arXiv:2011.04027v2 [math.OC] 3 Dec 2020

Sum-of-squares hierarchies for binary polynomial optimization

Lucas Slot[*] **Monique Laurent**[†]

December 4, 2020

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 $(\mathbb{Z}/q\mathbb{Z})^n$.

Keywords binary polynomial optimization · Lasserre hierarchy · sum-of-squares polynomials · Fourier analysis · Krawtchouk polynomials · polynomial kernels · semidefinite programming

AMS subject classification 90C09; 90C22; 90C26; 90C27; 90C30

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

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

In fact, 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 [12] 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., [1], [32], [22], [13], [36], [27]. In [16] it is shown that the sums-of-squares hierarchy of Lasserre [13] 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

[*]Centrum Wiskunde & Informatica (CWI), Amsterdam. lucas.slot@cwi.nl

[†]Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University. monique.laurent@cwi.nl

This work is supported by the European Union's Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No. 764759 (MINOA).

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 [20], and the links to the Unique Game Conjecture in [2]. For background about the sum-of-squares hierarchy applied to polynomial optimization over general semialgebraic sets we refer to [11], [25], [14], [18] 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 [11, 13] and Parrilo [25] 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:

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

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 problem (2) can be reformulated as a semidefinite program, one can compute $f_{(r)}$ efficiently (up to any precision) for fixed r (see [11, 25]). The bounds $f_{(r)}$ have finite convergence: $f_{(r)} = f_{\min}$ for $2r \geq n$ [13, 16]. In fact, it has been shown in [29] that the bound $f_{(r)}$ is exact already for $2r \geq n + d - 1$. That is,

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

In addition, it is shown in [29] 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 [8] 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) [17]. Similarly, the result (3) is tight when d is even and n is odd [10].

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 least roots of Krawtchouk polynomials (see Section 2).

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* (15) *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 [21] shows their asymptotic behaviour.

Theorem 2 ([21], 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 noise term in $O(1/n^{2/3})$, which vanishes as n tends to ∞. As an illustration, Figure 1 shows the function $\varphi(t)$.

1.2 A second hierarchy of bounds

In addition to the *lower* bound $f_{(r)}$, Lasserre [15] also defines an *upper* bound $f^{(r)} \geq f_{\min}$ on $f_{\min}$ as follows:

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

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 [15]. Furthermore, as shown in [15] 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 Section 5).

Essentially as a side result in the proof of our main Theorem 1, we can show the following analog of Theorem 1 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:*

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

where $C_d > 0$ is the constant of 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^n_{r+1} as was necessary for the analysis of $f_{(r)}$ in Theorem 1. Indeed, as will become clear later, the condition put on ξ^n_{r+1} follows from a technical argument (see Lemma 13), 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 4. *Fix $d \leq n$ and for $n, r \in \mathbb{N}$ write*

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

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:

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

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

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

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 4 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 [29].

Our results in Theorems 1 and 3 and Corollary 4 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 B, 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 [3]) 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 $\mathbb{B}^n$:

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

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 [8], positively answering a question posed in [17]. Extending the strategy of [8], the general case was settled in [29]. These exactness results are best possible when d is even and n is odd [10].

In [9], 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 $a_1x_1 + \ldots + a_nx_n \leq b$ 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 [24] 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 [30]. Specializing to the hypersphere S^{n-1}, better rates in $O(1/r)$ were shown in [26, 6], and recently improved to $O(1/r^2)$ in [7]. 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 [3] for an overview.

Unfortunately, it is hard to compare the asymptotic results above for general semi-algebraic sets with the known and our new results on the boolean cube. We indeed have to consider a different regime in the case of the boolean cube $\mathbb{B}^n$ since 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 [4], the unit ball and the simplex [33], and the unit box [5], 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 [33, 19]. 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 a broad overview of the main ideas that we use to show our results. Our broad strategy follows the one employed in [7] 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 it 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 $\mathbb{B}^n$, we find that $\tilde{f}$ is a sum-of-squares on $\mathbb{B}^n$ with degree at most $2r$, and thus that $f_{\min} - f_{(r)} \leq \delta$.

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 5. *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: $\|\mathbf{K}^{-1}\tilde{f} - \tilde{f}\|_\infty = \|\mathbf{K}^{-1}f - f\|_\infty \leq \delta\|f\|_\infty = \delta$. Therefore we obtain that $\mathbf{K}^{-1}\tilde{f}(x) \geq \tilde{f}(x) - \delta = f(x) \geq f_{\min} = 0$ 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 of $\mathbf{K}$ in terms of the polynomial u. More precisely, it turns out that the eigenvalues of $\mathbf{K}$ are given by the coefficients of the expansion of u^2 in the basis of *Krawtchouk polynomials*.

It remains then to find such 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.

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.

Organization

The rest of the paper is structured as follows. We review the necessary background on Fourier analysis on the boolean cube in Section 2. In Section 3, we recall a connection between the inner Lasserre hierarchy and the roots of orthogonal polynomials. Then, in Section 4, we give a proof of Theorem 1. Finally, in Section 5, we discuss how to generalize the proofs of Section 4 to obtain Theorem 3. We group the proofs of some technical results needed to prove Lemma 11 in Appendix A and, in Appendix B, we indicate how our arguments extend to the case of polynomial optimization over the q-ary hypercube $\{0,1,\ldots,q-1\}^n$ for $q > 2$.

2 Fourier analysis on the boolean cube

In this section, we cover some basic Fourier analysis on the boolean cube. For a general reference on Fourier analysis on (finite, abelian) groups, see e.g. [28, 35].

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*:

$$\chi_a(x) := (-1)^{x \cdot a} = \prod_{i : a_i = 1} (1 - 2x_i) \quad (a \in \mathbb{B}^n). \tag{10}$$

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

$$\langle \chi_a, \chi_b \rangle_\mu = \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} \chi_a(x)\chi_b(x) = \delta_{a,b} \quad \forall a, b \in \mathbb{B}^n. \tag{11}$$

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

$$p(x) = \sum_{a \in \mathbb{B}^n} \widehat{p}(a)\chi_a(x) \quad \forall x \in \mathbb{B}^n \tag{12}$$

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 permutations corresponding to bit-flips, of the form $x \in \mathbb{B}^n \mapsto x \oplus a \in \mathbb{B}^n$ for any $a \in \mathbb{B}^n$. 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 (12), 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

$$p = p_0 + p_1 + \cdots + p_d \quad (p_k \in H_k). \tag{13}$$

The polynomials $p_k \in H_k$ $(k = 0, \ldots, d)$ are known as the *harmonic components* of p and the decomposition (13) 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

$$X_k(x) := \sum_{|a|=k} \chi_a(x). \tag{14}$$

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:

$$\mathcal{K}_k^n(t) := \sum_{i=0}^{k} (-1)^i \binom{t}{i} \binom{n-t}{k-i} \tag{15}$$

(see, e.g. [34]). 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:

$$\langle \mathcal{K}_k^n, \mathcal{K}_{k'}^n \rangle_\omega := \frac{1}{2^n} \sum_{t=0}^{n} \mathcal{K}_k^n(t)\mathcal{K}_{k'}^n(t)w(t) = \delta_{k,k'} \binom{n}{k}. \tag{16}$$

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

Lemma 6. *Let $t \in [0:n]$ and choose $x, y \in \mathbb{B}^n$ so that $d(x, y) = t$. Then for any $0 \leq k \leq n$ we have:*

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(x)\chi_a(y). \tag{17}$$

In particular, we have:

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(x^t) = X_k(x^t), \tag{18}$$

where $x^t \in \mathbb{B}^n$ is given by $x_i^t = 1$ if $1 \leq i \leq t$ and $x_i^t = 0$ if $t + 1 \leq i \leq n$.

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

$$\sum_{|a|=k} \chi_a(x)\chi_a(y) = \sum_{i=0}^{k} (-1)^i \cdot \#\{|a| = k : a \cdot (x + y) = i\} = \sum_{i=0}^{k} (-1)^i \binom{t}{i}\binom{n-t}{k-i} = \mathcal{K}_k^n(t).$$

□

From this, we see that any polynomial $p \in \mathbb{R}[x]_d$ that is invariant under the action of $\mathrm{St}(0)$ 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:

$$\widehat{\mathcal{K}}_k^n(t) := \mathcal{K}_k^n(t)/\mathcal{K}_k^n(0) \quad (k \in \mathbb{N}). \tag{19}$$

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 7. *We have:*

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

for all $0 \leq k \leq n$ 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 x^t from Lemma 6. By (18) we have

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

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

Lemma 8. *We have:*

$$\begin{aligned} |\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t+1)| &\leq \frac{2k}{n}, \qquad (t = 0, 1, \ldots, n-1) \\ |\widehat{\mathcal{K}}_k^n(t) - 1| &\leq \frac{2k}{n} \cdot t \qquad (t = 0, 1, \ldots, n) \end{aligned} \tag{20}$$

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

Proof. Let $t \in [0 : n-1]$ and $0 < k \leq d$. Consider the elements $x^t, x^{t+1} \in \mathbb{B}^n$ from Lemma 6. Then we have:

$$\begin{aligned}|\mathcal{K}_k^n(t) - \mathcal{K}_k^n(t+1)| &\overset{(18)}{=} |\sum_{|a|=k} \chi_a(x^t) - \chi_a(x^{t+1})| \\ &\leq 2 \cdot \#\big\{a \in \mathbb{B}^n : |a| = k, a_{t+1} = 1\big\} = 2\binom{n-1}{k-1},\end{aligned}$$

where for the inequality we note that $\chi_a(x^t) = \chi_a(x^{t+1})$ 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 identity of (20). The second identity now follows from this using a summation argument and $\widehat{\mathcal{K}}_k^n(0) = 1$. □

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 associated kernel $K : \mathbb{B}^n \times \mathbb{B}^n \to \mathbb{R}$ defined by

$$K(x, y) := u^2(d(x, y)). \tag{21}$$

A kernel of the form (21) coincides with a polynomial on $\mathbb{B}^n$, 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 $\mathrm{Aut}(\mathbb{B}^n)$, in the sense that:

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

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

$$\mathbf{K}p(x) := \int_{\mathbb{B}^n} p(y)K(x, y)d\mu(y) = \frac{1}{2^n}\sum_{y \in \mathbb{B}^n} p(y)K(x, y). \tag{22}$$

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

$$u^2(t) = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(t) \quad (\lambda_i \in \mathbb{R}). \tag{23}$$

As we show now, the eigenvalues of the operator $\mathbf{K}$ are given precisely by the coefficients λ_i occurring in this expansion.

Theorem 9 (Funk-Hecke)**.** *Let $p \in \mathcal{R}[x]_d$ with harmonic decomposition $p = p_0 + p_1 + \cdots + p_d$. Then we have:*

$$\mathbf{K}p = \lambda_0 p_0 + \lambda_1 p_1 + \cdots + \lambda_d p_d. \tag{24}$$

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 $x \in \mathbb{B}^n$:

$$\begin{aligned}\mathbf{K}\chi_z(x) = \frac{1}{2^n}\sum_{y \in \mathbb{B}^n} \chi_z(y)u^2(d(x, y)) &\overset{(23)}{=} \frac{1}{2^n}\sum_{y \in \mathbb{B}^n} \chi_z(y)\sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n(d(x, y)) \\ &= \frac{1}{2^n}\sum_{i=0}^{2r} \lambda_i \sum_{y \in \mathbb{B}^n} \chi_z(y)\mathcal{K}_i^n(d(x, y)) \\ &\overset{(17)}{=} \sum_{i=0}^{2r} \lambda_i \sum_{y \in \mathbb{B}^n} \chi_z(y) \sum_{|a|=i} \chi_a(x)\chi_a(y) \\ &= \sum_{i=0}^{2r} \lambda_i \sum_{|a|=i} \Big(\sum_{y \in \mathbb{B}^n} \chi_z(y)\chi_a(y)\Big)\chi_a(x) \\ &\overset{(11)}{=} \frac{1}{2^n}\sum_{i=0}^{2r} \lambda_i \sum_{|a|=i} 2^n \delta_{z,a}\chi_a(x) \\ &= \lambda_{|z|}\chi_z(x).\end{aligned}$$

□

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 (23) of u^2 in the following way:

$$\lambda_i = \langle \mathcal{K}_i^n, u^2 \rangle_\omega \,/\, \|\mathcal{K}_i^n\|_\omega^2 = \langle \widehat{\mathcal{K}}_i^n, u^2 \rangle_\omega. \tag{25}$$

In addition, since in view of Lemma 7 we have $\widehat{\mathcal{K}}_i^n(t) \leq \widehat{\mathcal{K}}_0^n(t)$ for all $t \in [0:n]$, it folllows that

$$\lambda_i \leq \lambda_0 \quad \text{ for } 0 \leq i \leq 2r. \tag{26}$$

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:

$$g^{(r)} := g_{M,\nu}^{(r)} = \inf_{s \in \Sigma_r} \left\{ \int_M g \cdot s d\nu : \int_M s d\nu = 1 \right\} \tag{27}$$

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 [5] and further employed for this purpose, e.g., in [4, 33].

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 10 ([5])**.** *Let $M \subseteq \mathbb{R}$ be an interval and let ν be a measure supported on M with corresponding orthogonal polonomials $p_i \in \mathbb{R}[t]_i$ ($i \in \mathbb{N}$). Then the Lasserre inner bound $g^{(r)}$ (from (27)) of order r for the polynomial $g(t) = t$ can be reformulated as:*

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

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

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

$$g^{(r)} - g_{\min} \leq c \cdot \xi_{r+1}, \tag{28}$$

where ξ_{r+1} is the smallest root of the relevant 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 (21) and the corresponding linear operator $\mathbf{K}$ from (22). 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 (23) 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 Section 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).

Expressing $\|\mathbf{K}^{-1} - I\|$ 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 Appendix A.

Lemma 11. *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 12. *Assume that $\lambda_0 = 1$ and $\lambda_k \neq 0$ for $1 \leq k \leq d$. Then we have:*

$$\|\mathbf{K}^{-1} - I\| \leq \gamma_d \cdot \Lambda, \text{ where } \Lambda := \sum_{i=1}^{d} |\lambda_i^{-1} - 1|. \tag{29}$$

Proof. By assumption, the operator $\mathbf{K}$ is invertible and, in view of Funck-Hecke relation (24), its inverse is given by $\mathbf{K}^{-1}p = \sum_{i=0}^{d} \lambda_i^{-1} p_k$ for any $p = p_0 + p_1 + \cdots + p_d \in \mathcal{R}[x]_d$. Then we have:

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

where we use Lemma 11 for the last inequality. □

The expression Λ in (29) is difficult to analyze. Therefore, following [7], 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 (26)). Thus, Λ and $\tilde{\Lambda}$ are both minimized when the λ_i are close to 1. The following lemma makes this more precise.

Lemma 13. *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:

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

□

Optimizing the choice of the univariate polynomial u. In light of Lemma 13, and recalling (25), 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:

$$\inf_{u \in \mathbb{R}[t]_r} \left\{ \int g \cdot u^2 d\omega : \int u^2 d\omega = 1 \right\}, \text{ where } g(t) := d - \sum_{i=1}^{d} \widehat{\mathcal{K}}_i^n(t). \tag{30}$$

(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 (27) defining the inner Lasserre bound[3] 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 (30) is equal to $g^{(r)}$ and, using Lemma 13, we may conclude the following.

[3]Technically, the program (27) allows for the density to be a *sum of squares*, whereas the program (30) 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 (27) can in fact always be chosen to be a square [15].

Theorem 14. *Let g be as in* (30)*. 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 (30), and γ_d is the constant of Lemma 11.

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

Lemma 15. *We have:*

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

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

Proof. Using (20), we find for each $k \leq n$ that:

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

Therefore, we have:

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

□

Lemma 16. *We have:*

$$g^{(r)} - g_{\min} \leq d(d+1) \cdot (\xi^n_{r+1}/n), \tag{31}$$

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

Proof. This follows immediately from Lemma 15 and our observation (28) at the end of Section 3, noting that the Krawtchouk polynomials are indeed orthogonal w.r.t. the measure ω on $[0 : n]$ (cf. (16)). □

Putting things together, we thus prove our main result.

Theorem 17 (Restatement of Theorem 1)**.** *Fix $d \leq n$ and let $f \in \mathbb{R}[x]$ be a polynomial of degree d. Then we have:*

$$\frac{f_{\min} - f_{(r)}}{\|f\|_\infty} \leq 2\gamma_d \cdot d(d+1) \cdot (\xi^n_{r+1}/n),$$

whenever $d(d+1) \cdot (\xi^n_{r+1}/n) \leq 1/2$. Here, ξ^n_{r+1} is the smallest root of $\mathcal{K}^n_{r+1}$ and γ_d is the constant of Lemma 11.

Proof. Combining Theorem 14 and Lemma 16 we obtain $\|\mathbf{K}^{-1} - I\| \leq 2\gamma_d \cdot d(d+1) \cdot (\xi^n_{r+1}/n)$. As $\mathbf{K}(1) = 1$ we can use Lemma 5 to conclude the proof. We obtain Theorem 1 with $C_d := \gamma_d \cdot d(d+1)$. □

Note that the condition $d(d+1) \cdot \xi^n_{r+1} \leq 1/2$ follows from relation (31) (which itself refers back to condition $\tilde{\Lambda} \leq 1/2$ in Lemma 13).

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:

$$f^{(r)} := \min_{s \in \Sigma[x]_r} \left\{ \int_{\mathbb{B}^n} f(x) \cdot s(x) d\mu : \int_{\mathbb{B}^n} s(x) d\mu = 1 \right\} \geq f_{\min}. \tag{32}$$

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 (32) 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:

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

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

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

In other words, the behaviour of the *symmetrized* inner hierarchy $f_{\mathrm{sym}}^{(r)}$ 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 Section 3. First we find a linear upper estimator $\widehat{F}$ for F on $[0:n]$.

Lemma 18. *We have*

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

where γ_d is the same constant as in Lemma 11.

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 15:

$$F(t) = \sum_{i=0}^d \lambda_i (\widehat{\mathcal{K}}_i^n(t) - 1) \leq \sum_{i=0}^d |\lambda_i| |\widehat{\mathcal{K}}_i^n(t) - 1| \overset{(20)}{\leq} \max_i |\lambda_i| \cdot t \cdot \sum_{i=0}^d \frac{2i}{n} \leq \max_i |\lambda_i| \cdot t \cdot \frac{d(d+1)}{n}.$$

As $\|f\|_\infty = 1$, using Lemma 11, we can conclude that $|\lambda_i| = \max_{t \in [0:n]} |\lambda_i \widehat{\mathcal{K}}_i^n(t)| \leq \gamma_d$, which gives the desired result. □

We may thus conclude the following analysis for the inner bounds $f^{(r)}$; in comparison to the result in Lemma 16 we only have the additional constant factor γ_d.

Theorem 19 (Restatement of Theorem 3)**.** *Let f be a polynomial of degree d. Then we have:*

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

where ξ_{r+1}^n is the smallest root of $\mathcal{K}_{r+1}^n$ and and γ_d is the constant of Lemma 11.

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 $r \geq (n+d-1)/2$, the inner hierarchy is exact in general if and only if $r \geq n$. We give a short proof of this fact below, for reference.

Lemma 20. *Let f be a polynomial on $\mathbb{B}^n$. Then $f^{(r)} = f_{\min}$ for all $r \geq n$.*

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 21. *Let $f(x) = |x| = x_1 + \ldots + x_n$. 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 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).

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 13, showing instead that $\Lambda \leq \tilde{\Lambda}/(1 - \tilde{\Lambda})$, assuming now only that $\tilde{\Lambda} \leq 1$. Indeed, Lemma 13 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 γ_d appearing in Theorem 1 and Theorem 3, where we may set $C_d = d(d + 1)\gamma_d$. In Appendix A we show the existence of this constant γ_d, but the dependence on d we show there is quite bad. This dependence, however, seems to be mostly an artifact of our proof. As we explain in Appendix A, it is possible to compute explicit upper bounds on γ_d for small values of d. Table 1 lists some of these upper bounds, which appear much more reasonable than our theoretical guarantee would suggest.

d	1	2	3	4	5	6	7	8	9	10	11	12
γ_d	1.00	1.00	4.00	8.00	20.0	48.1	112	258	578	1306	2992	6377

Table 1: Upper bounds on γ_d. Values rounded to indicated precision.

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 explicitely for small values of r, n, thus allowing for a concrete estimate of the error of both Lasserre hierarchies via Theorem 1 and Theorem 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., [34] 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 explicitely mentioned there, this is the case for the analysis on S^{n-1} in [7] 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 [7] and for the boolean cube $\mathbb{B}^n$ here might be extended to these cases as well.

Although B^n and $[-1,1]^n$ have similar symmetric structure to S^{n-1} and $\mathbb{B}^n$, respectively, the accompanying Fourier analysis is significantly more complicated. In particular, a direct analog of the Funk-Hecke formula (24) is not available. New ideas are therefore needed to define the kernel $K(x,y)$ (cf. (8)) and analyze its eigenvalues.

The q-ary cube. The techniques we use on the *binary* cube $\mathbb{B}^n$ can be generalized naturally to the q-ary cube $(\mathbb{Z}/q\mathbb{Z})^n = \{0,1,\ldots,q-1\}^n$ for $q > 2$. In particular, the required Fourier analysis and connection to Krawtchouk polynomials carry over almost directly. As a result we are able to show close analogs of our results on $\mathbb{B}^n$ in this more general setting as well. We present this generalization in Appendix B; see, in particular, Theorem 36.

Acknowledgments

We wish to thank Sven Polak and Pepijn Roos Hoefgeest for several useful discussions.

A Proof of Lemma 11

In this section we give a proof of Lemma 11, 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 22. For $n \geq d \geq k \geq 0$ integers, we write:

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

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

$$\gamma_d \geq \rho(n,d) \text{ for all } n \in \mathbb{N}. \tag{33}$$

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 23. *We have*

$$\rho(n,d,k) = \sup_{p\in\mathcal{R}[x]_d^n} \{p_k(0) : \|p\|_\infty \leq 1\} \tag{34}$$

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 $\sigma \in \mathrm{Aut}(\mathbb{B}^n)$ 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 $\mathrm{Aut}(\mathbb{B}^n)$, 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 24. *In the program* (34) *we may restrict the optimization to polynomials of the form*

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

Proof. Let p be a feasible solution to (34). Consider the following polynomial $\widehat{p}$ obtained as symmetrization of p under action of $\mathrm{St}(0)$, the set of automorphism of $\mathbb{B}^n$ corresponding to the coordinate permutations:

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

Then $\|\widehat{p}\|_\infty \leq 1$ and $\widehat{p}_k(0) = p_k(0)$, so $\widehat{p}$ is still feasible for (34) and has the same objective value as p. Furthermore, for each i, $\widehat{p}_i$ is invariant under $\mathrm{St}(0)$, 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 (14)). □

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 25. *For any $n \geq d \geq k$ we have:*

$$\begin{aligned} \rho(n,d,k) = \max \quad & \lambda_k \\ s.t. \quad & -1 \leq \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^n(t) \leq 1 \quad (t = 0,1,\ldots,n). \end{aligned} \tag{P}$$

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 (P) 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 27 below, is in fact a polynomial. We first present the polynomial $\widehat{\mathcal{K}}_k^\infty(t)$ for small k as an illustration.

Example 26. We have:

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

Lemma 27. *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., [23]):

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

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

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

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 (P) remain feasible after increasing the dimension n.

Lemma 28. *Let $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_d)$ be a feasible solution to* (P) *for a certain $n \in \mathbb{N}$. Then it is also feasible to* (P) *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 (14):

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

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:

$$\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \frac{X_i^n}{\binom{n}{i}} \circ \sigma = \frac{X_i^{n+1}}{\binom{n+1}{i}}. \tag{35}$$

To see that (35) 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 (35) holds.

Assume λ is a feasible solution of (P) for a given value of n. Then, in view of (17), this means

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

Using (35) we obtain:

$$\begin{aligned} \Big| \sum_{i=0}^{d} \lambda_i \frac{X_i^{n+1}(x)}{\binom{n+1}{i}} \Big| &= \Big| \sum_{i=0}^{d} \lambda_i \cdot \frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \frac{X_i^n(\sigma(x))}{\binom{n}{i}} \Big| \\ &= \Big| \Big(\frac{1}{|\mathrm{St}(0)|} \sum_{\sigma \in \mathrm{St}(0)} \big(\sum_{i=0}^{d} \lambda_i \frac{X_i^n}{\binom{n}{i}} \big) \circ \sigma \Big)(x) \Big| \leq 1 \end{aligned}$$

for all $x \in \mathbb{B}^{n+1}$. Using (17) again, this shows that λ is a feasible solution of program (P) for $n+1$. □

Example 29. To illustrate the identity (35), 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 30. *For $d \geq k \in \mathbb{N}$, define the program:*

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

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 (P) for (n, d, k). We show that λ is feasible for (PL). 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 28 that λ is also a feasible solution of program (P) 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:

$$|\sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(t)| = \lim_{j \to \infty} |\sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{n_j}(n_j t)| \leq 1. \tag{36}$$

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 (36) holds for all $t \in [0,1]$. This shows that λ is feasible for (PL) and we thus have $\rho(n, d, k) \leq \rho(\infty, d, k)$, as desired. □

We can now immediately conclude the existence of a constant γ_d satisfying $\gamma_d \geq \rho(\infty, d, k)$ for all $0 \leq k \leq d$, which, combined with Lemma 30, thus gives the desired inequality $\gamma_d \geq \rho(n, d)$ as in (33). Indeed, for a polynomial $p \in \mathbb{R}[t]_d$, written as $p = \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(t)$, setting $\|p\| := \sum_{i=0}^{d} |\lambda_i|$ defines a norm on $\mathbb{R}[t]_d$. Since all norms on the $(d+1)$-dimensional space $\mathbb{R}[t]_d$ are equivalent there exists a constant γ_d such that $\|p\| \leq \gamma_d \|p\|_{\infty}$, which implies $\rho(\infty, d, k) \leq \gamma_d$.

Concrete estimations of γ_d. In what follows we present another argument, based on the Markov Brothers' inequality, which permits to give a more concrete estimation on the constant γ_d.

Proposition 31 (Markov Brothers' inequality [31])**.** *There exists a constant m_d, depending only on d, such that, for any univariate polynomial $p \in \mathbb{R}[t]_d$ with degree d, we have:*

$$\max_{0 \leq t \leq 1} |p^{(k)}(t)| \leq m_d \max_{0 \leq t \leq 1} |p(t)| \quad \textit{for all } k \leq d.$$

Furthermore, $m_d \leq 2^d d^{2d}$.

We now use the Markov Brothers' inequality to bound the parameter $\rho(\infty, d, d)$.

Lemma 32. *Let $d \in \mathbb{N}$, then we may bound:*

$$\rho(n, d, d) \leq \rho(\infty, d, d) \leq \frac{m_d}{2^d d!}$$

independent of n. Here m_d is the constant of Lemma 31.

Proof. In view of Lemma 30 we only need to show $\rho(\infty, d, d) \leq \frac{m_d}{2^d d!}$. Let λ be a feasible solution to (PL), meaning that $\max_{0 \leq t \leq 1} |\sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(t)| \leq 1$. Then, using Proposition 31, we obtain that

$$\Big| \frac{\partial^d}{\partial t^d} \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(0) \Big| \leq m_d.$$

As $\widehat{\mathcal{K}}_i^{\infty}$ has degree i, we have:

$$\frac{\partial^d}{\partial t^d} \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^{\infty}(0) = \lambda_d \frac{\partial^d}{\partial t^d} \widehat{\mathcal{K}}_d^{\infty}(0) = \lambda_d \cdot d! \cdot (-2)^d.$$

Here, for the last identity, we have used the fact that the leading coefficient of the polynomial $\widehat{\mathcal{K}}_d^{\infty}$ is $(-2)^d$ (recall Lemma 27). We can conclude that $\lambda_d \leq m_d/(2^d d!)$. □

We now show how to extend the bound from Lemma 32 on $\rho(n, d, d)$ to $\rho(n, d, k)$ for arbitrary $k \leq d$. This can be done in an iterative way as the proof of the next lemma shows. In principle, an explicit value for the constant γ_d could be extracted from the proof. We do not carry out the details, however, as it appears that this leads to a rather unwieldy expression.

Lemma 33. *For each $d, k \in \mathbb{N}$ fixed, there exists a constant γ_d, depending only on d, such that*

$$\rho(n, d, k) \leq \gamma_d \quad \textit{for all } n \in \mathbb{N}.$$

Proof. Let $p = p_0 + p_1 + \cdots + p_d \in \mathcal{R}[x]_d$ be an optimal solution to $\rho(n, d, d-1)$, i.e., $\|p\|_{\infty} \leq 1$ and $\|p_{d-1}\|_{\infty} = \rho(n, d, d-1)$. Then the polynomial $\tilde{p} = p - p_d$ has degree $d-1$ and

$$\frac{\|\tilde{p}\|_{\infty}}{1 + \rho(n, d, d)} \leq 1.$$

Therefore, we have

$$\rho(n, d-1, d-1) \geq \frac{\rho(n, d, d-1)}{1 + \rho(n, d, d)}.$$

In other words, we have

$$\rho(n, d, d-1) \leq (1 + \rho(n, d, d)) \rho(n, d-1, d-1). \tag{37}$$

By Lemma 32, this permits to bound $\rho(n, d, d-1)$ by a constant depending only on d, independent of n.

In the same way let us now consider an optimal solution $p \in \mathcal{R}[x]_d$ for $\rho(n, d, d-2)$, i.e., $\|p\|_\infty \leq 1$ and $\|p_{d-2}\|_\infty = \rho(n, d, d-2)$. Then the polynomial $\tilde{p} = p - p_d - p_{d-1}$ has degree $d-2$ and satisfies

$$\frac{\|\tilde{p}\|_\infty}{1 + \rho(n, d, d) + \rho(n, d, d-1)} \leq 1.$$

Therefore, we have

$$\rho(n, d-2, d-2) \geq \frac{\rho(n, d, d-2)}{1 + \rho(n, d, d) + \rho(n, d, d-1)},$$

giving

$$\rho(n, d, d-2) \leq (1 + \rho(n, d, d) + \rho(n, d, d-1))\rho(n, d-2, d-2).$$

By Lemma 32) and relation (37) this permits to bound $\rho(n, d, d-2)$ by a constant independent of n.

We may proceed inductively to bound $\rho(n, d, k)$ for any $k \leq d$ by a constant dependent only on d. □

Upper bounds on γ_d for small values of d. Finally, we show how to compute explicit upper bounds on γ_d for any fixed value of d which likely outperform the theoretical guarantee derived above.

For this consider again the program (PL). As it has infinitely many linear constraints, it is not possible to solve it using conventional methods. However, by considering only a finite subset of these constraints, indexed by some finite set $T \subseteq [0, 1]$, we obtain a tractable relaxation which provides an upper bound on $\rho(\infty, d, k)$:

$$\begin{aligned} \rho(\infty, d, k) \leq \max \quad & \lambda_k \\ s.t. \quad & -1 \leq \sum_{i=0}^{d} \lambda_i \widehat{\mathcal{K}}_i^\infty(t) \leq 1 \quad (t \in T \subseteq [0, 1]). \end{aligned} \tag{PLR}$$

Hence the optimum value of (PLR) provides an upper bound on $\rho(n, d)$ which depends only on d (and not on n) and thus can serve as upper bound on the constant γ_d. We selected the set $T = \{i/100 : 0 \leq i \leq 100\}$ and solved the linear program (PLR) for small $d = 1, \ldots, 12$. In this way we obtained the bounds for the constant γ_d shown in Table 1 above.

B 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 $\mathbb{Z}/qZ$ denotes the cyclic group of order q, so that $(\mathbb{Z}/q\mathbb{Z})^n = \mathbb{B}^n$ when $q = 2$.

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 36 below, which can be seen as an analog of Corollary 4 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 (10) 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:

$$X_i = \sum_{|a|=i} \chi_a \in H_i. \tag{38}$$

The Krawtchouk polynomials introduced in Section 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 (16), the Krawtchouk polynomials $\mathcal{K}_k^n$ $(0 \le k \le n)$ are pairwise orthogonal w.r.t. the discrete measure ω on $[0:n]$ given by:

$$\omega(t) = \frac{1}{q^n} \sum_{t=0}^{n} w(t)\delta_t, \text{ with } w(t) := (q-1)^t \binom{n}{t}. \tag{39}$$

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. (19)).

We have the following connection (cf. (18)) between the characters and the Krawtchouk polynomials:

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a|=k} \chi_a(x) = \mathcal{K}_k^n(i) \quad \text{for } x \in (\mathbb{Z}/q\mathbb{Z})^n \text{ with } |x| = i. \tag{40}$$

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. (17)):

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

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 9, and we obtain the Funk-Hecke formula (recall (24)): for any polynomial $p \in \mathcal{R}[x]_{(q-1)d}$ with Harmonic decomposition $p = p_0 + \ldots + p_d$,

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

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. (25)), 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 (39)). We show first that this function g again has a nice linear upper estimator.

Lemma 34. *We have:*

$$\begin{aligned} |\widehat{\mathcal{K}}_k^n(t) - \widehat{\mathcal{K}}_k^n(t+1)| &\le \frac{2k}{n}, \qquad (t = 0,1,\ldots,n-1) \\ |\widehat{\mathcal{K}}_k^n(t) - 1| &\le \frac{2k}{n}\cdot t \qquad (t = 0,1,\ldots,n) \end{aligned} \tag{41}$$

for all $k \le n$.

Proof. The proof is almost identical to that of Lemma 8. Let $t \in [0:n-1]$ and $0 < k \le d$. Consider the elements $x^t, x^{t+1} \in (\mathbb{Z}/q\mathbb{Z})^n$ from Lemma 6. Then we have:

$$\begin{aligned} |\mathcal{K}_k^n(t) - \mathcal{K}_k^n(t+1)| &\overset{(40)}{=} |\sum_{|a|=k} \chi_a(x^t) - \chi_a(x^{t+1})| \\ &\le 2\cdot \#\big\{a \in (\mathbb{Z}/q\mathbb{Z})^n : |a| = k, a_{t+1} \neq 0\big\} = 2\cdot(q-1)^k \cdot \binom{n-1}{k-1}, \end{aligned}$$

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

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

This shows the first identity of (41) and the second identity follows using a summation argument and $\widehat{\mathcal{K}}_k^n(0) = 1$. □

From Lemma 34 we obtain that the function $g(t) = d - \sum_{i=0}^{d} \widehat{\mathcal{K}}_i^n(t)$ admits the following linear upper estimator: $g(t) \le d(d+1)\cdot(t/n)$ for $t \in [0:n]$. Now the same arguments as used for the case $q = 2$ enable us to conclude:

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

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

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

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 we consider the order $(q-1)r$ of the outer Lasserre hierarchy, which corresponds to the degree $2(q-1)r$ of the kernel $K(x,y) = u^2(d(x,y))$ in each variable x and y. 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 35 ([21], Section 5)**.** *Fix $t \in [0, \frac{q-1}{q}]$. Then the smallest roots $\xi^n_{r,q}$ of the q-ary Krawtchouk polynomials $\mathcal{K}^n_{r,q}$ satisfy:*

$$\lim_{r/n \to t} \xi^n_{r,q}/n = \varphi_q(t) := \frac{q-1}{q} - \left(\frac{q-2}{q} \cdot t + \frac{2}{q} \sqrt{(q-1)t(1-t)} \right). \tag{42}$$

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 $\lim_{j\to\infty} \xi^{n_j}_{r_j,q}/n_j = \varphi_q(t)$.

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 35 the asymptotic analog of Theorem 2 (and not the exact bound on the root $\xi^n_{r,q}$ for any n). Therefore we have shown the following q-analog of Corollary 4.

Theorem 36. *Fix $d \le n$ and for $n, r \in \mathbb{N}$ write*

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

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

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

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

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

Here $\varphi_q(t)$ is the function defined in (42). 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 Figure 1.

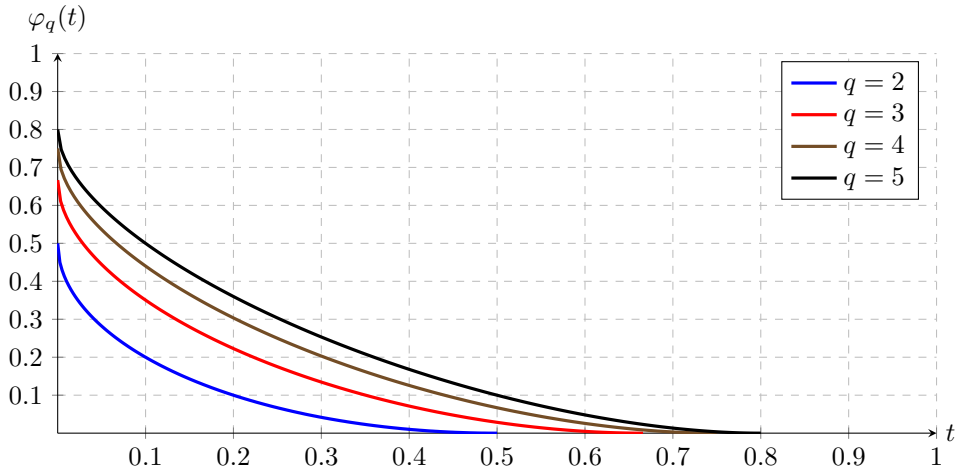

Figure 1: The function $\varphi_q(t)$ for several values of q. Note that the case $q = 2$ corresponds to the function $\varphi(t)$ of (5).

A generalization of Lemma 11. The arguments above omit a generalization of Lemma 11, 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:

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

Then, as in the binary case, we may set $C_d = d(d+1)\gamma_d$. The proof given in Appendix A 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:

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

Lemmas 23 and 24, 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[4]:

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

where $X_i = \sum_{|a|=i} \chi_a \in H_i$ is the zonal spherical function of degree $(q-1)i$ (cf. (38) and (14)). Using (40), we obtain a reformulation of $\rho(n,d,k)$ as an LP (cf. (P)):

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

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. (PL)):

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

As before, we have $\rho(n,d,k) \leq \rho(\infty,d,k)$, noting that (the proofs of) Lemma 28 and Lemma 30 may be applied directly to the case $q > 2$. From there, it suffices to show $\rho(\infty,d,k) < \infty$, which can also be argued exactly as in the case $q = 2$.

References

[1] E. Balas, S. Ceria and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0–1 programs. *Mathematical Programming*, 58:295–324, 1993.

[2] B. Barak and D. Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians (ICM)*, 2014.

[3] E. de Klerk and M. Laurent. *A Survey of Semidefinite Programming Approaches to the Generalized Problem of Moments and Their Error Analysis*, pp. 17–56. Springer International Publishing, 2019.

[4] E. de Klerk and M. Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. *Mathematical Programming*, 2020. https://doi.org/10.1007/s10107-019-01465-1

[5] E. de Klerk and M. Laurent. Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube. *Mathematics of Operations Research*, 45(1):86–98, 2020.

[6] A. C. Doherty and S. Wehner. Convergence of SDP hierarchies for polynomial optimization on the hypersphere. *arXiv preprint*, 2012.

[7] K. Fang and H. Fawzi. The sum-of-squares hierarchy on the sphere, and applications in quantum information theory. *Mathematical Programming*, 2020. https://doi.org/10.1007/s10107-020-01537-7

[8] H. Fawzi, J. Saunderson, and P. A. Parrilo. Sparse sums of squares on finite abelian groups and improved semidefinite lifts. *Mathematical Programming*, 160(1-2):149–191, 2016.

[9] A. R. Karlin, C. Mathieu, and C. T. Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. O. Günlük and Gerhard J. Woeginger (eds.) *Integer Programming and Combinatoral Optimization*, pp. 301–314, Springer Berlin, Heidelberg, 2011.

[10] A. Kurpisz, S. Leppänen, M. Mastrolilli. Tight Sum-of-Squares lower bounds for binary polynomial optimization problems. I. Chatzigiannakis et al. (eds.) *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, 78:1–14, 2016.

[11] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[12] J.B. Lasserre. A max-cut formulation of $0/1$ programs. *Operations Research Letters*, 44:158–164, 2016.

[4]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 $\langle p, \chi_a\rangle_\mu = \lambda_{|a|}\|\chi_a\|^2 = \lambda_{|a|}\|\chi_a^{-1}\|^2 = \langle p, \overline{\chi_a}\rangle_\mu = \overline{\langle p, \chi_a\rangle_\mu}$.

[13] J.B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. K. Aardal and B. Gerards (eds.) *Integer Programming and Combinatorial Optimization*, pp. 293–303, Springer Berlin, Heidelberg, 2001.

[14] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications.* Imperial College Press, London (2009)

[15] J.B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization*, 21(3):864–885, 2010.

[16] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28(3):470–496.

[17] M. Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of Operations Research*, 28(4):871–883, 2003.

[18] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging Applications of Algebraic Geometry,* Vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pages 157-270, 2009.

[19] M. Laurent and L. Slot. Near-optimal analysis of of Lasserre's univariate measure-based bounds for multivariate polynomial optimization. *Mathematical Programming*, 2020. https://doi.org/10.1007/s10107-020-01586-y

[20] J.R. Lee, P. Raghavendra and D. Steurer. Lower Bounds on the Size of Semidefinite Programming Relaxations. In STOC '15: Proceedings of the forty-seventh annual ACM symposium on Theory of Computing, pages 567–576, 2015.

[21] V. I. Levenshtein. Universal bounds for codes and designs. *Handbook of Coding Theory*, vol. 9, pp. 499–648, North-Holland, Amsterdam, 1998.

[22] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0–1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[23] F. Macwilliams and N. Sloane. The Theory of Error Correcting Codes, vol. 16 of *North-Holland Mathematical Library*, Elsevier, 1983.

[24] J. Nie and M. Schweighofer. On the complexity of Putinar's positivstellensatz. *Journal of Complexity*, 23(1):135–150, 2007.

[25] P. A. Parrilo. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, 2000. PhD thesis, California Institure of Technology.

[26] B. Reznick. Uniform denominators in Hilbert's seventeenth problem. *Mathematische Zeitschrift*, 220(1):75–97, 1995.

[27] T. Rothvoss. The Lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial, 2013.

[28] W. Rudin. Fourier Analysis on Groups. John Wiley & Sons, Ltd, 1990.

[29] S. Sakaue, A. Takeda, S. Kim, and N. Ito. Exact semidefinite programming relaxations with truncated moment matrix for binary polynomial optimization problems. *SIAM Journal on Optimization*, 27(1):565–582, 2017.

[30] M. Schweighofer. On the complexity of Schmüdgen's positivstellensatz. *Journal of Complexity*, 20(4):529–543, 2004.

[31] A. Shadrin. Twelve proofs of the Markov inequality. M. Drinov (ed.) *Approximation Theory: A Volume Dedicated to Borislav Bojanov*, pp. 233–298, Sofia, 2005.

[32] H.D. Sherali and W.P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

[33] L. Slot and M. Laurent. Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets. *Mathematical Programming*, 2020. https://doi.org/10.1007/s10107-020-01468-3

[34] G. Szegö. Orthogonal Polynomials. vol. 23 in *American Mathematical Society colloquium publications*. American Mathematical Society, 1959.

[35] A. Terras. Fourier Analysis on Finite Groups and Applications. *London Mathematical Society Student Texts*. Cambridge University Press, 1999.

[36] L. Tunçel. Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization, *Fields Institute Monograph*, 2010.