On the Sum-of-Squares Degree of Symmetric Quadratic Functions

Troy Lee[*1], **Anupam Prakash**[†2], **Ronald de Wolf**[‡3], **and Henry Yuen**[§4]

1 **School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore; and**
Centre for Quantum Technologies, Singapore
`troyjlee@gmail.com`

2 **School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore; and**
Centre for Quantum Technologies, Singapore
`aprakash@ntu.edu.sg`

3 **QuSoft, Amsterdam, The Netherlands; and**
CWI, Amsterdam, The Netherlands; and
University of Amsterdam, Amsterdam, The Netherlands
`rdewolf@cwi.nl`

4 **Massachusetts Institute of Technology, Cambridge, USA**
`hyuen@mit.edu`

Abstract

We study how well functions over the boolean hypercube of the form $f_k(x) = (|x|-k)(|x|-k-1)$ can be approximated by sums of squares of low-degree polynomials, obtaining good bounds for the case of approximation in ℓ_∞-norm as well as in ℓ_1-norm. We describe three complexity-theoretic applications: (1) a proof that the recent breakthrough lower bound of Lee, Raghavendra, and Steurer [19] on the positive semidefinite extension complexity of the correlation and TSP polytopes cannot be improved further by showing better sum-of-squares degree lower bounds on ℓ_1-approximation of f_k; (2) a proof that Grigoriev's lower bound on the degree of Positivstellensatz refutations for the knapsack problem is optimal, answering an open question from [12]; (3) bounds on the query complexity of quantum algorithms whose expected output approximates such functions.

1998 ACM Subject Classification G.1.2 Approximation

Keywords and phrases Sum-of-squares degree, approximation theory, Positivstellensatz refutations of knapsack, quantum query complexity in expectation, extension complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.17

[*] Supported by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

[†] Supported by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

[‡] Partially supported by the ERC Consolidator Grant QPROGRESS and by the European Commission FET-Proactive project Quantum Algorithms (QALGO) 600700.

[§] Supported by NSF grants 1218547 and 1452302 and a Simons Graduate Award in Theoretical Computer Science.

31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 17; pp. 17:1–17:31
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

1.1 Approximation of functions on the boolean hypercube by polynomials

Classical approximation theory studies how well a function $f : \mathbb{R} \to \mathbb{R}$ can be approximated by simpler functions, most commonly by polynomials of bounded degree. Approximation theory has found applications throughout complexity theory, for example in learning theory [21, 23], query complexity [22, 1], communication complexity [29, 30], and more.

An important special case is the investigation of the best approximation to a real boolean function $f : \{0,1\}^n \to \mathbb{R}$ in ℓ_∞-distance by a degree-d polynomial in the n variables $x_1, \ldots, x_n$. Nisan and Szegedy [22] initiated this study, showing that any polynomial that approximates the OR function with constant error in ℓ_∞-norm on $\{0,1\}^n$ has degree $\Omega(\sqrt{n})$. They also showed this bound is tight by constructing an $O(\sqrt{n})$-degree approximating polynomial for the OR function from a Chebyshev polynomial. Paturi [25] followed this by characterizing the approximate degree of *any* symmetric boolean function, i.e., any function $f : \{0,1\}^n \to \{0,1\}$ which only depends on the number of ones $|x|$ in the n-bit input x, and not on their locations. To get a feel for Paturi's theorem, consider the special case of a function $g_k : \{0,1\}^n \to \{0,1\}$ where $g_k(x) = 0$ unless $|x| = k$ in which case $g_k(x) = 1$. Paturi's theorem says that the $\frac{1}{4}$-error approximate degree of g_k, denoted by $\deg_{1/4}(g_k)$, is $\Theta(\sqrt{k(n-k)})$. Later, the ℓ_∞-approximate degree of symmetric boolean functions was characterized for all approximation errors ε by [28, 34]. Again in the special case of g_k, these results say that the degree of a polynomial that approximates g_k up to error $\varepsilon \geq 2^{-n}$ is $\Theta(\sqrt{k(n-k)} + \sqrt{n \log(1/\varepsilon)})$.

1.2 Our results on sum-of-squares approximation

Here we study the representation of non-negative functions on the boolean hypercube by *sums of squares* of polynomials. More precisely, a non-negative boolean function $f : \{0,1\}^n \to \mathbb{R}_+$ has an (exact) *degree-d sum-of-squares (sos) representation* if there exist degree-d polynomials $h_1, \ldots, h_r$ over the reals such that for all $x \in \{0,1\}^n$,

$$f(x) = h_1(x)^2 + \cdots + h_r(x)^2.$$

Let sos-deg(f) be the minimum d such that a non-negative function f has a degree-d sum-of-squares representation.[1] This sos degree is an important quantity that arises in the context of optimization and proof complexity, as also witnessed by our applications below.

The obvious fact that a sum of squares of polynomials is globally non-negative is remarkably useful. For example, for a graph $G = ([n], E)$, if $f(x_1, \ldots, x_n) = c - \sum_{(i,j)\in E}(x_i - x_j)^2$ has an sos representation on the boolean cube, then $c \geq \sum_{(i,j)\in E}(x_i - x_j)^2$ for all $x \in \{0,1\}^n$, and hence G has no cut of size larger than c. Moreover if f has a degree-d sos representation for small d, then this provides a small *certificate* (of size $n^{O(d)}$) that f has no cut of size larger than c. Such certificates can in fact be found by means of semidefinite programming; these observations are the basis of the semidefinite programming hierarchies of Lasserre and Parrilo [31, 18, 24] that have been the subject of intense study in approximation algorithms.

While exact sum-of-squares degree of functions on the boolean hypercube has been previously studied, there has been little work on the *approximation* of such functions by sos polynomials. This is the focus of our paper, and we prove a number of tight bounds on the

[1] Note that the degree of the polynomial representing f will actually be $2d$.

approximate sum-of-squares degree of functions on the hypercube. We consider two notions of approximation in this paper. The most familiar is ℓ_∞-approximation: an sos polynomial h ε-approximates a function $f : \{0,1\}^n \to \mathbb{R}_+$ in ℓ_∞-distance if $|f(x) - h(x)| \le \varepsilon$ for all $x \in \{0,1\}^n$. We let $\text{sos-deg}_\varepsilon(f, \ell_\infty)$ denote the minimum degree of an sos polynomial that ε-approximates f in ℓ_∞-distance. The other notion is ℓ_1-approximation: an sos polynomial h ε-approximates f in ℓ_1-distance if $\sum_{x\in\{0,1\}^n} |f(x) - h(x)| \le \varepsilon$, and we let $\text{sos-deg}_\varepsilon(f, \ell_1)$ denote the minimum degree of an sos polynomial that ε-approximates f in ℓ_1-distance. Note that $\varepsilon = \delta 2^n$ corresponds to *average* approximation error δ.[2]

For much of this paper we will focus on understanding the approximate sos degree of the symmetric quadratic functions $f_k : \{0,1\}^n \to \mathbb{R}_+$ defined as $f_k(x) = (|x| - k)(|x| - k - 1)$ for $k = 0, 1, 2, \ldots, n-1$. Our study of these functions is motivated by several reasons. First, these functions have a close connection to the proof complexity of the knapsack problem [12], and have recently been used to show lower bounds on semidefinite extension complexity [19]; we survey these two applications in Section 1.3 below. Furthermore, while the f_k may look very special, they are not too far from a general symmetric quadratic polynomial with real coefficients that is nonnegative on the hypercube, for the following reason. Any symmetric quadratic polynomial on the hypercube is of the form $p(|x|)$ for a quadratic univariate polynomial p. The polynomial p will have two roots. If the roots are complex they must come in a conjugate pair and p is already sos; if the roots are real, and not both ≤ 0 or $\ge n$, then they must lie in an interval $[k, k+1]$, for some $k \in \{0, 1, \ldots, n\}$, just as with f_k.

In our first set of results, we give lower and upper bounds on the ℓ_∞-approximate sos degree of the functions f_k.

▶ **Theorem 1.1** (ℓ_∞ sos approximations of f_k). *For all integers $n \ge 0$, $k \in \{1, \ldots, n-2\}$, and $\varepsilon = \varepsilon(n)$ satisfying $0 < \varepsilon < 1/50$, we have*

1. *$\text{sos-deg}_\varepsilon(f_k, \ell_\infty) = \Omega(\sqrt{k(n-k)})$*
2. *$\text{sos-deg}_\varepsilon(f_k, \ell_\infty) = O(\sqrt{k(n-k)} + \sqrt{n \log(1/\varepsilon)})$*

We expect the lower bound can be improved for the case of small ε to match the upper bound, but have been unable to show that so far. Observe that in the case of constant error, we obtain the tight bound of $\text{sos-deg}_{1/50}(f_k, \ell_\infty) = \Theta\left(\sqrt{k(n-k)}\right)$. While we are not aware of any previous work on ℓ_∞-approximate sos degree, techniques of Grigoriev [12] can be used to show that, for n odd, any degree-$(n-1)/2$ sos polynomial has error at least $\Omega(1/\log n)$ for approximating $f_{\lfloor n/2 \rfloor}$. This derivation is given in Appendix D.

The similarity between our ℓ_∞ bounds for f_k and Paturi's bound for the 0/1-valued functions g_k defined above is striking. For the upper bound, the connection can be seen as

[2] Also note that the existence of a degree-d sos approximation in either of these notions can be formulated as the feasibility of a semidefinite program of size polynomial in the domain size 2^n, as follows. For $x \in \{0,1\}^n$, let m_x be the column vector of dimension $D = \sum_{i=0}^d \binom{n}{i}$ indexed by sets $S \subseteq [n]$ of size $\le d$, with entry $m_{x,S} = \prod_{i\in S} x_i$. Let M_x be the $D \times D$ rank-1 matrix $m_x m_x^T$. Suppose $p(x) = \sum_{S:|S|\le d} p_S \prod_{i\in S} x_i$ is a multilinear polynomial of degree d, where with slight abuse of notation we use p also to denote the D-dimensional vector of real coefficients p_S. Then $p(x)$ is the inner product of p and m_x, so $p(x)^2 = \text{Tr}(pp^T M_x)$. Accordingly, every sos polynomial h of degree $\le d$ corresponds to a psd matrix Z such that $h(x) = \text{Tr}(ZM_x)$ for all $x \in \{0,1\}^n$ (Z can be written as $\sum_{i=1}^r p_i p_i^T$, so the rank r of Z would be the number of squared polynomials that h sums over). Hence the existence of an sos polynomial h of degree $\le d$ that ε-approximates f in ℓ_∞-distance, is equivalent to the existence of a psd matrix Z such that $|\text{Tr}(ZM_x) - f(x)| \le \varepsilon$ for all $x \in \{0,1\}^n$. The latter corresponds to the feasibility of an SDP with 2^n constraints, which (up to issues of precision) can be solved in time $2^{O(n)}$. However, we won't use this fact in this paper.

follows: we can construct an ε-approximation to f_k in ℓ_∞-distance by finding a univariate polynomial $e(z)$ such that $h(z) = (z-k)(z-k-1) + e(z)$ is globally nonnegative (i.e., $h(z) \geq 0$ for all $z \in \mathbb{R}$), and $|e(i)| \leq \varepsilon$ on integer points $i \in \{0, 1, \ldots, n\}$. As $h(z)$ is a globally nonnegative *univariate* polynomial, it is sos, and furthermore $h(|x|)$ is an ε-approximation to f_k. What are the requirements on e? It must be large enough to "cancel out" the negative values of $(z-k)(z-k-1)$ in the interval $(k, k+1)$, but small on all integer points $0, 1, \ldots, n$. This is very similar to looking for an ε-approximation of g_k, and techniques similar to those used by Paturi show that there is an e satisfying these requirements of degree $O(\sqrt{k(n-k)} + \sqrt{n}\log(1/\varepsilon))$. Note that this is slightly weaker than what Theorem 1.1 claims; we will soon discuss how to bring the $\log(1/\varepsilon)$ inside the square-root.

This upper bound argument shows that f_k can be approximated by a polynomial $h(|x|)$ where h is globally nonnegative. For the lower bound, it is not clear at all why the optimal approximating polynomial should be of this form. Any symmetric polynomial $f(x_1, \ldots, x_n)$ on the hypercube is of the form $f(x_1, \ldots, x_n) = p(|x|)$ for a univariate polynomial p. Even if f is sos, however, this does not mean that p will be globally nonnegative.

For the lower bound, we use an elegant recent result of Blekherman [2] that gives a characterization of the possible form of univariate polynomials p such that $f(x_1, \ldots, x_n) = p(|x|)$ when f is a sos and symmetric real-valued boolean function. This structural theorem allows us to reduce the analysis of the approximate sos degree of f_k to the approximate degree of a symmetric function on the boolean hypercube, for which we can apply Paturi's lower bound.

Interestingly, for small ε we can show a better upper bound than that given by the argument sketched above. To get the better upper bound of Theorem 1.1, we take advantage of a recent characterization of the sos degree of a non-negative real-valued boolean function as the quantum query complexity of computing that function *in expectation* (see Section 1.3.3). We explicitly design a quantum algorithm to approximately compute f_k in expectation with query complexity $O(\sqrt{k(n-k)} + \sqrt{n\log(1/\varepsilon)})$, which by the characterization implies the same upper bound on sos-$\deg_\varepsilon(f_k, \ell_\infty)$. This again parallels the situation for symmetric boolean-valued functions, where the tight upper bound of $O(\sqrt{k(n-k)} + \sqrt{n\log(1/\varepsilon)})$ on $\deg_\varepsilon(g_k)$ was first shown by the construction of a quantum query algorithm [34].

We also study sos ℓ_1-approximations of f_k:

▶ **Theorem 1.2** (sos ℓ_1-approximations of f_k)**.** *Let n be odd and $k = \lfloor n/2 \rfloor$. Then*

$$\textit{sos-deg}_{\delta 2^n}(f_k, \ell_1) \leq \left\lceil \frac{3\sqrt{n}}{\sqrt{2}\delta} \ln\left(\frac{1}{\delta}\right) \right\rceil ,$$

for any $8/\sqrt{2n} \leq \delta \leq 1/4$. For $k < 0.49n$, we have $\textit{sos-deg}_{\delta 2^n}(f_k, \ell_1) = O\left(\ln\left(\frac{1}{\delta}\right)\right)$.

The proof of this theorem follows the same plan sketched above for the upper bound in the ℓ_∞ case. We construct a low-degree univariate polynomial $e(z)$ such that $h(z) = (z-k)(z-k-1) + e(z)$ is globally nonnegative and $\varepsilon = \sum_{i=0}^n \binom{n}{i}|e(i)|$ is relatively small. Then $h(|x|)$ gives the desired sos approximation to f_k in ℓ_1-norm. We discuss the applications of this theorem to the lower bounds on semidefinite extension complexity of [19] below in Section 1.3.1.

1.3 Applications in complexity theory

Here we describe complexity-theoretic consequences of such sos bounds in three different settings.

1.3.1 Positive semidefinite extension complexity

The approximation of boolean functions by sos polynomials has played an important role in *inapproximability* results. Our first application is to the analysis of the positive semidefinite extension complexity of polytopes. The recent breakthrough work of Lee, Raghavendra and Steurer [19] showed that any semidefinite program whose feasible region projects to the *correlation polytope* must have size $2^{\tilde{\Omega}(n^{2/11})}$. By reduction this in turn implies a $2^{\tilde{\Omega}(n^{1/11})}$ lower bound for the polytope corresponding to the Traveling Salesman Problem on n-vertex graphs, showing (roughly speaking) that TSP cannot be solved by small semidefinite programs.

The argument of [19] shows that lower bounds on the degree of sos polynomials that approximate a function $f_k(x) = (|x| - k)(|x| - k - 1)$ in ℓ_1-distance on the boolean cube imply lower bounds on the semidefinite extension complexity of the correlation polytope. They build on the work of Grigoriev [12] to show that, for odd n and $k = \lfloor n/2 \rfloor$, any sum of squares of degree-$\lfloor n/2 \rfloor$ polynomials has ℓ_1-error at least $2^{n-2}/\sqrt{n}$ in approximating $f_{\lfloor n/2 \rfloor}$.[3] Our Theorem 1.2 shows that this bound is tight, up to logarithmic factors. Further, our upper bound on sos-$\deg_{\delta 2^n}(f_k, \ell_1)$ throughout the full range of error implies, roughly speaking, that the quantitative bounds of [19] cannot be improved simply by showing better sos degree lower bounds on f_k.

1.3.2 Proof complexity

Our second result is in proof complexity. Grigoriev and Vorobjov [14] introduced a proof system based on the Positivstellensatz [32]. We explain this proof system in the context of the knapsack problem. In this instance, the knapsack problem can be phrased as looking for a solution $x \in \mathbb{R}^n$ to the system of equations

$$\sum_{i=1}^{n} x_i - r = 0,\ x_1^2 - x_1 = 0, \ldots, x_n^2 - x_n = 0 \tag{1}$$

where $r \in \mathbb{R}$. When r is not an integer, this system obviously has no solution. One way to certify that there is no solution, is to find polynomials $g, g_1, \ldots, g_n$ and sos polynomial h such that

$$g(x) \cdot \left(\sum_{i=1}^{n} x_i - r \right) + \sum_{i=1}^{n} g_i(x) \cdot (x_i^2 - x_i) = 1 + h(x)\,. \tag{2}$$

Such a collection of polynomials constitutes a *Positivstellensatz refutation* of the statement that (1) has a solution: if $a \in \mathbb{R}^n$ satisfied $\sum_{i=1}^n a_i - r = 0$, and $a_i^2 - a_i = 0$ for $i = 1, \ldots, n$, then the left-hand side of (2) would evaluate to 0 on a, while the right-hand side would evaluate to $1 + h(a) \geq 1$, a contradiction.

Grigoriev and Vorobjov define the *Positivstellensatz refutation degree* of the system (1) as

$$\max \left\{ \deg \left(g(x) \cdot \left(\sum_{i=1}^{n} x_i - r \right) \right), \max_{i \in [n]} \{ \deg(g_i(x) \cdot (x_i^2 - x_i)) \}, \deg(h(x)) \right\},$$

[3] The initial version of [19] only claimed a lower bound on the ℓ_1-error of $\Omega(2^n/n^{3/2})$. However, their argument actually shows a bound of $\Omega(2^n/\sqrt{n})$ after a computational error is corrected. This improves the bound on the psd extension complexity of the correlation polytope from the $2^{\tilde{\Omega}(n^{2/13})}$, of their paper, to the $2^{\tilde{\Omega}(n^{2/11})}$ quoted here.

maximized over all sets of polynomials satisfying (2). Grigoriev [12] shows that if $k < r < k+1$ for a nonnegative integer $k < (n-3)/2$, then any Positivstellensatz refutation of (1) has degree at least $2k+4$. We provide a simple proof of this in Appendix C using Blekherman's theorem. Kurpisz et al. [17] independently also give an alternative proof of Grigoriev's lower bound, by showing a more general theorem that reduces the analysis of dual certificates for very symmetric sos proof systems (such as for knapsack) to the analysis of univariate polynomials.

Grigoriev's lower bound shows the weakness of the Positivstellensatz-based proof system: even to refute such easy instances it already needs polynomials of fairly high degree. We prove here that Grigoriev's lower bound is exactly tight, answering an open question from [12].

▶ **Theorem 1.3.** *Let $k < r < k+1$ for a nonnegative integer k. The Positivstellensatz refutation degree of* (1) *with this value of r is at most $2k+4$.*

1.3.3 Quantum query complexity of approximating a function in expectation

The third application is to quantum algorithms. Kaniewski et al. [16] observed a very close connection between the sos degree of a function $f : \{0,1\}^n \to \mathbb{R}_+$ and a variant of quantum query complexity: sos-deg(f) is *exactly equal* to the optimal query complexity among all quantum algorithms with non-negative outputs whose *expected* output on input x equals $f(x)$.[4] This model of query complexity in expectation is motivated by similar models of *communication* complexity that arose in the study of extension complexity of polytopes [9].

However, as [16] note, this model has some intrinsic interest and motivation as well. Suppose we want to approximate $F(x) = \sum_{i=1}^{m} f_i(x)$, where each f_i is a non-negative function of $x \in \{0,1\}^n$. Then we can just compute, for each i, a random variable whose expected value is $f_i(x)$ and then output the sum of those random variables. By linearity of expectation, the output will have the correct expectation $F(x)$. It will be tightly concentrated around its expectation if the individual random variables have a variance that is not too large. Thus in some cases it suffices to compute the $f_i(x)$ in expectation only, rather than to compute the values $f_i(x)$ themselves (which may be much more expensive). In this example, it is actually not even necessary to compute each $f_i(x)$ *exactly* in expectation. If the ith random variable has an expectation that is within ε_i of $f_i(x)$, then the expected value of our output is within $\sum_{i=1}^{m} \varepsilon_i$ of the correct value $F(x)$.

The same proofs that Kaniewski et al. [16] used to equate sos-deg(f) and quantum query complexity in expectation also work in the approximate case. For example, $\text{sos-deg}_\varepsilon(f, \ell_\infty)$ is the optimal query complexity among all quantum algorithms with non-negative outputs whose expected output on input x differs from $f(x)$ by at most ε, for every $x \in \{0,1\}^n$, and the analogous statements hold for approximation using the other norms. Accordingly, our above results about approximate sos degree immediately translate to results about quantum query complexity of algorithms that approximate f in expectation.

1.4 Organization

The rest of the paper is organized as follows. In Section 2, we prove our sos ℓ_∞-approximation bounds (Theorem 1.1). In Section 3 we prove upper bounds on the degree of sos ℓ_1-ap-

[4] To avoid potential confusion: for each fixed x the expectation is taken over the internal randomness of the algorithm; it is not an expectation over different inputs x.

proximations to $f_{\lfloor n/2 \rfloor}$ (Theorem 1.2), and show that the lower bound of [19] on the extension complexity of the correlation polytope cannot be improved by obtaining better sos ℓ_1-approximate degree lower bounds. In Section 4 we prove Theorem 1.3, showing tightness of Grigoriev's knapsack lower bound.

2 Sum-of-squares approximation in ℓ_∞-norm

In this section we give lower and upper bounds on the ℓ_∞-approximate sos degree of the function $f_k(x) = (|x| - k)(|x| - k - 1)$ to prove Theorem 1.1.

▶ Remark. Throughout this section, we will assume that $k \leq n/2$. Letting $\bar{x}$ denote the bitwise complement of x, we see that $f_k(x) = (|\bar{x}|-(n-k-1))(|\bar{x}|-(n-k)) = (|\bar{x}|-\ell)(|\bar{x}|-\ell-1)$ where $\ell = n-k-1$. Thus if $k > n/2$, then $\ell \leq n/2$ and $f_k(x) = f_\ell(\bar{x})$. For any $h : \{0,1\}^n \to \mathbb{R}$ the functions $h(x)$ and $h(\bar{x})$ have the same sos degree, so it suffices to work with f_k for $k \leq n/2$.

2.1 Lower bound preliminaries

The following lemma is implicit in Paturi [25].

▶ **Lemma 2.1** (Paturi [25]). *Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial and suppose that $0 \leq p(i) \leq c$ for all $i \in \{0,1,\ldots,n\}$. If $|p'(\alpha)| \geq \delta$ for some $0 \leq \alpha \leq n$, then $\deg(p) = \Omega(\frac{\delta}{c}\sqrt{\alpha(n-\alpha)})$.*

We give a simple, but convenient, application of this lemma to the case where p is bounded on $\{0,1,\ldots,n\}$, except possibly for a small set S near where p is known to be small.

▶ **Lemma 2.2.** *Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial, $S \subseteq \{0,1,\ldots,n\}$, and suppose that the following bounds are known.*

- *$|p(i)| \leq c$ for all $i \in \{0,1,\ldots,n\} \setminus S$, for a constant c.*
- *$p(\alpha) \leq \varepsilon$, for some $\alpha \in \{0,1,\ldots,\lfloor n/2 \rfloor\}$.*
- *$p(\beta) \geq a$ where $|\alpha - \beta| \leq d_1$, for a constant d_1.*
- *$\max_{i \in S} |i - \alpha| \leq d_2$, for a constant d_2.*

If $a \geq c > \varepsilon$, then $\deg(p) = \Omega(\sqrt{\alpha(n-\alpha)})$, where the constant in the $\Omega(.)$ depends on c, d_1, d_2.

Proof. If $|p(i)| \leq c$ for all $i \in S$, then applying Paturi's lemma directly we obtain a bound of

$$\Omega\left(\frac{a-\varepsilon}{d_1 c}\sqrt{(\alpha-d_1)(n-\alpha+d_1)}\right).$$

Otherwise, suppose the maximum of $|p(i)|$ over $i \in \{0,1,\ldots,n\}$ is attained at $j \in S$. Then the derivative of p is at least $(|p(j)| - \varepsilon)/d_2$. Applying Paturi's lemma in this case gives a bound of

$$\Omega\left(\frac{|p(j)|-\varepsilon}{d_2|p(j)|}\sqrt{(\alpha-d_2)(n-\alpha+d_2)}\right) \geq \Omega\left(\frac{1}{d_2}\left(1-\frac{\varepsilon}{c}\right)\sqrt{(\alpha-d_2)(n-\alpha+d_2)}\right). \quad ◀$$

We will also need the following elegant theorem of Blekherman [2]. Recall that a *symmetric* real-valued boolean function $f : \{0,1\}^n \to \mathbb{R}$ satisfies $f(x) = f(\pi(x))$ for all $x \in \{0,1\}^n$ and $\pi \in S_n$, where the permutation π acts as $\pi((x_1,\ldots,x_n)) = (x_{\pi(1)},\ldots,x_{\pi(n)})$. For any symmetric boolean function f of degree d, there is a univariate polynomial $\tilde{f}$ of degree d such that $f(x_1,\ldots,x_n) = \tilde{f}(x_1+\cdots+x_n)$.

▶ **Theorem 2.3** (Blekherman [2]). *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be a symmetric non-negative real-valued boolean function and $\tilde{f}$ a univariate polynomial such that $f(x_1, \ldots, x_n) = \tilde{f}(x_1 + \cdots + x_n)$. If f can be written as the sum of squares of n-variate polynomials of degree $d \leq n/2$, then we can write*

$$\begin{aligned}\tilde{f}(z) &= q_d(z) + z(n-z)q_{d-1}(z) + z(z-1)(n-z)(n-1-z)q_{d-2}(z) + \cdots \\ &\quad \cdots + z(z-1)\cdots(z-d+1)(n-z)(n-1-z)\cdots(n-d+1-z)q_0(z) \qquad (3)\end{aligned}$$

where each $q_t(z)$ is a univariate sos polynomial with sos-deg$(q_t) \leq t$.

In Appendix B we include a proof of Blekherman's theorem. In Appendix C, we use Blekherman's theorem to provide a simple proof of Grigoriev's lower bound [12] on the degree of Positivstellensatz refutations for the knapsack problem.

2.2 Lower bound for exact sos degree

To illustrate our proof technique, we first show how the above tools can be used to prove a bound of $\Omega(\sqrt{k(n-k)})$ on the exact sos degree of $f(x) = (|x|-k)(|x|-k-1)$. In the next section, we will extend this proof to also work for the approximate case.

▶ **Theorem 2.4.** *Let $f_k : \{0,1\}^n \to \mathbb{R}_+$ be defined as $f_k(x) = (|x|-k)(|x|-k-1)$ for an integer $1 \leq k \leq n-2$. Then sos-deg$(f_k) = \Omega(\sqrt{k(n-k)})$.*

Proof. Following Remark 2, if we show the theorem for $1 \leq k \leq n/2$, then it will also imply the theorem for $1 \leq k \leq n-2$. We thus assume $1 \leq k \leq n/2$.

Let d be the sos degree of f. We may assume $d \leq n/2$ as otherwise the theorem holds. Let $\tilde{f}$ be a univariate polynomial of degree $\leq 2d$ such that $\tilde{f}(x_1 + \cdots + x_n) = f(x_1, \ldots, x_n)$. Write $\tilde{f}(z) = g_1(z) + g_2(z)$ where $g_1(z) = q_d(z) + z(n-z)q_{d-1}(z) + \cdots + z(z-1)\cdots(z-(k-1))(n-z)(n-1-z)(n-(k-1)-z)q_{d-k}(z)$ is the first $k+1$ terms in the representation of $\tilde{f}$ of Theorem 2.3, and $g_2(z)$ is the remaining part of that representation.

Our first claim is that $(z-k)$ is a factor of both g_1 and g_2. Notice that $\tilde{f}(k) = g_1(k) + g_2(k) = 0$. Furthermore each term of g_1 and g_2 is nonnegative on integer points between 0 and n, which means that each individual term of g_1 and g_2 must evaluate to 0 at k.

Consider now a general term $z(z-1)\cdots(z-t)(n-z)(n-1-z)(n-t-z)q_{d-t-1}(z)$ of Blekherman's representation. If $t \geq k$ then this term obviously has a factor of $z-k$. If $t < k$ then the prefactor $z(z-1)\cdots(z-t)(n-z)(n-1-z)(n-t-z)$ is non-zero for $z = k$, so it must be the case that $q_{d-t-1}(k) = 0$. Since $q_{d-t-1}(z)$ is a univariate sum-of-squares polynomial, even $(z-k)^2$ divides $q_{d-t-1}(z)$.

By the choice of the breakpoint between g_1 and g_2, this shows that $(z-k)^2$ is a factor of g_1 and $z-k$ is a factor of g_2. By the same argument, $(z-(k+1))^2$ is also a factor of g_1, and $(z-(k+1))$ is a factor of g_2.

In light of this, we can write $g_1(z) = (z-k)(z-k-1)h_1(z)$, $g_2(z) = (z-k)(z-k-1)h_2(z)$ so that

$$(z-k)(z-k-1) = \tilde{f}(z) = (z-k)(z-k-1)(h_1(z) + h_2(z)).$$

This means that $h_1(i) + h_2(i) = 1$ for all $i \in \{0,1,\ldots,n\} \setminus \{k, k+1\}$. Furthermore, $h_1(k) = h_1(k+1) = 0$ as h_1 still has roots at $k, k+1$ (as g_1 had double roots there), and $h_2(i) = 0$ for $i \in \{0, \ldots, k-1\}$ because each term in h_2 includes the prefactor $z(z-1)\cdots(z-k+1)$. Combining these observations with the fact that $h_1(i) \geq 0, h_2(i) \geq 0$ for all $i \in \{0,1,\ldots,n\} \setminus \{k, k+1\}$ gives the following:

1. $0 \leq h_1(i) \leq 1$ for all $i \in \{0, 1, \dots, n\}$.
2. $h_1(i) = 1$ for $i \in \{0, \dots, k-1\}$.
3. $h_1(k) = h_1(k+1) = 0$.

Applying Lemma 2.2 to h_1 now gives the desired result. ◀

2.3 Lower bound for ℓ_∞-approximate sos degree

Now we show the lower bound of Theorem 1.1.

▶ **Theorem 2.5.** *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be defined as $f(x) = (|x| - k)(|x| - k - 1)$ for some integer $1 \leq k \leq n-2$. Then sos-deg$_{1/50}(f, \ell_\infty) = \Omega(\sqrt{k(n-k)})$.*

Proof. We now describe how the above proof can be modified to work for ℓ_∞-approximate sum-of-squares degree. We again assume $1 \leq k \leq n/2$. Suppose that $h : \{0,1\}^n \to \mathbb{R}$ is a sum of squares of degree-$d \leq n/2$ polynomials that satisfies $|h(x) - f(x)| \leq \varepsilon$ for all $x \in \{0,1\}^n$, for some $\varepsilon < 1/4$ to be determined later. Let $\tilde{h}$ be the univariate polynomial such that $\tilde{h}(x_1 + \cdots + x_n) = h(x_1, \dots, x_n)$. Note that $\tilde{h}$ satisfies $|\tilde{h}(i) - (i-k)(i-k-1)| \leq \varepsilon$ for all $i \in \{0, 1, \dots, n\}$. We again use Blekherman's theorem to decompose $\tilde{h}(z) = g_1(z) + g_2(z)$ where, as before, $g_1(z) = q_d(z) + z(n-z)q_{d-1}(z) + \cdots + z(z-1)\cdots(z-k+1)(n-z)(n-1-z)(n-k+1-z)q_{d-k}(z)$. The polynomial g_1 has the following properties.

1. $g_1(i) = \tilde{h}(i)$ for $i \in \{0, 1, \dots, k\}$, because all terms of g_2 are zero on these points.
2. $g_1(i) \leq \tilde{h}(i)$ for $i \in \{0, 1, \dots, n\}$. This follows as $g_2(i)$ is nonnegative on integer points in $\{0, 1, \dots, n\}$.
3. $g_1(i) \geq 0$ for $i \in [k-1, n-k+1]$. Each term of g_1 is nonnegative in this interval because the prefactor is.

We will consider two cases based on the value of $g_1(k+3/2)$. First consider the case $g_1(k+3/2) > \varepsilon$. In this case, consider a point $\alpha \in \text{argmin}_z\{g_1(z) : k-1 \leq z \leq k+3/2\}$. Let $g_1(\alpha) = \delta$. By item (3) above and as $g_1(k-1), g_1(k+3/2) > \varepsilon$ and $g_1(k), g_1(k+1) \leq \varepsilon$ we have $0 \leq \delta \leq \varepsilon$ and also $g_1'(\alpha) = 0$.

Now consider the function $p_1 = g_1 - \delta$. As $p_1(\alpha) = p_1'(\alpha) = 0$ it follows that p_1 has a double root at α. Define q_1 by $p_1(z) = (z-\alpha)^2 q_1(z)$. Note that q_1 has the following properties.

1. $q_1(i) \leq 6 + \varepsilon$ for $i \in \{0, 1, \dots, n\} \setminus \{k-1, k, k+1, k+2\}$.
2. $q_1(k-1) \geq \frac{2-2\varepsilon}{9}$.
3. As either $|\alpha - k| \geq 1/2$ or $|\alpha - k - 1| \geq 1/2$ we have either $q_1(k) \leq 4\varepsilon$ or $q_1(k+1) \leq 4\varepsilon$.

Applying Lemma 2.2 then gives the desired lower bound in this case as long as $\varepsilon < 1/19$.

Now we consider the second case, that $g_1(k+3/2) \leq \varepsilon$. In this case, we modify g_1 by adding a function that is shaped like a “smile.” Let $p_1(z) = g_1(z) + 8\varepsilon(x-k-1)(x-k-2)$. Note that p_1 satisfies $p_1(k+1) \geq 0$, $p_1(k+3/2) \leq -\varepsilon$, and $p_1(k+2) \geq 0$. Thus $p_1(z)$ has two roots α, β in $[k+1, k+2]$, with $\alpha \leq \beta$. Let $p_1(z) = (z-\alpha)(z-\beta)r_1(z)$. Then r_1 satisfies the following properties.

1. $r_1(i) \leq 2$ for $i \in \{0, 1, \dots, n\} \setminus \{k+1, k+2\}$.
2. $r_1(k-1) \geq \frac{2}{9} + 5\varepsilon$.
3. $|r_1(k)| \leq 16\varepsilon$.

Applying Lemma 2.2 then gives the desired lower bound as long as $\varepsilon < 2/99$. ◀

2.4 Upper bound for ℓ_∞-approximate sos degree

In this section we show that the lower bound in Theorem 2.5 is tight. To do this, we use the characterization of the sos degree of a function $f : \{0,1\}^n \to \mathbb{R}_+$ as the quantum query complexity of computing f in expectation [16]. In this model, a quantum algorithm A makes a number T of quantum queries to the hidden input x, and outputs a non-negative real number. We say that the algorithm A *computes f in expectation* if the expected value of the output of the algorithm A on input x is exactly equal to $f(x)$. We will use $\mathrm{QE}(f)$ to denote the minimum number of quantum queries T needed by such an algorithm to compute f in expectation. Kaniewski et al. [16] show that $\mathrm{QE}(f)$ exactly captures the sos degree of f:

▶ **Theorem 2.6** ([16]). *Let $f : \{0,1\}^n \to \mathbb{R}_+$. Then $\mathrm{QE}(f) = \textit{sos-deg}(f)$.*

Thus, in order to prove an upper bound on the (approximate) sos degree of a function f, it suffices to construct a quantum query algorithm that (approximately) computes f in expectation. The only knowledge of quantum query complexity needed to understand the algorithm is Theorem 2.6 above, and the existence of the following quantum algorithms, all of which are variants of Grover search.

- *Regular Grover* [15, 4]: If $|x| \geq t$ then there is a quantum algorithm (depending on t) using $O(\sqrt{n/t})$ queries that finds an i such that $x_i = 1$ with probability at least 1/2.
- *ε-error Grover* [5]: There is a quantum algorithm using $O(\sqrt{n \log(1/\varepsilon)})$ queries that finds an i such that $x_i = 1$ with probability at least $1 - \varepsilon$ if $|x| \geq 1$.
- *Exact Grover* [4]: If $|x| = t$ then there is a quantum algorithm (depending on t) using $O(\sqrt{n/t})$ queries that finds an i such that $x_i = 1$ *with certainty.*

The algorithm consists of three subroutines, which we now describe. We begin with the simplest procedure, SAMPLE(x, S), which motivates the basic plan of the algorithm.

Algorithm 1 Given $x \in \{0,1\}^n$ and $S \subseteq [n]$, samples two entries of x outside of S

```
procedure SAMPLE$(x, S)$
    Randomly choose $i \neq j \in [n] \setminus S$. Output $x_i x_j \cdot (n - |S|)(n - |S| - 1)$
end procedure
```

▶ **Claim 2.7.** *The procedure SAMPLE(x, S) makes two queries and the expected value of its output is $(|x| - |S|)(|x| - |S| - 1)$.*

The procedure SAMPLE suggests the following high-level idea for an algorithm for computing $f_k(x) = (|x| - k)(|x| - k - 1)$. For simplicity we describe the high-level idea for the case where $k \leq n/2$ and where we want to compute a constant-error ε-approximation of f_k, in expectation.

First we try to find a set S of k ones in x assuming that $|x| > 2k$, using a procedure HIGH. If we find such a set S then we run SAMPLE(x, S) and output $f(x)$ exactly, in expectation. If the procedure HIGH fails to find such a set S, then we run a procedure LOW. This uses exact Grover search to determine the Hamming weight of x with certainty if $|x| \leq 2k$. Once we know the Hamming weight of x we can correctly output $f(x)$, deterministically. Both the procedures HIGH and LOW can be done with $O(\sqrt{kn})$ queries, in the constant error ε case. The only case where the algorithm may err is if $|x| > 2k$ but the procedure HIGH fails to find k ones in x. The most subtle part of the algorithm is tuning the parameters such that this error is at most ε in expectation. We now describe the procedures HIGH and LOW.

▶ **Lemma 2.8.** *Fix δ and let $M = \max\{k, \lceil \log(1/\delta) \rceil\}$. Suppose that $|x| \geq t > 2k$. Then procedure HIGH(x, t, δ) makes $O(M\sqrt{n/t})$ queries and returns a set S with $|S| = k$ and $x_i = 1$ for all $i \in S$ with probability at least $1 - \delta$.*

Algorithm 2 Find k ones in x with probability $1-\delta$, assuming $|x| \geq t > 2k$

```
procedure HIGH($x, t, \delta$)
    $S = \emptyset$
    $\ell = 1$
    while $\ell \leq 5\max(k, \lceil\log(1/\delta)\rceil)$ and $|S| < k$ do
        $\ell \leftarrow \ell + 1$
        Grover search assuming $|x| \geq t/2$.
        if find $x_i = 1$ then $S \leftarrow S \cup i$, $x \leftarrow x \setminus x_i$
        end if
    end while
    return $S$
end procedure
```

Proof. As each Grover search requires $O(\sqrt{n/t})$ queries, in total the procedure makes $O(M\sqrt{n/t})$ queries. Let us now estimate the probability that it exits without finding a set S of size k.

As we are given that initially $|x| \geq t > 2k$, if less than k ones are found then throughout the algorithm there remain at least $t/2$ ones in x. Thus each run of Grover has probability of success at least $1/2$. The probability to have fewer than k successes among the $5M$ runs is therefore at most

$$\frac{1}{2^{5M}}\sum_{i=0}^{k-1}\binom{5M}{i} \leq 2^{-(1-H(k/5M))5M} \leq 2^{-M} \leq 2^{-\log(1/\delta)} = \delta,$$

where $H(\cdot)$ denotes binary entropy, and we used that $1 - H(k/5M) \geq 1 - H(1/5) \geq 1/5$. ◀

Next we give the algorithm LOW.

Algorithm 3 Outputs $(|x|-k)(|x|-k-1)$ with certainty if $|x| \leq t$

```
procedure LOW($x, t$)
    $S = \emptyset$
    for $i = t$ to $1$ do
        Exact Grover search assuming $|x| = i$
        if find $x_i = 1$ then $S \leftarrow S \cup i$, $x \leftarrow x \setminus x_i$
        end if
    end for
    Output $(|S|-k)(|S|-k-1)$.
end procedure
```

▶ **Claim 2.9.** *If $|x| \leq t$, then LOW(x,t) outputs $(|x|-k)(|x|-k-1)$ and makes $O(\sqrt{tn})$ queries.*

Proof. The number of queries is

$$\sum_{i=1}^{t} O(\sqrt{n/i}) = O(\sqrt{tn})\,.$$

Next we show that if $|x| \leq t$, then LOW(x,t) will find all of the ones in x (this is similar to [11]). Initially the index $i = t$ and thus $i \geq |x|$. This invariant is maintained throughout the algorithm. If ever $i = |x|$ then we will find all the remaining ones in x as our guess for the number of ones is always correct after this point. On the other hand, if the algorithm terminates with $i = 1 > |x|$ then we have found all the ones in the original input x. ◀

With these procedures in place, we can describe the main algorithm and prove its correctness.

Algorithm 4 Main

```
procedure MAIN($x, \varepsilon$)
    $m = \max(k, \lceil \log(1/\varepsilon) \rceil)$
    for $i = 1$ to $\lfloor \log(n/m) \rfloor$ do
        $t \leftarrow 2^i m$
        $\delta \leftarrow \varepsilon/(4t^2)$
        $S$=HIGH($x, t, \delta$)
        if $|S| = k$ then
            SAMPLE($x, S$)
            Exit
        end if
    end for
    LOW($x, 2m$)
end procedure
```

▶ **Theorem 2.10.** *For every $x \in \{0,1\}^n$, the expected value of Main(x, ε) differs from $(|x| - k)(|x| - k - 1)$ by at most ε. The algorithm makes at most $O(\sqrt{kn} + \sqrt{n \log(1/\varepsilon)})$ queries.*

Proof. Following Remark 2 we may assume that $k \leq n/2$. First we verify the stated complexity of the algorithm. Note that by definition of m in the main Algorithm 4, it suffices to show that the algorithm makes $O(\sqrt{nm})$ queries. By Claim 2.9 the call to LOW$(x, 2m)$ makes $O(\sqrt{mn})$ queries, and by Claim 2.7 there are at most 2 queries made by SAMPLE as this is called at most once. Finally, the number of queries in the call to HIGH when $t = 2^i m$ and $\delta = \varepsilon/(4t^2)$ is at most

$$O\left(k\sqrt{\frac{n}{2^i m}} + \log(2^{2i+2}m^2/\varepsilon)\sqrt{\frac{n}{2^i m}}\right) = O\left(\sqrt{\frac{kn}{2^i}} + \sqrt{\frac{n\log(1/\varepsilon)}{2^i}} + \log(2^{2i+2}m^2)\sqrt{\frac{n}{2^i m}}\right)$$

where we have used the fact that $m \geq k$ and $m \geq \log(1/\varepsilon)$. The sum of the first two terms over $i \geq 1$ is $O(\sqrt{kn} + \sqrt{n\log(1/\varepsilon)})$ as desired. As for the sum of the third term, we have

$$\sum_{i \geq 1} O\left(\log(2^{2i+2}m^2)\sqrt{\frac{n}{2^i m}}\right) = O\left(\log(m)\sqrt{\frac{n}{m}}\right) = O(\sqrt{n}).$$

We now verify correctness. If $|x| \leq 2m$ then the algorithm will output $(|x| - k)(|x| - k - 1)$ in expectation exactly: if k ones are found in x by a call to HIGH then this will be done by SAMPLE, otherwise all ones in x will be found with certainty by LOW, which will then output correctly. If $|x| > 2m$ and a call to HIGH succeeds in finding k ones in x, the algorithm will also output $(|x| - k)(|x| - k - 1)$ exactly, in expectation. Let p be the probability that this does not happen, i.e., that the output on x is given by the procedure LOW. Then the expected value of the output on x is

$$(1 - p)(|x| - k)(|x| - k - 1) + p \cdot \mathbb{E}[\mathrm{LOW}(x, 2m)],$$

and the deviation from the desired output $(|x|-k)(|x|-k-1)$ is

$$p \cdot (\mathbb{E}[\mathrm{LOW}(x,2m)] - (|x|-k)(|x|-k-1)) .$$

Now $\mathrm{LOW}(x,2m)$ will always output a value $0 \leq (\ell-k)(\ell-k-1)$ for some $\ell \in [2m]$, which is always at most the correct value $(|x|-k)(|x|-k-1)$ as $|x| \geq 2m > k$. Therefore the largest difference between these is when $\mathrm{LOW}(x,2m)$ outputs 0, giving

$$|p \cdot (\mathbb{E}[\mathrm{LOW}(x,2m)] - (|x|-k)(|x|-k-1))| \leq p \cdot (|x|-k)(|x|-k-1) \leq p \cdot |x|^2 .$$

We now finally upper bound this error by giving an upper bound on p. Let i and $t = 2^i m$ be such that $t \leq |x| < 2t$. For this value of t and $\delta = \varepsilon/(4t^2)$ the call to $\mathrm{HIGH}(x,t,\delta)$ fails to find a set S of size k with probability at most $\delta \leq \varepsilon/|x|^2$. Thus $p \cdot |x|^2 \leq \delta \cdot |x|^2 \leq \varepsilon$, as desired. ◀

By Theorem 2.6, the characterization of sos degree in terms of quantum query complexity in expectation (Theorem 2.10) gives the upper bound in Theorem 1.1

3 Sum-of-squares approximation in ℓ_1-norm

In this section, we show upper bounds on the sos degree of polynomials to approximate f_k in ℓ_1-norm. In this section we focus on the case where k is $\lfloor n/2 \rfloor$. When $k < 0.49n$ the function f_k is quite easy to approximate in ℓ_1-norm: there is an sos polynomial of degree $O(\ln(1/\delta))$ which gives a $\delta 2^n$-approximation. We omit the details. [5] Our main result on ℓ_1-approximation is the following.

▶ **Theorem 3.1.** *Let n be odd and $k = \lfloor n/2 \rfloor$. Then for any $8/\sqrt{2n} \leq \delta \leq 1/4$*

$$\textit{sos-deg}_{\delta 2^n}(f_k, \ell_1) \leq \left\lceil \frac{3\sqrt{n}}{\sqrt{2}\delta} \ln\left(\frac{1}{\delta}\right) \right\rceil .$$

Lee, Raghavendra, and Steurer [19], building on work of Grigoriev [12], show that in this case $\text{sos-deg}_{2^n/\sqrt{n}}(f,\ell_1) \geq (n-1)/2$. This lower bound was then plugged into their general theorem to lift ℓ_1-approximate sos degree lower bounds to lower bounds on semidefinite extension complexity. By taking $\delta = 3\ln(n)/\sqrt{2n}$, Theorem 3.1 shows that this lower bound on the ℓ_1-error is tight, up to a logarithmic factor. Also, taking δ to be a small additive constant shows that there is a degree-$O(\sqrt{n})$ sos polynomial which, on average, disagrees with f_k by only a small constant. Taken as a whole, Theorem 3.1 implies that the quantitative bounds on the semidefinite extension complexity of the correlation polytope of [19] cannot be improved simply by improving the sos degree lower bounds on the f_k. We now describe the connection to [19] in greater detail.

3.1 The theorem of Lee, Raghavendra, and Steurer

For a function $f : \{0,1\}^n \to \mathbb{R}$ and an integer $N \geq n$, let $M_N^f : \binom{N}{n} \times \{0,1\}^N \to [0,1]$ be the matrix where $M_N^f(S,x) = f(x|_S)$. The *pattern matrix* of f, introduced in the work of

[5] One way to construct such an sos polynomial is to construct a polynomial e as mentioned after Theorem 1.2 from a classical sampling algorithm: query $O(\ln(1/\delta))$ randomly chosen input bits; output some large number if the observed ratio of 1s is very close to k/n, output 0 otherwise. This induces an sos polynomial with the right properties.

Sherstov [29], is a submatrix of M_N^f. The main theorem of Lee, Raghavendra, and Steurer is the following statement. Here a *degree-d pseudo-density* D is a function $D : \{0,1\}^n \to \mathbb{R}$ such that $\mathbb{E}_x[D(x)] = 1$ and $\mathbb{E}_x[D(x)g(x)^2] \geq 0$ for all polynomials g of degree at most $d/2$ on the boolean cube, with the expectation over a uniformly random $x \in \{0,1\}^n$. We use $\|D\|_\infty$ to denote $\max_{x\in\{0,1\}^n} |D(x)|$.

▶ **Theorem 3.2** **([19]).** *Let $f : \{0,1\}^n \to [0,1]$. If there exists an $\varepsilon \in (0,1]$ and a degree-d pseudo-density $D : \{0,1\}^n \to \mathbb{R}$ satisfying $\mathbb{E}_x[D(x)f(x)] < -\varepsilon$, then for every $N \geq 2n$*

$$\mathrm{rk}_{\mathrm{psd}}(M_N^f) \geq \left(\frac{c\varepsilon N}{dn^2\|D\|_\infty \log n}\right)^{d/4} \left(\frac{\varepsilon}{\|D\|_\infty}\right)^{3/2} \sqrt{\mathbb{E}_x f(x)}\,,$$

where $c > 0$ is a universal constant.

We do not formally define here the positive semidefinite (psd) rank of a matrix (denoted by $\mathrm{rk}_{\mathrm{psd}}$ above), but remark that psd rank lower bounds are equivalent to semidefinite extension complexity lower bounds. Lee, Raghavendra, and Steurer proved Theorem 3.2 en route to their breakthrough result on superpolynomial size lower bounds for semidefinite programming relaxations of hard optimization problems.

The more pertinent aspect of Theorem 3.2 to us is the role of the degree-d pseudo-density D. Note that, once we fix the degree of the pseudo-density, the bound only depends on the ratio $\mathbb{E}_x[D(x)f(x)]/\|D\|_\infty$. The largest such ratio that a degree-d pseudo-density can achieve is closely related to the best ℓ_1-approximation of f by degree-$d/2$ sos polynomials. The following claim follows from strong duality of semidefinite programming.

▶ **Claim 3.3.** *Let $f : \{0,1\}^n \to \mathbb{R}$. Then sos-deg$_{\delta 2^n}(f, \ell_1) > d$ if and only if there exists a "witness" function $\psi : \{0,1\}^n \to \mathbb{R}$ satisfying $\mathbb{E}_x[f(x)\psi(x)] > \delta$, and $\mathbb{E}_x[p^2(x)\psi(x)] \leq 0$ for all polynomials p of degree at most d, and $\|\psi\|_\infty = 1$.*

For n odd and $k = \lfloor n/2 \rfloor$ Lee et al. [19], building on work of Grigoriev [12], show that there is a degree-$(n-1)$ pseudo-density D such that $\mathbb{E}_x[D(x)f_{\lfloor n/2 \rfloor}]/\|D\|_\infty < -\frac{1}{4\sqrt{n}}$. Plugging $f = f_{\lfloor n/2 \rfloor}/n^2$ (with this normalization the range is in $[0,1]$) into Theorem 3.2 gives a lower bound of $2^{\widetilde{\Omega}(N^{2/11})}$ on the psd rank of M_N^f for $N = \widetilde{O}(n^{11/2})$. As M_N^f is a submatrix of the slack matrix of the correlation polytope, this gives the desired lower bound on the semidefinite extension complexity of the correlation polytope.

In light of Claim 3.3, if sos-deg$_{\delta 2^n}(f, \ell_1) \leq d$, then there can be no degree-$2d$ pseudo-density with $\mathbb{E}_x[D(x)f(x)]/\|D\|_\infty < -\delta$. The ℓ_1-approximate sos degree upper bounds of Theorem 3.1 therefore imply the non-existence of pseudo-densities with good properties for Theorem 3.2. It can be verified that $2^{\widetilde{\Omega}(N^{2/11})}$ is in fact the best quantitative bound that Theorem 3.2 can show on $\mathrm{rk}_{\mathrm{psd}}(M_N^{f_k})$ over all the functions f_k and tradeoffs between δ and sos-deg$_{\delta 2^n}(f_k, \ell_1)$.

3.2 Proof of Theorem 3.1

Throughout this proof we set $f = f_{\lfloor n/2 \rfloor}$. The main idea of the proof of Theorem 3.1 is to construct a univariate polynomial p such that $h(z) = (z - \lfloor n/2 \rfloor)(z - \lceil n/2 \rceil) + p(z)$ is globally nonnegative (and therefore sos) and $\sum_{i=0}^n \binom{n}{i}|p(i)|$, which is the ℓ_1-error of $h(|x|)$ in approximating f, is reasonably small. We will construct p using Chebyshev polynomials. Similar constructions to what we need have been done before, see for example [28]; as our requirements are somewhat specific, however, we do the construction from scratch.

Let T_d be the Chebyshev polynomial of degree d. We first recall some basic facts about Chebyshev polynomials [27].

▶ **Fact 3.4.** *Let $T_d(z)$ be the Chebyshev polynomial of degree d. Then*

1. $|T_d(z)| \leq 1$ *for* $z \in [-1, 1]$.
2. $T_d(z) = \frac{1}{2}\left((z - \sqrt{z^2-1})^d + (z + \sqrt{z^2-1})^d\right)$.
3. $T_{d+1}(z) = 2zT_d(z) - T_{d-1}(z)$.
4. $T_d(z)$ *is monotonically increasing for* $z \geq 1$, *and if* d *is even* $T_d(z) \geq 1$ *for* $z \in \mathbb{R} \setminus [-1, 1]$.

▶ **Theorem 3.5.** *Let $n \geq 1$ be an integer, $\varepsilon \in (0, 1/4]$ an error parameter, and let $a \in \mathbb{R}$ satisfy $1/\sqrt{2} \leq a \leq \sqrt{n}/8$. There is a polynomial p of degree at most*

$$\left\lceil \frac{3n}{4\sqrt{2}\,a} \ln\left(\frac{1}{2\varepsilon}\right) \right\rceil + 1$$

with the following properties:

1. $p(z) \geq \frac{1}{4} - z^2$ *for all* z.
2. $|p(z)| \leq \varepsilon$ *for* $z \in [-\frac{n}{2}, -a] \cup [a, \frac{n}{2}]$.
3. $|p(z)| \leq 2$ *for* $z \in [-\frac{n}{2}, -\frac{1}{2}] \cup [\frac{1}{2}, \frac{n}{2}]$.

Proof. Note that we require in particular that $p(0) \geq 1/4$. Roughly speaking, p should have a ‘peak’ around 0 and then quickly calm down and be bounded on either side of this peak once $|z| \geq a$. The difficulty in constructing p is that its peak is *in between* the intervals on which is bounded. To get around this, we note that is suffices to let $p(z) = \varepsilon q(z^2)$, where q has the properties

1. $q(z) \geq \frac{1}{4\varepsilon} - \frac{z}{\varepsilon}$ for $z \geq 0$.
2. $|q(z)| \leq 1$ for $z \in [a^2, \frac{n^2}{4}]$.
3. $|q(z)| \leq \frac{2}{\varepsilon}$ for $z \in [\frac{1}{4}, \frac{n^2}{4}]$.

Now $q(z)$ is ripe for a construction with Chebyshev polynomials, and this is what we do. For notational convenience, let $L = n/2$. Define the mapping $s(z) = -2(z - a^2)/(L^2 - a^2) + 1$ that takes the interval $[a^2, L^2]$ to $[-1, 1]$. Note that this mapping takes L^2 to -1 and a^2 to 1. Let T_d be the Chebyshev polynomial of *even* degree d (to be chosen later) and define

$$q(z) = T_d(s(z)) .$$

As $|T_d(z)| \leq 1$ for $z \in [-1, 1]$ by Fact 3.4 item (1), it follows that $q(z)$ satisfies condition (2).

We now turn to item (1) and handle the easy cases first. For $z \geq \frac{1}{4}$ we have $\frac{1}{4\varepsilon} - \frac{z}{\varepsilon} \leq 0$, so in this region we just need to check that $q(z)$ is not too negative. If $z \in [\frac{1}{4}, a^2]$, then $s(z) \geq 1$ and therefore $q(z) \geq 1$. Likewise, as we take d to be even, $q(z) \geq 1$ for $z \geq L^2$. For $z \in [a^2, L^2]$, we have $|q(z)| \leq 1$. Thus item (1) will be satisfied in this region so long as $a^2 \geq \frac{1}{4} + \varepsilon$. This holds as in the theorem statement $\varepsilon \leq 1/4$ and $a \geq 1/\sqrt{2}$.

With these easy cases taken care of, we turn to verify the first item for $z \in [0, \frac{1}{4}]$. To do this it suffices to choose d such that $q(1/4) \geq \frac{1}{4\varepsilon}$ as $q(z)$ is monotonically decreasing in the interval $[0, 1/4]$, since $T_d(y)$ is monotonically increasing for $y \geq 1$ by Fact 3.4 item (4). This condition is at odds with item (3). As the maximum of $q(z)$ in the interval $[1/4, L^2]$ is attained at $z = 1/4$, we can simultaneously satisfy item (3) by ensuring $q(1/4) \leq \frac{2}{\varepsilon}$. Thus we choose $d = d^*$ to be the least even number such that

$$q(1/4) = T_{d^*}\left(1 + \frac{2(a^2 - 1/4)}{L^2 - a^2}\right) \geq \frac{1}{4\varepsilon} .$$

By this choice, item (1) is now satisfied. To verify item (3), we use Fact 3.4 item (3) to see the inequality $T_{s+2}(z) \leq 4z^2 T_s(z)$, valid for $z \geq 1$. Applying this we have

$$T_{d^*}\left(1 + \frac{2(a^2 - 1/4)}{L^2 - a^2}\right) \leq \frac{1}{\varepsilon}\left(1 + \frac{2(a^2 - 1/4)}{L^2 - a^2}\right)^2 \leq \frac{2}{\varepsilon} ,$$

as $T_{d^*-2}\left(1 + \frac{2(a^2-1/4)}{L^2-a^2}\right) < \frac{1}{4\varepsilon}$ by definition, and $a \leq \sqrt{n}/8$.

Finally, we upper bound d^*. Let $\mu = 2(a^2 - 1/4)/(L^2 - a^2)$. We want $T_d(1+\mu) \geq \frac{1}{4\varepsilon}$. Using the fact that $T_d(1+\mu) \geq (1/2)(1+\sqrt{2\mu})^d$ for $\mu \geq 0$ by Fact 3.4 item (2), it suffices to take $d \geq \ln(\frac{1}{2\varepsilon})/\ln(1+\sqrt{2\mu})$.

As $\ln(1+y) \geq 2y/(2+y)$ for $y \geq 0$ and $\sqrt{2\mu} \leq 1$, it suffices to take $d \geq 3\ln(\frac{1}{2\varepsilon})/(2\sqrt{2\mu})$. Since $a \geq 1/\sqrt{2}$, and therefore $a^2 - 1/4 \geq a^2/2$, we have $\mu \geq a^2/L^2 = 4a^2/n^2$. Hence there is a d such that $T_d(1+\mu) \geq \frac{1}{4\varepsilon}$ satisfying

$$d \leq \left\lceil \frac{3n}{4\sqrt{2}\,a} \ln\left(\frac{1}{2\varepsilon}\right) \right\rceil .$$

We add 1 in the theorem statement for the additional requirement that the degree is even. ◀

Proof of Theorem 3.1. Fix $8/\sqrt{2n} \leq \delta \leq 1/4$, and let $\varepsilon = \delta/2$ and $a = \varepsilon\sqrt{n}/4$. Note that $1/\sqrt{2} \leq a \leq \sqrt{n}/8$ with these choices. Thus by Theorem 3.5, there is a polynomial p of degree at most $\lceil \frac{6\sqrt{n}}{\sqrt{2}\delta} \ln(\frac{1}{\delta}) \rceil + 1$ satisfying the three conditions of Theorem 3.5 with this value of a, ε.

Let $g(z) = (z - n/2)^2 - 1/4 + p(z - n/2)$ be a univariate polynomial, and consider the approximation to f given by $g(|x|)$. By construction g is globally nonnegative and thus (as it is univariate) is a sum of squares of polynomials of degree at most $\lceil \frac{3\sqrt{n}}{\sqrt{2}\delta} \ln(\frac{1}{\delta}) \rceil$. Let us examine the ℓ_1-error of the function $g(|x|)$ in approximating f. We divide the error into two cases: the error on strings whose Hamming weight is at most $n/2 - a$ or at least $n/2 + a$ (type I), and those whose Hamming weight is in the interval $[n/2 - a, n/2 + a]$ (type II).

As p is bounded by ε for $z \in [-n/2, -a] \cup [a, n/2]$ the ℓ_1-error over type I inputs is at most $\varepsilon \cdot 2^n$. The number of type II inputs is at most $(2a/\sqrt{n})2^n$, and the error on each is at most 2 as $p(z) \leq 2$ for $z \in [-n/2, n/2]$. Thus the total ℓ_1-error is $2^n \left(\varepsilon + \frac{4a}{\sqrt{n}}\right) = 2^n \cdot 2\varepsilon = \delta 2^n$. ◀

4 Proof complexity: Positivstellensatz refutations

Say that we have a system of polynomial equalities

$$f_1 = \cdots = f_m = 0,\ x_1^2 - x_1 = \cdots = x_n^2 - x_n = 0 \tag{4}$$

where each $f_i \in \mathbb{R}[x_1, \ldots, x_n]$. Because of the presence of the equalities $x_i^2 - x_i = 0$ (which force $x_i \in \{0,1\}$), this is referred to as the *boolean* setting.

The Positivstellensatz [32] implies that the system (4) has no common solutions in $\mathbb{R}^n$ if and only if there are polynomials $g_1, \ldots, g_{m+n} \in \mathbb{R}[x_1, \ldots, x_n]$ and a sos polynomial h such that

$$\sum_{i=1}^{m} f_i g_i + \sum_{i=1}^{n} (x_i^2 - x_i) g_{m+i} = 1 + h\,. \tag{5}$$

Grigoriev and Vorobjov [14] define a proof system based on this principle.

▶ **Definition 4.1.** A *Positivstellensatz refutation* of the system (4) is given by a set of polynomials $\{g_1, \ldots, g_{m+n}, h\}$ which satisfy (5) and where h is a sum of squares. The *degree* of this refutation is

$$\max\{\deg(h), \max_{i\in[m]} \deg(f_i g_i), \max_{i\in[n]} \deg((x_i^2 - x_i) g_{m+i})\}\,.$$

By the Positivstellensatz, this proof system is sound and complete: a system is unsatisfiable if and only if it has a refutation of a certain degree. One may view the degree of a refutation as a measure of complexity.

4.1 Knapsack

The knapsack system is given by the equations

$$f = \sum_i x_i - r = 0,\ x_j^2 - x_j = 0 \text{ for } j = 1, \ldots, n\,. \tag{6}$$

If r is not an integer then this system has no solution: Grigoriev [12] shows the following theorem.

▶ **Theorem 4.2** (Grigoriev [12]). *Let $0 \leq k \leq (n-3)/2$ be an integer. If $k < r < n-k$, then any Positivstellensatz refutation of the system (6) has degree at least $2k+4$.*

We provide a simple proof of this in Appendix C using Blekherman's theorem.

Note that the equations for non-integer r correspond to a trivially easy (and obviously unsatisfiable) instance of the knapsack problem, where all items have weight 1. As mentioned in the introduction, this shows the weakness of the Positivstellensatz-based proof system: even to refute such easy instances it already needs polynomials of fairly high degree.

Grigoriev asked if this upper bound of $2k+4$ was tight. Later work of Grigoriev et al. [13] showed that the proof technique of [12] could not show a larger lower bound than $2k+4$. We show that there actually exist Positivstellensatz refutations of (6) of degree $2k+4$.

▶ **Theorem 4.3.** *Let $0 \leq k \leq n/2$ be an integer. For $k < r < k+1$, the system (6) has a Positivstellensatz refutation of degree $2k+4$.*

Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $|\mathbf{x}| = \sum_{i=1}^n x_i$. A key role in the proof will be played by the polynomials

$$A_k(\mathbf{x}) = |\mathbf{x}|(|\mathbf{x}|-1)(|\mathbf{x}|-2)\cdots(|\mathbf{x}|-k+1)\,.$$

The function A_k can be computed with k queries by a natural extension of the Sampling Algorithm 1 and thus can be written as a sum-of-squares on the boolean cube of total degree $2k$. We go ahead and record this formally in the next lemma. Recall that the kth elementary symmetric polynomial is defined as

$$e_k(x_1, \ldots, x_n) = \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k}\,.$$

▶ **Lemma 4.4.** *There exist polynomials $g_i(\mathbf{x})$ of degree at most $2k-2$ such that*

$$A_k(\mathbf{x}) = \sum_{i=1}^n (x_i^2 - x_i) g_i(\mathbf{x}) + (k!) e_k(x_1^2, \ldots, x_n^2)\,.$$

We give the proof of this in Appendix A.

Proof of Theorem 4.3. Rearranging Equation (5), we are looking for functions $g, g_1, \ldots, g_n$ of low degree and a low-degree sum-of-squares h such that

$$g(\mathbf{x})(|\mathbf{x}| - r) - 1 = h + \sum_i g_i(\mathbf{x})(x_i^2 - x_i)\,.$$

Notice that, for any g, the left-hand side will be negative when $|\mathbf{x}| = r$. By Lemma 4.4, A_{k+2} is of the form of the right-hand side. Since A_{k+2} has degree $2k+4$, and is also negative when $|\mathbf{x}| = r$, we try to find a polynomial $g(\mathbf{x})$ of degree at most $2k+3$ such that

$$g(\mathbf{x})\,(|\mathbf{x}| - r) - b = A_{k+2}(\mathbf{x})$$

for a positive constant b. Dividing g and A_{k+2} by b will then give us the required solution. Let $b = -r(r-1)\cdots(r-k)(r-k-1) > 0$. Then $|\mathbf{x}| - r$ divides $A_{k+2}(\mathbf{x}) + b$ and we can write $A_{k+2}(\mathbf{x}) + b = g(\mathbf{x})\,(|\mathbf{x}| - r)$ for some polynomial g of degree $2k+3$. ◀

5 Future work

We list a few questions for future work:

- Can we improve the lower bound of Theorem 1.1 for small ε? To match the upper bound for all k, it would suffice to show that $\text{sos-deg}_{\varepsilon}(f_1, \ell_\infty) = \Omega(\sqrt{n \log(1/\varepsilon)})$, which is very plausible by analogy with what is known for the n-bit OR function.
- Can we extend our results to all symmetric quadratic functions, or to even larger classes of symmetric functions?
- Can we find more applications of Blekherman's theorem (Theorem 2.3), in complexity theory, in quantum computing, or in optimization? Kurpisz et al. [17, Section 5] used their general reduction to univariate polynomials (already mentioned in Section 1.3.2), to show that strengthening the knapsack polytope with Wolsey's "Knapsack Covering Inequalities" and applying nearly $\log n$ rounds of the Lasserre hierarchy does not produce an SDP with integrality gap below $2 - o(1)$ (which is the integrality gap of the natural LP relaxation). Similar results may be obtainable using Blekherman's theorem.

Acknowledgments. We thank Mario Szegedy for early discussions on lower bounding the sos degree of f_1, Srinivasan Arunachalam for many useful discussions and comments, Greg Blekherman for discussions about [2], and Adam Kurpisz for alerting us to [17]. We thank the authors of the software SDPT3 [33] and YALMIP [20], which we used to compute the error of sos polynomials in approximating f_k on small examples. TL would like to thank Savaroo Ranch for their hospitality and BBQ during the course of this research.

References

1 R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98.

2 G. Blekherman. Symmetric sums of squares on the hypercube. *Manuscript in preparation*, 2015.

3 G. Blekherman and C. Riener. Symmetric nonnegative forms and sums of squares, 2012. arXiv:1205.3102.

4 G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *AMS Contemporary Mathematics Series*, pages 53–74. AMS, 2002. quant-ph/0005055.

5 H. Buhrman, R. Cleve, R. de Wolf, and Ch. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of 40th IEEE FOCS*, pages 358–368, 1999. cs.CC/9904019.

6 T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli. *Harmonic analysis on finite groups: representation theory, Gelfand pairs and Markov chains*, volume 108. Cambridge University Press, 2008.

7 A. Childs and T. Lee. Optimal quantum adversary lower bounds for ordered search. In *Proceedings of 35th International Colloquium on Automata, Languages and Programming (ICALP'08)*, pages 869–880, 2008.

8 Y. Filmus and E. Mossel. Harmonicity and invariance on slices of the boolean cube. In *Proceedings of 31st CCC*, 2016. arXiv:1507.02713.

9 S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM*, 16(2), 2015. arxiv/1111.0837. Earlier version (under a different name) in STOC'12.

10 C. Godsil. Association schemes, 2010. Lecture notes available on `http://www.math.uwaterloo.ca/~cgodsil/pdfs/assoc2.pdf`, version 1.

11 M. de Graaf and R. de Wolf. On quantum versions of the Yao principle. In *Proceedings of 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS'02)*, pages 347–358, 2002. quant-ph/0109070.

12 D. Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001.

13 D. Grigoriev, E. Hirsch, and D. Pasechnik. Complexity of semialgebraic proofs. *Moscow Mathematical Journal*, 2(4):647–679, 2002.

14 D. Grigoriev and N. Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113:153–160, 2001.

15 L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996. quant-ph/9605043.

16 J. Kaniewski, T. Lee, and R. de Wolf. Query complexity in expectation. In *Proceedings of 42nd International Colloquium on Automata, Languages and Programming (ICALP'15)*, pages 761–772, 2015. arXiv:1411.7280.

17 A. Kurpisz, S. Leppänen, and M. Mastrolilli. Sum-of-squares hierarchy lower bounds for symmetric formulations. In *Proceedings of 18th IPCO*, 2016. arXiv:1407.1746.

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

19 J. R. Lee, P. Raghavendra, and D. Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of 47th ACM STOC*, pages 567–576, 2015.

20 J. Löfberg. YALMIP: A toolbox for modeling and optimization in Matlab. In *Proceedings of the CACSD conference*, 2004.

21 M. Minsky and S. Papert. *Perceptrons*. MIT press, Cambridge, MA, 1968. Second, expanded edition 1988.

22 N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

23 R. O'Donnell and R. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.

24 P. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

25 R. Paturi. On the degree of polynomials that approximate symmetric boolean functions. In *Proceedings of 24th ACM STOC*, pages 468–474, 1992.

26 M. Petkovšek, H. Wilf, and D. Zeilberger. *"A=B"*. Springer Science & Business Media, 1997.

27 T. Rivlin. *An introduction to the approximation of functions*. Dover publications, inc., 1969.

28 A. Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. *Computational Complexity*, 18:219–247, 2009.

29 A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

30 Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum information and computation*, 9(5,6):444–460, 2009. URL: `http://www.rintonpress.com/xxqic9/qic-9-56/0444-0460.pdf`.

31 N. Z. Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23:695–700, 1987.

32 G. Stengle. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207:87–97, 1974.

33 K.C. Toh, M.J. Todd, and R.H. Tutuncu. SDPT3 – a Matlab software package for semidefinite programming. *Optimization methods and software*, 11:545–581, 1999.

34 R. de Wolf. A note on quantum algorithms and the minimal degree of epsilon-error polynomials for symmetric functions. *Quantum Information and Computation*, 8:943–950, 2008.

A Proof of Lemma 4.4

Recall that

$$A_k(x_1, \ldots, x_n) = |\mathbf{x}|(|\mathbf{x}| - 1) \cdots (|\mathbf{x}| - k + 1) .$$

We first prove two claims.

▶ **Claim A.1.** *There exist polynomials $g_i(\mathbf{x})$ of degree at most $k-1$ such that*

$$A_k(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_i^2 - x_i) g_i(\mathbf{x}) + (k!) e_k(\mathbf{x}) .$$

Proof. We prove the claim by induction on k. When $k = 1$ then $A_1 = e_1$, and the claim follows by setting all g_i to be 0.

Now suppose the claim is true up to k. Then, using the induction hypothesis to rewrite A_k,

$$\begin{aligned} A_{k+1}(\mathbf{x}) = A_k(\mathbf{x}) \cdot (e_1(\mathbf{x}) - k) &= \left(\sum_{i=1}^{n} (x_i^2 - x_i) g_i(\mathbf{x}) + k! e_k(\mathbf{x}) \right) (e_1(\mathbf{x}) - k) \\ &= \sum_{i=1}^{n} (x_i^2 - x_i) h_i(\mathbf{x}) + k! e_k(\mathbf{x})(e_1(\mathbf{x}) - k) , \end{aligned}$$

where each $h_i(\mathbf{x}) = g_i(\mathbf{x}) \cdot (e_1(\mathbf{x}) - k)$ is of degree at most k. We now focus on

$$e_k(\mathbf{x})(e_1(\mathbf{x}) - k) = \sum_{\substack{S \subseteq [n] \\ |S| = k}} \prod_{i \in S} x_i \cdot (e_1(\mathbf{x}) - k) .$$

A term in this sum corresponding to the subset S can be rewritten as

$$\prod_{i \in S} x_i \left(\sum_{i \in S} x_i + \sum_{i \notin S} x_i - k \right) = \sum_{i \in S} (x_i^2 - x_i) \prod_{j \in S, j \neq i} x_j + \sum_{i \notin S} x_i \prod_{j \in S} x_j$$

Summing over all terms of this form gives $(k+1)! e_{k+1}(\mathbf{x}) + \sum_i (x_i^2 - x_i) \cdot f_i(\mathbf{x})$, where $f_i(\mathbf{x})$ is of degree at most $k-1$, proving the claim. ◀

To complete the proof, we now need to show that $e_d(\mathbf{x})$ is a sum of squares of total degree $2d$ modulo the ideal $\langle x_1^2 - x_1, \ldots, x_n^2 - x_n \rangle$. To do this, it suffices to show the same for $\prod_{i=1}^{d} x_i$, which we do in the next claim.

▶ **Claim A.2.** *Fix a natural number d. Then there are polynomials $g_i \in \mathbb{R}[x_1, \ldots, x_d]$ for $i = 1, \ldots, d$ such that*

$$\prod_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} x_i = \sum_{i=1}^{d} (x_i^2 - x_i) g_i(x_1, \ldots, x_d) ,$$

and each g_i is of degree at most $2d - 2$.

Proof. We write $x_1^2 x_2^2 \cdots x_d^2 - x_1 x_2 \cdots x_d$ as a telescoping sum. We use the convention that the product over the empty set is 1.

$$\begin{aligned}\prod_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} x_i &= \sum_{j=1}^{d} \left(\prod_{i<j} x_i \prod_{i\geq j} x_i^2 - \prod_{i\leq j} x_i \prod_{i>j} x_i^2 \right) \\ &= \sum_{j=1}^{d} (x_j^2 - x_j) \prod_{i<j} x_i \prod_{i>j} x_i^2\end{aligned}$$

This is of the desired form, and it can be seen that each multiplier of $x_j^2 - x_j$ is of degree at most $2d-2$. ◀

We put these claims together to prove Lemma 4.4, which we restate here.

▶ **Lemma A.3.** *There exist polynomials $g_i(\mathbf{x})$ of degree at most $2k-2$ such that*

$$A_k(\mathbf{x}) = \sum_{i=1}^{n} (x_i^2 - x_i) g_i(\mathbf{x}) + (k!) e_k(x_1^2, \ldots, x_n^2)\,.$$

Proof. By Claim A.1 we can write $A_k(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_i^2 - x_i) g_i(\mathbf{x}) + (k!) e_k(\mathbf{x})$ where each $g_i(\mathbf{x})$ is of degree at most $k-1$. Now by Claim A.2

$$e_k(\mathbf{x}) = e_k(x_1^2, \ldots, x_n^2) + \sum_{i=1}^{n} (x_i^2 - x_i) \cdot f_i(\mathbf{x})\,,$$

where each $f_i(\mathbf{x})$ is of degree at most $2k-2$. This proves the lemma. ◀

B Blekherman's theorem

Blekherman and Riener [3] made a general study of the relationship between symmetric nonnegative forms and symmetric sums of squares. Subsequently, Blekherman [2] considered the special case of polynomials that are nonnegative on the hypercube, and gave a very useful decomposition of such polynomials. We include a proof here of a special case of his theorem.

The technique used for our proof is a novel decomposition of functions on the hypercube using the kernels of certain differential operators. A similar decomposition was independently discovered by Filmus and Mossel [8] who use it to prove an invariance principle for low-degree functions on slices of the boolean hypercube.

Let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. The ideal $\mathcal{I} := \langle x_i^2 - x_i : i \in [n] \rangle$ consists of polynomials that are identically zero on the hypercube $\mathcal{H} = \{0,1\}^n$. Let $L_t = \mathbb{R}[x]_t/\mathcal{I}$ be the space of degree-t homogeneous multilinear polynomials on n variables. The $\binom{n}{t}$ monomials x^S where $S \subseteq [n], |S| = t$, form a basis for L_t. The correspondence between set S and monomial x^S can be used to map degree-t polynomials to linear combinations of t-subsets of $[n]$. Here polynomial $p(x)$ corresponds to the $\binom{n}{t}$-dimensional vector of the coefficients of its monomials, say in lexicographic order.

Let $M_t = \mathbb{R}[x]_{\leq t}/\mathcal{I}$ denote the space of n-variate polynomials of degree at most t on the hypercube. Given $x \in \mathbb{R}^n$, the sum $\sum_{i\in[n]} x_i$ is denoted by $|x|$.

B.1 Kernels of the operators W_t

For $t \geq 0$, define the linear operator W_t that acts by summing over the partial derivatives of a degree-t polynomial,

$$W_t p(x) = \left(\sum_{i \in [n]} \frac{\partial}{\partial x_i} \right) p(x)\,. \tag{7}$$

For $t \geq 1$ the operator $W_t : L_t \to L_{t-1}$ is represented by a matrix with rows and columns indexed by $S, T \subset [n]$ with $|S| = t-1$ and $|T| = t$ respectively, and entry $(W_t)_{S,T} = 1$ if $S \subset T$ and 0 otherwise. The adjoint operator $W_t^T : L_{t-1} \to L_t$ acts as multiplication by $|x| - t + 1$ on each degree-$(t-1)$ monomial (and by linear extension on all of L_{t-1}),

$$W_t^T x^S = \sum_{i \notin S} x^{S \cup \{i\}} = x^S(|x| - t + 1)\,. \tag{8}$$

Note that we used the hypercube constraints $x_i^2 = x_i$ to derive the second equality in Equation (8). Our goal in this section is to bound the dimension of $\mathrm{Ker}(W_t)$ and find an explicit basis for these spaces.

We relate $\mathrm{Ker}(W_t)$ to the eigenspaces of the Johnson graphs. The Johnson graph $J(n,t)$ has $\binom{n}{t}$ vertices corresponding to the t-subsets $S \subset [n], |S| = t$, with subsets S, T connected by an edge if and only if $|S \cap T| = t - 1$. The adjacency matrix of $J(n,t)$ is denoted by $A_J(n,t)$. The following lemma computes the spectrum of $A_J(n,t)$; it can be found in Godsil's notes [10], but we include a proof here as these notes are no longer online.

▶ **Theorem B.1.** *The eigenvalues of $A_J(n,t)$ are $t(n-t) - i(n+1-i)$ with multiplicity $\binom{n}{i} - \binom{n}{i-1}$ for $i = \{0, 1, \ldots, t\}$ and $t \leq (n+1)/2$.*

Proof. We proceed by induction on n and t. For the base case, note that the Johnson graph $J(n,1)$ is the complete graph on n vertices. The corresponding adjacency matrix $A_J(n,1)$ has eigenvalue -1 with multiplicity $(n-1)$ and $(n-1)$ with multiplicity 1, thus the theorem is true for $n = 1$.

We obtain the spectrum of $A_J(n,t)$ in terms of the spectrum of $A_J(n,t-1)$. Computing the entries of $W_t^T W_t$ and $W_t W_t^T$ it follows that,

$$\begin{aligned} W_t^T W_t &= tI + A_J(n,t) \\ W_t W_t^T &= (n-t+1)I + A_J(n,t-1)\,. \end{aligned} \tag{9}$$

The non-zero eigenspaces of $W_t^T W_t$ correspond to those of $W_t W_t^T$, so if v is an eigenvector for $A_J(n,t-1)$ with eigenvalue λ_i then $W_t^T v$ is an eigenvector for $A_J(n,t)$ with eigenvalue $\lambda_i + n - 2t + 1$. By the induction hypothesis, $(t-1)(n-t+1) - i(n+1-i)$ is an eigenvalue for $A_J(n,t-1)$ with multiplicity $\binom{n}{i} - \binom{n}{i-1}$ for $i \in [t-1]$. Adding $n - 2t + 1$, it follows that $t(n-t) - i(n+1-i)$ is an eigenvalue for $A_J(n,t)$ with the same multiplicity.

The induction hypothesis also implies that $W_t W_t^T = (n-t+1)I + A_J(n,t-1)$ has rank $\binom{n}{t-1}$ as it is positive semidefinite and the smallest eigenvalue is $n - 2t + 2 > 0$. Hence the $\binom{n}{t}$-dimensional matrix $W_t^T W_t$ has rank $\binom{n}{t-1}$, so its kernel has dimension $\binom{n}{t} - \binom{n}{t-1}$. This implies that $A_J(n,t)$ has an eigenspace of dimension $\binom{n}{t} - \binom{n}{t-1}$ with eigenvalue $-t = t(n-t) - t(n+1-t)$. ◀

The following corollary computes the dimension of $\mathrm{Ker}(W_t)$.

▶ **Lemma B.2.** *$Dim(Ker(W_t)) = \binom{n}{t} - \binom{n}{t-1}$ for $t \leq (n+1)/2$.*

Proof. $\mathrm{Dim}(\mathrm{Ker}(W_t)) = \binom{n}{t} - \mathrm{rank}(W_t^T)$ by definition, and from the above proof it follows that $\mathrm{rank}(W_t^T) = \mathrm{rank}(W_t W_t^T) = \binom{n}{t-1}$ for $t \leq (n+1)/2$. ◀

We next compute an explicit basis for $\mathrm{Ker}(W_t)$, viewed as a subspace of L_t. We recall the notion of a standard Young tableau of shape $(n-t,t)$ to describe the basis.

▶ **Definition B.3.** A standard Young tableau $\mathcal{U}$ of shape $(n-t,t)$ is an arrangement of $[n]$ in an array with two rows of size $n-t$ and t respectively, such that each row and column is sorted in ascending order.

The basis for $\mathrm{Ker}(W_t)$ described by the following theorem will be used for computations in the following sections. Note that polynomials $p_{\mathcal{U}}$ in this basis evaluate to 0 for all $x \in \{0,1\}^n$ with $|x| \in \{0,1,\ldots,t-1\} \cup \{n, n-1, \ldots, n-t+1\}$.

▶ **Theorem B.4.** *For $t \leq n/2$ and $\mathcal{A} = (a(1), a(2), \ldots, a(2t))$ an array of distinct elements $a(i) \in [n]$, define the polynomial $p_{\mathcal{A}}(x) := \prod_{i\in[t]}(x_{a(2i-1)} - x_{a(2i)})$.*

The polynomials $p_{\mathcal{U}}(x)$, where $(u(2i-i), u(2i))$ for $i \in [t]$ are the entries of the i-th column of a standard $(n-t,t)$ Young tableau $\mathcal{U}$, form a basis for Ker(W_t).

Proof. We first show that for all $|\mathcal{A}| = 2t$, the degree-t polynomial $p_{\mathcal{A}}(x)$ belongs to the kernel of W_t. Computing the partial derivatives of $p_{\mathcal{A}}(x)$,

$$\frac{\partial}{\partial x_j} p_{\mathcal{A}}(x) = \begin{cases} p_{\mathcal{A}}(x)/(x_{a(2i-1)} - x_{a(2i)}) \text{ if } j = a(2i-1) \\ -p_{\mathcal{A}}(x)/(x_{a(2i-1)} - x_{a(2i)}) \text{ if } j = a(2i) \\ 0 \text{ otherwise} \end{cases} \tag{10}$$

Summing over the partial derivatives and using Equation (7) it follows that $W_t p_{\mathcal{A}}(x) = 0$.

The set of polynomials $\{p_{\mathcal{A}}(x) : |\mathcal{A}| = 2t\}$ is not linearly independent. The straightening algorithm for Young tableaux (see for example Section 10.5 in [6]) shows that the polynomials $p_{\mathcal{U}}(x)$ where $(u(2i-i), u(2i))$ for $i \in [t]$ are entries of the i-th column of a standard $(n-t,t)$ Young tableau $\mathcal{U}$ form a basis for $\mathrm{Span}\{p_{\mathcal{A}}(x) : |\mathcal{A}| = 2t\}$. A simple counting argument or the hook length formula [6] shows that the number of such $\mathcal{U}$ is $\binom{n}{t} - \binom{n}{t-1}$. These $p_{\mathcal{U}}$ together thus span a space of dimension $\binom{n}{t} - \binom{n}{t-1}$, which is $\mathrm{Dim}(\mathrm{Ker}(W_t))$ by Lemma B.2. Hence the $p_{\mathcal{U}}$ form a basis for $\mathrm{Ker}(W_t)$. ◀

B.2 Polynomial decompositions

The action of the operators W_t yields the decomposition $L_t = \mathrm{Ker}(W_t) \oplus \mathrm{Im}(W_t^T)$. Applying this decomposition iteratively we obtain the following theorem,

▶ **Theorem B.5.** *A polynomial $p(x) \in L_t$ can be decomposed as*

$$\begin{aligned} p(x) &= p_t(x) + (|x|-t+1)p_{t-1}(x) + \cdots + (|x|-t+1)\cdots(|x|-1)|x|p_0(x) \\ &= p_t(x) + \sum_{i=1}^{t} p_{t-i}(x) \prod_{j=1}^{i} (|x|-t+j) \end{aligned} \tag{11}$$

where $p_{t-i}(x) \in$ Ker(W_{t-i}).

Proof. We proceed by induction on t. For the base case $t = 0$, observe that a degree-0 polynomial belongs to $\mathrm{Ker}(W_0)$. For the inductive step, a polynomial $p(x) \in L_t$ can be written as $p_t(x) + q(x)$ where $p_t(x) \in \mathrm{Ker}(W_t)$ and $q(x) \in \mathrm{Im}(W_t^T)$. The action of W_t^T on polynomials in L_{t-1} is described by Equation (8): for all $g(x) \in L_{t-1}$ we have

$$W_t^T g(x) = (|x| - t + 1)g(x)\,. \tag{12}$$

As $q(x) \in \mathrm{Im}(W_t^T)$, it can be factored as $q(x) = (|x| - t + 1)h(x)$ where $h(x) \in L_{t-1}$. The result follows using the induction hypothesis for $g(x)$. ◀

Applying the above theorem to the subspaces L_j ($j \in \{0, \dots, t\}$) that are contained in M_t and collecting the terms corresponding to $\mathrm{Ker}(W_j)$, we obtain the following decomposition for polynomials in M_t.

▶ **Corollary B.6.** *A polynomial $p(x) \in M_t$ can be decomposed as $p(x) = \sum_{j=0}^{t} q_j(x)$, where*

$$q_j(x) = \sum_{0 \le i \le t-j} |x|^i p_{ij}(x) \tag{13}$$

such that each $p_{ij}(x) \in Ker(W_j)$.

Proof. A polynomial $p(x) \in M_t$ can be written as $p(x) = \sum_{i=0}^{t} p_i(x)$ where $p_i(x) \in L_i$ is the homogeneous degree-i component of p. Applying Theorem B.5 to each $p_i(x)$ and collecting all the terms over the Equations (11) with prefactors in $\mathrm{Ker}(W_j)$, we obtain a decomposition $p(x) = \sum_{j\in[t]} q'_j(x)$ such that

$$\begin{aligned} q'_j(x) = p'_{jj}(x) + p'_{j+1,j}(x)(|x| - j) + p'_{j+2,j}(x)(|x| - j)(|x| - j + 1) + \cdots \\ \cdots + p'_{t,j}(x)(|x| - j)(|x| - j + 1) \cdots (|x| - t + 1)\,. \end{aligned} \tag{14}$$

Note that the indices in the above equation increase, because $\mathrm{Ker}(W_j)$ occurs in the decompositions of $p_i(x)$ for $i \ge j$. Let $p_{ij}(x)$ be the coefficient of $|x|^i$ for $0 \le i \le t - j$ in the above expression. This $p_{ij}(x)$ is a linear combination of polynomials $p'_{ij}(x) \in \mathrm{Ker}(W_j)$ and therefore also lies in $\mathrm{Ker}(W_j)$. The decomposition in Equation (13) follows. ◀

B.3 Symmetrization and Blekherman's theorem

The symmetric group S_n acts on the polynomial ring M_n by permuting the indices of the monomials. The subspace of symmetric polynomials in M_n that are invariant under the action of S_n is denoted by Λ_n. The operator $\mathrm{Sym} : M_n \to \Lambda_n$ maps a polynomial to its symmetrization,

$$\mathrm{Sym}(p)(x) := \frac{1}{n!} \sum_{\sigma \in S_n} p(\sigma x)\,. \tag{15}$$

The symmetrization of degree k monomials evaluates to a univariate polynomial in $|x|$ over M_n.

▶ **Lemma B.7.** *Let $m_k(x) = x_1 x_2 \cdots x_k$ be a degree-k monomial, then the following identity is true in the ring M_n,*

$$Sym(m_k)(x) = \frac{|x|(|x| - 1) \cdots (|x| - k + 1)}{n(n-1) \cdots (n - k + 1)}\,. \tag{16}$$

Proof. We proceed by induction on k, for $k = 1$ the result is clearly true. Let x^S be an arbitrary degree k monomial. There are $k!(n-k)!$ permutations $\sigma \in S_n$ such that $\sigma(x_1 x_2 \cdots x_k) = x^S$, thus $\mathrm{Sym}(m_k)(x)$ evaluates to

$$\mathrm{Sym}(m_k)(x) = \frac{1}{\binom{n}{k}} \sum_{|S|=k} x^S \,. \tag{17}$$

In order to to express $\mathrm{Sym}(m_k)(x)$ in terms of $\mathrm{Sym}(m_{k-1})(x)$, we write the above equation in terms of the operator W_k^t from Section B.1,

$$\mathrm{Sym}(m_k)(x) = \frac{1}{\binom{n}{k}} W_k^T \left(\frac{1}{k} \sum_{|U|=k-1} x^U \right) = \frac{(|x|-k+1)}{(n-k+1)} Sym(m_{k-1})(x) \,. \tag{18}$$

The second equality follows from Equation (8) and the expression for $\mathrm{Sym}(m_{k-1})$ in Equation (17). The result follows from the induction hypothesis. ◀

The above lemma shows that $\mathrm{Sym}(p)$ for polynomials $p \in M_n$ can be viewed as a univariate polynomial in $|x|$ by extending the mapping given by Lemma B.7 to all $p \in M_n$. We denote the univariate polynomial thus obtained by $\mathrm{Sym}^{uni}(p)$ to disambiguate from the multivariate polynomial in Equation (15).

We can define an inner product on polynomials $p, q \in L_t$ by treating them as vectors of coefficients: if $p(x) = \sum_{|S|=t} p_S x^S$ and $q(x) = \sum_{|S|=t} q_S x^S$ then

$$\langle p|q \rangle := \sum_{S \subseteq [n], |S|=t} p_S q_S \,. \tag{19}$$

The symmetrization of the product of polynomials in $\mathrm{Ker}(W_t)$ can be expressed in terms of this inner product.

▶ **Lemma B.8.** *If $p, q \in Ker(W_t)$ for some $t \leq n/2$, then:*

$$Sym(pq)(x) = \langle p|q \rangle \frac{(n-2t)!}{n!} \prod_{0 \leq i < t} (|x|-i)(n-|x|-i) \,. \tag{20}$$

Proof. Theorem B.4 shows that $\mathrm{Ker}(W_t)$ has a basis consisting of polynomials $p_{\mathcal{U}}(x)$ such that $p_{\mathcal{U}}(x) = 0$ for all $x \in \{0,1\}^n$ with $|x| \in \{0, 1, \ldots, t-1\} \cup \{n, n-1, \ldots, n-t+1\}$. Consider such an x. Evaluating $\mathrm{Sym}(pq)$ at x using Equation (15) by expanding p and q in the basis given by Theorem B.4, it follows that $\mathrm{Sym}^{uni}(pq)(\alpha) = 0$ for all $\alpha \in \{0, 1, \ldots, t-1\} \cup \{n, n-1, \ldots, n-t+1\}$. Lemma B.7 shows that $\mathrm{Sym}(pq)(x)$ is a univariate polynomial Sym^{uni} in $|x|$ of degree at most $2t$, hence

$$\mathrm{Sym}(pq)(x) = \lambda \prod_{0 \leq i < t} (|x|-i)(n-|x|-i) \,. \tag{21}$$

for some $\lambda \in \mathbb{R}$. Below we determine λ by evaluating $\mathrm{Sym}(pq)$ for $x \in \{0,1\}^n$ such that $|x| = t$.

We compute $\mathrm{Sym}(pq)(x)$ by evaluating the sum $\sum_{\sigma \in S_n} p(\sigma x) q(\sigma x)$ in Equation (15). As p, q are homogeneous degree-t polynomials, for each x with $|x| = t$ there is a unique $S \subset [n]$, $|S| = t$, such that $p(x) = p_S$ and $q(x) = q_S$. In other words, x sets exactly one degree-t monomial x^S to 1 and all others to 0. There are $t!(n-t)!$ different $\sigma \in S_n$ such that $\sigma(x)$ sets the same monomial to 1. The symmetrization $\mathrm{Sym}(pq)(x)$ therefore evaluates to

$$\mathrm{Sym}(pq)(x) = \frac{1}{n!} \sum_{\sigma \in S_n} p(\sigma x) q(\sigma x) = \frac{t!(n-t)!}{n!} \sum_{|S|=t} p_S q_S \,. \tag{22}$$

$\mathrm{Sym}(pq)(x)$ also evaluates to $\lambda \prod_{0 \le i < t}(t-i)(n-t-i)$. Equating the two expressions we have:

$$\lambda t! \prod_{0 \le i < t}(n-t-i) = \frac{\langle p|q \rangle t!(n-t)!}{n!} \tag{23}$$

which implies $\lambda = \frac{\langle p|q \rangle (n-2t)!}{n!}$, and the theorem follows. ◀

We next show that the symmetrization of the product of polynomials $p \in \mathrm{Ker}(W_t), q \in \mathrm{Ker}(W_{t'})$ evaluates to 0 if $t \neq t'$. The following lemma is used for the proof in Lemma B.10.

▶ **Lemma B.9.** *If $p(x) = \prod_{i \in [k]}(x_i - x_{i+1})q(x)$ for some odd k, and $q(x)$ is a polynomial that does not depend on variables $x_1, \dots, x_{k+1}$, then Sym$(p) = 0$.*

Proof. It suffices to show that $\mathrm{Sym}(p)(x) = 0$ for all $x \in \{0,1\}^n$, because a multilinear polynomial that is 0 on the hypercube is identically equal to 0. Define the involution $\sigma \to \overline{\sigma}$ on S_n by setting $\overline{\sigma}(i) = \sigma(i+1)$ if $i \in [k+1]$ is odd, $\overline{\sigma}(i) = \sigma(i-1)$ if $i \in [k+1]$ is even, and $\overline{\sigma}(i) = \sigma(i)$ for $i > k+1$. It follows that $\overline{\sigma}$ is an involution as $k+1$ is even and it acts by swapping the pairs $(\sigma(2i-1), \sigma(2i))$ for $i \in [(k+1)/2]$. This involution partitions S_n into pairs $(\sigma, \overline{\sigma})$, and hence

$$\mathrm{Sym}(p)(x) = \frac{1}{n!} \sum_{(\sigma, \overline{\sigma})} (p(\sigma x) + p(\overline{\sigma} x)) = 0. \tag{24}$$

The second equality follows as $p(\overline{\sigma} x) = -p(\sigma x)$ for all $x \in \{0,1\}^n$ and $\sigma \in S_n$. ◀

▶ **Lemma B.10.** *If $p \in$ Ker(W_t) and $q \in$ Ker$(W_{t'})$ for $n/2 \ge t > t'$, then Sym$(pq) = 0$.*

Proof. It suffices to prove the statement for polynomials $p = p_{\mathcal{U}}$ and $q = q_{\mathcal{V}}$ belonging to the bases for $\mathrm{Ker}(W_t)$ and $\mathrm{Ker}(W_{t'})$ constructed in Theorem B.4. The arrays $\mathcal{U}, \mathcal{V}$ define matchings $M(\mathcal{U}) = \bigcup_{i \in [t]}(u(2i-1), u(2i))$ and $M(\mathcal{V}) = \bigcup_{i \in [t']}(v(2i-1), v(2i))$ on $[n]$ of size t and t' respectively. The product $p_{\mathcal{U}} q_{\mathcal{V}} = \prod_{(a,b) \in M(\mathcal{U}) \cup M(\mathcal{V})}(x_a - x_b)$. If $M(\mathcal{U}) \cup M(\mathcal{V})$ contains an odd-length path as an induced subgraph, then we can use Lemma B.9 to conclude that $\mathrm{Sym}(p_{\mathcal{U}} q_{\mathcal{V}}) = 0$.

It suffices to show that the union of two matchings of different sizes contains an odd-length path as an induced subgraph. The connected components of a union of two distinct matchings on $[n]$ either form even-length cycles or paths. Color the edges in $M(\mathcal{U})$ red and the edges in $M(\mathcal{V})$ blue. The number of red edges t is greater than blue edges t', so there must be at least one connected component that is an odd-length path, as even-length paths and cycles have an equal number of red and blue edges. ◀

The preceding lemmas allow us to give a proof of Blekherman's result [2] on the symmetrization of sum-of-squares polynomials on the hypercube.

▶ **Theorem B.11** (Blekherman). *The symmetrization of the square of polynomial $p \in M_t$ for $t \le n/2$ can be decomposed as*

$$Sym(p^2)(x) = \sum_{j=0}^{t} p_{t-j}(|x|) \left(\prod_{0 \le i < j} (|x|-i)(n-|x|-i) \right) \tag{25}$$

where p_{t-j} is a univariate polynomial that is the sum of squares of polynomials of degree at most $t-j$.

Proof. Consider the representation of the polynomial $p(x) = \sum_{j=0}^{t} q_j(x)$ given by Corollary B.6,

$$q_j(x) = \sum_{0 \le k \le t-j} |x|^k p_{kj}(x) \tag{26}$$

where the polynomials $p_{kj}(x) \in \mathrm{Ker}(W_j)$. Lemma B.10 shows that $\mathrm{Sym}(p_{kj} p_{k'j'}) = 0$ if $j \neq j'$, hence $\mathrm{Sym}(p^2)$ can be decomposed as

$$\mathrm{Sym}(p^2) = \sum_{j=0}^{t} \mathrm{Sym}\left(q_j^2\right) . \tag{27}$$

Expanding the term $\mathrm{Sym}\left(q_j^2\right)$ using Lemma B.8, we have

$$\begin{aligned}\sum_{0 \le k,l \le t-j} \mathrm{Sym}(|x|^{k+l} p_{kj} p_{lj}) &= c \left(\prod_{0 \le i < j} (|x| - i)(n - |x| - i) \right) \sum_{0 \le k,l \le t-j} \langle p_{kj} | p_{lj} \rangle |x|^{k+l} \\ &= c \left(\prod_{0 \le i < j} (|x| - i)(n - |x| - i) \right) \mathbf{x}^T P \mathbf{x}\end{aligned} \tag{28}$$

where c is a constant independent of $|x|$, $P \in \mathbb{R}^{(t-j+1)\times(t-j+1)}$ is the matrix with entries $P_{kl} = \langle p_{kj} | p_{lj} \rangle$, and $\mathbf{x} \in \mathbb{R}^{t-j+1}$ is the vector with entries $(1, |x|, |x|^2, \ldots, |x|^{t-j})$. The matrix P is positive semidefinite, hence the polynomial $p_{t-j}(|x|)$ corresponding to the quadratic form $\mathbf{x}^t P \mathbf{x}$ is a sum of squares of polynomials in $|x|$ of degree at most $t - j$. The theorem follows. ◀

Note that the proof is constructive, as it provides a way to compute the terms in the decomposition by projecting onto the eigenspaces W_t of the Johnson scheme. For example, the first term $p_t(|x|)$ in (25) is in fact $\mathrm{Sym}^{uni}(p)^2$ as the Sym operator maps $\mathrm{Ker}(W_j)$ to 0 for all $j > 0$.

▶ **Corollary B.12.** *The polynomial $p_t(|x|)$ in Theorem B.11 is $Sym^{uni}(p)^2$.*

A symmetric function f that is the sum of squares of polynomials of degree $d \le n/2$ is a sum of terms $\mathrm{Sym}(p^2)$ for n-variate polynomials p of degree $d \le n/2$. Applying Theorem B.11 for $t = d$ we obtain Blekherman's result as stated in Theorem 2.3.

Note that Theorem B.11 applies to the setting where $deg(p(x)) \le n/2$, this suffices for our applications. Blekherman's theorem in [2] is valid for all degrees modulo the ideal $I = \langle \prod_{0 \le i \le n} (|x| - i) \rangle$.

C Grigoriev's knapsack lower bound

We now see how Blekherman's theorem can be easily used to reprove Grigoriev's lower bound on the degree of Positivstellensatz refutations of knapsack (Theorem 4.2). A Positivstellensatz refutation of the knapsack system of equations (1) with parameter r consists of polynomials $g, g_1, \ldots, g_n$ and a sos polynomial h such that

$$g(x) \cdot \left(\sum_{i=1}^{n} x_i - r \right) + \sum_{i=1}^{n} g_i(x) \cdot (x_i^2 - x_i) = 1 + h(x) . \tag{29}$$

▶ **Theorem C.1** (Grigoriev [12]). *Let $0 \le k \le (n-3)/2$ be an integer. If $k < r < n-k$, then any Positivstellensatz refutation of the knapsack system of equations with parameter r, as in Equation (29), has degree at least $2k+4$.*

Proof. Grigoriev constructs a functional $\mathcal{G}_r : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ such that: when $\mathcal{G}_r$ is applied to the left-hand side of Equation (29) it evaluates to 0, provided that the total degree of the left-hand side is at most n; and when $\mathcal{G}_r$ is applied to the right-hand side of Equation (29) it is at least 1, provided the total degree of the right-hand side is at most $2k+2$. This leads to a contradiction, hence constructing such a functional $\mathcal{G}_r$ suffices to prove that a Positivstellensatz refutation must have degree at least $\min\{n, 2k+4\}$. The theorem then follows, with the additional observation that if $\min\{n, 2k+4\} = n$ is odd, then we can actually obtain a lower bound of $n+1$ (since any sum-of-squares polynomial must have even degree).

The functional $\mathcal{G}_r$ is first defined on the quotient ring $\mathcal{A} = \mathbb{R}[x_1, \dots, x_n]/\langle x_1^2 - x_1, \dots, x_n^2 - x_n\rangle$. For $p \in \mathcal{A}$ define

$$\mathcal{G}_r(p) = \mathrm{Sym}^{uni}(p)(r).$$

In other words, $\mathcal{G}_r$ looks at the univariate polynomial formed from the symmetrization of p over the symmetric group, and evaluates it at the point r. Explicitly, for a monomial $x_S = \prod_{i \in S} x_i$ with $|S| = t$ we see by Lemma B.7 that $\mathcal{G}_r(x_S) = B_t$ where

$$B_t = \frac{r(r-1)\cdots(r-t+1)}{n(n-1)\cdots(n-t+1)}. \tag{30}$$

For $p \in \mathbb{R}[x_1, \dots, x_n]$ let $\bar{p}$ be its canonical multilinear representative in $\mathcal{A}$. The definition of the functional $\mathcal{G}_r$ is extended from $\mathcal{A}$ to the polynomial ring by letting $\mathcal{G}_r(p) := \mathcal{G}_r(\bar{p})$ for $p \in \mathbb{R}[x_1, \dots, x_n]$.

Grigoriev's theorem now follows from the following four observations about $\mathcal{G}_r$:

1. $\mathcal{G}_r\left(g(x) \cdot (\sum_i x_i - r)\right) = 0$ for all polynomials g with $\deg(g) < n$. It suffices to show this for $g(x) = x_S = \prod_{i \in S} x_i$ for some $S \subsetneq [n]$ with $|S| = t < n$. In this case, by Equation (30), $\mathcal{G}_r(x_S(\sum_i x_i - r)) = (n-t)B_{t+1} + (t-r)B_t = 0$.
2. $\mathcal{G}_r\left(g_i(x)(x_i^2 - x_i)\right) = 0$ for all polynomials g_i. This is because the canonical multilinear representative of $g_i(x)(x_i^2 - x_i)$ in the quotient ring $\mathcal{A}$ is the constant-0 polynomial, and $\mathrm{Sym}^{uni}(0)(r) = 0$.
3. $\mathcal{G}_r(1) = 1$ for all values of r. The symmetrization of the constant-1 polynomial is itself, and the constant-1 polynomial always evaluates to 1.
4. $\mathcal{G}_r\left(p^2(x)\right) \ge 0$ if p is a polynomial of degree at most $k+1$. By Blekherman's Theorem 2.3, if $p \in \mathcal{A}$ and $d = \deg(p)$ then

$$\begin{aligned}\mathrm{Sym}^{uni}(p^2)(x) = q_d(x) + x(n-x)q_{d-1}(x) + x(x-1)(n-x)(n-1-x)q_{d-2}(x) + \cdots \\ + x(x-1)\cdots(x-d+1)(n-x)(n-1-x)\cdots(n-d+1-x)q_0(r).\end{aligned} \tag{31}$$

   It follows that $\mathrm{Sym}^{uni}(p^2)(x) \ge 0$ for $x \in [d-1, n-d+1]$. Thus if $k < r < n-k$, then $\mathcal{G}_r(p^2) \ge 0$ for any p of degree $\le k+1$. By linearity this extends to any h that is a sum of squares of polynomials of degree $\le k+1$.

The first two observations imply that the left-hand side of Equation (29) evaluates to 0 under $\mathcal{G}_r$ (provided the total degree of the left-hand side is at most n), while the last two observations imply that the right-hand side evaluates to at least 1 (provided the total degree on the right is at most $2k+2$). ◀

D Application of Grigoriev's bound to ℓ_∞-error sos degree

Let n be odd and let $f = f_{\lfloor n/2 \rfloor}$, that is $f(x) = (|x| - n/2)^2 - 1/4$. Our Theorem 1.1 gives that any sos polynomial approximating f with ℓ_∞-error at most $1/50$, needs degree $\Omega(n)$. The functional $\mathcal{G}_r$ defined by Grigoriev (see discussion above Equation (30)) can be used to show an incomparable result: any sos polynomial of degree $(n-1)/2$ has error at least $\Omega(1/\log n)$ in approximating f in ℓ_∞-norm.

▶ **Theorem D.1.** *Let n be odd and $f : \{0,1\}^n \to \mathbb{R}$ be defined as $f(x) = (|x| - n/2)^2 - 1/4$. Any sos polynomial of degree $(n-1)/2$ has error at least*

$$(1 - O(1/n)) \frac{\pi}{4} \frac{1}{\ln((n+1)/2) + \gamma + \ln(16)}$$

in approximating f in ℓ_∞ norm. Here $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

Proof. Let $h : \{0,1\}^n \to \mathbb{R}$ be a sos polynomial of degree $(n-1)/2$ approximating f with ℓ_∞-error ϵ. Write $h(x) = f(x) + e(x)$ where e is the function of "errors" satisfying $|e(x)| \leq \epsilon$ for all $x \in \{0,1\}^n$. Let $\delta_y : \{0,1\}^n \to \{0,1\}$ be the delta function on the boolean cube, where $\delta_y(x) = 1$ if and only if $x = y$. Recall that $\mathcal{G}_{n/2}(f) = \mathrm{Sym}^{uni}(f)(n/2) = (n/2 - n/2)^2 - 1/4 = -1/4$. By linearity of $\mathcal{G}_{n/2}$ we have

$$\begin{aligned}\mathcal{G}_{n/2}(f+e) = -1/4 + \mathcal{G}_{n/2}(e) = -1/4 + \mathcal{G}_{n/2}\left(\sum_{y \in \{0,1\}^n} e(y)\delta_y\right) \\ \leq -1/4 + \epsilon \sum_{y \in \{0,1\}^n} |\mathcal{G}_{n/2}(\delta_y)| \,.\end{aligned}$$

On the other hand, $\mathcal{G}_{n/2}(f+e) \geq 0$ as $f+e$ is a sum-of-squares of polynomials of degree at most $(n-1)/2$ (property 4 in the proof of Theorem C.1). Thus

$$\epsilon \geq \left(4 \sum_{y \in \{0,1\}^n} |\mathcal{G}_{n/2}(\delta_y)|\right)^{-1} . \tag{32}$$

The main part of the proof will be to evaluate this sum.

Let $L_i : \mathbb{R} \to \mathbb{R}$ be the degree-n polynomial uniquely defined by

$$L_i(z) = \begin{cases} 1 & z = i \\ 0 & z \in \{0,1,2,\ldots,n\} \setminus \{i\} \end{cases} .$$

Then we see that $\mathcal{G}_{n/2}(\delta_y) = L_{|y|}(n/2)/\binom{n}{|y|}$, and so

$$\sum_{y \in \{0,1\}^n} |\mathcal{G}_{n/2}(\delta_y)| = \sum_{k=0}^{n} |L_k(n/2)| \,.$$

To do this sum, let us first simplify the summand

$$\begin{aligned}|L_k(n/2)| &= \frac{\prod_{a=0,a\neq k}^{n} |n/2-a|}{\prod_{a=0,a\neq k}^{n} |k-a|} \\ &= \frac{\prod_{a=0}^{n} |n/2-a|}{k!(n-k)!|n/2-k|} \\ &= \frac{1}{2^{n+1}} \frac{n!!n!!}{k!(n-k)!|n/2-k|} \\ &= \frac{n!}{2^{2n-1}\left(\frac{n-1}{2}\right)!^2} \binom{n}{k} \frac{1}{|n-2k|} \\ &= \frac{n}{2^{2n-1}} \binom{n-1}{(n-1)/2} \binom{n}{k} \frac{1}{|n-2k|}\,,\end{aligned}$$

where $n!!$ is defined as $\prod_{j=0}^{\lceil n/2\rceil-1}(n-2j)$, and we used $n! = n!! \cdot 2^{(n-1)/2} \cdot ((n-1)/2)!$ for odd n in the penultimate equality.

For what comes next, it will be more convenient to express $|L_k(n/2)|$ in terms of $m = (n-1)/2$. In this way, we obtain an expression defined for all m, rather than just odd n.

$$|L_k(m+1/2)| = \frac{2m+1}{2^{4m+1}} \binom{2m}{m} \binom{2m+1}{k} \frac{1}{|2m-2k+1|}$$

Let $A(m)$ denote the sum over $k = 0, \ldots, n = 2m+1$, which is

$$A(m) = \frac{2m+1}{2^{4m+1}} \binom{2m}{m} \sum_{k=0}^{2m+1} \frac{1}{|2m-2k+1|} \binom{2m+1}{k}.$$

By symmetry of the binomial coefficients we can multiply by 2 and sum over only half of them, thereby removing the absolute values.

$$A(m) = \frac{2m+1}{4^{2m}} \binom{2m}{m} \sum_{\ell=0}^{m} \frac{1}{2\ell+1} \binom{2m+1}{\ell+m+1}$$

Now we look at the difference between consecutive $A(m)$:

▶ **Claim D.2.**

$$A(m+1) - A(m) = \left(\frac{\binom{2(m+1)}{m+1}}{4^{m+1}}\right)^2$$

Proof. It is somewhat cumbersome to verify this claim directly. We take the following approach. Let $A(m) = B(m)C(m)$, where

$$B(m) = \frac{2m+1}{4^{2m}} \binom{2m}{m}, \qquad C(m) = \sum_{k=0}^{m} \frac{1}{2k+1} \binom{2m+1}{k+m+1}.$$

Note that

$$\frac{B(m+1)}{B(m)} = \frac{2m+3}{8(m+1)}.$$

Since $B(0) = 1$, this resolves to

$$B(m+1) = \frac{(2m+3)!!}{8^{m+1}(m+1)!} = \frac{2m+3}{4^{2(m+1)}} \binom{2(m+1)}{m+1}.$$

By Zeilberger's algorithm [26] we find a recurrence satisfied by the summand of $C(m)$.

$$\frac{2m+3}{2k+1}\binom{2m+3}{k+m+2} - \frac{8(m+1)}{2k+1}\binom{2m+1}{k+m+1} = \binom{2(m+1)}{m+k+1} - \binom{2(m+1)}{m+k+2}.$$

Summing this recurrence over $k = 0, \dots, m+1$ we find

$$(2m+3)C(m+1) - 8(m+1)C(m) = \binom{2(m+1)}{m+1}.$$

This means that

$$A(m+1) - \underbrace{\frac{8(m+1)}{2m+3}B(m+1)}_{B(m)}\,C(m) = \frac{B(m+1)}{2m+3}\binom{2(m+1)}{m+1},$$

and in turn

$$A(m+1) - A(m) = \frac{1}{4^{2(m+1)}}\binom{2(m+1)}{m+1}^2.$$

◀

As $A(0) = 1$ this gives

$$A(m) = \sum_{i=0}^{m}\left(\frac{\binom{2i}{i}}{4^i}\right)^2.$$

Luckily, the latter sum has already been asymptotically evaluated in the study of the quantum adversary bound for the ordered search problem [7]. There it is shown that

$$\sum_{i=0}^{N}\left(\frac{\binom{2i}{i}}{4^i}\right)^2 = \frac{1}{\pi}\left(\ln(N+1) + \gamma + \ln(16)\right) + O(1/N),$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

This gives

$$\begin{aligned}\sum_{y\in\{0,1\}^n} |\mathcal{G}_{n/2}(\delta_y)| = A((n-1)/2) &= \sum_{i=0}^{(n-1)/2}\left(\frac{\binom{2i}{i}}{4^i}\right)^2 \\ &= \frac{1}{\pi}\left(\ln((n+1)/2) + \gamma + \ln(16)\right) + O(1/n).\end{aligned}$$

Plugging this into Equation (32) gives the theorem. ◀