explicitly, Theorem 9 allows us to give exact values of the constant γ_d appearing in our main results (see Table 1). Using the following identity:

$$\sum_{i=0}^{d} |t_{d,i}| = \frac{1}{2}(1 + \sqrt{2})^d + \frac{1}{2}(1 - \sqrt{2})^d \leq (1 + \sqrt{2})^d,$$

we are also able to concretely estimate:

$$\gamma_d \leq \max_{k \leq d} \rho(\infty, d, k) \leq (1 + \sqrt{2})^d.$$

7 Concluding remarks

Summary We have shown a theoretical guarantee on the quality of the sum-of-squares hierarchy $f_{(r)} \leq f_{\min}$ for approximating the minimum of a polynomial f of degree d over the boolean cube $\mathbb{B}^n$. As far as we are aware, this is the first such analysis that applies to values of r smaller than $(n + d)/2$, i.e., when the hierarchy is not exact. Additionally, our guarantee applies to a second, measure-based hierarchy of bounds $f^{(r)} \geq f_{\min}$. Our result may therefore also be interpreted as bounding the range $f^{(r)} - f_{(r)}$. Our analysis also applies to polynomial optimization over the cube $\{\pm 1\}^n$ (by a simple change of variables), over the q-ary cube (see Appendix A) and in the setting of matrix-valued polynomials (see Appendix B).

Analysis for small values of r A limitation of Theorem 1 is that the analyis of $f_{(r)}$ applies only for choices of d, r, n satisfying $d(d + 1)\xi_{r+1}^n \leq 1/2$. One may partially avoid this limitation by proving a slightly sharper version of Lemma 6, showing instead that $\Lambda \leq \tilde{\Lambda}/(1 - \tilde{\Lambda})$, assuming now only that $\tilde{\Lambda} \leq 1$. Indeed, Lemma 6 is a special case of this result, assuming that $\tilde{\Lambda} \leq 1/2$ to obtain $\Lambda \leq 2\tilde{\Lambda}$. Nevertheless, our methods exclude values of r outside of the regime $r = \Omega(n)$.

The constant γ_d The strength of our results depends in large part on the size of the constant C_d appearing in Theorem 1 and Theorem 3, where we may set $C_d = d(d + 1)\gamma_d$. Recall that γ_d is defined in Lemma 5 as a constant for which $\|p_k\|_\infty \leq \gamma_d \|p\|_\infty$ for any polynomial $p = p_0 + p_1 + \ldots + p_d$ of degree d and $k \leq d$ on $\mathbb{B}^n$, independently of the dimension n. In Sect. 6 we have shown the existence of such a constant. Furthermore, we have shown there that we may choose $\gamma_d \leq (1 + \sqrt{2})^d$, and have given an explicit expression for the smallest possible value of γ_d in terms of the coefficients of Chebyshev polynomials. Table 1 lists these values for small d.

Computing extremal roots of Krawtchouk polynomials Although Theorem 2 provides only an asymptotic bound on the least root ξ_r^n of $\mathcal{K}_r^n$, it should be noted that ξ_r^n can be computed explicitly for small values of r, n, thus allowing for a concrete estimate of the error of both Lasserre hierarchies via Theorems 1 and 3, respectively. Indeed, as is well-known, the root ξ_{r+1}^n is equal to the smallest eigenvalue of the $(r+1) \times (r+1)$ matrix A (aka Jacobi matrix), whose entries are given by $A_{i,j} = \langle t\widehat{\mathcal{K}}_i^n(t), \widehat{\mathcal{K}}_j^n(t)\rangle_\omega$ for $i, j \in \{0, 1, \ldots, r\}$. See, e.g., [41] for more details.

Connecting the hierarchies Our analysis of the *outer* hierarchy $f_{(r)}$ on $\mathbb{B}^n$ relies essentially on knowledge of the the *inner* hierarchy $f^{(r)}$. Although not explicitly mentioned there, this is the case for the analysis on S^{n-1} in [11] as well. As the behaviour of $f^{(r)}$ is generally quite well understood, this suggests a potential avenue for proving further results on $f_{(r)}$ in other settings.

For instance, the inner hierarchy $f^{(r)}$ is known to converge at a rate in $O(1/r^2)$ on the unit ball B^n or the unit box $[-1, 1]^n$, but matching results on the outer hierarchy $f_{(r)}$ are not available. The question is thus whether the strategy used for the hypersphere S^{n-1} in [11] and for the boolean cube $\mathbb{B}^n$ here might be extended to these cases as well.

A difficulty is that the sets B^n and $[-1, 1]^n$ have a more complicated symmetric structure than S^{n-1} and $\mathbb{B}^n$, respectively. In particular, the group actions have uncountably many orbits in these cases, and a direct analog of the Funk–Hecke formula (28) is not available. New ideas are therefore needed to define the kernel $K(x, y)$ (cf. 8) and analyze its eigenvalues.

A The *q*-ary cube

In this section, we indicate how our results for the boolean cube $\mathbb{B}^n$ may be extended to the q-ary cube $(\mathbb{Z}/q\mathbb{Z})^n = \{0, 1, \ldots, q - 1\}^n$ when $q > 2$ is a fixed integer. Here $\mathbb{Z}/q\mathbb{Z}$ denotes the cyclic group of order q, so that $(\mathbb{Z}/q\mathbb{Z})^n = \mathbb{B}^n$ when $q = 2$. The lower bound $f_{(r)}$ for the minimum of a polynomial f over $(\mathbb{Z}/q\mathbb{Z})^n$ is defined analogously to the case $q = 2$; namely we set

$$f_{(r)} := \sup_{\lambda \in \mathbb{R}} \left\{ f(x) - \lambda \text{ is a sum-of-squares of degree at most } 2r \text{ on } (\mathbb{Z}/q\mathbb{Z})^n \right\},$$

where the condition means that $f(x) - \lambda$ agrees with a sum of squares $s \in \Sigma[x]_{2r}$ for all $x \in (\mathbb{Z}/q\mathbb{Z})^n$ or, alternatively, that $f - \lambda - s$ belongs to the ideal generated by the polynomials $x_i(x_i - 1) \ldots (x_i - q + 1)$ for $i \in [n]$. Similarly, the upper bound $f^{(r)}$ is defined as in (7) after equipping $(\mathbb{Z}/q\mathbb{Z})^n$ with the uniform measure μ. The parameters $f_{(r)}$ and $f^{(r)}$ may again be computed by solving a semidefinite program of size polynomial in n for fixed $r, q \in \mathbb{N}$, see [23].

As before $d(x, y)$ denotes the Hamming distance and $|x|$ denotes the Hamming weight (number of nonzero components). Note that, for $x, y \in (\mathbb{Z}/q\mathbb{Z})^n$, $d(x, y)$ can again be expressed as a polynomial in x, y, with degree $q - 1$ in each of x and y.

We will prove Theorem 11 below, which can be seen as an analog of Corollary 1 for $(\mathbb{Z}/q\mathbb{Z})^n$. The general structure of the proof is identical to that of the case $q = 2$. We therefore only give generalizations of arguments as necessary. For reasons that will become clear later, it is most convenient to consider the sum-of-squares bound $f_{(r)}$ on the minimum $f_{\min}$ of a polynomial f with degree at most $(q - 1)d$ over $(\mathbb{Z}/q\mathbb{Z})^n$, where $d \leq n$ is fixed.

Fourier analysis on $(\mathbb{Z}/q\mathbb{Z})^n$ and Krawtchouk polynomials Consider the space

$$\mathcal{R}[x] := \mathbb{C}[x]/(x_i(x_i - 1) \ldots (x_i - q + 1) : i \in [n])$$

consisting of the polynomials on $(\mathbb{Z}/q\mathbb{Z})^n$ with complex coefficients. We equip $\mathcal{R}[x]$ with its natural inner product

$$\langle f, g \rangle_\mu = \int_{(\mathbb{Z}/q\mathbb{Z})^n} f(x)\overline{g(x)}d\mu(x) = \frac{1}{q^n} \sum_{x \in (\mathbb{Z}/q\mathbb{Z})^n} f(x)\overline{g(x)},$$

where μ is the uniform measure on $(\mathbb{Z}/q\mathbb{Z})^n$. The space $\mathcal{R}[x]$ has dimension $|(\mathbb{Z}/q\mathbb{Z})^n| = q^n$ over $\mathbb{C}$ and it is spanned by the polynomials of degree up to $(q - 1)n$. The reason we now need to work with polynomials with complex coefficients is that the characters have complex coefficients when $q > 2$.

Let $\psi = e^{2\pi i/q}$ be a primitive q-th root of unity. For $a \in (\mathbb{Z}/q\mathbb{Z})^n$, the associated *character* $\chi_a \in \mathcal{R}[x]$ is defined by:

$$\chi_a(x) = \psi^{a \cdot x} \quad (x \in (\mathbb{Z}/q\mathbb{Z})^n).$$

So (14) is indeed the special case of this definition when $q = 2$. The set of all characters $\{\chi_a : a \in (\mathbb{Z}/q\mathbb{Z})^n\}$ forms an orthogonal basis for $\mathcal{R}[x]$ w.r.t. the above inner product $\langle \cdot, \cdot \rangle_\mu$. A character χ_a can be written as a polynomial of degree $(q-1) \cdot |a|$ on $(\mathbb{Z}/q\mathbb{Z})^n$, i.e., we have $\chi_a \in \mathcal{R}[x]_{(q-1)|a|}$ for all $a \in (\mathbb{Z}/q\mathbb{Z})^n$.

As before, we have the direct sum decomposition into pairwise orthogonal subspaces:

$$\mathcal{R}[x] = H_0 \perp H_1 \perp \cdots \perp H_n,$$

where H_i is spanned by the set $\{\chi_a : |a| = i\}$ and $H_i \subseteq \mathbb{R}[x]_{(q-1)i}$. The components H_i are invariant and irreducible under the action of $\mathrm{Aut}((\mathbb{Z}/q\mathbb{Z})^n)$, which is generated by the coordinate permutations and the action of $\mathrm{Sym}(q)$ on individual coordinates. Hence any $p \in \mathcal{R}[x]$ of degree at most $(q - 1)d$ can be (uniquely) decomposed as:

$$p = p_0 + p_1 + \cdots + p_d \quad (p_i \in H_i).$$

As before $\mathrm{St}(0) \subseteq \mathrm{Aut}((\mathbb{Z}/q\mathbb{Z})^n)$ denotes the stabilizer of $0 \in (\mathbb{Z}/q\mathbb{Z})^n$, which is generated by the coordinate permutations and the permutations in $\mathrm{Sym}(q)$ fixing 0

in $\{0, 1, \ldots, q - 1\}$ at any individual coordinate. We note for later reference that the subspace of H_i invariant under action of $\mathrm{St}(0)$ is of dimension one, and is spanned by the zonal spherical function:

$$X_i = \sum_{|a|=i} \chi_a \in H_i. \tag{46}$$

The Krawtchouk polynomials introduced in Sect. 2 have the following generalization in the q-ary setting:

$$\mathcal{K}_k^n(t) = \mathcal{K}_{k,q}^n(t) := \sum_{i=0}^{k} (-1)^i (q - 1)^{k-i} \binom{t}{i} \binom{n - t}{k - i}.$$

Analogously to relation (20), the Krawtchouk polynomials $\mathcal{K}_k^n$ ($0 \leq k \leq n$) are pairwise orthogonal w.r.t. the discrete measure ω on $[0 : n]$ given by:

$$\omega(t) = \frac{1}{q^n} \sum_{t=0}^{n} w(t)\delta_t, \quad \text{with } w(t) := (q - 1)^t \binom{n}{t}. \tag{47}$$

To be precise, we have:

$$\sum_{t=0}^{n} \mathcal{K}_k^n(t)\mathcal{K}_{k'}^n(t)(q - 1)^t \binom{n}{t} = \delta_{k,k'}(q - 1)^k \binom{n}{k}.$$

As $\mathcal{K}_k^n(0) = (q - 1)^k \binom{n}{k} = \|\mathcal{K}_k^n\|_\omega^2$, we may normalize $\mathcal{K}_k^n$ by setting:

$$\widehat{\mathcal{K}}_k^n(t) := \mathcal{K}_k^n(t)/\mathcal{K}_k^n(0) = \mathcal{K}_k^n(t)/\|\mathcal{K}_k^n\|_\omega^2,$$

so that $\widehat{\mathcal{K}}_k^n$ satisfies $\max_{t=0}^n \widehat{\mathcal{K}}_k^n(t) = \widehat{\mathcal{K}}_k^n(0) = 1$ (cf. 23).

We have the following connection (cf. 22) between the characters and the Krawtchouk polynomials:

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a|=k} \chi_a(x) = \mathcal{K}_k^n(i) \quad \text{for } x \in (\mathbb{Z}/q\mathbb{Z})^n \text{ with } |x| = i. \tag{48}$$

Note that for all $a, x, y \in (\mathbb{Z}/q\mathbb{Z})^n$, we have:

$$\chi_a^{-1}(x) = \overline{\chi_a(x)} = \chi_a(-x), \quad \chi_a(x)\chi_a(y) = \chi_a(x + y).$$

Hence, for any $x, y \in (\mathbb{Z}/q\mathbb{Z})^n$, we also have (cf. 21):

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a|=k} \chi_a(x)\overline{\chi_a(y)} = \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a|=k} \chi_a(x - y) = \mathcal{K}_k^n(i) \quad \text{when } d(x, y)$$
$$= |x - y| = i.$$

Invariant kernels In analogy to the binary case $q = 2$, for a degree r univariate polynomial $u \in \mathbb{R}[t]_r$ we define the associated polynomial kernel $K(x, y) := u^2(d(x, y))$ $(x, y \in (\mathbb{Z}/q\mathbb{Z})^n)$ and the associated kernel operator:

$$\mathbf{K} : p \mapsto \mathbf{K}p(x) = \int_{(\mathbb{Z}/q\mathbb{Z})^n} \overline{p(y)} K(x, y) d\mu(y)$$

$$= \frac{1}{q^n} \sum_{y \in (\mathbb{Z}/q\mathbb{Z})^n} \overline{p(y)} K(x, y) \quad (p \in \mathcal{R}[x]).$$

Note that $K(x, y)$ is a polynomial on $(\mathbb{Z}/q\mathbb{Z})^n$ with degree $2r(q - 1)$ in each of x and y. Let us decompose the univariate polynomial $u(t)^2$ in the Krawtchouk basis as

$$u(t)^2 = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t).$$

Then the kernel operator $\mathbf{K}$ acts as follows on characters: for $z \in (\mathbb{Z}/q\mathbb{Z})^n$,

$$\mathbf{K}\chi_z = \lambda_{|z|}\chi_z,$$

which can be seen by retracing the proof of Theorem 6, and we obtain the Funk–Hecke formula (recall (28)): for any polynomial $p \in \mathcal{R}[x]_{(q-1)d}$ with Harmonic decomposition $p = p_0 + \cdots + p_d$,

$$\mathbf{K}p = \lambda_0 p_0 + \cdots + \lambda_d p_d.$$

Performing the analysis It remains to find a univariate polynomial $u \in \mathbb{R}[t]$ of degree r with $u^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t)$ for which $\lambda_0 = 1$ and the other scalars λ_i are close to 1. As before (cf. 29), we have:

$$\lambda_i = \langle \mathcal{K}_i^n, u^2 \rangle_\omega / \|\mathcal{K}_i^n\|_\omega^2 = \langle \widehat{\mathcal{K}}_i^n, u^2 \rangle_\omega.$$

So we would like to minimize $\sum_{i=1}^{2r}(1 - \lambda_i)$. We are therefore interested in the inner Lasserre hierarchy applied to the minimization of the function $g(t) = d - \sum_{i=0}^d \widehat{\mathcal{K}}_i^n(t)$ on the set $[0 : n]$ (equipped with the measure ω from (47)). We show first that this function g again has a nice linear upper estimator.

Lemma 18 *We have:*

$$|\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t + 1)| \leq \frac{2k}{n}, \qquad (t = 0, 1, \ldots, n - 1)$$

$$|\widehat{\mathcal{K}}_k^n(t) - 1| \leq \frac{2k}{n} \cdot t \qquad (t = 0, 1, \ldots, n) \tag{49}$$

for all $k \leq n$.

Proof The proof is almost identical to that of Lemma 4. Let $t \in [0 : n-1]$ and $0 < k \leq d$. Consider the elements $1^t 0^{n-t}, 1^{t+1} 0^{n-t-1} \in (\mathbb{Z}/q\mathbb{Z})^n$ from Lemma 2. Then we have:

$$|\mathcal{K}_k^n(t) - \mathcal{K}_k^n(t+1)| \overset{(48)}{=} |\sum_{|a|=k} \chi_a(1^t 0^{n-t}) - \chi_a(1^{t+1} 0^{n-t-1})|$$

$$\leq 2 \cdot \#\{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a| = k, a_{t+1} \neq 0\}$$

$$= 2 \cdot (q-1)^k \cdot \binom{n-1}{k-1},$$

where for the inequality we note that $\chi_a(1^t 0^{n-t}) = \chi_a(1^{t+1} 0^{n-t-1})$ if $a_{t+1} = 0$. As $\mathcal{K}_k^n(0) = (q-1)^k \binom{n}{k}$, this implies that:

$$|\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t+1)| \leq 2 \cdot \binom{n-1}{k-1} / \binom{n}{k} = \frac{2k}{n}.$$

This shows the first inequality of (49). The second inequality follows using the triangle inequality, a telescope summation argument and the fact that $\widehat{\mathcal{K}}_k^n(0) = 1$. $\qquad\square$

From Lemma 18 we obtain that the function $g(t) = d - \sum_{i=0}^d \widehat{\mathcal{K}}_i^n(t)$ admits the following linear upper estimator: $g(t) \leq d(d+1) \cdot (t/n)$ for $t \in [0 : n]$. Now the same arguments as used for the case $q = 2$ enable us to conclude:

$$f^{((q-1)r)} - f_{\min} \leq C_d \cdot \xi_{r+1,q}^n / n$$

and, when $d(d+1)\xi_{r+1,q}^n / n \leq 1/2$,

$$f_{\min} - f_{((q-1)r)} \leq 2C_d \cdot \xi_{r+1,q}^n / n.$$

Here C_d is a constant depending only on d and $\xi_{r+1,q}^n$ is the least root of the Krawtchouk polynomial $\mathcal{K}_{r+1,q}^n$. Note that as the kernel $K(x, y) = u^2(d(x, y))$ is of degree $2(q-1)r$ in x (and y), we are only able to analyze the corresponding levels $(q-1)r$ of the hierarchies. We come back below to the question on how to show the existence of the above constant C_d.

But first we finish the analysis. Having shown analogs of Theorem 1 and Theorem 3 in this setting, it remains to state the following more general version of Theorem 2, giving information about the smallest roots of the q-ary Krawtchouk polynomials.

Theorem 10 ([26], Section 5) *Fix $t \in [0, \frac{q-1}{q}]$. Then the smallest roots $\xi_{r,q}^n$ of the q-ary Krawtchouk polynomials $\mathcal{K}_{r,q}^n$ satisfy:*

$$\lim_{r/n \to t} \xi_{r,q}^n / n = \varphi_q(t) := \frac{q-1}{q} - \left(\frac{q-2}{q} \cdot t + \frac{2}{q}\sqrt{(q-1)t(1-t)}\right). \quad (50)$$

Here the above limit means that, for any sequences $(n_j)_j$ and $(r_j)_j$ of integers such that $\lim_{j\to\infty} n_j = \infty$ and $\lim_{j\to\infty} r_j/n_j = t$, we have $\lim_{j\to\infty} \xi_{r_j,q}^{n_j} / n_j = \varphi_q(t)$.
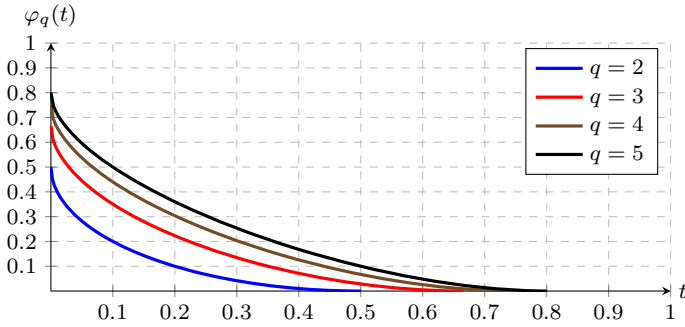
Fig. 1 The function $\varphi_q(t)$ for several values of q. Note that the case $q = 2$ corresponds to the function $\varphi(t)$ of (5)

Note that for $q = 2$ we have $\varphi_q(t) = \frac{1}{2} - \sqrt{t(1 - t)}$, which is the function $\varphi(t)$ from (5). To avoid technical details we only quote in Theorem 10 the asymptotic analog of Theorem 2 (and not the exact bound on the root $\xi_{r,q}^n$ for any n). Therefore we have shown the following q-analog of Corollary 1.

Theorem 11 *Fix $d \leq n$ and for $n, r \in \mathbb{N}$ write*

$$E_{(r)}(n) := \sup_{f \in \mathbb{R}[x]_{(q-1)d}} \left\{ f_{\min} - f_{(r)} : \|f\|_\infty = 1 \right\},$$

$$E^{(r)}(n) := \sup_{f \in \mathbb{R}[x]_{(q-1)d}} \left\{ f^{(r)} - f_{\min} : \|f\|_\infty = 1 \right\}.$$

There exists a constant $C_d > 0$ (depending also on q) such that, for any $t \in [0, \frac{q-1}{q}]$, we have:

$$\lim_{r/n \to t} E^{((q-1)r)}(n) \leq C_d \cdot \varphi_q(t)$$

and, if $d(d + 1) \cdot \varphi_q(t) \leq 1/2$, then we also have:

$$\lim_{r/n \to t} E_{((q-1)r)}(n) \leq 2 \cdot C_d \cdot \varphi_q(t).$$

Here $\varphi_q(t)$ is the function defined in (50). Recall that the limit notation $r/n \to t$ means that the claimed convergence holds for any sequences $(n_j)_j$ and $(r_j)_j$ of integers such that $\lim_{j \to \infty} n_j = \infty$ and $\lim_{j \to \infty} r_j/n_j = t$.

For reference, the function $\varphi_q(t)$ is shown for several values of q in Fig. 1.

A generalization of Lemma 5 The arguments above omit a generalization of Lemma 5, which is instrumental to show the existence of the constant C_d claimed above. In other words, we still need to show that if $p : (\mathbb{Z}/q\mathbb{Z})^n \to \mathbb{R}$ is a polynomial of degree $(q - 1)d$ on $(\mathbb{Z}/q\mathbb{Z})^n$ with harmonic decomposition $p = p_0 + \ldots + p_d$, there then exists a constant $\gamma_d > 0$ (independent of n) such that:

$$\|p_i\|_\infty \leq \gamma_d \|p\|_\infty \text{ for all } 0 \leq i \leq d.$$

Then, as in the binary case, we may set $C_d = d(d+1)\gamma_d$. The proof given in Sect. 6 for the case $q = 2$ applies almost directly to the general case, and we only generalize certain steps as required. So consider again the parameters:

$$\rho(n, d, k) := \sup\{\|p_k\|_\infty : p = p_0 + p_1 + \ldots + p_d \in \mathbb{R}[x]_{(q-1)d}, \|p\|_\infty \leq 1\}, \text{ and}$$
$$\rho(n, d) := \max_{0 \leq k \leq d} \rho(n, d, k).$$

Lemmas 12 and 13, which show that the optimum solution p to $\rho(n, d, k)$ may be assumed to be invariant under $\mathrm{St}(0) \subseteq \mathrm{Aut}((\mathbb{Z}/q\mathbb{Z})^n)$, clearly apply to the case $q > 2$ as well. That is to say, we may assume p is of the form:[2]

$$p(x) = \sum_{i=0}^{d} \lambda_i X_i(x) \quad (\lambda_i \in \mathbb{R})$$

where $X_i = \sum_{|a|=i} \chi_a \in H_i$ is the zonal spherical function of degree $(q-1)i$ (cf. 46 and 18). Using (48), we obtain a reformulation of $\rho(n, d, k)$ as an LP (cf. 41):

$$\rho(n, d, k) = \max \quad \lambda_k$$
$$s.t. \quad -1 \leq \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_{i,q}^n(t) \leq 1 \quad (t = 0, 1, \ldots, n). \tag{51}$$

For $k \in \mathbb{N}$, let $\widehat{\mathcal{K}}_k^\infty(t) := \lim_{n \to \infty} \widehat{\mathcal{K}}_k^n(nt) = \left(1 - \frac{q}{q-1}t\right)^k$ and consider the program (cf. 43):

$$\rho(\infty, d, k) := \max \quad \lambda_k$$
$$s.t. \quad -1 \leq \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^\infty(t) \leq 1 \quad (t \in [0, 1]). \tag{52}$$

As before, we have $\rho(n, d, k) \leq \rho(\infty, d, k)$, noting that (the proofs of) Lemmas 16 and 17 may be applied directly to the case $q > 2$. From there, it suffices to show $\rho(\infty, d, k) < \infty$, which can be argued in an analogous way to the case $q = 2$.

B Matrix-valued polynomials

In this section, we show how the arguments used for the proofs of our main results in Theorems 1 and 3 may be applied in the setting of matrix-valued polynomials, thereby proving Theorems 4 and 5.

---

[2] Note that as p is assumed to be real-valued, the coefficients λ_i must be real. Indeed, for each $a \in (\mathbb{Z}/q\mathbb{Z})^n$, we have $\langle p, \chi_a \rangle_\mu = \lambda_{|a|} \|\chi_a\|^2 = \lambda_{|a|} \|\chi_a^{-1}\|^2 = \langle p, \overline{\chi_a} \rangle_\mu = \overline{\langle p, \chi_a \rangle_\mu}.$

Recall that $\mathcal{S}^k$ is the space of $k \times k$ real symmetric matrices and $\mathcal{S}^k[x] \subseteq \mathbb{R}^{k \times k}[x] = \mathbb{R}[x]^{k \times k}$ is the space of n-variate polynomials whose coefficients lie in $\mathcal{S}^k$. Given a polynomial matrix $F \in \mathcal{S}^k[x]$ we consider the matrix-valued polynomial optimization problem:

$$F_{\min} := \min_{x \in \mathbb{B}^n} \lambda_{\min}(F(x)), \tag{53}$$

for which we have the outer Lasserre hierarchy:

$$F_{(r)} := \sup_{\lambda \in \mathbb{R}} \left\{ F(x) - \lambda \cdot I = S(x) \text{ on } \mathbb{B}^n \text{ for some } S \in \Sigma_r^{k \times k} \right\}, \tag{54}$$

and the inner Lasserre hierarchy:

$$F^{(r)} := \inf_{S \in \Sigma_r^{k \times k}} \left\{ \int_{\mathbb{B}^n} \text{Tr}\big(F(x)S(x)\big) d\mu(x) : \int_{\mathbb{B}^n} \text{Tr}\big(S(x)\big) d\mu(x) = 1 \right\}. \tag{55}$$

Here, the set $\Sigma_r^{k \times k}$ consists of all sum-of-squares polynomial matrices $S \in \mathcal{S}^k[x]$, of the form:

$$S(x) = \sum_i U_i(x) U_i(x)^\top \quad (U_i \in \mathbb{R}^{k \times m}[x], \ \deg U_i \le r, \ m \in \mathbb{N}).$$

The outer hierarchy (proof of Theorem 4) We generalize the outline of Sect. 1.5 to the matrix-valued setting. Let $F \in \mathcal{S}^k[x]$ be the polynomial matrix of degree d to be optimized, and assume w.l.o.g. that $0 \le \|F\|_\infty \le 1$. Here, and throughout this section, $\|F\|_\infty := \max_{x \in \mathbb{B}^n} \|F(x)\|$ is the largest absolute value of an eigenvalue of $F(x)$ over $\mathbb{B}^n$. A kernel K of the form $K(x, y) = u^2(d(x, y))$ with $u \in \mathbb{R}[t]_r$ (cf. 8) induces a linear operator $\mathbf{K}$ on $\mathcal{S}^k[x]$ by:

$$\mathbf{K}P(x) := \int_{\mathbb{B}^n} P(y) K(x, y) d\mu(y) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} P(y) K(x, y) \quad (P \in \mathcal{S}^k[x]).$$

If $P(x) \succeq 0$ for all $x \in \mathbb{B}^n$, then the polynomial $\mathbf{K}P$ is a sum-of-squares polynomial matrix of degree at most $2r$ on $\mathbb{B}^n$. Indeed, we then have:

$$\mathbf{K}P(x) = \frac{1}{2^n} \sum_{y \in \mathbb{B}^n} U_y(x) U_y(x)^\top, \quad \text{where } U_y(x) = u(d(x, y)) \sqrt{P(y)}.$$

Given $\delta > 0$ to be determined later, set $\tilde{F} = F + \delta I$. Assuming that $\mathbf{K}$ is non-singular, we can write $F = \mathbf{K}(\mathbf{K}^{-1} \tilde{F})$. Therefore, assuming that $\mathbf{K}^{-1} \tilde{F}$ is positive semidefinite over $\mathbb{B}^n$, we find that $F + \delta I$ is a sum-of-squares polynomial matrix of degree $2r$ on $\mathbb{B}^n$, and thus that $F_{\min} - F_{(r)} \le \delta$.

To guarantee positive semidefiniteness of $\tilde{F}$, it suffices to ensure that (cf. 9):

$$\|\mathbf{K}^{-1} - I\| := \sup_{P \in \mathcal{S}^k[x]_d} \frac{\|\mathbf{K}^{-1}P - P\|_\infty}{\|P\|_\infty} \leq \delta.$$

Indeed, as the smallest eigenvalue of $\tilde{F}(x)$ is at least δ for each $x \in \mathbb{B}^n$, the smallest eigenvalue of $\mathbf{K}^{-1}F(x)$ must then be at least zero.

As in the case of scalar-valued polynomials, the eigenvalues of $\mathbf{K}$ are given by the coefficients λ_i in the expansion $u^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t)$. Indeed, if $P \in \mathcal{S}^k[x]$ is a polynomial matrix of degree d then we may decompose it into harmonic components entry-wise to obtain $P(x) = \sum_{i=0}^{d} P_i(x)$ and (cf. Theorem 6):

$$\mathbf{K}P(x) = \sum_{i=0}^{d} \lambda_i P_i(x).$$

It remains to express the quantity $\|\mathbf{K}^{-1} - I\|$ in terms of the eigenvalues λ_i of $\mathbf{K}$, after which the proof proceeds as in the case of scalar polynomials. For this, note that:

$$\|\mathbf{K}^{-1}P - P\|_\infty = \|\sum_{i=1}^{d}(\lambda_i^{-1} - 1)P_i\|_\infty \leq \sum_{i=1}^{d} |\lambda_i^{-1} - 1| \|P_i\|_\infty$$

$$\leq \sum_{i=1}^{d} |\lambda_i^{-1} - 1| \cdot \gamma_d \|P\|_\infty.$$

where γ_d is the constant of Lemma 5 (cf. 34). The last inequality relies on the following generalization of Lemma 5, whose proof here is essentially as given in [11].

Lemma 19 *Let $P(x) = \sum_{i=0}^{d} P_i(x)$ be a polynomial matrix of degree d, decomposed into harmonic components. If γ_d is the constant of Lemma 5, we then have:*

$$\|P_i\|_\infty \leq \gamma_d \|P\|_\infty \quad \text{for all } i \leq d.$$

Proof For any matrix $M \in \mathcal{S}^k$, its spectral norm is $\|M\| = \max_{y \in \mathbb{R}^k}\{|y^\top M y| : \|y\| = 1\}$. Therefore, we have:

$$\|P\|_\infty = \max_{x \in \mathbb{B}^n} \max_{\|y\|=1} |y^\top P(x)y| \quad \text{and} \quad \|P_i\|_\infty = \max_{x \in \mathbb{B}^n} \max_{\|y\|=1} |y^\top P_i(x)y|.$$

For fixed y, the function $p^y : x \mapsto y^\top P(x)y$ is a (scalar) polynomial on $\mathbb{B}^n$ of degree d, whose harmonic components are given by $p_i^y : x \mapsto y^\top P_i(x)y$. Therefore, we may invoke Lemma 5 to bound:

$$\max_{x \in \mathbb{B}^n} |p_i^y(x)| \leq \gamma_d \max_{x \in \mathbb{B}^n} |p^y(x)| \text{ for all } \|y\| = 1,$$

and conclude that $\|P_i\|_\infty \leq \gamma_d \|P\|_\infty$. □

The inner hierarchy (proof of Theorem 5) We generalize the arguments of Sect. 5 to the matrix-valued setting. Let F again be the polynomial matrix of degree d to be optimized, and assume w.l.o.g. that $0 \leq \|F\|_\infty \leq 1$ and that the minimum in the optimization problem (53) is attained at 0, i.e., that $F_{\min} = \lambda_{\min}(F(0))$.

As in the scalar case, we work to reduce problem (55) to a (now matrix-valued) instance of the inner hierarchy in one variable. Note first that $F^{(r)} \leq F_{\text{sym}}^{(r)}$ for each $r \in \mathbb{N}$, where $F_{\text{sym}}^{(r)}$ is obtained by restricting the optimization in (55) to polynomial matrices $S(x)$ of the form $S(x) = U(|x|)$. Writing $\widehat{F}$ for the univariate polynomial matrix satisfying

$$\widehat{F}(|x|) = \frac{1}{|\text{St}(0)|} \sum_{\sigma \in \text{St}(0)} F \circ \sigma(x) \quad (x \in \mathbb{B}^n),$$

we find (cf. 38):

$$F_{\text{sym}}^{(r)} = \min_{U \in \Sigma_r^{k \times k}[t]} \left\{ \int_{[0:n]} \text{Tr}(\widehat{F}(t)U(t)) d\omega(t) : \int_{[0:n]} \text{Tr}(U(t)) d\omega(t) = 1 \right\}. \quad (56)$$

It remains to analyze the program (56). We first give a linear upper estimator for $\widehat{F}$ (cf. Lemma 9).

Lemma 20 *For all $t \in [0 : n]$, we have:*

$$\widehat{F}(t) \preceq \widehat{G}(t) := d(d+1) \cdot \gamma_d \cdot t/n \cdot I + C,$$

where $C = \widehat{F}(0)$ is a constant matrix with $\lambda_{\min}(C) = 0$.

Proof We may write $\widehat{F}(t) = \sum_{i=0}^d \Lambda_i \widehat{\mathcal{K}}_i^n(t)$ for certain $\Lambda_i \in \mathcal{S}^k$. We then have:

$$\sum_{i=0}^d \Lambda_i \widehat{\mathcal{K}}_i^n(t) = \sum_{i=0}^d \Lambda_i (\widehat{\mathcal{K}}_i^n(t) - 1) + \sum_{i=0}^d \Lambda_i$$

$$\preceq \max_{i=0}^d \|\Lambda_i\|_\infty \cdot I \cdot \sum_{i=0}^d |1 - \widehat{\mathcal{K}}_i^n(t)| + \sum_{i=0}^d \Lambda_i$$

$$\preceq d(d+1) \cdot \gamma_d \cdot t/n \cdot I + \sum_{i=0}^d \Lambda_i,$$

making use of Lemma 19 and (24) for the final inequality. It remains to note that $\sum_{i=0}^d \Lambda_i = \widehat{F}(0)$, and that $\lambda_{\min}(\widehat{F}(0)) = 0$ by assumption. $\qquad\square$

As $\widehat{F}(t) \preceq \widehat{G}(t)$ for all $t \in [0 : n]$, we have $F_{\text{sym}}^{(r)} \leq \widehat{G}_{[0:n],\omega}^{(r)}$, where:

$$\widehat{G}_{[0:n],\omega}^{(r)} = \min_{U \in \Sigma_r^{k \times k}[t]} \left\{ \int_{[0:n]} \text{Tr}(\widehat{G}(t)U(t)) d\omega(t) : \int_{[0:n]} \text{Tr}(U(t)) d\omega(t) = 1 \right\}$$

is the inner Lasserre hierarchy for G computed on $[0 : n]$ w.r.t. the measure ω. To conclude the argument, we prove the following generalization of Theorem 7 (see also Remark 1).

Corollary 3 *Let $G(t) = ct \cdot I + C$ be a linear matrix-valued polynomial with $c > 0$ and $\lambda_{\min}(C) = 0$. Then we have:*

$$G^{(r)}_{[0:n],\omega} \leq c \cdot \xi^n_{r+1},$$

where ξ^n_{r+1} is the least root of the degree $r + 1$ Krawtchouk polynomial.

Proof Let u be a unit eigenvector for C corresponding to (one of its) zero eigenvalues. Then for any univariate sum-of-squares polynomial $s \in \Sigma_r$, the matrix-valued polynomial $U(t) = s(t)uu^\top$ is a sum-of-squares polynomial matrix of degree $2r$. Furthermore, for such a U we have:

$$\int_{[0:n]} \mathrm{Tr}\big(G(t)U(t)\big)d\omega(t) = \int_{[0:n]} ct \cdot s(t)d\omega(t)$$

and

$$\int_{[0:n]} \mathrm{Tr}\big(U(t)\big)d\omega(t) = \int_{[0:n]} s(t)d\omega(t).$$

Therefore, writing $g(t) = ct$, and making use of Theorem 7 and Remark 1, we have:

$$G^{(r)}_{[0:n],\omega} \leq \inf_{s \in \Sigma_r} \left\{ \int_{[0:n]} ct \cdot s(t)d\omega(t) : \int_{[0:n]} s(t)d\omega(t) = 1 \right\} = g^{(r)}_{[0:n],\omega} = c \cdot \xi^n_{r+1}.$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

References

1. Alon, N., Naor, A.: Approximating the cut-norm via Grothendieck's inequality. In: 36th Annual ACM Symposium on Theory of Computing, pp. 72–80 (2004)
2. Arora, S., Berger, E., Hazan, E., Kindler, G., Safra, M.: On non-approximability for quadratic programs. In: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, pp. 206–215 (2005)
3. Balas, E., Ceria, S., Cornuéjols, G.: A lift-and-project cutting plane algorithm for mixed 0–1 programs. Math. Program. **58**, 295–324 (1993)
4. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Mach. Learn. **46**(1–3), 89–113 (2004)
5. Barak, B., Steurer, D.: Sum-of-squares proofs and the quest toward optimal algorithms. In Proceedings of International Congress of Mathematicians (ICM) (2014)
6. Charikar, M., Wirth, A.: Maximizing quadratic programs: Extending Grothendieck's inequality. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, pp. 54–60 (2004)
7. de Klerk, E., Laurent, M.: A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In: Araujo C., Benkart G., Praeger C., Tanbay B. (eds), World Women in Mathematics 2018. Association for Women in Mathematics Series, vol. 20, pp. 17–56. Springer, Cham (2019)

8. de Klerk, E., Laurent, M.: Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. Math. Program. (2020). https://doi.org/10.1007/s10107-019-01465-1
9. de Klerk, E., Laurent, M.: Worst-case examples for Lasserre's measure based hierarchy for polynomial optimization on the hypercube. Math. Oper. Res. **45**(1), 86–98 (2020)
10. Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere (2012). arXiv:1210.5048
11. Fang, K., Fawzi, H.: The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01537-7
12. Fawzi, H., Saunderson, J., Parrilo, P.A.: Sparse sums of squares on finite abelian groups and improved semidefinite lifts. Math. Program. **160**(1–2), 149–191 (2016)
13. Goemans, M., Williamson, D.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. Assoc. Comput. Mach. **42**(6), 1115–1145 (1995)
14. Karlin, A.R., Mathieu, C., Nguyen, C.T.: Integrality gaps of linear and semi-definite programming relaxations for knapsack. In: Günlük, O., Woeginger, G.J. (eds.) Integer Programming and Combinatoral Optimization, pp. 301–314. Springer, Berlin, Heidelberg (2011)
15. Kurpisz, A., Leppänen, S., Mastrolilli, M.: Tight Sum-of-Squares lower bounds for binary polynomial optimization problems. In: Chatzigiannakis, I., et al. (eds) 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016) vol. 78, pp. 1–14 (2016)
16. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
17. Lasserre, J.B.: A max-cut formulation of 0/1 programs. Oper. Res. Lett. **44**, 158–164 (2016)
18. Lasserre, J.B.: An explicit exact sdp relaxation for nonlinear 0–1 programs. In: Aardal, K., Gerards, B. (eds.) Integer Programming and Combinatorial Optimization, pp. 293–303, Springer, Berlin, Heidelberg (2001)
19. Lasserre, J.B.: Moments. Positive Polynomials and Their Applications. Imperial College Press, London (2009)
20. Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. SIAM J. Optim. **21**(3), 864–885 (2010)
21. Laurent, M.: A comparison of the Sherali–Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496 (2003)
22. Laurent, M.: Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. Math. Oper. Res. **28**(4), 871–883 (2003)
23. Laurent, M.: Semidefinite representations for finite varieties. Math. Program. **109**, 1–26 (2007)
24. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry, Vol. 149 of IMA Volumes in Mathematics and its Applications, pp. 157–270. Springer (2009)
25. Lee, J.R., Raghavendra, P., Steurer, D.: Lower bounds on the size of semidefinite programming relaxations. In: STOC '15: Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, pp. 567–576 (2015)
26. Levenshtein, V.I.: Universal bounds for codes and designs. In: Handbook of Coding Theory, vol. 9, pp. 499–648. North-Holland, Amsterdam (1998)
27. Lovász, L., Schrijver, A.: Cones of matrices and set-functions and 0–1 optimization. SIAM J. Optim. **1**, 166–190 (1991)
28. Macwilliams, F., Sloane, N.: The Theory of Error Correcting Codes, volume 16 of North-Holland Mathematical Library. Elsevier (1983)
29. Natanson, I.P.: Constructive Function Theory, Vol. I Uniform Approximation (1964)
30. Nie, J., Schweighofer, M.: On the complexity of Putinar's positivstellensatz. J. Complex. **23**(1), 135–150 (2007)
31. O'Donnell, R.: SOS is not obviously automatizable, even approximately. In: 8th Innovations in Theoretical Computer Science Conference vol. 59, pp. 1–10 (2017)
32. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D thesis, California Institure of Technology (2000)
33. Raghavendra, P., Weitz, B.: On the bit complexity of sum-of-squares proofs. In: 44th International Colloquium on Automata, Languages, and Programming, vol. 80, pp. 1–13 (2017)
34. Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Math. Z. **220**(1), 75–97 (1995)
35. Rothvoss, T.: The Lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (2013)

36. Sakaue, S., Takeda, A., Kim, S., Ito, N.: Exact semidefinite programming relaxations with truncated moment matrix for binary polynomial optimization problems. SIAM J. Optim. **27**(1), 565–582 (2017)
37. Scherer, C.W., Hol, C.W.J.: Matrix sum-of-squares relaxations for robust semi-definite programs. Math. Program. **107**, 189–211 (2006)
38. Schweighofer, M.: On the complexity of Schmüdgen's positivstellensatz. J. Complex. **20**(4), 529–543 (2004)
39. Sherali, H.D., Adams, W.P.: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. SIAM J. Discret. Math. **3**, 411–430 (1990)
40. Slot, L., Laurent, M.: Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01468-3
41. Szegö, G.: Orthogonal Polynomials. vol. 23 in American Mathematical Society colloquium publications. American Mathematical Society (1959)
42. Slot, L., Laurent, M.: Near-optimal analysis of of Lasserre's univariate measure-based bounds for multivariate polynomial optimization. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01586-y
43. Terras, A.: Fourier Analysis on Finite Groups and Applications. London Mathematical Society Student Texts, Cambridge University Press, Cambridge (1999)
44. Tunçel, L.: Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization. Fields Institute Monograph (2010)
45. Vallentin, F.: Semidefinite programs and harmonic analysis. arXiv:0809.2017 (2008)