Sum-of-Squares Hierarchies for Binary Polynomial Optimization

Lucas Slot[1( )] and Monique Laurent[1,2]

[1] Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
{lucas,monique}@cwi.nl
[2] Tilburg University, Tilburg, The Netherlands

Abstract. We consider the sum-of-squares hierarchy of approximations for the problem of minimizing a polynomial f over the boolean hypercube $^n = \{0,1\}^n$. This hierarchy provides for each integer r $\mathbb{N}$ a lower bound $f_{r)}$ on the minimum $f_{\min}$ of f, given by the largest scalar for which the polynomial $f -$ is a sum-of-squares on ^{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 $[0, 1/2]$, we can show that this worst-case error in the regime $r \approx t \cdot n$ is of the order $1/2 - \overline{t(1-t)}$ as n tends to ∞. Our proof combines classical Fourier analysis on ^{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

1 Introduction

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

$$f_{\min} := \min_{x \quad n} f(x). \tag{1}$$

This optimization problem is NP-hard in general, already for $d = 2$. 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:

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

M. Singh and D. P. Williamson (Eds.): IPCO 2021, LNCS 12707, pp. 43–57, 2021.
https://doi.org/10.1007/978-3-030-73879-2_4

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

$$(G) = \max_{x \quad ^V} \quad x_i - \quad x_i x_j .$$
$$\qquad\qquad\qquad\qquad\quad {}_{i \ \ V} \qquad\quad {}_{\{i,j\} \ \ E}$$

One may replace the boolean cube $^n = \{0, 1\}^n$ by the discrete cube $\{\pm 1\}^n$, in which case maximizing a quadratic polynomial $x^T A x$ 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 $^n \cap P$ 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). For this, 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 ^{n} where the constraints can be linear or polynomial; see, e.g., [3,18,27,33,36,40]. In [21] it is shown that the sum-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 ^{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 semi-algebraic sets we refer to [16,19,23,30] 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 [30] as a tool to produce tractable lower bounds for polynomial optimization problems. When applied to problem (1) it provides for any integer $r \quad \mathbb{N}$ a lower bound $f_{(r)} \leq f_{\min}$ on $f_{\min}$, given by:

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

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 $p_i \quad \mathbb{R}[x]_r$. In program (2), the condition

'$f(x)-\lambda$ is a sum-of-squares of degree at most $2r$ on $\{0,1\}^n$' means that there exists a sum-of-squares polynomial $s \in \Sigma_r$ such that $f(x)-\lambda = s(x)$ for all $x \in \{0,1\}^n$, or, equivalently, that the polynomial $f-\lambda-s$ belongs to the ideal generated by the polynomials $x_1-x_1^2, \dots, x_n-x_n^2$.

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 [16,30]. In the case of unconstrained boolean optimization, the resulting semidefinite program is known to have an optimum solution with small coefficients (see [29] and [31]). 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]. In fact, it has been shown in [34] 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 [34] 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 $f_{(r)}$ when $2r < n+d-1$. The following is our main result, which expresses the error of the bound $f_{(r)}$ in terms of the least roots of Krawtchouk polynomials.

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 (11) 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 $\|f\|_\infty := \max_{x \in \{0,1\}^n} |f(x)|$.

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], Sect. 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 ∞.

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:

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

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

Essentially as a side result in the proof of our main Theorem 1, we get the following analog of Theorem 1 for the upper bounds $f^{(r)}$.

Theorem 3. *Fix $d \leq n$ and let $f \quad \mathbb{R}[x]$ be a polynomial of degree d. Then, for any $r, n \quad \mathbb{N}$ 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.

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, this condition on ξ_{r+1}^n follows from a technical argument which is not required in the proof of Theorem 3 (namely, the condition $\Lambda \leq 1/2$ above relation (16) in Sect. 2.3).

1.3 Asymptotic Analysis for Both Hierarchies

The results above imply 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 $d \leq n$ and for $n, r \quad \mathbb{N}$ write*

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

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

Let C_d be the constant of Theorem 1 and let $\varphi(t)$ be the function from 5). Then, for any $t \quad [0, 1/2]$, 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 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 [34].

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 $\ ^n = \{0,1\}^n$, as can easily be seen by applying a change of variables $x \quad \{0,1\} \mapsto 2x - 1 \quad \{\pm 1\}$. In addition, our results extend to the case of polynomial optimization over the q-ary cube $\{0, 1, \ldots, q-1\}^n$ for $q > 2$ (see the extended version of this work in [38]).

Clearly, we may also obtain upper (resp., lower) bounds on the *maximum* $f_{\max}$ of f over $\ ^n$ by using $f_{(r)}$ (resp., $f^{(r)}$) applied to $-f$. To avoid possible confusion we will also refer to $f_{(r)}$ as the *outer* Lasserre hierarchy, whereas we will refer to $f^{(r)}$ as the *inner* Lasserre hierarchy. This terminology (borrowed from [7]) is motivated by the following observations. One can reformulate $f_{\min}$ via optimization over the set $\mathcal{M}$ of Borel measures on $\ ^n$:

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

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 $s \quad \Sigma_r$ with respect to a given fixed measure μ, then we obtain the bound $f^{(r)}$. On the other hand, any $\nu \quad \mathcal{M}$ gives a linear functional $L_\nu : p \quad \mathbb{R}[x]_{2r} \mapsto \int_{\ ^n} p(x)d\nu(x)$ which is nonnegative on sum-of-squares on $\ ^n$. 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 $f_{(r)}$ from (2) are 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 [34]. These exactness results are best possible for d even and n 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 coe cients, 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 $r \geq 2$. 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 [28] 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 [35]. Specializing to the hypersphere S^{n-1}, better rates in $O(1/r)$ were shown in [10,32], 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 ^{n} mentioned above all apply in the asymptotic regime where the dimension n is fixed and $r \to \infty$. This makes it di cult to compare them directly to our new results. Indeed, we have to consider a different regime in the case of the boolean cube ^{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 ^{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 [37], 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 [24,37]. Again, these results are all asymptotic in r, and thus hard to compare directly to our analysis on ^{n}.

1.5 Overview of the Proof

We give here a broad 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 on ^{n}, thus allowing an analysis of $f_{\min} - f_{(r)}$;
2. using classical *Fourier analysis* on the boolean cube ^{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 $f^{(r)}$. In the next section we will sketch in some more detail how our proof articulates along these four main steps.

2 Sketch of Proof

Here we sketch the main arguments needed to prove Theorem 1. It turns out that the proof for Theorem 3 follows essentially from some of these arguments. For a complete detailed proof we refer to the extended version [38] of this work. This section is organized along the four main steps outlined in Sect. 1.5.

2.1 The Polynomial Kernel Technique

Let $f \quad \mathbb{R}[x]_d$ 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 $\quad ^n$ at $0 \quad\quad ^n$ 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 $\quad ^n$ of the form:

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

where $u \quad \mathbb{R}[t]_r$ 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 $\quad ^n$ by:

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

Recall that μ is the uniform probability distribution on $\quad ^n$. An easy but important observation is that, if p is nonnegative on $\quad ^n$, then $\mathbf{K}p$ is a sum-of-squares (on $\quad ^n$) of degree at most $2r$. We use this observation 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 $\quad ^n$, we find that $\tilde{f}$ is a sum-of-squares on $\quad ^n$ 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 $\quad ^n$ is to select the operator $\mathbf{K}$ in such a way that $\mathbf{K}(1) = 1$ and

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

We collect this as a lemma for further reference.

Lemma 1. *If the kernel operator* $\mathbf{K}$ *associated to* $u \quad \mathbb{R}[t]_r$ *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 $\;^n$. □

In light of Lemma 1, we want to choose $u \quad \mathbb{R}[t]_r$ in such a way that the operator $\mathbf{K}^{-1}$ (and thus $\mathbf{K}$) is 'close to the identity operator' in a certain sense. In other words, we want the *eigenvalues* of $\mathbf{K}$ to be as close as possible to 1.

2.2 Fourier Analysis on $\;^n$ and the Funk-Hecke Formula

As kernels of the form (8) are invariant under the symmetries of $\;^n$, 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 coe cients of the expansion of u^2 in the basis of *Krawtchouk polynomials*. This link is known as the *Funk-Hecke formula* (cf. Theorem 4 below).

The Character Basis. Consider the space $\mathcal{R}[x]$ of polynomials on $\;^n$, defined as the quotient of $\mathbb{R}[x]_r$ under the relation $p \sim q$ if $p(x) = q(x)$ for all $x \quad ^n$. Equip this space $\mathcal{R}[x]$ with the inner product: $\langle p, q\rangle_\mu = \int p(x)q(x)d\mu(x)$, where μ is the uniform probability measure on $\;^n$. W.r.t. this inner product the space $\mathcal{R}[x]$ has an orthonormal basis given by the set of *characters:*

$$\chi_a(x) := (-1)^{a\cdot x} \quad (a \quad ^n).$$

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

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

then each H_k is an irreducible, Aut(^{n})-invariant subspace of $\mathcal{R}[x]$ of dimension $\binom{n}{k}$. 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 $p \quad \mathcal{R}[x]_d$ (in a unique way) as

$$p = p_0 + p_1 + \cdots + p_d \quad (p_k \quad H_k). \tag{10}$$

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

$$\mathcal{K}_k^n(t) := \sum_{i=0}^{k} (-1)^i \binom{t}{i} \binom{n-t}{k-i} \tag{11}$$

(see, e.g. [39]). Here $\binom{t}{i} := t(t-1)\dots(t-i+1)/i!$. The Krawtchouk polynomials form an orthogonal basis for $\mathbb{R}[t]$ with respect to the inner product $\langle \cdot, \cdot \rangle_\omega$ given by the following discrete probability measure on the set $[0:n] = \{0, 1, \dots, n\}$:

$$\omega := \frac{1}{2^n} \sum_{t=0}^{n} \binom{n}{t} \delta_t. \tag{12}$$

The following 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 $x, y \in {}^n$ so that $d(x, y) = t$. Then for any $0 \leq k \leq n$ we have:*

$$\mathcal{K}_k^n(t) = \sum_{|a|=k} \chi_a(x)\chi_a(y). \tag{13}$$

Using Lemma 2, one is then able to show the Funk-Hecke formula.

Theorem 4 (Funk-Hecke). *Given $u \in \mathbb{R}[t]_r$, decompose u^2 in the basis of Krawtchouk polynomials as $u^2 = \sum_{i=0}^{2r} \lambda_i \mathcal{K}_i^n$ and consider the kernel operator $\mathbf{K}$ associated to u via (8). For any $p \in \mathcal{R}[x]_d$ with harmonic decomposition $p = p_0 + p_1 + \dots + p_d$ as in (10), we have:*

$$\mathbf{K}p = \lambda_0 p_0 + \lambda_1 p_1 + \dots + \lambda_d p_d. \tag{14}$$

2.3 Optimizing the Choice of the Univariate Polynomial u

Recall that in light of Lemma 1 we wish to bound the quantity $\|\mathbf{K}^{-1} - I\|$ from (9). To define such $\mathbf{K}$ we need to suitably select the polynomial $u \in \mathbb{R}[t]_r$. Assume we choose $u \in \mathbb{R}[t]_r$ such that $u^2 = \sum_{i=0}^{r} \lambda_i \mathcal{K}_i^n$ with $\lambda_0 = 1$ and $\lambda_i \neq 0$ for all i. Then, for any $p = p_0 + \dots + p_d$ with $\|p\|_\infty = 1$, we find that

$$\|\mathbf{K}^{-1}p - p\|_\infty \leq \sum_{i=1}^{d} |1 - \lambda_i^{-1}| \cdot \|p_i\|_\infty \leq \gamma_d \sum_{i=1}^{d} |1 - \lambda_i^{-1}|, \tag{15}$$

where $\gamma_d > 0$ is a constant depending only on d. The left most inequality follows after an application of the Funk-Hecke formula (14). The right most inequality is a result of the following technical lemma (see [38] for the proof).

Lemma 3. *There exists a constant $\gamma_d > 0$, depending only on d, such that for any $p = p_0 + p_1 + \dots + p_d \in \mathcal{R}[x]_d$, we have:*

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

The key fact here is that the constant γ_d in Lemma 3 does not depend on the dimension n.

The quantity $\sum_{i=1}^{d} |1-\lambda_i^{-1}|$ in (15) is still difficult to analyze. Following [11], we therefore consider the following 'linearized' version instead:

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

It turns out that, as long as $\Lambda \le 1/2$, we have $\sum_{i=1}^{d} |1-\lambda_i^{-1}| \le 2\Lambda$, implying:

$$\|\mathbf{K}^{-1} - I\| \le 2\gamma_d \cdot \Lambda. \tag{16}$$

Recall that the Krawtchouk polynomials are orthogonal w.r.t. the inner product $\langle \cdot, \cdot \rangle_\omega$, where ω is the discrete probability measure on $[0:n]$ of (12). Therefore, we may express the scalars λ_i as:

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

We thus wish to find a univariate polynomial $u \in \mathbb{R}[t]_r$ for which:

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

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

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

$$\inf_{u \in \mathbb{R}[t]_r} \left\{ \Lambda := \int g \cdot u^2 d\omega : \int u^2 d\omega = 1 \right\}, \text{ where } g(t) := d - \sum_{i=1}^{d} \widehat{\mathcal{K}}_i^n(t). \tag{17}$$

We recognize this program to be the analog of the program (7), where we now consider the inner Lasserre bound of order r for the minimum $g_{\min} = g(0) = 0$ of the polynomial g over the set $[0:n]$, computed with respect to the measure $d\omega(t) = 2^{-n}\binom{n}{t}$ on $[0:n]$. Hence the optimal value of (17) is equal to $g^{(r)}$ and, using (16), we may conclude the following result, which tells us how to select the polynomial u (and thus $\mathbf{K}$).

Theorem 5. *Let g be as in (17). Assume that $g^{(r)} - g_{\min} \le 1/2$. Then there exists a polynomial $u \in \mathbb{R}[t]_r$ such that $\lambda_0 = 1$ and*

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

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 (17), and γ_d is the constant of Lemma 3.

2.4 The Inner Lasserre Hierarchy and Orthogonal Polynomials

In order to finish the proof of Theorem 1, it now remains to analyze the range $g^{(r)} - g_{\min}$ for the polynomial g in (17).

We recall a technique that may be used to perform such an analysis, which was developed in [9] and further employed for this purpose, e.g., in [8,37]. We present it here for the special case for optimization over $[0 : n]$ w.r.t. the measure ω, but it actually applies to univariate optimization w.r.t. general measures.

First, we observe that we may replace g by a suitable *upper estimator* $\widehat{g}$ which satisfies $\widehat{g}_{\min} = g_{\min}$ and $\widehat{g}(t) \geq g(t)$ for all t $[0 : n]$. Indeed, then we have:

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

Next, we use the following crucial link to the roots of Krawtchouk polynomials. This is a special case of a result by de Klerk and Laurent [9], applied to optimization over the set $[0 : n]$ equipped with the measure ω, so that the corresponding orthogonal polynomials are given by the Krawtchouk polynomials $\mathcal{K}_k^n$.

Theorem 6 [9]**).** *Suppose $\widehat{g}(t) = ct$ is a linear polynomial with $c > 0$. Then the Lasserre inner bound $\widehat{g}^{(r)}$ of order r for minimization of $\widehat{g}(t)$ on $[0 : n]$ w.r.t. the measure ω can be reformulated in terms of the smallest root ξ_{r+1}^n of $\mathcal{K}_{r+1}^n$ as:*

$$\widehat{g}^{(r)} = c \cdot \xi_{r+1}^n.$$

The upshot is that if we can upper bound the function g in (17) by some linear polynomial $\widehat{g}(t) = ct$ with $c > 0$, we then find:

$$g^{(r)} - g_{\min} \leq \widehat{g}^{(r)} - \widehat{g}_{\min} \leq c \cdot \xi_{r+1}^n.$$

Indeed, g can be upper bounded on $[0 : n]$ by its linear approximation at $t = 0$:

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

This inequality can be obtained by combining basic identities and inequalities concerning Krawtchouk polynomials (see [38]). We have thus shown that:

$$g^{(r)} - g_{\min} \leq d(d+1) \cdot (\xi_{r+1}^n / n). \tag{18}$$

We have now gathered all the tools required to prove Theorem 1.

Theorem 7 Restatement of Theorem 1). *Fix $d \leq n$ and let f $\mathbb{R}[x]$ be a polynomial of degree d. Then we have:*

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

whenever $d(d+1) \cdot (\xi_{r+1}^n / n) \leq 1/2$. Here, ξ_{r+1}^n is the smallest root of $\mathcal{K}_{r+1}^n$ and γ_d is the constant of Lemma 3.

Proof. Combining Theorem 5 with (18), we find that we may choose $u \quad \mathbb{R}[t]_r$ such that $\lambda_0 = 1$ and:

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

Using the Funk-Hecke formula (14), we see that $\lambda_0 = 1$ implies that $\mathbf{K}(1) = 1$. We may thus use Lemma 1 to conclude the proof, obtaining Theorem 1 with $C_d := \gamma_d \cdot d(d+1)$. □

3 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 $\ ^n$. 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 $f^{(r)} - f_{(r)}$.

A limitation of the present work is that no information is gained for low levels of the hierarchy, when r is fixed and the dimension n grows. Indeed, our results apply only in the regime $r \approx t \cdot n$, where $t \quad [0, 1/2]$ is a fixed fraction. They are therefore of limited practical value, as computation beyond the first few levels of the hierarchy is currently infeasible.

Our analysis also applies to polynomial optimization over the cube $\{\pm 1\}^n$ (by a simple change of variables). Furthermore, the techniques we use on the *binary* cube $\ ^n$ generalize naturally to the q-ary cube $(\mathbb{Z}/q\mathbb{Z})^n = \{0, 1, \ldots, q-1\}^n$ for $q > 2$. As a result we are able to show close analogs of our results on $\ ^n$ in this more general setting as well. We present this generalization in the expanded version [38] of this work.

The Constant $\ _d$. 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 [38] we show the existence of this constant γ_d, but the resulting dependence on d there is quite bad. This dependence, however, seems to be mostly an artifact of our proof. As we explain in [38], 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.

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

d	1	2	3	4	5	6	7	8	9	10	11	12
γ_d	1.00	2.00	4.00	8.00	20.0	48.1	112	258	578	1306	2992	6377

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 $i, j \quad \{0, 1, \ldots, r\}$. See, e.g., [39] for more details.

Connecting the Hierarchies. Our analysis of the *outer* hierarchy $f_{(r)}$ on $\ ^n$ relies essentially on knowledge of the *inner* hierarchy $f^{(r)}$. Although not explicitely 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 $\ ^n$ 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 $\ ^n$, respectively, the accompanying Fourier analysis is significantly more complicated. In particular, a direct analog of the Funk-Hecke formula (14) is not available. New ideas are therefore needed to define the kernel $K(x, y)$ (cf. (8)) and analyze its eigenvalues.

Acknowledgments. This work is supported by the European Union's Framework Programme for Research and Innovation Horizon 2020 under the Marie Sklodowska-Curie Actions Grant Agreement No. 764759 (MINOA). We wish to thank Sven Polak and Pepijn Roos Hoefgeest for several useful discussions, as well as the anonymous referees for their helpful suggestions.

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. 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.E., Tanbay, B. (eds.) World Women in Mathematics 2018. AWMS, vol. 20, pp. 17–56. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21170-7_1
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. arXiv:1210.5048v2 (2013)
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). https://doi.org/10.1007/s10107-015-0977-z
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 semidefinite programming relaxations for knapsack. In: Günlük, O., Woeginger, G.J. (eds.) IPCO 2011. LNCS, vol. 6655, pp. 301–314. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20807-2_24
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.) IPCO 2001. LNCS, vol. 2081, pp. 293–303. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45535-3_23
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.: Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (eds.) Emerging Applications of Algebraic Geometry. The IMA Volumes in Mathematics and Its Applications, vol. 149, pp. 157–270. Springer, New York (2009). https://doi.org/10.1007/978-0-387-09686-5_7

24. Laurent, M., Slot, L.: 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
25. Lee, J.R., Raghavendra, P., Steurer, D.: Lower bounds on the size of semidefinite programming relaxations. In: STOC 2015: 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. Nie, J., Schweighofer, M.: On the complexity of Putinar's positivstellensatz. J. Complex. **23**(1), 135–150 (2007)
29. O'Donnell, R.: SOS is not obviously automatizable, even approximately. In: 8th Innovations in Theoretical Computer Science Conference, vol. 59, pp. 1–10 (2017)
30. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology (2000)
31. 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)
32. Reznick, B.: Uniform denominators in Hilbert's seventeenth problem. Mathematische Zeitschrift **220**(1), 75–97 (1995)
33. Rothvoss, T.: The Lasserre hierarchy in approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial (2013)
34. 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)
35. Schweighofer, M.: On the complexity of Schmüdgen's positivstellensatz. J. Complex. **20**(4), 529–543 (2004)
36. Sherali, H.D., Adams, W.P.: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. SIAM J. Discrete Math. **3**, 411–430 (1990)
37. 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
38. Slot, L., Laurent, M.: Sum-of-squares hierarchies for binary polynomial optimization. arXiv:2011.04027 (2020)
39. Szegö, G.: Orthogonal Polynomials. American Mathematical Society Colloquium Publications, vol. 23. American Mathematical Society (1959)
40. Tunçel, L.: Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization, Fields Institute Monograph. American Mathematical Society, Providence (2010)