

# Weak Decoupling, Polynomial Folds, and Approximate Optimization over the Sphere

Vijay Bhattiprolu\*   Mrinalkanti Ghosh†   Venkatesan Guruswami‡  
 Euiwoong Lee§   Madhur Tulsiani¶

## Abstract

We consider the following basic problem: given an  $n$ -variate degree- $d$  homogeneous polynomial  $f$  with real coefficients, compute a unit vector  $x \in \mathbb{R}^n$  that maximizes  $|f(x)|$ . Besides its fundamental nature, this problem arises in diverse contexts ranging from tensor and operator norms to graph expansion to quantum information theory. The homogeneous degree 2 case is efficiently solvable as it corresponds to computing the spectral norm of an associated matrix, but the higher degree case is NP-hard.

We give approximation algorithms for this problem that offer a trade-off between the approximation ratio and running time: in  $n^{O(q)}$  time, we get an approximation within factor  $O_d((n/q)^{d/2-1})$  for arbitrary polynomials,  $O_d((n/q)^{d/4-1/2})$  for polynomials with non-negative coefficients, and  $O_d(\sqrt{m/q})$  for sparse polynomials with  $m$  monomials. The approximation guarantees are with respect to the optimum of the level- $q$  sum-of-squares (SoS) SDP relaxation of the problem (though our algorithms do not rely on actually solving the SDP). Known polynomial time algorithms for this problem rely on “decoupling lemmas.” Such tools are not capable of offering a trade-off like our results as they blow up the number of variables by a factor equal to the degree. We develop new decoupling tools that are more efficient in the number of variables at the expense of less structure in the output polynomials. This enables us to harness the benefits of higher level SoS relaxations. Our decoupling methods also work with “folded polynomials,” which are polynomials with polynomials as coefficients. This allows us to exploit easy substructures (such as quadratics) by considering them as coefficients in our algorithms.

We complement our algorithmic results with some polynomially large integrality gaps for  $d$ -levels of the SoS relaxation. For general polynomials this follows from known results for *random* polynomials, which yield a gap of  $\Omega_d(n^{d/4-1/2})$ . For polynomials with non-negative coefficients, we prove an  $\tilde{\Omega}(n^{1/6})$  gap for the degree 4 case, based on a novel distribution of 4-uniform hypergraphs. We establish an  $n^{\Omega(d)}$  gap for general degree  $d$ , albeit for a slightly weaker (but still very natural) relaxation. Toward this, we give a method to lift a level-4 solution matrix  $M$  to a higher level solution, under a mild technical condition on  $M$ .

From a structural perspective, our work yields worst-case convergence results on the performance of the sum-of-squares hierarchy for polynomial optimization. Despite the popularity of SoS in this context, such results were previously only known for the case of  $q = \Omega(n)$ .

\*Supported by NSF CCF-1422045 and CCF-1526092. vpb@cs.cmu.edu

†Supported by NSF CCF-1254044 mkgghosh@ttic.edu

‡Supported in part by NSF grant CCF-1526092. guruswami@cmu.edu

§Supported by a Samsung Fellowship, Simons award, and NSF CCF-1526092. euiwoonl@cs.cmu.edu

¶Supported by NSF CCF-1254044 madhurt@ttic.edu

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Our Algorithmic Results . . . . .	2
1.2	Connection to sum-of-squares hierarchy . . . . .	4
1.3	Related Previous and Recent Works . . . . .	6
1.4	Organization of the Paper . . . . .	8
<b>2</b>	<b>Preliminaries and Notation</b>	<b>8</b>
<b>3</b>	<b>Overview of Proofs and Techniques</b>	<b>9</b>
3.1	Warmup: $(n^2/q^2)$ -Approximation . . . . .	10
3.1.1	$n^2$ -Approximation using level-4 relaxation . . . . .	10
3.1.2	$(n^2/q^2)$ -Approximation Assuming Theorem 1.5 . . . . .	10
3.1.3	Reduction to Optimization of Multi-linear Polynomials . . . . .	10
3.1.4	Bounding $\ g\ _2$ via a new weak decoupling lemma . . . . .	12
3.2	Exploiting Easy Substructures via Folding and Improved Approximations . . . . .	13
3.2.1	$n$ -Approximation using Degree-4 SoS . . . . .	13
3.2.2	$n/q$ -approximation using Degree- $q$ SoS . . . . .	14
3.3	Lower Bounds for Polynomials with Non-negative Coefficients . . . . .	15
3.3.1	Degree-4 Lower Bound for Polynomials with Non-Negative Coefficients . . . . .	15
3.3.2	Lifting Stable Degree-4 Lower Bounds . . . . .	16
<b>4</b>	<b>Additional Preliminaries and the SoS Hierarchy</b>	<b>17</b>
4.1	Pseudoexpectations and Moment Matrices . . . . .	17
4.1.1	Constrained Pseudoexpectations . . . . .	18
4.2	Matrix Representations of Polynomials and relaxations of $\ f\ _2$ . . . . .	18
4.2.1	Properties of relaxations obtained from constrained pseudoexpectations . . . . .	18
4.3	An additional operation on folded polynomials . . . . .	20
<b>5</b>	<b>Results for Polynomials in <math>\mathbb{R}_d[x]</math> and <math>\mathbb{R}_d^+[x]</math></b>	<b>20</b>
5.1	Reduction to Multilinear Polynomials . . . . .	20
5.1.1	Upper Bounding $\Lambda(f)$ in terms of $\Lambda(F_{2\alpha})$ . . . . .	21
5.1.2	Lower Bounding $\ f\ _2$ in terms of $\ F_{2\alpha}\ _2$ (non-negative coefficients) . . . . .	21
5.1.3	Lower Bounding $\ f\ _2$ in terms of $\ F_{2\alpha}\ _2$ (general case) . . . . .	22
5.2	$(n/q)^{d/4}$ -Approximation for Non-negative Coefficient Polynomials . . . . .	26
5.3	$(n/q)^{d/2}$ -Approximation for General Polynomials . . . . .	27
5.4	$\sqrt{m/q}$ -Approximation for $m$ -sparse polynomials . . . . .	27
<b>6</b>	<b>Approximating 2-norms via Folding</b>	<b>28</b>

6.1	Preliminaries . . . . .	28
6.2	Reduction to Multilinear Folded Polynomials . . . . .	28
6.2.1	Relating $\Lambda_C(f)$ to $\Lambda_C(F_{2\alpha})$ . . . . .	29
6.3	Relating Evaluations of $f$ to Evaluations of $F_{2\alpha}$ . . . . .	29
6.4	Bounding $\Lambda_C()$ of Multilinear Folded Polynomials . . . . .	30
6.4.1	General Folds: Bounding $\Lambda()$ in terms of $\Lambda()$ of the "worst" fold . . . . .	30
6.4.2	Non-Negative Coefficient Folds: Relating SoS Value to the SoS Value of the $d_1/2$ -collapse . . . . .	31
6.5	$(n/q)^{d/4-1/2}$ -Approximation for Non-negative Coefficient Polynomials . . . . .	33
6.6	$(n/q)^{d/2-1}$ -Approximation for General Polynomials . . . . .	34
6.7	Algorithms . . . . .	35
6.7.1	Non-negative coefficient polynomials . . . . .	35
6.7.2	General Polynomials . . . . .	36
<b>7</b>	<b>Constant Level Lower Bounds for Polynomials with Non-negative Coefficients</b> . . . . .	<b>36</b>
7.1	Upper Bound on $\ f\ _2$ . . . . .	37
7.1.1	Reduction to counting shattered cliques . . . . .	37
7.1.2	Counting edges and triangles . . . . .	38
7.1.3	Bounding 4-clique Density . . . . .	39
7.2	Lower Bound on $\Lambda(f)$ . . . . .	40
7.2.1	A simple lower bound on $\ f\ _{sp}$ . . . . .	40
7.2.2	Lower bound for the stronger relaxation computing $\Lambda(f)$ . . . . .	40
7.2.3	Bounding the smallest eigenvalue via the trace method . . . . .	42
7.2.4	Analyzing contributing subgraphs . . . . .	44
<b>8</b>	<b>Lifting <math>\ \cdot\ _{sp}</math> lower bounds to higher levels</b> . . . . .	<b>47</b>
8.1	Gap between $\ \cdot\ _{sp}$ and $\ \cdot\ _2$ for Non-Neg. Coefficient Polynomials . . . . .	47
8.2	Tetris Theorem . . . . .	48
8.3	Lifting Stable Degree-4 Lower Bounds . . . . .	49
8.4	Proof of Tetris Theorem . . . . .	49
8.4.1	Hypergraphical Matrix . . . . .	50
8.4.2	Kronecker Products of Hypergraphical Matrices . . . . .	50
8.4.3	Hypergraphical Matrices under Permutation . . . . .	51
8.4.4	Proof of Eq. (8.3) . . . . .	52
<b>9</b>	<b>Open problems</b> . . . . .	<b>53</b>
<b>A</b>	<b>Oracle Lower Bound</b> . . . . .	<b>57</b>
<b>B</b>	<b>Maximizing <math> f(x) </math> vs. <math>f(x)</math></b> . . . . .	<b>58</b>

# 1. Introduction

We study the problem of optimizing homogeneous polynomials over the unit sphere. Formally, given an  $n$ -variate degree- $d$  homogeneous polynomial  $f$ , the goal is to compute

$$\|f\|_2 := \sup_{\|x\|=1} |f(x)| \tag{1.1}$$

When  $f$  is a homogeneous polynomial of degree 2, this problem is equivalent computing the spectral norm of an associated symmetric matrix  $M_f$ . For higher degree  $d$ , it defines a natural higher-order analogue of the eigenvalue problem for matrices. The problem also provides an important testing ground for the development of new spectral and semidefinite programming (SDP) techniques, and techniques developed in the context of this problem have had applications to various other constrained settings [HLZ10, Lau09, Las09].

Besides being a natural and fundamental problem in its own right, it has connections to widely studied questions in many other areas. In quantum information theory [BH13, BKS14], the problem of computing the optimal success probability of a protocol for Quantum Merlin-Arthur games can be thought of as optimizing certain classes of polynomials over the unit sphere. The problem of estimating the  $2 \rightarrow 4$  norm of an operator, which is equivalent to optimizing certain homogeneous degree-4 polynomials over the sphere, is known to be closely related to the Small Set Expansion Hypothesis (SSEH) and the Unique Games Conjecture (UGC) [BBH<sup>+</sup>12, BKS14]. The polynomial optimization problem is also very relevant for natural extensions of spectral problems, such as low-rank decomposition and PCA, to the case of tensors [BKS15, GM15, MR14, HSS15]. Frieze and Kannan [FK08] (see also [BV09]) also established a connection between the problem of approximating the spectral norm of a tensor (or equivalently, computing  $\|f\|_2$  for a polynomial  $f$ ), and finding planted cliques in random graphs.

The problem of polynomial optimization has been studied<sup>1</sup> over various compact sets [Las09, DK08], and is natural to ask how well polynomial time algorithms can *approximate* the optimum value over a given compact set (see [DK08] for a survey). While the maximum of a degree- $d$  polynomial over the simplex admits a PTAS for every fixed  $d$  [dKLP06], the problem of optimizing even a degree 3 polynomial over the hypercube does not admit any approximation better than  $2^{(\log n)^{1-\epsilon}}$  (for arbitrary  $\epsilon > 0$ ) assuming NP cannot be solved in time  $2^{(\log n)^{O(1)}}$  [HV04].

The approximability of polynomial optimization on the sphere is poorly understood in comparison. It is known that the maximum of a degree- $d$  polynomial can be approximated within a factor of  $n^{d/2-1}$  in polynomial time [HLZ10, So11]. On the hardness side, Nesterov [Nes03] gave a reduction from Maximum Independent Set to optimizing a homogeneous cubic polynomial over  $S^{n-1}$ . Formally, given a graph  $G$ , there exists a homogeneous cubic polynomial  $f(G)$  such that  $\sqrt{1 - \frac{1}{\alpha(G)}} = \max_{\|x\|=1} f(x)$ . Combined with the hardness of Maximum Independent Set [Hås96], this rules out an FPTAS for optimization over the unit sphere. Assuming the Exponential Time Hypothesis, Barak et al. [BBH<sup>+</sup>12] proved that computing  $2 \rightarrow 4$  norm of a matrix, a special case when  $f$  is a degree-4 homogeneous polynomial, is hard to approximate within a factor  $\exp(\log^{1/2-\epsilon}(n))$  for any  $\epsilon > 0$ .

---

<sup>1</sup>In certain cases, the problem studied is not to maximize  $|f|$ , but just  $f(x)$ . While the two problems are equivalent for homogeneous polynomials of odd degree, some subtle issues arise when considering polynomials of even degree. We compare the two notions in Appendix B.

Optimization over  $\mathbb{S}^{n-1}$  has been given much attention in the optimization community, where for a fixed number of variables  $n$  and degree  $d$  of the polynomial, it is known that the estimates produced by  $q$  levels a certain hierarchy of SDPs (Sum of Squares) get arbitrarily close to the true optimal solution as  $q$  increases (see [Las09] for various applications). We refer the reader to the recent work of Doherty and Wehner [DW12] and de Klerk, Laurent, and Sun [dKLS14] and references therein for more information on convergence results. These algorithms run in time  $n^{O(q)}$ , which is polynomial for constant  $q$ . Unfortunately, known convergence results often give a non-trivial bound only when the  $q$  is linear in  $n$ .

In computer science, much attention has been given to the sub-exponential runtime regime (i.e.  $q \ll n$ ) since many of the target applications such as SSE, QMA and refuting random CSPs are of considerable interest in this regime. In addition to the polytime  $n^{d/2-1}$ -approximation for general polynomials [HLZ10, So11], approximation guarantees have been proved for several special cases including  $2 \rightarrow q$  norms [BBH<sup>+</sup>12], polynomials with non-negative coefficients [BKS14], some polynomials that arise in quantum information theory [BKS17, BH13], and random polynomials [RRS16, BGL16]. Hence there is considerable interest in tightly characterizing the approximation guarantee achievable using sub-exponential time.

In this paper, we develop general techniques to design and analyze algorithms for polynomial optimization over the sphere. The sphere constraint is one of the simplest constraints for polynomial optimization and thus is a good testbed for techniques. Indeed, we believe these techniques will also be useful in understanding polynomial optimization for other constrained settings.

In addition to giving an analysis the problem for arbitrary polynomials, these techniques can also be adapted to take advantage of the structure of the input polynomial, yielding better approximations for several special cases such as polynomials with non-negative coefficients, and sparse polynomials. Previous polynomial time algorithms for polynomial optimization work by reducing the problem to diameter estimation in convex bodies [So11] and seem unable to utilize structural information about the (class of) input polynomials. Development of a method which can use such information was stated as an open problem by Khot and Naor [KN08] (in the context of  $\ell_\infty$  optimization).

Our approximation guarantees are with respect to the optimum of the well-studied Lasserre/sum-of-squares (SoS) semidefinite programming relaxation. Such SDPs are the most natural tool to bound the optima of polynomial optimization problems, and our results shed light on the efficacy of higher levels of the SoS hierarchy to deliver better approximations to the optimum. We discuss the SoS connection in Section 1.2, but first turn to stating our approximation guarantees.

## 1.1. Our Algorithmic Results

For a homogeneous polynomial  $h$  of even degree  $q$ , a matrix  $M_h \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  is called a matrix representation of  $h$  if  $(x^{\otimes q/2})^T \cdot M_h \cdot x^{\otimes q/2} = h(x) \quad \forall x \in \mathbb{R}^n$ . Next we define the quantity,

$$\Lambda(h) := \inf \left\{ \sup_{\|z\|_2=1} z^T M_h z \mid M \text{ is a representation of } h \right\}. \quad (1.2)$$

Let  $h_{\max}$  denote  $\sup_{\|x\|=1} h(x)$ . Clearly,  $h_{\max} \leq \Lambda(h)$ , i.e.  $\Lambda(h)$  is a relaxation of  $h_{\max}$ . However, this does not imply that  $\Lambda(h)$  is a relaxation of  $\|h\|_2$ , since it can be the case that

$h_{\max} \neq \|h\|_2$ . To remedy this, one can instead consider  $\sqrt{\Lambda(h^2)}$  which is a relaxation of  $\|h\|_2$ , since  $(h^2)_{\max} = \|h^2\|_2$ . More generally, for a degree- $d$  homogeneous polynomial  $f$  and an integer  $q$  divisible by  $2d$ , we have the upper estimate

$$\|f\|_2 \leq \Lambda(f^{q/d})^{d/q}$$

The following result shows that  $\Lambda(f^{q/d})^{d/q}$  approximates  $\|f\|_2$  within polynomial factors, and also gives an algorithm to approximate  $\|f\|_2$  with respect to the upper bound  $\Lambda(f^{q/d})^{d/q}$ . In the statements below and the rest of this section,  $O_d(\cdot)$  and  $\Omega_d(\cdot)$  notations hide  $2^{O(d)}$  factors. Our algorithmic results are as follows:

**Theorem 1.1.** *Let  $f$  be an  $n$ -variate homogeneous polynomial of degree- $d$ , and let  $q \leq n$  be an integer divisible by  $2d$ . Then,*

$$\begin{aligned} \text{Arbitrary } f: & \quad \left(\Lambda(f^{q/d})\right)^{d/q} \leq O_d\left((n/q)^{d/2-1}\right) \cdot \|f\|_2 \\ f \text{ with Non-neg. Coefficients:} & \quad \left(\Lambda_C(f^{q/d})\right)^{d/q} \leq O_d\left((n/q)^{d/4-1/2}\right) \cdot \|f\|_2 \\ f \text{ with Sparsity } m: & \quad \left(\Lambda(f^{q/d})\right)^{d/q} \leq O_d\left(\sqrt{m/q}\right) \cdot \|f\|_2. \end{aligned}$$

(where  $\Lambda_C(\cdot)$  is a related efficiently computable quantity that we define in Section 4.2)

Furthermore, there is a deterministic algorithm that runs in  $n^{O(q)}$  time and returns  $x$  such that

$$|f(x)| \geq \frac{\Lambda(f^{q/d})^{d/q}}{O_d(c(n, d, q))}$$

where  $c(n, d, q)$  is  $(n/q)^{d/2-1}$ ,  $(n/q)^{d/4-1/2}$  and  $\sqrt{m/q}$  respectively, for each of the above cases (the inequality uses  $\Lambda_C(\cdot)$  in the case of polynomials with non-negative coefficients).

**Remark 1.2.** *Interestingly, our deterministic algorithms only involve computing the maximum eigenvectors of  $n^{O(q)}$  different matrices in  $\mathbf{R}^{n \times n}$ , and actually don't require computing  $\Lambda(f^{q/d})^{d/q}$  (even though this quantity can also be computed in  $n^{O(q)}$  time by the sum-of-squares SDP; see Section 1.2). The quantity  $\Lambda(f^{q/d})^{d/q}$  is only used in the analysis.*

**Remark 1.3.** *If  $m = n^{\rho-d}$  for  $\rho < 1/3$ , then for all  $q \leq n^{1-\rho}$ , the  $\sqrt{m/q}$ -approximation for sparse polynomials is better than the  $(n/q)^{d/2-1}$  arbitrary polynomial approximation.*

**Remark 1.4.** *In cases where  $\|f\|_2 = f_{\max}$  (such as when  $d$  is odd or  $f$  has non-negative coefficients), the above result holds whenever  $q$  is even and divisible by  $d$ , instead of  $2d$ .*

A key technical ingredient en route establishing the above results is a method to reduce the problem for arbitrary polynomials to a list of *multilinear* polynomial problems (over the same variable set). We believe this to be of independent interest, and describe its context and abstract its consequence (Theorem 1.5) next.

Let  $M_g$  be a matrix representation of a degree- $q$  homogeneous polynomial  $g$ , and let  $K = (I, J) \in [n]^{q/2} \times [n]^{q/2}$  have all distinct elements. Observe that there are  $q!$  distinct entries of  $M_g$  including  $K$  across which, one can arbitrarily assign values and maintain the property

of representing  $g$ , as long as the sum across all  $q!$  entries remains the same (specifically, this is the set of all permutations of  $K$ ). In general for  $K' = (I', J') \in [n]^{q/2} \times [n]^{q/2}$ , we define the orbit of  $K'$  denoted by  $\mathcal{O}(K')$ , as the set of permutations of  $K'$ , i.e. the number of entries to which 'mass' from  $M_g[I', J']$  can be moved while still representing  $g$ .

As  $q$  increases, the orbit sizes of the entries increase, and to show better bounds on  $\Lambda(f^{q/d})$ , one must exploit these additional "degrees of freedom" in representations of  $f^{q/d}$ . However, a big obstacle is that the orbit sizes of different entries can range anywhere from 1 to  $q!$ , two extremal examples being  $((1, \dots, 1), (1, \dots, 1))$  and  $((1, \dots, q/2), (q/2 + 1, \dots, q))$ . This makes it hard to exploit the additional freedom afforded by growing  $q$ . Observe that if  $g$  were multilinear, all matrix entries corresponding to non-zero coefficients have a span of  $q!$  and indeed it turns out to be easier to analyze the approximation factor in the multilinear case as a function of  $q$  since the representations of  $g$  can be highly symmetrized. However, we are still faced with the problem of  $f^{q/d}$  being highly non-multilinear. The natural symmetrization strategies that work well for multilinear polynomials fail on general polynomials, which motivates the following result:

**Theorem 1.5 (Informal version of Theorem 5.13).** *For even  $q$ , let  $g(x)$  be a degree- $q$  homogeneous polynomial. Then there exist multilinear polynomials  $g_1(x), \dots, g_m(x)$  of degree at most  $q$ , such that*

$$\frac{\Lambda(g)}{\|g\|_2} \leq 2^{O(q)} \cdot \max_{i \in [m]} \frac{\Lambda(g_i)}{\|g_i\|_2}$$

By combining Theorem 1.5 (or an appropriate generalization) with the appropriate analysis of the multilinear polynomials induced by  $f^{q/d}$ , we obtain the aforementioned results for various classes of polynomials.

**Weak decoupling lemmas.** A common approach for reducing to the multilinear case is through more general "decoupling" or "polarization" lemmas (see e.g., Lemma 5.6), which also have variety of applications in functional analysis and probability [DIPG12]. However, such methods increase the number of variables to  $nq$ , which would completely nullify any advantage obtained from the increased degrees of freedom.

Our proof of Theorem 1.5 (and its generalizations) requires only a decoupling with somewhat weaker properties than given by the above lemmas. However, we need it to be very efficient in the number of variables. In analogy with "weak regularity lemmas" in combinatorics, which trade structural control for complexity of the approximating object, we call these results "weak decoupling lemmas" (see Section 3.1.4 and Lemma 5.12). They provide a milder form of decoupling but only increase the number of variables to  $2n$  (independently of  $q$ ).

We believe these could be more generally applicable; in particular to other constrained settings of polynomial optimization as well as in the design of sub-exponential algorithms. Our techniques might also be able to yield a full tradeoff between the number of variables and quality of decoupling.

## 1.2. Connection to sum-of-squares hierarchy

The *Sum of Squares Hierarchy* (SoS) is one of the canonical and well-studied approaches to attack polynomial optimization problems. Algorithms based on this framework are parametrized by the degree or level  $q$  of the SoS relaxation. For the case of optimization of

a homogenous polynomial  $h$  of even degree  $q$  (with some matrix representation  $M_h$ ) over the unit sphere, the level  $q$  SoS relaxes the non-convex program of maximizing  $(x^{\otimes q/2})^T \cdot M_h \cdot x^{\otimes q/2} = h(x)$  over  $x \in \mathbb{R}^n$  with  $\|x\|_2 = 1$ , to the semidefinite program of maximizing  $\text{Tr}(M_h^T X)$  over all positive semidefinite matrices  $X \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  with  $\text{Tr}(X) = 1$ . (This is a relaxation because  $X = x^{\otimes q/2}(x^{\otimes q/2})^T$  is psd with  $\text{Tr}(X) = \|x\|_2^q$ .)

It is well known (see for instance [Lau09]) that the quantity  $\Lambda(h)$  from (1.2) is the dual value of this SoS relaxation. Further, strong duality holds for the case of optimization on the sphere and therefore  $\Lambda(h)$  equals the optimum of the SoS SDP and can be computed in time  $n^{O(q)}$ . (See Section 4.1 for more detailed SoS preliminaries.) In light of this, our results from Theorem 1.1 can also be viewed as a convergence analysis of the SoS hierarchy for optimization over the sphere, as a function of the number of levels  $q$ . Such results are of significant interest in the optimization community, and have been studied for example in [DW12, dKLS14] (see Section 1.3 for a comparison of results).

**SoS Lower Bounds.** While the approximation factors in our upper bounds of Theorem 1.1 are modest, there is evidence to suggest that this is inherent.

When  $h$  is a degree- $q$  polynomial with *random* i.i.d  $\pm 1$  coefficients, it was shown in [BGL16] that there is a constant  $c$  such that w.h.p.  $\left(\frac{n}{q^{c+o(1)}}\right)^{q/4} \leq \Lambda(h) \leq \left(\frac{n}{q^{c-o(1)}}\right)^{q/4}$ . On the other hand,  $\|h\|_2 \leq O(\sqrt{nq \log q})$  w.h.p. Thus the ratio between  $\Lambda(h)$  and  $\|h\|_2$  can be as large as  $\Omega_q(n^{q/4-1/2})$ .

Hopkins et al. [HKP<sup>+</sup>17] recently proved that degree- $d$  polynomials with random coefficients achieve a degree- $q$  SoS gap of roughly  $(n/q^{O(1)})^{d/4-1/2}$  (provided  $q > n^\epsilon$  for some constant  $\epsilon > 0$ ). This is also a lower bound on the ratio between  $\Lambda(f^{q/d})^{d/q}$  and  $\|f\|_2$  for the case of *arbitrary* polynomials. Note that this lower bound is roughly square root of our upper bound from Theorem 1.1. Curiously, our upper bound for the case of polynomials with non-negative coefficients essentially matches this lower bound for random polynomials.

**Non-Negative Coefficient Polynomials.** In this paper, we give a new lower bound construction for the case of non-negative polynomials, To the best of our knowledge, the only previous lower bound for this problem, was known through Nesterov's reduction [DK08], which only rules out a PTAS. We give the following polynomially large lower bound. The gap applies for random polynomials associated with a novel distribution of 4-uniform hypergraphs, and is analyzed using subgraph counts in a random graph.

**Theorem 1.6.** *There exists an  $n$  variate degree-4 homogeneous polynomial  $f$  with non-negative coefficients such that*

$$\|f\|_2 \leq (\log n)^{O(1)} \quad \text{and} \quad \Lambda(f) \geq \tilde{\Omega}(n^{1/6}).$$

For larger degree  $t$ , we prove an  $n^{\Omega(t)}$  gap between  $\|h\|_2$  and a quantity  $\|h\|_{sp}$  that is closely related to  $\Lambda(h)$ . Specifically,  $\|h\|_{sp}$  is defined by replacing the largest eigenvalue of matrix representations  $M_h$  of  $h$  in (1.2) by the *spectral norm*  $\|M_h\|_2$ . (See Figure 4.2 for a formal definition.) Note that  $\|h\|_{sp} \geq \max\{\Lambda(h), \Lambda(-h)\}$ . Like  $\Lambda(\cdot)$ ,  $\|\cdot\|_{sp}$  suggests a natural hierarchy of relaxations for the problem of approximating  $\|h\|_2$ , obtained by computing  $\|h^{q/t}\|_{sp}^{t/q}$  as the  $q$ -th level of the hierarchy.



We prove a lower bound of  $n^{q/24} / (q \cdot \log n)^{O(q)}$  on  $\|f^{q/4}\|_{sp}$  where  $f$  is as in Theorem 1.6. This not only gives  $\|\cdot\|_{sp}$  gaps for the degree- $q$  optimization problem on polynomials with non-negative coefficients, but also an  $n^{1/6} / (q \log n)^{O(1)}$  gap on higher levels of the aforementioned  $\|\cdot\|_{sp}$  hierarchy for optimizing degree-4 polynomials with non-negative coefficients. Formally we show:

**Theorem 1.7.** *Let  $g := f^{q/4}$  where  $f$  is the degree-4 polynomial as in Theorem 1.6. Then*

$$\frac{\|g\|_{sp}}{\|g\|_2} \geq \frac{n^{q/24}}{(q \log n)^{O(q)}}.$$

Our lower bound on  $\|f^{q/4}\|_{sp}$  is based on a general tool that allows one to “lift” level-4  $\|\cdot\|_{sp}$  gaps, that meet one additional condition, to higher levels. While we derive final results only for the weaker relaxation  $\|\cdot\|_{sp}$ , the underlying structural result can be used to lift SoS lower bounds (i.e. gaps for  $\Lambda(\cdot)$ ) as well, provided the SoS solution matrix  $X$  satisfies PSD-ness of two other matrices of appropriately related shapes to  $X$  (Corollary 8.6) — this inspired us to name our tool “Tetris theorem.” Recently, the insightful pseudo-calibration approach [BHK<sup>+</sup>16] has provided a recipe to give higher level SoS lower bounds for certain *average-case* problems. We believe our lifting result might similarly be useful in the context of *worst-case* problems, where in order to get higher degree lower bounds, it suffices to give lower bounds for constant degree SoS with some additional structural properties.

### 1.3. Related Previous and Recent Works

Polynomial optimization is a vast area with several previous results. Below, we collect the results most relevant for comparison with the ones in this paper, grouped by the class of polynomials. Please see the excellent monographs [Lau09, Las09] for a survey.

**Arbitrary Polynomials.** For general homogeneous polynomials of degree- $d$ , an  $O_d(n^{d/2-1})$  approximation was given by He et al. [HLZ10], which was improved to  $O_d((n/\log n)^{d/2-1})$  by So [So11]. The convergence of SDP hierarchies for polynomial optimization was analyzed by Doherty and Wehner [DW12]. However, their result only applies to relaxations given by  $\Omega(n)$  levels of the SoS hierarchy (Theorem 7.1 in [DW12]). Thus, our results can be seen as giving an interpolation between the polynomial time algorithms obtained by [HLZ10, So11] and the exponential time algorithms given by  $\Omega(n)$  levels of SoS, although the bounds obtained by [DW12] are tighter (by a factor of  $2^{O(d)}$ ) for  $q = \Omega(n)$  levels.

For the case of arbitrary polynomials, we believe a tradeoff between running time and approximation quality similar to ours can also be obtained by considering the tradeoffs for the results of Brieden et al. [BGK<sup>+</sup>01] used by So [So11]. However, to the best of our knowledge, this is not published. In particular, So uses the techniques of Khot and Naor [KN08] to reduce degree- $d$  polynomial optimization to  $d - 2$  instances of the problem of optimizing the  $\ell_2$  diameter of a convex body. This is solved by [BGK<sup>+</sup>01], who give an  $O((n/k)^{1/2})$  approximation in time  $2^k \cdot n^{O(1)}$ . We believe this can be combined with proof of So, to yield a  $O_d((n/q)^{d/2-1})$  approximation in time  $2^q$ . We note here that the method of Khot and Naor [KN08] cannot be improved further (up to polylog) for the case  $d = 3$  (see Appendix A). Our results for the case of arbitrary polynomials show that similar bounds can also be obtained by a very generic algorithm given by the SoS hierarchy. Moreover, the general techniques developed here are versatile and demonstrably

applicable to various other cases (like polynomials with non-negative coefficients, sparse polynomials, worst-case sparse PCA) where no alternate proofs are available. The techniques of [KN08, So11] are oblivious to the structure in the polynomials and it appears to be unlikely that similar results can be obtained by using diameter estimation techniques.

**Polynomials with Non-negative Coefficients.** The case of polynomials with non-negative coefficients was considered by Barak, Kelner, and Steurer [BKS14] who proved that the relaxation obtained by  $\Omega(d^3 \cdot \log n / \varepsilon^2)$  levels of the SoS hierarchy provides an  $\varepsilon \cdot \|f\|_{BKS}$  additive approximation to the quantity  $\|f\|_2$ . Here, the parameter we denote by  $\|f\|_{BKS}$  corresponds to a relaxation for  $\|f\|_2$  that is weaker than the one given by  $\|f\|_{sp}$ .<sup>2</sup> Their results can be phrased as showing that a relaxation obtained by  $q$  levels of the SoS hierarchy gives an approximation ratio of

$$1 + \left( \frac{d^3 \cdot \log n}{q} \right)^{1/2} \cdot \frac{\|f\|_{BKS}}{\|f\|_2}.$$

Motivated by connections to quantum information theory, they were interested in the special case where  $\|f\|_{BKS} / \|f\|_2$  is bounded by a constant. However, this result does not imply strong multiplicative approximations outside of this special case since in general  $\|f\|_{BKS}$  and  $\|f\|_2$  can be far apart. In particular, we are able to establish that there exist polynomials  $f$  with non-neg. coefficients such that  $\|f\|_{BKS} / \|f\|_2 \geq n^{d/24}$ . Moreover we conjecture that the worst-case gap between  $\|f\|_{BKS}$  and  $\|f\|_2$  for polynomials with non-neg. coefficients is as large as  $\tilde{\Omega}_d((n/d)^{d/4-1/2})$  (note that the conjectured  $(n/d)^{d/4-1/2}$  gap for non-negative coefficient polynomials is realizable using arbitrary polynomials, i.e. it was established in [BGL16] that polynomials with i.i.d.  $\pm 1$  coefficients achieve this gap w.h.p.).

Our results show that  $q$  levels of SOS gives an  $(n/q)^{d/4-1/2}$  approximation to  $\|f\|_2$  which has a better dependence on  $q$  and consequently, converges to a constant factor approximation after  $\Omega(n)$  levels.

**2-to-4 norm.** It was proved in [BKS14] that for any matrix  $A$ ,  $q$  levels of the SoS hierarchy approximates  $\|A\|_{2 \rightarrow 4}^4 = \|\|Ax\|_4\|_2^4$  (i.e. the fourth power of the 2-to-4-norm) within a factor of

$$1 + \left( \frac{\log n}{q} \right)^{1/2} \cdot \frac{\|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2}{\|A\|_{2 \rightarrow 4}^4}.$$

Brandao and Harrow [BH15] also gave a nets based algorithm with runtime  $2^q$  that achieves the same approximation as above. Here again, the cases of interest were those matrices for which  $\|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2$  and  $\|A\|_{2 \rightarrow 4}^4$  are at most constant apart.

We would like to bring attention to an open problem in this line of work. It is not hard to show that for an  $m \times n$  matrix  $A$  with i.i.d. Gaussian entries,  $\|A\|_{2 \rightarrow 2}^2 = \Theta(m+n)$ ,  $\|A\|_{2 \rightarrow \infty}^2 = \Theta(n)$ , and  $\|A\|_{2 \rightarrow 4}^2 = \Theta(m+n^2)$  which implies the worst case approximation factor achieved above is  $\Omega(n/\sqrt{q})$  when we take  $m = \Omega(n^2)$ .

Our result for arbitrary polynomials of degree-4, achieves an approximation factor of  $O(n/q)$  after  $q$  levels of SoS which implies that the current best known approximation 2-to-4 norm is oblivious to the structure of the 2-to-4 polynomial and seems to suggest that

---

<sup>2</sup>Specifically,  $\|f\|_{BKS}$  minimizes the spectral norm over a smaller set of matrix representations of  $f$  than  $\|f\|_{sp}$  which allows all matrix representations.

this problem can be better understood for arbitrary tall matrices. For instance, can one get a  $\sqrt{m}/q$  approximation for  $(m \times n)$  matrices (note that [BH15] already implies a  $\sqrt{m}/q$ -approximation for all  $m$ , and our result implies a  $\sqrt{m}/q$ -approximation when  $m = \Omega(n^2)$ ).

**Random Polynomials.** For the case when  $f$  is a degree- $d$  homogeneous polynomial with i.i.d. random  $\pm 1$  coefficients [BGL16, RRS16] showed that degree- $q$  SoS certifies an upper bound on  $\|f\|_2$  that is with high probability at most  $\tilde{O}((n/q)^{d/4-1/2}) \cdot \|f\|_2$ . Curiously, this matches our approximation guarantee for the case of *arbitrary* polynomials with non-negative coefficients. This problem was also studied for the case of sparse random polynomials in [RRS16] motivated by applications to refuting random CSPs.

## 1.4. Organization of the Paper

We cover some preliminaries in Section 2 and provide an overview of our proofs and techniques in Section 3. Section 4 provides details of various relaxations used in this paper, and their duals in terms of the Sum-of-Squares hierarchy. We first give a basic version of the reduction from general to multilinear polynomials in Section 5.1, which only obtains a weaker result (without the additive term in the exponent). Section 6 gives a generalization of this reduction, which yields Theorem 1.1. We prove an SoS lower bound for degree-4 polynomials with non-negative coefficients in Section 7. In Section 8, we provide a general technique for lifting lower bounds for the slightly weaker relaxation given by  $\|f\|_{sp}$ , to relaxation higher level relaxations.

## 2. Preliminaries and Notation

**Polynomials.** We use  $\mathbb{R}_d[x]$  to denote the set of all homogeneous polynomials of degree (exactly)  $d$ . Similarly,  $\mathbb{R}_d^+[x]$  is used to denote the set of polynomials with non-negative coefficients. All polynomials considered in this paper will be  $n$ -variate and homogeneous (with  $x$  denoting the set of  $n$  variables  $x_1, \dots, x_n$ ) unless otherwise stated.

A multi-index is defined as sequence  $\alpha \in \mathbb{N}^n$ . We use  $|\alpha|$  to denote  $\sum_{i=1}^n \alpha_i$  and  $\mathbb{N}_d^n$  (resp.  $\mathbb{N}_{\leq d}^n$ ) to denote the set of all multi-indices  $\alpha$  with  $|\alpha| = d$  (resp.  $|\alpha| \leq d$ ). Thus, a polynomial  $f \in \mathbb{R}_d[x]$  can be expressed in terms of its coefficients as

$$f(x) = \sum_{\alpha \in \mathbb{N}_d^n} f_\alpha \cdot x^\alpha,$$

where  $x^\alpha$  is used to denote the monomial corresponding to  $\alpha$ . A polynomial is multilinear if  $\alpha \leq \mathbb{1}$  whenever  $f_\alpha \neq 0$ , where  $\mathbb{1}$  denotes the multi-index  $1^n$ . We use the notation  $\alpha^r$  to denote the vector  $(\alpha_1^r, \dots, \alpha_n^r)$  for  $r \in \mathbb{R}$ . In general, with the exception of absolute-value, any scalar function when applied to a vector/multi-index returns the vector obtained by applying the function entry-wise. We also use  $\circ$  to denote the Hadamard (entry-wise) product of two vectors.

To save the additive constant terms in the exponent of our results, we will need to extract the “quadratic part” of a given polynomial, and use the fact that eigenvalue problems are easy for quadratic polynomials. We thus define the following polynomials where the coefficients themselves may be polynomials (in the same variables).

**Definition 2.1 (Folded Polynomials).** A degree- $(d_1, d_2)$  folded polynomial  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  is

defined to be a polynomial of the form

$$f(x) = \sum_{\alpha \in \mathbb{N}_{d_1}^n} \bar{f}_\alpha(x) \cdot x^\alpha,$$

where each  $\bar{f}_\alpha(x) \in \mathbb{R}_{d_2}[x]$  is a homogeneous polynomial of degree  $d_2$ . Folded polynomials over  $\mathbb{R}^+$  are defined analogously.

- We refer to the polynomials  $\bar{f}_\alpha$  as the folds of  $f$  and the terms  $x^\alpha$  as the monomials in  $f$ .
- A folded polynomial can also be used to define a degree  $d_1 + d_2$  polynomial by multiplying the monomials with the folds (as polynomials in  $\mathbb{R}[x]$ ). We refer to this polynomial in  $\mathbb{R}_{d_1+d_2}[x]$  as the unfolding of  $f$ , and denote it by  $U(f)$ .
- For a degree  $(d_1, d_2)$ -folded polynomial  $f$  and  $r \in \mathbb{N}$ , we take  $f^r$  to be a degree- $(r \cdot d_1, r \cdot d_2)$  folded polynomial, obtained by multiplying the folds as coefficients.

**Matrices.** For  $k \in \mathbb{N}$ , we will consider  $n^k \times n^k$  matrices  $M$  with real entries. All matrices considered in this paper should be taken to be symmetric (unless otherwise stated). We index entries of the matrix  $M$  as  $M[I, J]$  by tuples  $I, J \in [n]^k$ .

A tuple  $I = (i_1, \dots, i_k)$  naturally corresponds to a multi-index  $\alpha(I) \in \mathbb{N}_k^n$  with  $|\alpha(I)| = k$ , i.e.  $\alpha(I)_j = |\{\ell \mid i_\ell = j\}|$ . For a tuple  $I \in [n]^k$ , we define  $\mathcal{O}(I)$  the set of all tuples  $J$  which correspond to the same multi-index i.e.,  $\alpha(I) = \alpha(J)$ . Thus, any multi-index  $\alpha \in \mathbb{N}_k^n$ , corresponds to an equivalence class in  $[n]^k$ . We also use  $\mathcal{O}(\alpha)$  to denote the class of all tuples corresponding to  $\alpha$ .

Note that a matrix of the form  $(x^{\otimes k})(x^{\otimes k})^T$  has many additional symmetries, which are also present in solutions to programs given by the SoS hierarchy. To capture this, consider the following definition:

**Definition 2.2 (SoS-Symmetry).** A matrix  $M$  which satisfies  $M[I, J] = M[K, L]$  whenever  $\alpha(I) + \alpha(J) = \alpha(K) + \alpha(L)$  is referred to as SoS-symmetric.

**Remark.** It is easily seen that every homogeneous polynomial has a unique SoS-Symmetric matrix representation.

### 3. Overview of Proofs and Techniques

In the interest of clarity, we shall present all techniques for the special case where  $f$  is an arbitrary degree-4 homogeneous polynomial. We shall further assume that  $\|f\|_2 = f_{\max}$  just so that  $\Lambda(f)$  is a relaxation of  $\|f\|_2$ . Summarily, the goal of this section is to give an overview of an  $O(n/q)$ -approximation of  $\|f\|_2$ , i.e.

$$\Lambda\left(f^{q/4}\right)^{4/q} \leq O(n/q) \cdot \|f\|_2.$$

Many of the high level ideas remain the same when considering higher degree polynomials and special classes like polynomials with non-negative coefficients, or sparse polynomials.

### 3.1. Warmup: $(n^2/q^2)$ -Approximation

We begin with seeing how to analyze constant levels of the  $\Lambda(\cdot)$  relaxation and will then move onto higher levels in the next section. The level-4 relaxation actually achieves an  $n$ -approximation, however we will start with  $n^2$  as a warmup and cover the  $n$ -approximation a few sections later.

#### 3.1.1. $n^2$ -Approximation using level-4 relaxation

We shall establish that  $\Lambda(f) \leq O(n^2) \cdot \|f\|_2$ . Let  $M_f$  be the SoS-symmetric representation of  $f$ , let  $x_{i_1}x_{i_2}x_{i_3}x_{i_4}$  be the monomial whose coefficient in  $f$  has the maximum magnitude, and let  $B$  be the magnitude of this coefficient. Now by Gershgorin circle theorem, we have  $\Lambda(f) \leq \|M_f\|_2 \leq n^2 \cdot B$ .

It remains to establish  $\|f\|_2 = \Omega(B)$ . To this end, define the decoupled polynomial  $\mathcal{F}(x, y, z, t) := (x \otimes y)^T \cdot M_f \cdot (z \otimes t)$  and define the decoupled two-norm as

$$\|\mathcal{F}\|_2 := \sup_{\|x\|, \|y\|, \|z\|, \|t\|=1} \mathcal{F}(x, y, z, t).$$

It is well known that  $\|f\|_2 = \Theta(\|\mathcal{F}\|_2)$  (see Lemma 3.1). Thus, we have,

$$\|f\|_2 = \Omega(\|\mathcal{F}\|_2) \geq \Omega(|\mathcal{F}(e_{i_1}, e_{i_2}, e_{i_3}, e_{i_4})|) = \Omega(B) = \Omega(\Lambda(f)/n^2).$$

In order to better analyze  $\Lambda(f^{q/4})^{4/q}$  we will need to introduce some new techniques.

#### 3.1.2. $(n^2/q^2)$ -Approximation Assuming Theorem 1.5

We will next show that  $\Lambda(f^{q/4})^{4/q} \leq O(n^2/q^2) \cdot \|f\|_2$  (for  $q$  divisible by 4). In fact, one can show something stronger, namely that for every homogeneous polynomial  $g$  of degree- $q$ ,  $\Lambda(g) \leq 2^{O(q)} \cdot (n/q)^{q/2} \cdot \|g\|_2$  which clearly implies the above claim (also note that for the target  $O(n^2/q^2)$ -approximation to  $\|f\|_2$ , losses of  $2^{O(q)}$  in the estimate of  $\|g\|_2$  are negligible, while factors of the order  $q^{\Omega(q)}$  are crucial).

Given the additional freedom in choice of representation (due to the polynomial having higher degree), a first instinct would be to completely symmetrize, i.e. take the SoS-symmetric representation of  $g$ , and indeed this works for multilinear  $g$  (see Theorem 5.16 for details).

However, the above approach of taking the SoS-symmetric representation breaks down when the polynomial is non-multilinear. To circumvent this issue, we employ Theorem 1.5 which on combining with the aforementioned multilinear polynomial result, yields that for every homogeneous polynomial  $g$  of degree- $q$ ,  $\Lambda(g) \leq (n/q)^{q/2} \cdot \|g\|_2$ . The proofs of Theorem 1.5 and its generalizations (that will be required for the  $n/q$  approximation), are quite non-trivial and are the most technically involved sections of our upper bound results. We shall next give an outline of the proof of Theorem 1.5.

#### 3.1.3. Reduction to Optimization of Multi-linear Polynomials

One of the main techniques we develop in this work, is a way of reducing the optimization problem for general polynomials to that of multi-linear polynomials, which *does not increase the number of variables*. While general techniques for reduction to the multi-linear case

have been widely used in the literature [KN08, HLZ10, So11] (known commonly as decoupling/polarization techniques), these reduce the problem to optimizing a multi-linear polynomial in  $n \cdot d$  variables (when the given polynomial  $h$  is of degree  $d$ ). Below is one example:

**Lemma 3.1 ([HLZ10]).** *Let  $\mathcal{A}$  be a SoS-symmetric  $d$ -tensor and let  $h(x) := \langle \mathcal{A}, x^{\otimes d} \rangle$ . Then  $\|h\|_2 \geq 2^{-O(d)} \cdot \max_{\|x^i\|=1} \langle \mathcal{A}, x^1 \otimes \dots \otimes x^d \rangle$ .*

Since we are interested in the improvement in approximation obtained by considering  $f^{q/4}$  for a large  $q$ , applying these would yield a multi-linear polynomial in  $n \cdot q$  variables. For our analysis, this increase in variables exactly cancels the advantage we obtain by considering  $f^{q/4}$  instead of  $f$  (i.e., the advantage obtained by using  $q$  levels of the SoS hierarchy).

We can uniquely represent a homogeneous polynomial  $g$  of degree  $q$  as

$$g(x) = \sum_{|\alpha| \leq q/2} x^{2\alpha} \cdot G_{2\alpha}(x) = \sum_{r=0}^{q/2} \sum_{|\alpha|=r} x^{2\alpha} \cdot G_{2\alpha}(x) = \sum_{r=0}^{q/2} g_r(x), \quad (3.1)$$

where each  $G_{2\alpha}$  is a multi-linear polynomial and  $g_r(x) := \sum_{|\alpha|=r} x^{2\alpha} \cdot G_{2\alpha}(x)$ . We reduce the problem to optimizing  $\|G_{2\alpha}\|_2$  for each of the polynomials  $G_{2\alpha}$ . More formally, we show that

$$\frac{\Lambda(g)}{\|g\|_2} \leq \max_{\alpha \in \mathbb{N}_{\leq q/2}^n} \frac{\Lambda(G_{2\alpha})}{\|G_{2\alpha}\|_2} \cdot 2^{O(q)} \quad (3.2)$$

As a simple and immediate example of its applicability, (3.2) provides a simple proof of a polytime constant factor approximation for optimization over the simplex (actually this case is known to admit a PTAS [dKLP06, dKLS15]). Indeed, observe that a simplex optimization problem for a degree- $q/2$  polynomial in the variable vector  $y$  can be reduced to a sphere optimization by substituting  $y_i = x_i^2$ . Now since every variable present in a monomial has even degree in that monomial, each  $G_{2\alpha}$  is constant, which implies a constant factor approximation (dependent on  $q$ ) on applying (3.2).

Returning to our overview of the proof, note that given representations of each of the polynomials  $G_{2\alpha}$ , each of the polynomials  $g_r$  can be represented as a block-diagonal matrix with one block corresponding to each  $\alpha$ . Combining this with triangle inequality and the fact that the maximum eigenvalue of a block-diagonal matrix is equal to the maximum eigenvalue of one of the blocks, gives the following inequality:

$$\Lambda(g) \leq (1 + q/2) \cdot \max_{\alpha \in \mathbb{N}_{\leq q/2}^n} \Lambda(G_{2\alpha}). \quad (3.3)$$

We can further strengthen (3.3) by averaging the "best" representation of  $G_{2\alpha}$  over  $|\mathcal{O}(\alpha)|$  diagonal-blocks which all correspond to  $x^{2\alpha}$ . This is the content of Lemma 5.2 wherein we show

$$\Lambda(g) \leq (1 + q/2) \cdot \max_{\alpha \in \mathbb{N}_{\leq q/2}^n} \frac{\Lambda(G_{2\alpha})}{|\mathcal{O}(\alpha)|}. \quad (3.4)$$

Since  $|\mathcal{O}(\alpha)|$  can be as large as  $q^{\Omega(q)}$ , the above strengthening is crucial. We then prove the following inequality, which shows that the decomposition in Eq. (3.1) not only gives a block-diagonal decomposition for matrix representations of  $g$ , but can in fact be thought of as a "block-decomposition" of the tensor corresponding to  $g$  (with regards to computing

$\|g\|_2$ ). Just as the maximum eigenvalue of a block-diagonal matrix is at least the maximum eigenvalue of a block, we show that

$$\|g\|_2 \geq 2^{-O(q)} \cdot \max_{\alpha \in \mathbb{N}_{\leq q/2}^n} \frac{\|G_{2\alpha}\|_2}{|\mathcal{O}(\alpha)|}. \quad (3.5)$$

The above inequality together with (3.4), implies (3.2).

### 3.1.4. Bounding $\|g\|_2$ via a new weak decoupling lemma

Recall that the expansion of  $g(x)$  in Eq. (3.1), contains the term  $x^{2\alpha} \cdot G_{2\alpha}(x)$ . The key part of proving the bound in (3.5) is to show the following “weak decoupling” result for  $x^{2\alpha}$  and  $G_{2\alpha}$ .

$$\forall \alpha \quad \|g\|_2 \geq \max_{\|y\|=\|x\|=1} y^{2\alpha} \cdot G_{2\alpha}(x) \cdot 2^{-O(q)} = \max_{\|y\|=1} y^{2\alpha} \cdot \|G_{2\alpha}\|_2 \cdot 2^{-O(q)}.$$

The proof of (3.5) can then be completed by considering the unit vector  $y := \sqrt{\alpha}/\sqrt{|\alpha|}$ , i.e.  $y := \sum_{i \in [n]} \frac{\sqrt{\alpha_i}}{\sqrt{|\alpha|}} \cdot e_i$ . A careful calculation shows that  $y^{2\alpha} \geq 2^{-O(q)}/|\mathcal{O}(\alpha)|$  which finishes the proof.

The primary difficulty in establishing the above decoupling is the possibility of cancellations. To see this, let  $x^*$  be the vector realizing  $\|G_{2\alpha}\|_2$  and substitute  $z = (x^* + y)$  into  $g$ . Clearly,  $y^{2\alpha} \cdot G_{2\alpha}(x^*)$  is a term in the expansion of  $g(z)$ , however there is no guarantee that the other terms in the expansion don't cancel out this value. To fix this our proof relies on multiple delicate applications of the first-moment method, i.e. we consider a complex vector random variable  $Z(x^*, y)$  that is a function of  $x^*$  and  $y$ , and argue about  $\mathbb{E}[|g(Z)|]$ .

**The base case of  $\alpha = 0^n$ .** We first consider the base case with  $\alpha = 0^n$ , where we define  $y^{2\alpha} = 1$ . This amounts to showing that for every homogeneous polynomial  $h$  of degree  $t$ ,  $\|h\|_2 \geq \|h_m\|_2 \cdot 2^{-O(t)}$  where  $h_m$  is the restriction of  $h$  to it's multilinear monomials.

Given the optimizer  $x^*$  of  $\|h_m\|_2$ , let  $z$  be a random vector such that each  $Z_i = x_i^*$  with probability  $p$  and  $Z_i = 0$  otherwise. Then,  $\mathbb{E}[h(Z)]$  is a *univariate* degree- $t$  polynomial in  $p$  with the coefficient of  $p^t$  equal to  $h_m(x^*)$ . An application of Chebyshev's extremal polynomial inequality (Lemma 5.5) then gives that there exists a value of the probability  $p$  such that

$$\|h\|_2 \geq \mathbb{E}[|h(Z)|] \geq |\mathbb{E}[h(Z)]| \geq 2^{-O(t)} \cdot |h_m(x^*)| = 2^{-O(t)} \cdot \|h_m\|_2.$$

For the case of general  $\alpha$ , we first pass to the *complex version* of  $\|g\|_2$  defined as

$$\|g\|_2^c := \sup_{z \in \mathbb{C}^n, \|z\|=1} |g(z)|.$$

We use another averaging argument together with an application of the polarization lemma (Lemma 3.1) to show that we do not loose much by considering  $\|g\|_2^c$ . In particular,  $\|g\|_2 \leq \|g\|_2^c \leq 2^{O(q)} \cdot \|g\|_2$ .

**The case of  $g = g_r$ .** In this case, the problem reduces to showing that for all  $\alpha \in \mathbb{N}_r^n$  and for all  $y \in \mathbb{S}^{n-1}$ ,

$$\|g_r\|_2^c \geq y^{2\alpha} \cdot \|G_{2\alpha}\|_2 \cdot 2^{-O(q)}.$$

Fix any  $\alpha \in \mathbb{N}_r^n$ , and let  $\omega \in \mathbb{C}^n$  be a complex vector random variable, such that  $\omega_i$  is an independent and uniformly random  $(2\alpha_i + 1)$ -th root of unity. Let  $\Xi$  be a random  $(q - 2r + 1)$ -th root of unity, and let  $x^*$  be the optimizer of  $\|G_{2\alpha}\|_2$ . Let  $Z := \omega \circ y + \Xi \cdot x^*$ , where  $\omega \circ y$  denotes the coordinate-wise product. Observe that for any  $\alpha', \gamma$  such that  $|\alpha'| = r$ ,  $|\gamma| = q - 2r$ ,  $\gamma \leq \mathbb{1}$ ,

$$\mathbb{E} \left[ \prod_i \omega_i \cdot \Xi \cdot Z^{2\alpha' + \gamma} \right] = \begin{cases} y^{2\alpha} \cdot (x^*)^\gamma & \text{if } \alpha' = \alpha \\ 0 & \text{otherwise} \end{cases}$$

By linearity, this implies  $\mathbb{E} [\prod_i \omega_i \cdot \Xi \cdot g_r(Z)] = y^{2\alpha} \cdot G_{2\alpha}(x^*)$ . The claim then follows by noting that

$$\|g_r\|_2^c \geq \mathbb{E} \|g_r(Z)\| = \mathbb{E} \left[ \left| \prod_i \omega_i \cdot \Xi \cdot g_r(Z) \right| \right] \geq \left| \mathbb{E} \left[ \prod_i \omega_i \cdot \Xi \cdot g_r(Z) \right] \right| \geq y^{2\alpha} \cdot \|G_{2\alpha}\|_2.$$

**The general case.** The two special cases considered here correspond to the cases when we need to extract a specific  $g_r$  (for  $r = 0$ ), and when we need to extract a fixed  $\alpha$  from a given  $g_r$ . The argument for the general case uses a combination of the arguments for both these cases. Moreover, to get an  $O(n/q)$  approximation, we also need versions of such decoupling lemmas for folded polynomials to take advantage of “easy substructures” as described next.

## 3.2. Exploiting Easy Substructures via Folding and Improved Approximations

To obtain the desired  $n/q$ -approximation to  $\|f\|_2$ , we need to use the fact that the problem of optimizing quadratic polynomials can be solved in polynomial time, and moreover that SoS captures this. More generally, in this section we consider the problem of getting improved approximations when a polynomial contains “easy substructures”. It is not hard to obtain improved guarantees when considering constant levels of SoS. The second main technical contribution of our work is in giving sufficient conditions under which higher levels of SoS improve on the approximation of constant levels of SoS, when considering the optimization problem over polynomials containing “easy substructures”.

As a warmup, we shall begin with seeing how to exploit easy substructures at constant levels by considering the example of degree-4 polynomials that trivially “contain” quadratics.

### 3.2.1. $n$ -Approximation using Degree-4 SoS

Given a degree-4 homogeneous polynomial  $f$  (assume  $f$  is multilinear for simplicity), we consider a degree-(2,2) folded polynomial  $h$ , whose unfolding yields  $f$ , chosen so that  $\max_{\|y\|=1} \|h(y)\|_2 = \Theta(\|f\|_2)$  (recall that an evaluation of a folded polynomial returns a polynomial, i.e., for a fixed  $y$ ,  $h(y)$  is a quadratic polynomial in the indeterminate  $x$ ). Such an  $h$  always exists and is not hard to find based on the SoS-symmetric representation of  $f$ . Also recall,

$$h(x) = \sum_{|\beta|=2, \beta \leq \mathbb{1}} \bar{h}_\beta(x) \cdot x^\beta,$$



where each  $\bar{h}_\beta$  is a quadratic polynomial (the aforementioned phrase “easy substructures” is referencing the folds:  $\bar{h}_\beta$  which are easy to optimize). Now by assumption we have,

$$\|f\|_2 \geq \max_{|\beta|=2, \beta \leq \mathbf{1}} \|h(\beta/\sqrt{2})\|_2 = \max_{|\beta|=2, \beta \leq \mathbf{1}} \|\bar{h}_\beta\|_2/2.$$

We then apply the block-matrix generalization of Gershgorin circle theorem to the SoS-symmetric matrix representation of  $f$  to show that

$$\Lambda(f) \leq \|f\|_{sp} \leq n \cdot \max_{|\beta|=2, \beta \leq \mathbf{1}} \|\bar{h}_\beta\|_{sp} = n \cdot \max_{|\beta|=2, \beta \leq \mathbf{1}} \|\bar{h}_\beta\|_2,$$

where the last step uses the fact that  $\bar{h}_\beta$  is a quadratic, and  $\|\cdot\|_{sp}$  is a tight relaxation of  $\|\cdot\|_2$  for quadratics. This yields the desired  $n$ -approximation.

### 3.2.2. $n/q$ -approximation using Degree- $q$ SoS

Following the cue of the  $n^2/q^2$ -approximation, we derive the desired  $n/q$  bound by proving a folded-polynomial analogue of every claim in the previous section (including the multilinear reduction tools), a notable difference being that when we consider a power  $f^{q/4}$  of  $f$ , we need to consider degree- $(q - 2q/4, 2q/4)$  folded polynomials, since we want to use the fact that any *product of  $q/4$  quadratic polynomials* is “easy” for SoS (in contrast to Section 3.2.1 where we only used the fact quadratic polynomials are easy for SoS). We now state an abstraction of the general approach we use to leverage the tractability of the folds.

**Conditions for Exploiting “Easy Substructures” at Higher Levels of SoS.** Let  $d := d_1 + d_2$  and  $f := U(h)$  where  $h$  is a degree- $(d_1, d_2)$  folded polynomial that satisfies

$$\sup_{\|y\|=1} \|h(y)\|_2 = \Theta_d(\|f\|_2).$$

Implicit in Section 6, is the following theorem we believe to be of independent interest:

**Theorem 3.2.** *Let  $h, f$  be as above, and let*

$$\Gamma := \min \left\{ \frac{\Lambda(p)}{\|p\|_2} \mid p(x) \in \text{span}(\bar{h}_\beta \mid \beta \in \mathbb{N}_{d_2}^n) \right\}.$$

*Then for any  $q$  divisible by  $2d$ ,  $\Lambda(f^{q/d})^{d/q} \leq O_d(\Gamma \cdot (n/q)^{d_1/2}) \cdot \|f\|_2$ .*

In other words, if degree- $d_2$  SoS gives a good approximation for every polynomial in the subspace spanned by the folds of  $h$ , then higher levels of SoS give an improving approximation that exploits this. In this work, we only apply the above with  $\Gamma = 1$ , where exact optimization is easy for the space spanned by the folds.

While we focus on general polynomials for the overview, let us remark that in the case of polynomials with non-negative coefficients, the approximation factor in Theorem 3.2 becomes  $O_d(\delta \cdot (n/q)^{d_1/4})$ .

### 3.3. Lower Bounds for Polynomials with Non-negative Coefficients

#### 3.3.1. Degree-4 Lower Bound for Polynomials with Non-Negative Coefficients

We discuss some of the important ideas from the proof of Theorem 1.6. The lower bound proved by a subset of the authors in [BGL16] proves a large ratio  $\frac{\Lambda(f)}{\|f\|_2}$  by considering a random polynomial  $f$  where each coefficient of  $f$  is an independent (Gaussian) random variable with bounded variance. The most natural adaptation of the above strategy to degree-4 polynomials with non-negative coefficients is to consider a random polynomial  $f$  where each coefficient  $f_\alpha$  is independently sampled such that  $f_\alpha = 1$  with probability  $p$  and  $f_\alpha = 0$  with probability  $1 - p$ . However, this construction fails for every choice of  $p$ . If we let  $A \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the natural matrix representation of  $f$  (i.e., each coefficient  $f_\alpha$  is distributed uniformly among the corresponding entries of  $A$ ), the Perron-Frobenius theorem shows that  $\|A\|_2$  is less than the maximum row sum  $\max(\tilde{O}(n^2p), 1)$  of  $M$ , which is also an upper bound on  $\Lambda(f)$ . However, we can match this bound by (within constant factors) choosing  $x = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$  when  $p \geq 1/n^2$ . Also, when  $p < 1/n^2$ , we can take any  $\alpha$  with  $f_\alpha = 1$  and set  $x_i = 1/2$  for all  $i$  with  $\alpha_i > 0$ , which achieves a value of  $1/16$ .

We introduce another natural distribution of random non-negative polynomials that bypasses this problem. Let  $G = (V, E)$  be a random graph drawn from the distribution  $G_{n,p}$  (where we choose  $p = n^{-1/3}$  and  $V = [n]$ ). Let  $\mathcal{C} \subseteq \binom{V}{4}$  be the set of 4-cliques in  $G$ . The polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in \mathcal{C}} x_{i_1} x_{i_2} x_{i_3} x_{i_4}.$$

Instead of trying  $\Theta(n^4)$   $p$ -biased random bits, we use  $\Theta(n^2)$  of them. This limited independence bypasses the problem above, since the rows of  $A$  now have significantly different row sums:  $\Theta(n^2p)$  rows that correspond to an edge of  $G$  will have row sum  $\Theta(n^2p^5)$ , and all other rows will be zeros. Since these  $n^2p$  rows (edges) are chosen independently from  $\binom{[n]}{2}$ , they still reveal little information that can be exploited to find a  $n$ -dimensional vector  $x$  with large  $f(x)$ . However, the proof requires a careful analysis of the trace method (to bound the spectral norm of an “error” matrix).

It is simple to prove that  $\|f\|_{sp} \geq \Omega(\sqrt{n^2p^5}) = \Omega(n^{1/6})$  by considering the Frobenius norm of the  $n^2p \times n^2p$  principal submatrix, over any matrix representation (indeed,  $A$  is the minimizer). To prove  $\Lambda(f) \geq \tilde{\Omega}(n^{1/6})$ , we construct a moment matrix  $M$  that is SoS-symmetric, positive semidefinite, and has a large  $\langle M, A \rangle$  (see the dual form of  $\Lambda(f)$  in Section 4). It turns out that the  $n^2p \times n^2p$  submatrix of  $A$  shares spectral properties of the adjacency matrix of a random graph  $G_{n^2p, p^4}$ , and taking  $M := c_1A + c_2I$  for some identity-like matrix  $I$  proves  $\Lambda(f) \geq \tilde{\Omega}(n^{1/6})$ . An application of the trace method is needed to bound  $c_2$ .

To upper bound  $\|f\|_2$ , we first observe that  $\|f\|_2$  is the same as the following natural combinatorial problem up to an  $O(\log^4 n)$  factor: find four sets  $S_1, S_2, S_3, S_4 \subseteq V$  that maximize

$$\frac{|\mathcal{C}_G(S_1, S_2, S_3, S_4)|}{\sqrt{|S_1||S_2||S_3||S_4|}}$$

where  $|\mathcal{C}_G(S_1, S_2, S_3, S_4)|$  is the number of 4-cliques  $\{v_1, \dots, v_4\}$  of  $G$  with  $v_i \in S_i$  for  $i = 1, \dots, 4$ . The number of copies of a fixed subgraph  $H$  in  $G_{n,p}$ , including its tail bound, has

been actively studied in probabilistic combinatorics [Vu01, KV04, JOR04, Cha12, DK12a, DK12b, LZ16], though we are interested in bounding the 4-clique density of *every* 4-tuple of subsets simultaneously. The previous results give a strong enough tail bound for a union bound, to prove that the optimal value of the problem is  $O(n^2 p^6 \cdot \log^{O(1)} n)$  when  $|S_1| = \dots = |S_4|$ , but this strategy inherently does not work when the set sizes become significantly different. However, we give a different analysis for the above asymmetric case, showing that the optimum is still no more than  $O(n^2 p^6 \cdot \log^{O(1)} n)$ .

### 3.3.2. Lifting Stable Degree-4 Lower Bounds

For a degree- $t$  ( $t$  even) homogeneous polynomial  $f$ , note that  $\max\{|\Lambda(f)|, |\Lambda(-f)|\}$  is a relaxation of  $\|f\|_2$ .  $\|f\|_{sp}$  is a slightly weaker (but still quite natural) relaxation of  $\|f\|_2$  given by

$$\|f\|_{sp} := \inf \{ \|M\|_2 \mid M \text{ is a matrix representation of } f \}.$$

As in the case of  $\Lambda(f)$ , for a (say) degree-4 polynomial  $f$ ,  $\|f^{q/4}\|_{sp}^{4/q}$  gives a hierarchy of relaxations for  $\|f\|_2$ , for increasing values of  $q$ .

We give an overview of a general method of “lifting” certain “stable” low degree gaps for  $\|\cdot\|_{sp}$  to gaps for higher levels with at most  $q^{O(1)}$  loss in the gap. While we state our techniques for lifting degree-4 gaps, all the ideas are readily generalized to higher levels. We start with the observation that the dual of  $\|f\|_{sp}$  is given by the following “nuclear norm” program. Here  $M_f$  the canonical matrix representation of  $f$ , and  $\|X\|_{S_1}$  is the Schatten 1-norm (nuclear norm) of  $X$ , which is the sum of it’s singular values.

$$\begin{aligned} & \text{maximize} && \langle M_f, X \rangle \\ & \text{subject to :} && \|X\|_{S_1} = 1 \\ & && X \text{ is SoS symmetric} \end{aligned}$$

Now let  $X$  be a solution realizing a gap of  $\delta$  between  $\|f\|_{sp}$  and  $\|f\|_2$ . We shall next see how assuming reasonable conditions on  $X$  and  $M_f$ , one can show that  $\|f^{q/4}\|_{sp} / \|f^{q/4}\|_2$  is at least  $\delta^{q/4} / q^{O(q)}$ .

In order to give a gap for the program corresponding to  $\|f^{q/4}\|_{sp}$ , a natural choice for a solution is the symmetrized version of the matrix  $X^{\otimes q/4}$  normalized by its Schatten-1 norm i.e., for  $Y = X^{\otimes q/4}$ , we take

$$Z := \frac{Y^S}{\|Y^S\|_{S_1}} \quad \text{where} \quad Y^S[K] = \mathbb{E}_{\pi \in S_q} [Y[\pi(K)]] \quad \forall K \in [n]^q.$$

To get a good gap, we are now left with showing that  $\|Y^S\|_{S_1}$  is not too large. Note that symmetrization can drastically change the spectrum of a matrix as for different permutations  $\pi$ , the matrices  $Y^\pi[K] := Y[\pi(K)]$  can have very different ranks (while  $\|Y\|_F = \|Y^\pi\|_F$ ). In particular, symmetrization can increase the maximum eigenvalue of a matrix by polynomial factors, and thus one must carefully count the number of such large eigenvalues in order to get a good upper bound on  $\|Y^S\|_{S_1}$ . Such an upper bound is a consequence of a structural result about  $Y^S$  that we believe to be of independent interest.

To state the result, we will first need some notation. For a matrix  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  let  $T \in \mathbb{R}^{[n]^4}$  denote the tensor given by,  $T[i_1, i_2, i_3, i_4] = M[(i_1, i_2), (i_3, i_4)]$ . Also for any non-negative integers  $x, y$  satisfying  $x + y = 4$ , let  $M_{x,y} \in \mathbb{R}^{[n]^x \times [n]^y}$  denote the (rectangular) matrix given by,  $M[(i_1, \dots, i_x), (j_1, \dots, j_y)] = T[i_1, \dots, i_x, j_1, \dots, j_y]$ . Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4 SoS-Symmetric matrix, let  $M_A := M_{1,3} \otimes M_{4,0} \otimes M_{1,3}$ , let  $M_B := M_{1,3} \otimes M_{3,1}$ , let  $M_C := M$  and let  $M_D := \text{Vec}(M) \text{Vec}(M)^T = M_{0,4} \otimes M_{4,0}$ .

We show (see Theorem 8.4) that  $(M^{\otimes q/4})^S$  can be written as the sum of  $2^{O(q)}$  terms of the form:

$$C(a, b, c, d) \cdot P \cdot (M_A^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d}) \cdot P$$

where  $12a + 8b + 4c + 8d = q$ ,  $P$  is a matrix with spectral norm 1 and  $C(a, b, c, d) = 2^{O(q)}$ . This implies that controlling the spectrum of  $M_A, M_B, M$  and  $M_D$  is sufficient to control on the spectrum of  $(M^{\otimes q/4})^S$ .

Using this result with  $M := X$ , we are able to establish that if  $X$  satisfies the additional condition of  $\|X_{1,3}\|_{S_1} \leq 1$  (note that we already know  $\|X\|_{S_1} \leq 1$ ), then  $\|Y^S\|_{S_1} = 2^{O(q)}$ . Thus  $Z$  realizes a  $\langle M_f^{\otimes q/4}, Y^S \rangle / 2^{O(q)}$  gap for  $\|f^{q/4}\|_{sp}$ . On composing this result with the degree-4 gap from the previous section, we obtain an  $\|\cdot\|_{sp}$  gap of  $n^{q/24} / (q \cdot \log n)^{O(q)}$  for degree- $q$  polynomials with non-neg. coefficients. We also show the  $q$ -th level  $\|\cdot\|_{sp}$  gap for degree-4 polynomials with non-neg. coefficients is  $\tilde{\Omega}(n^{1/6}) / q^{O(1)}$ .

Even though we only derive results for the weaker relaxation  $\|\cdot\|_{sp}$ , the structural result above can be used to lift “stable” low-degree SoS lower bounds as well (i.e. gaps for  $\Lambda(\cdot)$ ), albeit with a stricter notion of stability (see Corollary 8.6). However, the problem of finding such stable SoS lower bounds remains open.

There are by now quite a few results giving near-tight lower bounds on the performance of higher level SoS relaxations for *average-case* problems [BHK<sup>+</sup>16, KMOW17, HKP<sup>+</sup>17]. However, there are few examples in the literature of matching SoS upper/lower bounds on *worst-case* problems. We believe our lifting result might be especially useful in such contexts, where in order to get higher degree lower bounds, it suffices to give stable lower bounds for constant degree SoS.

## 4. Additional Preliminaries and the SoS Hierarchy

### 4.1. Pseudoexpectations and Moment Matrices

Let  $\mathbb{R}[x]_{\leq q}$  be the vector space of polynomials with real coefficients in variables  $x = (x_1, \dots, x_n)$ , of degree at most  $q$ . For an even integer  $q$ , the degree- $q$  pseudo-expectation operator is a linear operator  $\tilde{\mathbf{E}} : \mathbb{R}[x]_{\leq q} \mapsto \mathbb{R}$  such that

1.  $\tilde{\mathbf{E}}[1] = 1$  for the constant polynomial 1.
2.  $\tilde{\mathbf{E}}[p_1 + p_2] = \tilde{\mathbf{E}}[p_1] + \tilde{\mathbf{E}}[p_2]$  for any polynomials  $p_1, p_2 \in \mathbb{R}[x]_{\leq q}$ .
3.  $\tilde{\mathbf{E}}[p^2] \geq 0$  for any polynomial  $p \in \mathbb{R}[x]_{\leq q/2}$ .

The pseudo-expectation operator  $\tilde{\mathbf{E}}$  can be described by a moment matrix  $\hat{\mathbf{M}} \in \mathbb{R}^{\mathbb{N}_{\leq q/2}^n \times \mathbb{N}_{\leq q/2}^n}$  defined as  $\hat{\mathbf{M}}[\alpha, \beta] = \tilde{\mathbf{E}}[x^{\alpha+\beta}]$  for  $\alpha, \beta \in \mathbb{N}_{\leq q/2}^n$ .

For each fixed  $t \leq q/2$ , we can also consider the principal minor of  $\widehat{M}$  indexed by  $\alpha, \beta \in \mathbb{N}_t^n$ . This also defines a matrix  $M \in \mathbb{R}^{[n]^t \times [n]^t}$  with  $M[I, J] = \widetilde{\mathbf{E}} \left[ x^{\alpha(I) + \alpha(J)} \right]$ . Note that this new matrix  $M$  satisfies  $M[I, J] = M[K, L]$  whenever  $\alpha(I) + \alpha(J) = \alpha(K) + \alpha(L)$ . Recall that a matrix in  $\mathbb{R}^{[n]^t \times [n]^t}$  with this symmetry is said to be SoS-symmetric.

We will use the following facts about the operator  $\widetilde{\mathbf{E}}$  given by the SoS hierarchy.

**Claim 4.1 (Pseudo-Cauchy-Schwarz [BKS14]).**  $\widetilde{\mathbf{E}} [p_1 p_2] \leq (\widetilde{\mathbf{E}} [p_1^2] \widetilde{\mathbf{E}} [p_2^2])^{1/2}$  for any  $p_1, p_2$  of degree at most  $q/2$ .

#### 4.1.1. Constrained Pseudoexpectations

For a system of polynomial constraints  $C = \{f_1 = 0, \dots, f_m = 0, g_1 \geq 0, \dots, g_r \geq 0\}$ , we say  $\widetilde{\mathbf{E}}_C$  is a pseudoexpectation operator respecting  $C$ , if in addition to the above conditions, it also satisfies

1.  $\widetilde{\mathbf{E}}_C [p \cdot f_i] = 0, \forall i \in [m]$  and  $\forall p$  such that  $\deg(p \cdot f_i) \leq q$ .
2.  $\widetilde{\mathbf{E}}_C [p^2 \cdot \prod_{i \in S} g_i] \geq 0, \forall S \subseteq [r]$  and  $\forall p$  such that  $\deg(p^2 \cdot \prod_{i \in S} g_i) \leq q$ .

It is well-known that such constrained pseudoexpectation operators can be described as solutions to semidefinite programs of size  $n^{O(q)}$  [BS14, Lau09]. This hierarchy of semidefinite programs for increasing  $q$  is known as the SoS hierarchy.

#### 4.2. Matrix Representations of Polynomials and relaxations of $\|f\|_2$

For a homogeneous polynomial  $f$  of even degree  $d$ , we say a matrix  $M \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  is a degree- $d$  matrix representation of  $f$  if for all  $x$ ,  $f(x) = (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2}$ . Recall that we consider the semidefinite program for optimizing the quantity  $\Lambda(f)$ , which is a relaxation for  $\|f\|_2$  when  $f \geq 0$ . Let  $M_f \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$  denote the unique SoS-symmetric matrix representation of  $f$ . Figure 4.1 gives the primal and dual forms of the relaxation computing  $\Lambda(f)$ . It is easy to check that strong duality holds in this case, since the solution  $\widetilde{\mathbf{E}}_C [x^\alpha] = (1/\sqrt{n})^{|\alpha|}$  for all  $\alpha \in \mathbb{N}_{\leq d}^n$  is strictly feasible and in the relative interior of the domain. Thus, the objective values of the two programs are equal.

We will also consider a weaker relaxation of  $\|f\|_2$ , which we denote by  $\|f\|_{sp}$ . A somewhat weaker version of this was used as the reference value in the work of [BKS14]. Figure 4.2 gives the primal and dual forms of this relaxation.

We will also need to consider constraint sets  $C = \{\|x\|_2^2 = 1, x^{\beta_1} \geq 0, \dots, x^{\beta_m} \geq 0\}$ . We refer to the non-negativity constraints here as moment non-negativity constraints. When considering the maximum of  $\widetilde{\mathbf{E}}_C [f]$ , for constraint sets  $C$  containing moments non-negativity constraints in addition to  $\|x\|_2^2 = 1$ , we refer to the optimum value as  $\Lambda_C(f)$ . Note that the maximum is still taken over degree- $d$  pseudoexpectations. Also, strong duality still holds in this case since  $\widetilde{\mathbf{E}}_C [x^\alpha] = (1/\sqrt{n})^{|\alpha|}$  is still a strictly feasible solution.

##### 4.2.1. Properties of relaxations obtained from constrained pseudoexpectations

We use the following claim, which is an easy consequence of the fact that the sum-of-squares algorithm can produce a certificate of optimality (see [OZ13]). In particular, if

<u>Primal</u> $\Lambda(f) := \inf \left\{ \sup_{\ z\ =1} z^T M z \mid M \in \mathbb{S}^{n^{d/2}}(\mathbb{R}), (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2} = f(x) \forall x \in \mathbb{R}^n \right\}$	
<u>Dual I</u> maximize $\langle M_f, X \rangle$ subject to : $\text{Tr}(X) = 1$ $X$ is SoS symmetric $X \succeq 0$	<u>Dual II</u> maximize $\tilde{\mathbf{E}}_C[f]$ subject to : $\tilde{\mathbf{E}}_C$ is a degree- $d$ pseudoexpectation $\tilde{\mathbf{E}}_C$ respects $C \equiv \{\ x\ _2^d = 1\}$

Figure 4.1: Primal and dual forms for the relaxation computing  $\Lambda(f)$

<u>Primal</u> $\ f\ _{sp} := \inf \left\{ \ M\ _2 \mid M \in \mathbb{S}^{n^{d/2}}(\mathbb{R}), (x^{\otimes d/2})^T \cdot M \cdot x^{\otimes d/2} = f(x) \forall x \in \mathbb{R}^n \right\}$	
<u>Dual</u> maximize $\langle M_f, X \rangle$ subject to : $\ X\ _{S_1} = 1$ $X$ is SoS symmetric	

Figure 4.2: Primal and dual forms for the relaxation computing  $\|f\|_{sp}$

$\max_{\tilde{\mathbf{E}}_C} \tilde{\mathbf{E}}_C[f] = \Lambda_C(f)$  for a degree- $q_1$  pseudoexpectation operator respecting  $C$  containing  $\|x\|_2^2 = 1$  and moment non-negativity constraints for  $\beta_1, \dots, \beta_m$ , then for every  $\lambda > \Lambda_C(f)$ , we have that  $\lambda - f$  can be certified to be positive by showing that  $\lambda - f \in \Sigma_C^{(q_1)}$ . Here  $\Sigma_C^{(q_1)}$  is the set of all expressions of the form

$$\lambda - f = \sum_j p_j \cdot (\|x\|_2^2 - 1) + \sum_{S \subseteq [m]} h_S(x) \cdot \prod_{i \in S} x^{\beta_i},$$

where each  $h_S$  is a sum of squares of polynomials and the degree of each term is at most  $q_1$ .

**Lemma 4.2.** *Let  $\Lambda_C(f)$  denote the maximum of  $\tilde{\mathbf{E}}_C[f]$  over all degree- $d$  pseudoexpectation operators respecting  $C$ . Then, for a pseudoexpectation operator of degree  $d'$  (respecting  $C$ ) and a polynomial  $p$  of degree at most  $(d' - d)/2$ , we have that*

$$\tilde{\mathbf{E}}_C[p^2 \cdot f] \leq \tilde{\mathbf{E}}_C[p^2] \cdot \Lambda_C(f).$$

*Proof.* As described above, for any  $\lambda > \Lambda_C(f)$ , we can write  $\lambda - f = g$  for  $g \in \Sigma_C^{(d)}$ . Since the degree of each term in  $p^2 \cdot g$  is at most  $d'$ , we have by the properties of pseudoexpectation operators (of degree  $d'$ ) that

$$\lambda \cdot \tilde{\mathbf{E}}_C[p^2] - \tilde{\mathbf{E}}_C[p^2 \cdot f] = \tilde{\mathbf{E}}_C[p^2 \cdot (\lambda - f)] = \tilde{\mathbf{E}}_C[p^2 \cdot g] \geq 0. \quad \blacksquare$$

The following monotonicity claim for non-negative coefficient polynomials will come in handy in later sections.

**Lemma 4.3.** *Let  $C$  be a system of polynomial constraints containing  $\{\forall \beta \in \mathbb{N}_t^n, x^\beta \geq 0\}$ . Then for any non-negative coefficient polynomials  $f$  and  $g$  of degree  $t$ , and such that  $f \geq g$  (coefficient-wise, i.e.  $f - g$  has non-negative coefficients), we have  $\Lambda_C(f) \geq \Lambda_C(g)$ .*

*Proof.* For any pseudo-expectation operator  $\tilde{\mathbf{E}}_C$  respecting  $C$ , we have  $\tilde{\mathbf{E}}_C[f - g] \geq 0$  because of the moment non-negativity constraints and by linearity.

So let  $\tilde{\mathbf{E}}_C$  be a pseudo-expectation operator realizing  $\Lambda_C(g)$ . Then we have,

$$\Lambda_C(f) \geq \tilde{\mathbf{E}}_C[f] = \tilde{\mathbf{E}}_C[g] + \tilde{\mathbf{E}}_C[f - g] = \Lambda_C(g) + \tilde{\mathbf{E}}_C[f - g] \geq 0. \quad \blacksquare$$

### 4.3. An additional operation on folded polynomials

We define the following operation (and it's folded counterpart) which in the case of a multilinear polynomial corresponds (up to scaling) to the sum of a row of the SOS symmetric matrix representation of the polynomial. This will be useful in our result for non-negative polynomials.

**Definition 4.4 (Collapse).** *Let  $f \in \mathbb{R}_d[x]$  be a polynomial. The  $k$ -collapse of  $f$ , denoted as  $C_k(f)$  is the degree  $d - k$  polynomial  $g$  given by,*

$$g(x) = \sum_{\gamma \in \mathbb{N}_{d-k}^n} g_\gamma \cdot x^\gamma \quad \text{where} \quad g_\gamma = \sum_{\alpha \in \mathbb{N}_k^n} f_{\gamma+\alpha}.$$

*For a degree- $(d_1, d_2)$  folded polynomial  $f$ , we define  $C_k(f)$  similarly as the degree- $(d_1 - k, d_2)$  folded polynomial  $g$  given by,*

$$g(x) = \sum_{\gamma \in \mathbb{N}_{d_1-k}^n} \bar{g}_\gamma(x) \cdot x^\gamma \quad \text{where} \quad \bar{g}_\gamma = \sum_{\alpha \in \mathbb{N}_k^n} \bar{f}_{\gamma+\alpha}.$$

## 5. Results for Polynomials in $\mathbb{R}_d[x]$ and $\mathbb{R}_d^+[x]$

### 5.1. Reduction to Multilinear Polynomials

**Lemma 5.1.** *Any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$  has a unique representation of the form*

$$\sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} F_{2\alpha}(x) \cdot x^{2\alpha}$$

*where for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ ,  $F_{2\alpha}$  is a homogeneous multilinear degree- $(d - 2|\alpha|)$  polynomial.*

We would like to approximate  $\|f\|_2$  by individually approximating  $\|F_{2\alpha}\|_2$  for each multilinear polynomial  $F_{2\alpha}$ . This section will establish the soundness of this goal.

### 5.1.1. Upper Bounding $\Lambda(f)$ in terms of $\Lambda(F_{2\alpha})$

We first bound  $\Lambda(f)$  in terms of  $\max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \Lambda(F_{2\alpha})$ . The basic intuition is that any matrix  $M_f$  such that  $(x^{\otimes(d/2)})^T \cdot M_f \cdot x^{\otimes(d/2)}$  for all  $x$  (called a matrix representation of  $f$ ) can be written as a sum of matrices  $M_{t,f}$  for each  $t \leq d/2$ , each of which is block-diagonal matrix with blocks corresponding to matrix representations of the polynomials  $M_{F_{2\alpha}}$  for each  $\alpha$  with  $|\alpha| = 2t$ .

**Lemma 5.2.** *Consider any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$ . We have,*

$$\Lambda(f) \leq \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{|\mathcal{O}(\alpha)|} (1 + d/2)$$

*Proof.* We shall start by constructing an appropriate matrix representation  $M_f$  of  $f$  that will give us the desired upper bound on  $\Lambda(f)$ . To this end, for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ , let  $M_{F_{2\alpha}}$  be the matrix representation of  $F_{2\alpha}$  realizing  $\Lambda(F_{2\alpha})$ . For any  $0 \leq t \leq d/2$ , we define  $M_{(t,f)}$  so that for any  $\alpha \in \mathbb{N}_t^n$  and  $I \in \mathcal{O}(\alpha)$ ,  $M_{(t,f)}[I, I] := M_{F_{2\alpha}} / |\mathcal{O}(\alpha)|$ , and  $M_{(t,f)}$  is zero everywhere else. Now let  $M_f := \sum_{t \in [d/2]} M_{(t,f)}$ . As for validity of  $M_f$  as a representation of  $f$  we have,

$$\begin{aligned} \langle M_f, x^{\otimes d/2} (x^{\otimes d/2})^T \rangle &= \sum_{0 \leq t \leq \frac{d}{2}} \langle M_{(t,f)}, x^{\otimes d/2} (x^{\otimes d/2})^T \rangle \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} \sum_{I \in \mathcal{O}(\alpha)} \langle M_{(|\alpha|,f)}[I, I], x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle x^{2\alpha} \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} \sum_{I \in \mathcal{O}(\alpha)} \frac{1}{|\mathcal{O}(\alpha)|} \langle M_{F_{2\alpha}}, x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle x^{2\alpha} \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} x^{2\alpha} \cdot \langle M_{F_{2\alpha}}, x^{\otimes(d/2-|\alpha|)} (x^{\otimes(d/2-|\alpha|)})^T \rangle \\ &= \sum_{\alpha \in \mathbb{N}_{\leq d/2}^n} F_{2\alpha}(x) x^{2\alpha} \\ &= f(x) \end{aligned}$$

Now observe that  $M_{(t,f)}$  is a block-diagonal matrix (up to simultaneous permutation of it's rows and columns). Thus we have  $\|M_{(t,f)}\| \leq \max_{\alpha \in \mathbb{N}_t^n} \|M_{F_{2\alpha}}\| / |\mathcal{O}(\alpha)|$ . Thus on applying triangle inequality, we obtain  $\|M_f\| \leq \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} (1 + d/2) \|M_{F_{2\alpha}}\| / |\mathcal{O}(\alpha)|$  ■

### 5.1.2. Lower Bounding $\|f\|_2$ in terms of $\|F_{2\alpha}\|_2$ (non-negative coefficients)

We first bound  $\|f\|_2$  in terms of  $\max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \|F_{2\alpha}\|_2$ , when every coefficient of  $f$  is non-negative. If  $x^*$  is the optimizer of  $F_{2\alpha}$ , then it is easy to see that  $x^* \geq 0$ . Setting  $y = x^* + \frac{\sqrt{\alpha}}{|\alpha|}$  ensures that  $\|y\|_2 \leq 2$  and  $f(y)$  is large, since  $f(y)$  recovers a significant fraction (up to a  $2^{O(d)} \cdot |\mathcal{O}(\alpha)|$  factor) of  $F_{2\alpha}(x^*)$ .

**Lemma 5.3.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial with non-negative coefficients. Consider any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ . Then*

$$\|f\|_2 \geq \frac{\|F_{2\alpha}\|_2}{2^{O(d)} |\mathcal{O}(\alpha)|}.$$



*Proof.* Consider any  $0 \leq t \leq d/2$ , and any  $\alpha \in \mathbb{N}_t^n$ . Let  $x_\alpha^* := \operatorname{argmax} \|F_{2\alpha}\|_2$  (note  $x_\alpha^*$  must be non-negative). Let

$$y^* := x_\alpha^* + \frac{\sqrt{\alpha}}{\sqrt{t}}$$

and let  $x^* := y^* / \|y^*\|$ . The second term is a unit vector since  $\|\sqrt{\alpha}\|_2^2 = t$ . Thus  $\|y^*\| = \Theta(1)$  since  $y^*$  is the sum of two unit vectors. This implies  $f(x^*) \geq f(y^*)/2^{O(d)}$ . Now we have,

$$\begin{aligned} f(y^*) &= \sum_{\beta \in \mathbb{N}_{\leq d/2}^n} F_{2\beta}(y^*) \cdot (y^*)^{2\beta} && \text{(by Lemma 5.1)} \\ &\geq F_{2\alpha}(y^*) \cdot (y^*)^{2\alpha} && \text{(by non-negativity of coefficients)} \\ &\geq F_{2\alpha}(y^*) \frac{1}{t^t} \prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell^{\alpha_\ell} && (y^* \geq \frac{\sqrt{\alpha}}{\sqrt{t}} \text{ entry-wise)} \\ &\geq F_{2\alpha}(y^*) \frac{1}{2^{O(t)} t!} \prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell^{\alpha_\ell} \\ &\geq F_{2\alpha}(y^*) \frac{\prod_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell!}{2^{O(t)} t!} \\ &\geq F_{2\alpha}(y^*) \frac{1}{2^{O(t)} |\mathcal{O}(\alpha)|} \\ &\geq F_{2\alpha}(x^*) \frac{1}{2^{O(t)} |\mathcal{O}(\alpha)|} && (y^* \text{ is entry-wise at least } x^*) \\ &= \frac{\|F_{2\alpha}\|_2}{2^{O(t)} |\mathcal{O}(\alpha)|}. \end{aligned}$$

This completes the proof. ■

**Theorem 5.4.** *Consider any homogeneous  $n$ -variate degree- $d$  polynomial  $f(x)$  with non-negative coefficients. Then*

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2}.$$

*Proof.* Combining Lemma 5.2 and Lemma 5.3 yields the claim. ■

We will next generalize Theorem 5.4 by proving a more general version of Lemma 5.3.

### 5.1.3. Lower Bounding $\|f\|_2$ in terms of $\|F_{2\alpha}\|_2$ (general case)

We lower bound  $\|f\|_2$  in terms of  $\|F_{2\alpha}\|_2$  for all polynomials. We will first recollect and establish some polynomial identities that will be used in the proof of the generalized version of Lemma 5.3 (i.e. Lemma 5.11).

#### Polynomial Identities

**Lemma 5.5 (Chebyshev's Extremal Polynomial Inequality).** *Let  $p(x)$  be a univariate degree- $d$  polynomial and let  $c_d$  be its leading coefficient. Then we have,  $\max_{x \in [0,1]} |p(x)| \geq 2|c_d|/4^d$ .*

**Lemma 5.6 ([HLZ10]).** Let  $x^1, x^2, \dots, x^d \in \mathbb{R}^n$  be arbitrary, let  $\mathcal{A} \in \mathbb{R}^{[n]^d}$  be a SoS-symmetric  $d$ -tensor, and let  $\xi_1, \dots, \xi_d$  be independent Rademacher random variables. Then

$$\mathbb{E} \left[ \prod_{i \in [d]} \xi_i \langle \mathcal{A}, (\xi_1 x^1 + \dots + \xi_d x^d)^{\otimes d} \rangle \right] = d! \langle \mathcal{A}, x^1 \otimes \dots \otimes x^d \rangle.$$

This lemma implies:

**Lemma 5.7 ([HLZ10]).** Let  $\mathcal{A}$  be a SoS-symmetric  $d$ -tensor and let  $f(x) := \langle \mathcal{A}, x^{\otimes d} \rangle$ . Then

$$\|f\|_2 \geq \frac{1}{2^{O(d)}} \max_{\|x^i\|=1} \langle \mathcal{A}, x^1 \otimes \dots \otimes x^d \rangle.$$

**Lemma 5.8.** Let  $f$  be an  $n$ -variate degree- $d$  homogeneous polynomial. Let  $\|f\|_2^\zeta := \max_{\substack{z \in \mathbb{C}^n \\ \|z\|=1}} |f(z)|$ ,

then

$$\frac{\|f\|_2^\zeta}{2^{O(d)}} \leq \|f\|_2 \leq \|f\|_2^\zeta.$$

*Proof.* Let  $\mathcal{A}$  be the SoS-symmetric tensor representing  $f$ . Let  $z^* = a^* + ib^*$  be the complex unit vector realizing  $f(z^*) = \|f\|_2^\zeta$ . Then we have,

$$\begin{aligned} f(z^*) &= \langle \mathcal{A}, (z^*)^{\otimes d} \rangle \\ &= \langle \mathcal{A}, (a^* + ib^*)^{\otimes d} \rangle \\ &= \sum_{c^1, \dots, c^d \in \{a^*, ib^*\}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle \\ \Rightarrow \operatorname{Re}(f(z^*)) &= \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\}, \\ |\{j | c^j = b^*\}| \bmod 4 = 0}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle - \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\}, \\ |\{j | c^j = b^*\}| \bmod 4 = 2}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle, \\ \operatorname{Im}(f(z^*)) &= \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\}, \\ |\{j | c^j = b^*\}| \bmod 4 = 1}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle - \sum_{\substack{c^1, \dots, c^d \in \{a^*, b^*\}, \\ |\{j | c^j = b^*\}| \bmod 4 = 3}} \langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle \end{aligned}$$

which implies that there exists  $c^1, \dots, c^d \in \{a^*, b^*\}$  such that  $|\langle \mathcal{A}, \bigotimes_{j \in [d]} c^j \rangle| \geq \|f\|_2^\zeta / 2^{O(d)}$ . Lastly, applying Lemma 5.7 implies the claim.  $\blacksquare$

### Some Probability Facts

**Lemma 5.9.** Let  $X_1, \dots, X_k$  be i.i.d. Bernoulli( $p$ ) random variables. Then for any  $t_1, \dots, t_k \in \mathbb{N}$ ,  $\mathbb{E} \left[ X_1^{t_1} \dots X_k^{t_k} \right] = p^k$ .

**Lemma 5.10.** Let  $\zeta$  be a uniformly random  $p$ -th root of unity. Then for any  $t \in [p-1]$ ,  $\mathbb{E} [\zeta^t] = 0$ . Also, clearly  $\mathbb{E} [\zeta^p] = 1$ .

We finally lower bound  $\|f\|_2$  in terms of  $F_{2\alpha}$ . Fix  $\alpha \in \mathbb{N}_{\leq d/2}^n$  and, let  $x^*$  be the optimizer of  $F_{2\alpha}$ . Setting  $y = x^* + \frac{\sqrt{\alpha}}{|\alpha|}$  as in the non-negative coefficient case does not work since terms from  $F_{2\beta}$  may be negative. We bypass this issue by first lower bounding  $\|f\|_2^\zeta$  in terms of  $F_{2\alpha}$  and using Lemma 5.8. For  $\|f\|_2^\zeta$ , we use random roots of unity and Bernoulli random variables, together with Lemma 5.5, to extract nonzero contribution only from the monomials that are multiples of  $x^\alpha$  times multilinear parts.

**Lemma 5.11.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial. Then for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ ,*

$$\|f\|_2 \geq \frac{\|F_{2\alpha}\|_2}{2^{O(d)} |\mathcal{O}(\alpha)|}.$$

*Proof.* Fix any  $\alpha \in \mathbb{N}_{\leq d/2}^n$ , let  $t := |\alpha|$  and let  $k := d - 2t$ . For any  $i \in [n]$ , let  $\zeta_i$  be an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and let  $\Xi$  be an independent and uniformly randomly chosen  $(k + 1)$ -th root of unity.

Let  $\bar{x} := \operatorname{argmax}\|F_{2\alpha}\|_2$ . Let  $p \in [0, 1]$  be a parameter to be fixed later, let  $b_1, \dots, b_n$  be i.i.d. Bernoulli( $p$ ) random variables, let  $\zeta := (\zeta_1, \dots, \zeta_n)$ ,  $b := (b_1, \dots, b_n)$  and finally let

$$z := \Xi \cdot b \circ \frac{1}{2\alpha + \mathbf{1}} \circ \bar{x} + \frac{\zeta \circ \sqrt{\alpha}}{\sqrt{t}}.$$

Since  $\sum_{\ell \in \mathcal{S}(\alpha)} \alpha_\ell = t$  and roots of unity have magnitude one,  $z$  has length  $O(1)$ . Now consider any fixed  $\gamma \in \{0, 1\}_k^n$ . We have,

$$\begin{aligned} & \mathbb{E} \left[ z^{2\alpha + \gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \\ &= \text{coefficient of } \Xi^k \cdot \prod_{i \in [n]} \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} [z^{2\alpha + \gamma}] \quad (\text{by Lemma 5.10}) \\ &= \text{coefficient of } \Xi^k \cdot \prod_{i \in [n]} \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} \left[ \prod_{i \in [n]} \left( \zeta_i \cdot \frac{\sqrt{\alpha_i}}{\sqrt{t}} + \Xi \cdot \frac{b_i \cdot \bar{x}_i}{2\alpha_i + 1} \right)^{2\alpha_i + \gamma_i} \right] \\ &= \prod_{i \in [n]} \text{coefficient of } \Xi^{\gamma_i} \cdot \zeta_i^{2\alpha_i} \text{ in } \mathbb{E} \left[ \left( \zeta_i \cdot \frac{\sqrt{\alpha_i}}{\sqrt{t}} + \Xi \cdot \frac{b_i \cdot \bar{x}_i}{2\alpha_i + 1} \right)^{2\alpha_i + \gamma_i} \right] \quad (\text{since } \gamma \in \{0, 1\}_k^n) \\ &= p^k \cdot \prod_{i \in \mathcal{S}(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot \bar{x}_i^{\gamma_i} \quad (\text{by Lemma 5.9}) \\ &= p^k \cdot \bar{x}^\gamma \cdot \prod_{i \in \mathcal{S}(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \end{aligned}$$

Thus we have,

$$\begin{aligned} & \mathbb{E} \left[ f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \\ &= \sum_{\beta \in \mathbb{N}_d^n} f_\beta \cdot \mathbb{E} \left[ z^\beta \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \\ &= \sum_{\substack{\beta \in \mathbb{N}_d^n \\ \beta \geq 2\alpha}} f_\beta \cdot \mathbb{E} \left[ z^\beta \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \quad (\text{by Lemma 5.10}) \\ &= \sum_{\gamma \in \{0, 1\}_k^n} f_{2\alpha + \gamma} \cdot \mathbb{E} \left[ z^{2\alpha + \gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] + \sum_{\substack{\gamma \in \mathbb{N}_k^n \\ \gamma \not\leq \mathbf{1}}} f_{2\alpha + \gamma} \cdot \mathbb{E} \left[ z^{2\alpha + \gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \end{aligned}$$

$$= \sum_{\gamma \in \{0,1\}_k^n} f_{2\alpha+\gamma} \cdot \mathbb{E} \left[ z^{2\alpha+\gamma} \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] + r(p) \quad (\text{by Lemma 5.9})$$

where  $r(p)$  is some univariate polynomial in  $p$ , s.t.  $\deg(r) < k$

$$\begin{aligned} &= \sum_{\gamma \in \{0,1\}_k^n} f_{2\alpha+\gamma} \cdot p^k \cdot \bar{x}^\gamma \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} + r(p) \\ &= p^k \cdot F_{2\alpha}(\bar{x}) \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} + r(p) \quad (\text{where } \deg(r) < k) \end{aligned}$$

Lastly we have,

$$\begin{aligned} \|f\|_2 &\geq \|f\|_2^c \cdot 2^{-O(d)} && \text{by Lemma 5.8} \\ &\geq \max_{p \in [0,1]} \mathbb{E} [|f(z)|] \cdot 2^{-O(d)} && (\|z\| = O(1)) \\ &= \max_{p \in [0,1]} \mathbb{E} \left[ \left| f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right| \right] \cdot 2^{-O(d)} \\ &\geq \max_{p \in [0,1]} \left| \mathbb{E} \left[ f(z) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] \right| \cdot 2^{-O(d)} \\ &\geq |F_{2\alpha}(\bar{x})| \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot 2^{-O(d)} && (\text{by Chebyshev: Lemma 5.5}) \\ &= \|F_{2\alpha}\|_2 \cdot \prod_{i \in S(\alpha)} \frac{\alpha_i^{\alpha_i}}{t^{\alpha_i}} \cdot 2^{-O(d)} \\ &\geq \frac{\|F_{2\alpha}\|_2}{|\Theta(\alpha)|} \cdot 2^{-O(d)} \end{aligned}$$

This completes the proof. ■

In fact, the proof of Lemma 5.11 yields a more general result:

**Lemma 5.12 (Weak Decoupling).** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial. Then for any  $\alpha \in \mathbb{N}_{\leq d/2}^n$  and any unit vector  $y$ ,*

$$\|f\|_2 \geq y^{2\alpha} \cdot \|F_{2\alpha}\|_2 \cdot 2^{-O(d)}.$$

We are finally able to establish the multilinear reduction result that is the focus of this section.

**Theorem 5.13.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  (for even  $d$ ) polynomial. Then*

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \max_{\alpha \in \mathbb{N}_{\leq d/2}^n} \frac{\Lambda(F_{2\alpha})}{\|F_{2\alpha}\|_2}.$$

*Proof.* Combining Lemma 5.2 and Lemma 5.11 yields the claim. ■

## 5.2. $(n/q)^{d/4}$ -Approximation for Non-negative Coefficient Polynomials

**Theorem 5.14.** Consider any homogeneous multilinear  $n$ -variate degree- $d$  polynomial  $f(x)$  with non-negative coefficients. We have,

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4}}{d^{d/4}}.$$

*Proof.* Let  $M_f$  be the SoS-symmetric matrix representation of  $f$ . Let  $I^* = (i_1, \dots, i_{d/2}) \in [n]^{d/2}$  be the multi-index of any row of  $M_f$  with maximum row sum. Let  $S_I$  for  $I \in [n]^{d/2}$ , denote the sum of the row  $I$  of  $M_f$ . By Perron-Frobenius theorem,  $\|M_f\| \leq S_{I^*}$ . Thus  $\Lambda(f) \leq S_{I^*}$ .

We next proceed to bound  $\|f\|_2$  from below. To this end, let  $x^* := y^* / \|y^*\|$  where,

$$y^* := \frac{\mathbb{1}}{\sqrt{n}} + \frac{1}{\sqrt{d/2}} \sum_{i \in I^*} e_i$$

Since  $f$  is multilinear,  $I^*$  has all distinct elements, and so the second term in the definition of  $y^*$  is of unit length. Thus  $\|y^*\| = \Theta(1)$ , which implies that  $\|f\|_2 \geq f(x^*) \geq f(y^*)/2^{O(d)}$ . Now we have,

$$\begin{aligned} f(y^*) &= ((y^*)^{\otimes d/2})^T M_f (y^*)^{\otimes d/2} \\ &\geq \sum_{I \in \mathcal{O}(I^*)} \frac{1}{(nd)^{d/4}} e_{I(1)}^T \otimes \dots \otimes e_{I(d/2)}^T M_f \mathbb{1}^{\otimes d/2} \quad (\text{by non-negativity of } M_f) \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{1}{(nd)^{d/4}} e_I^T M_f \mathbb{1} \quad (\in \mathbb{R}^{[n]^{d/2}}) \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{S_I}{(nd)^{d/4}} \\ &= \sum_{I \in \mathcal{O}(I^*)} \frac{S_{I^*}}{(nd)^{d/4}} \quad (\text{by SoS-symmetry of } M_f) \\ &= \frac{(d/2)! S_{I^*}}{(nd)^{d/4}} \quad (|\mathcal{O}(I^*)| = (d/2)! \text{ by multilinearity of } f) \\ &\geq \frac{d^{d/4} S_{I^*}}{n^{d/4} 2^{O(d)}} \geq \frac{d^{d/4} \Lambda(f)}{n^{d/4} 2^{O(d)}}. \end{aligned}$$

This completes the proof. ■

**Theorem 5.15.** Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial with non-negative coefficients. Then for any even  $q$  such that  $d$  divides  $q$ ,

$$\frac{(\Lambda(f^{q/d}))^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4}}{q^{d/4}}.$$

*Proof.* Applying Theorem 5.4 to  $f^{q/d}$  and combining this with Theorem 5.14 yields the claim. ■

### 5.3. $(n/q)^{d/2}$ -Approximation for General Polynomials

**Theorem 5.16.** Consider any homogeneous multilinear  $n$ -variate degree- $d$  (for even  $d$ ) polynomial  $f(x)$ . We have,

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2}}{d^{d/2}}.$$

*Proof.* Let  $M_f$  be the SoS-symmetric matrix representation of  $f$ , i.e.

$$M_f[I, J] = \frac{f_{\alpha(I)+\alpha(J)}}{|\mathcal{O}(\alpha(I) + \alpha(J))|}.$$

By the Gershgorin circle theorem, we can bound  $\|M_f\|_2$ , and hence  $\Lambda(f)$  by  $n^{d/2} \cdot (\max_{\beta} |f_{\beta}| / d!)$ . Here, we use the multilinearity of  $f$ . On the other hand for a multilinear polynomial, using  $x = \beta / \sqrt{|\beta|}$  (where  $|\beta| = d$  by multilinearity), gives  $\|f\|_2 \geq d^{-d/2} \cdot |f_{\beta}|$ . Thus, we easily get

$$\Lambda(f) \leq \frac{d^{d/2}}{d!} \cdot n^{d/2} \cdot \|f\|_2 = 2^{O(d)} \frac{n^{d/2}}{d^{d/2}}. \quad \blacksquare$$

**Theorem 5.17.** Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial, and assume that  $2d$  divides  $q$ . Then

$$\frac{(\Lambda(f^{q/d}))^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2}}{q^{d/2}}.$$

*Proof.* Applying Theorem 5.13 to  $f^{q/d}$  and combining this with Theorem 5.16 yields the claim. ■

### 5.4. $\sqrt{m/q}$ -Approximation for $m$ -sparse polynomials

**Lemma 5.18.** Consider any homogeneous multilinear  $n$ -variate degree- $d$  (for even  $d$ ) polynomial  $f(x)$  with  $m$  non-zero coefficients. We have,

$$\frac{\Lambda(f)}{\|f\|_2} \leq 2^{O(d)} \sqrt{m}.$$

*Proof.* Let  $M_f$  be the SoS-symmetric matrix representation of  $f$ , i.e.

$$M_f[I, J] = \frac{f_{\alpha(I)+\alpha(J)}}{|\mathcal{O}(\alpha(I) + \alpha(J))|}.$$

Now  $\Lambda(f) \leq \|M_f\| \leq \|M_f\|_F$ . Thus we have,

$$\begin{aligned} \|M_f\|_F^2 &= \sum_{I, J \in [n]^{d/2}} M_f[I, J]^2 \\ &= \sum_{\beta \in \{0,1\}_d^n} \frac{f_{\beta}^2}{|\mathcal{O}(\beta)|} \\ &= \sum_{\beta \in \{0,1\}_d^n} \frac{f_{\beta}^2}{d!} \end{aligned}$$

$$\leq \frac{m}{d!} \cdot \max_{\beta} |f_{\beta}|$$

On the other hand, since  $f$  is multilinear, using  $x = \beta / \sqrt{|\beta|}$  (where  $|\beta| = d$  by multilinearity), implies  $\|f\|_2 \geq d^{-d/2} \cdot |f_{\beta}|$  for any  $\beta$ . This implies the claim.  $\blacksquare$

**Theorem 5.19.** *Let  $f(x)$  be a homogeneous  $n$ -variate degree- $d$  polynomial with  $m$  non-zero coefficients, and assume that  $2d$  divides  $q$ . Then*

$$\frac{(\Lambda(f^{q/d}))^{d/q}}{\|f\|_2} \leq 2^{O(d)} \sqrt{m/q}.$$

*Proof.* Combining Theorem 5.13 and Lemma 5.18, yields that for any degree- $q$  homogeneous polynomial  $g$  with sparsity  $\bar{m}$ , we have

$$\frac{(\Lambda(g))}{\|g\|_2} \leq 2^{O(q)} \sqrt{\bar{m}}.$$

Lastly, taking  $g = f^{q/d}$  and observing that the sparsity of  $g$  is at most  $\binom{m}{q/d}$  implies the claim.  $\blacksquare$

## 6. Approximating 2-norms via Folding

### 6.1. Preliminaries

Recall that we call a folded polynomial multilinear if all its monomials are multilinear. In particular, there's no restriction on the folds of the polynomial.

**Lemma 6.1 (Folded Analogue of Lemma 5.1).**

Let  $(\mathbb{R}_{d_2}[x])_{d_1}[x] \ni f(x) := \sum_{\beta \in \mathbb{N}_{d_1}^n} \bar{f}_{\beta}(x) \cdot x^{\beta}$  be a  $(d_1, d_2)$ -folded polynomial.  $f$  can be written as

$$\sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} F_{2\alpha}(x) \cdot x^{2\alpha}$$

where for any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$ ,  $F_{2\alpha}(x)$  is a multilinear  $(d_1 - 2|\alpha|, d_2)$ -folded polynomial.

*Proof.* Simply consider the folded polynomial

$$F_{2\alpha}(x) = \sum_{\gamma \in \{0,1\}_{d_1-2|\alpha|}^n} \overline{(F_{2\alpha})}_{\gamma} \cdot x^{\gamma}$$

where  $\overline{(F_{2\alpha})}_{\gamma} = \bar{f}_{2\alpha+\gamma}$ .  $\blacksquare$

### 6.2. Reduction to Multilinear Folded Polynomials

Here we will prove a generalized version of Lemma 5.2, which is a generalization in two ways; firstly it allows for folds instead of just coefficients, and secondly it allows a more general set of constraints than just the hypersphere since we will need to add some additional non-negativity constraints for the case of non-negative coefficient polynomials (so that  $\Lambda_C(\cdot)$  satisfies monotonicity over NNC polynomials which will come in handy later).

Recall that  $\Lambda_C(\cdot)$  is defined in Section 4.2 and that  $\|f\|_2$  and  $\Lambda_C(f)$  for a folded polynomial  $f$ , are applied to the unfolding of  $f$ .

### 6.2.1. Relating $\Lambda_C(f)$ to $\Lambda_C(F_{2\alpha})$

#### Lemma 6.2 (Folded Analogue of Lemma 5.2).

Let  $C$  be a system of polynomial constraints of the form  $\{\|x\|_2^2 = 1\} \cup C'$  where  $C'$  is a moment non-negativity constraint set. Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. We have,

$$\Lambda_C(f) \leq \max_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\alpha})}{|\mathcal{O}(\alpha)|} (1 + d_1/2)$$

*Proof.* Consider any degree- $(d_1 + d_2)$  pseudo-expectation operator  $\tilde{\mathbf{E}}_C$ . We have,

$$\begin{aligned} \tilde{\mathbf{E}}_C[f] &= \sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \tilde{\mathbf{E}}_C[F_{2\alpha}(x) \cdot x^{2\alpha}] && \text{(by Lemma 6.1)} \\ &\leq \sum_{\alpha \in \mathbb{N}_{\leq d_1/2}^n} \tilde{\mathbf{E}}_C[x^{2\alpha}] \cdot \Lambda_C(F_{2\alpha}) && \text{(by Lemma 4.2)} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[x^{2\alpha}] \cdot \Lambda_C(F_{2\alpha}) \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[|\mathcal{O}(\alpha)|x^{2\alpha}] \cdot \frac{\Lambda_C(F_{2\alpha})}{|\mathcal{O}(\alpha)|} \\ &\leq \sum_{0 \leq t \leq \frac{d_1}{2}} \sum_{\alpha \in \mathbb{N}_t^n} \tilde{\mathbf{E}}_C[|\mathcal{O}(\alpha)|x^{2\alpha}] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} && (\tilde{\mathbf{E}}_C[x^{2\alpha}] \geq 0) \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \tilde{\mathbf{E}}_C \left[ \sum_{\alpha \in \mathbb{N}_t^n} |\mathcal{O}(\alpha)|x^{2\alpha} \right] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \tilde{\mathbf{E}}_C[\|x\|_2^{2t}] \cdot \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \sum_{0 \leq t \leq \frac{d_1}{2}} \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} \\ &= \max_{\beta \in \mathbb{N}_{\leq d_1/2}^n} \frac{\Lambda_C(F_{2\beta})}{|\mathcal{O}(\beta)|} (1 + d_1/2) \quad \blacksquare \end{aligned}$$

### 6.3. Relating Evaluations of $f$ to Evaluations of $F_{2\alpha}$

Here we would like to generalize Lemma 5.3 and Lemma 5.11 to allow folds, however for technical reasons related to decoupling of the domain of the folds from the domain of the monomials of a folded polynomial, we instead generalize claims implicit in the proofs of Lemma 5.3 and Lemma 5.11.

Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. Recall that an evaluation of a folded polynomial treats the folds as coefficients and only substitutes values in the monomials of the folded polynomial. Thus for any fixed  $y \in \mathbb{R}^n$ ,  $f(y)$  (sometimes denoted by  $(f(y))(x)$  for contextual clarity) is a degree- $d_2$  polynomial in  $x$ , i.e.  $f(y) \in \mathbb{R}_{d_2}[x]$ .



**Lemma 6.3 (Folded Analogue of Lemma 5.3).**

Let  $f \in (\mathbb{R}_{d_2}^+[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial whose folds have non-negative coefficients. Then for any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$  and any  $y \geq 0$ ,

$$\left( f \left( y + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} \right) \right) (x) \geq \frac{(F_{2\alpha}(y))(x)}{|\mathcal{O}(\alpha)|} \cdot 2^{-O(d_1)}$$

where the ordering is coefficient-wise.

*Proof.* Identical to the proof of Lemma 5.3. ■

**Lemma 6.4 (Folded Analogue of Lemma 5.11).**

Let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a  $(d_1, d_2)$ -folded polynomial. Consider any  $\alpha \in \mathbb{N}_{\leq d_1/2}^n$  and any  $y$ , and let

$$z := \Xi \cdot y \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{|\alpha|}}$$

where  $\Xi$  is an independent and uniformly randomly chosen  $(d_1 - 2|\alpha| + 1)$ -th root of unity, and for any  $i \in [n]$ ,  $\zeta_i$  is an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and  $b_i$  is an independent Bernoulli( $p$ ) random variable ( $p$  is an arbitrary parameter in  $[0, 1]$ ). Then

$$\mathbb{E} \left[ (f(z))(x) \cdot \Xi \cdot \prod_{i \in [n]} \zeta_i \right] = p^{d_1 - 2|\alpha|} \cdot \frac{(F_{2\alpha}(y))(x)}{|\mathcal{O}(\alpha)|} \cdot 2^{-O(d_1)} + r(p)$$

where  $r(p)$  is a univariate polynomial in  $p$  with degree less than  $d_1 - 2|\alpha|$  (and whose coefficients are in  $\mathbb{R}_{d_2}[x]$ ).

*Proof.* This follows by going through the proof of Lemma 5.11 for every fixed  $x$ . ■

## 6.4. Bounding $\Lambda_C()$ of Multilinear Folded Polynomials

Here we bound  $\Lambda_C()$  of a multilinear folded polynomial in terms of properties of the polynomial that are inspired by treating the folds as coefficients and generalizing the coefficient-based approximations for regular (non-folded) polynomials from Theorem 5.16 and Theorem 5.14.

### 6.4.1. General Folds: Bounding $\Lambda()$ in terms of $\Lambda()$ of the "worst" fold

Here we will give a folded analogue of the proof of Theorem 5.16 wherein we used Gershgorin-Circle theorem to bound SOS value in terms of the max-magnitude-coefficient.

**Lemma 6.5 (Folded Analogue of Gershgorin Circle Bound on Spectral Radius).** For even  $d_1, d_2$ , let  $d = d_1 + d_2$ , let  $f \in (\mathbb{R}_{d_2}[x])_{d_1}[x]$  be a multilinear  $(d_1, d_2)$ -folded polynomial. We have,

$$\Lambda(f) \leq 2^{O(d)} \frac{n^{d_1/2}}{d_1^{d_1}} \max_{\gamma \in \{0,1\}_{d_1}^n} \|\bar{f}_\gamma\|_{sp}.$$

*Proof.* Since  $\Lambda(f) \leq \|f\|_{sp}$ , it is sufficient to bound  $\|f\|_{sp}$ .

Let  $M_{\bar{f}_\gamma}$  be the matrix representation of  $\bar{f}_\gamma$  realizing  $\|\bar{f}_\gamma\|_{sp}$ . Let  $M_f$  be an  $[n]^{d_1/2} \times [n]^{d_1/2}$  block matrix with  $[n]^{d_2/2} \times [n]^{d_2/2}$  size blocks, where for any  $I, J \in [n]^{d_1/2}$  the block of  $M_f$  at index  $(I, J)$  is defined to be  $\frac{1}{d_1!} \cdot M_{\bar{f}_{\alpha(I)+\alpha(J)}}$ . Clearly  $M_f$  (interpreted as an  $[n]^{d/2} \times [n]^{d/2}$ ) is a matrix representation of the unfolding of  $f$  since  $f$  is a multilinear folded polynomial. Lastly, applying Block-Gershgorin circle theorem to  $M_f$  and upper bounding the sum of spectral norms over a block row by  $n^{d_1/2}$  times the max term implies the claim.  $\blacksquare$

#### 6.4.2. Non-Negative Coefficient Folds: Relating SoS Value to the SoS Value of the $d_1/2$ -collapse

Observe that in the case of a multilinear degree- $d$  polynomial, the  $d/2$ -collapse corresponds (up to scaling) to the sum of a row of the SOS symmetric matrix representation of the polynomial. We will next develop a folded analogue of the proof of Theorem 5.14 wherein we employed Perron-Frobenius theorem to bound SOS value in terms of the  $d/2$ -collapse.

The proof here however, is quite a bit more subtle than in the general case above. This is because one can apply the block-matrix analogue of Gershgorin theorem (due to Feingold et al. [FV<sup>+</sup>62]) to a matrix representation of the folded polynomial (whose spectral norm is an upper bound on  $\Lambda(\cdot)$ ) in the general case. Loosely speaking, this corresponds to bounding  $\Lambda(f)$  in terms of

$$\max_{\gamma \in \{0,1\}_k^n} \sum_{\theta \in \{0,1\}_k^n} \Lambda(\bar{f}_{\gamma+\theta})$$

where  $k = d_1/2$ . This however is not enough in the nnc case as in order to win the  $1/2$  in the exponent, one needs to relate  $\Lambda_C(f)$  to

$$\max_{\gamma \in \{0,1\}_k^n} \Lambda \left( \sum_{\theta \in \{0,1\}_k^n} \bar{f}_{\gamma+\theta} \right).$$

This however, cannot go through Block-Gershgorin since it is **not** true that the spectral norm of a non-negative block matrix is upper bounded by the max over rows of the spectral norm of the sum of blocks in that row. It instead, can only be upper bounded by the max over rows of the sum of spectral norms of the blocks in that row.

To get around this issue, we skip the intermediate step of bounding  $\Lambda_C(f)$  by the spectral norm of a matrix and instead prove the desired relation directly through the use of pseudoexpectation operators. This involved first finding a pseudo-expectation based proof of Gershgorin/Perron-Frobenius bound on spectral radius that generalizes to folded polynomials in the right way.

**Lemma 6.6 (Folded analogue of Perron-Frobenius Bound on Spectral Radius).** *For even  $d_1 = 2k$ , let  $f \in (\mathbb{R}_{d_2}^+[x])_{d_1}[x]$  be a multilinear  $(d_1, d_2)$ -folded polynomial whose folds have non-negative coefficients. Let  $C$  be the system of polynomial constraints given by  $\{\|x\|_2^2 = 1; \forall \beta \in \mathbb{N}_{d_2}^n, x^\beta \geq 0\}$ . We have,*

$$\Lambda_C(f) \leq \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!}$$

where

$$\bar{g}_\gamma(x) := \overline{C_k(f)_\gamma} = \sum_{\substack{\theta \leq \mathbf{1} - \gamma \\ \theta \in \mathbb{N}_k^n}} \bar{f}_{\gamma + \theta}(x).$$

*Proof.* Consider any pseudo-expectation operator  $\tilde{\mathbf{E}}_C$  of degree at least  $d_1 + d_2$ . Note that since  $\tilde{\mathbf{E}}_C$  satisfies  $\{\forall \beta \in \mathbb{N}_{d_2}^n, x^\beta \geq 0\}$ , by linearity  $\tilde{\mathbf{E}}_C$  must also satisfy  $\{h \geq 0\}$  for any  $h \in \mathbb{R}_{d_2}^+[x]$  - a fact we will use shortly.

Since  $f$  is a multilinear folded polynomial,  $\bar{f}_\alpha$  is only defined when  $0 \leq \alpha \leq \mathbf{1}$ . If  $\alpha \not\leq \mathbf{1}$ , we define  $\bar{f}_\alpha := 0$ . We have,

$$\begin{aligned} \tilde{\mathbf{E}}_C[f] &= \sum_{\alpha \in \{0,1\}_{d_1}^n} \tilde{\mathbf{E}}_C[\bar{f}_\alpha \cdot x^\alpha] && (f \text{ is a multilinear folded polynomial}) \\ &= \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C[\bar{f}_{\alpha(I) + \alpha(J)} \cdot x^I x^J] \cdot \frac{1}{d_1!} && (\text{by multilinearity}) \\ &\leq \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C\left[\bar{f}_{\alpha(I) + \alpha(J)} \cdot \frac{(x^I)^2 + (x^J)^2}{2}\right] \cdot \frac{1}{d_1!} && (\tilde{\mathbf{E}}_C \text{ satisfies } \bar{f}_\alpha \geq 0) \\ &= \sum_{I \in [n]^k} \sum_{J \in [n]^k} \tilde{\mathbf{E}}_C[\bar{f}_{\alpha(I) + \alpha(J)} \cdot (x^I)^2] \cdot \frac{1}{d_1!} \\ &= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2 \cdot \sum_{J \in [n]^k} \bar{f}_{\alpha(I) + \alpha(J)}\right] \cdot \frac{1}{d_1!} \\ &= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2 \cdot \sum_{\substack{\theta \leq \mathbf{1} - \alpha(I) \\ \theta \in \mathbb{N}_k^n}} \bar{f}_{\alpha(I) + \theta}\right] \cdot \frac{k!}{d_1!} && (\text{by multilinearity}) \\ &= \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2 \cdot g_{\alpha(I)}\right] \cdot \frac{k!}{d_1!} \\ &\leq \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2\right] \cdot \Lambda_C(\bar{g}_{\alpha(I)}) \cdot \frac{1}{k!} && (\text{by Lemma 4.2}) \\ &\leq \sum_{I \in [n]^k} \tilde{\mathbf{E}}_C\left[(x^I)^2\right] \cdot \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!} && (\tilde{\mathbf{E}}_C\left[(x^I)^2\right] \geq 0) \\ &= \tilde{\mathbf{E}}_C\left[\|x\|_2^{d_1}\right] \cdot \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!} \\ &= \max_{\gamma \in \{0,1\}_k^n} \Lambda_C(\bar{g}_\gamma) \cdot \frac{1}{k!} \end{aligned}$$

We are finally equipped to prove the main results of this section. ■

## 6.5. $(n/q)^{d/4-1/2}$ -Approximation for Non-negative Coefficient Polynomials

**Theorem 6.7.** Consider any  $f \in \mathbb{R}_d^+[x]$  for  $d \geq 2$ , and any  $q$  divisible by  $2d$ . Let  $C$  be the system of polynomial constraints given by  $\{\|x\|_2^2 = 1; \forall \beta \in \mathbb{N}_{2q/d}^n, x^\beta \geq 0\}$ . Then we have,

$$\frac{\Lambda_C(f^{q/d})^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/4-1/2}}{q^{d/4-1/2}}.$$

*Proof.* Let  $h$  be any  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and whose folds have non-negative coefficients and let  $s$  be the  $(\bar{q}, 2q/d)$ -folded polynomial given by  $h^{q/d}$  where  $\bar{q} := (d-2)q/d$ . Finally, consider any  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  and let  $S_{2\alpha}$  be the multilinear component of  $s$  as defined in Lemma 6.1. We will establish that for any  $\gamma \in \{0, 1\}_k^n$  (where  $k := \bar{q}/2 - |\alpha|$ ),

$$\|f\|_2^{q/d} \geq \frac{2^{-O(q)} \cdot \Lambda_C(\overline{C_{\bar{q}/2-|\alpha|}(S_{2\alpha})}_\gamma)}{(\bar{q}/2 - |\alpha|)^{\bar{q}/4-|\alpha|/2} \cdot |\Theta(\alpha)| \cdot n^{\bar{q}/4-|\alpha|/2}} \quad (6.1)$$

which on combining with the application of Lemma 6.2 to  $s$  and its composition with Lemma 6.6, yields the claim. To elaborate, we apply Lemma 6.2 to  $s$  with  $d_1 = \bar{q}, d_2 = 2q/d$  and then for every  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  we apply Lemma 6.6 with  $d_1 = \bar{q} - 2|\alpha|, d_2 = 2q/d$ , to get

$$\Lambda_C(f^{q/d}) = \Lambda_C(s) \leq 2^{O(q)} \cdot \max_{\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n} \max_{\gamma \in \{0, 1\}_{\bar{q}/2-|\alpha|}^n} \frac{\Lambda_C(\overline{C_{\bar{q}/2-|\alpha|}(S_{2\alpha})}_\gamma)}{(\bar{q}/2 - |\alpha|)! \cdot |\Theta(\alpha)|}$$

which on combining with Eq. (6.1) yields the claim.

It remains to establish Eq. (6.1). So fix any  $\alpha, \gamma$  satisfying the above conditions. Let  $t := |\alpha|$  and let  $k := \bar{q}/2 - |\alpha|$ . Clearly  $\|f\|_2 \geq f(y/\|y\|_2)$  where  $y := a + z$ , and

$$z := \frac{\mathbf{1}}{\sqrt{n}} + \frac{\gamma}{\sqrt{k}} + \frac{\sqrt{\alpha}}{\sqrt{t}}$$

and  $a$  is the unit vector that maximizes the quadratic polynomial

$$(h(z))(x).$$

Since  $\|y\|_2 = O(1)$ ,  $\|f\|_2 \geq f(y)/2^{O(d)}$ . Now clearly by non-negativity we have,

$$f(y) \geq (h(z))(a) = \|h(z)\|_2$$

Thus we have,

$$\begin{aligned} \|f\|_2^{q/d} &\geq \|(h(z))(x)\|_2^{q/d} \cdot 2^{-O(q)} \\ &= \|h(z)^{q/d}(x)\|_2 \cdot 2^{-O(q)} \\ &= \Lambda_C(h(z)^{q/d}(x)) \cdot 2^{-O(q)} && \text{(SOS exact on powered quadratics)} \\ &= \Lambda_C(s(z)(x)) \cdot 2^{-O(q)} \\ &\geq \Lambda_C(S_{2\alpha}(\mathbf{1}/\sqrt{n} + \gamma/\sqrt{k})(x)) \cdot \frac{2^{-O(q)}}{|\Theta(\alpha)|} && \text{(by Lemma 4.3 and Lemma 6.3)} \end{aligned}$$

$$\begin{aligned} &\geq \frac{\Lambda_C(\overline{C_k(S_{2\alpha})}_\gamma)}{k^{k/2} \cdot n^{k/2}} \cdot \frac{2^{-O(q)}}{|\mathcal{O}(\alpha)|} && \text{(by Lemma 4.3, and} \\ & && S_{2\alpha}(\frac{\mathbb{1}}{\sqrt{n}} + \frac{\gamma}{\sqrt{k}}) \geq \overline{C_k(S_{2\alpha})}_\gamma \text{ coefficient-wise)} \end{aligned}$$

which completes the proof since we've established Eq. (6.1).  $\blacksquare$

## 6.6. $(n/q)^{d/2-1}$ -Approximation for General Polynomials

**Theorem 6.8.** Consider any  $f \in \mathbb{R}_d^+[x]$  for  $d \geq 2$ , and any  $q$  divisible by  $2d$ . Then we have,

$$\frac{\Lambda(f^{q/d})^{d/q}}{\|f\|_2} \leq 2^{O(d)} \frac{n^{d/2-1}}{q^{d/2-1}}.$$

*Proof.* Let  $h$  be the unique  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and such that for any  $\beta \in \mathbb{N}_{d-2}^n$ , the fold  $\bar{h}_\beta$  of  $h$  is equal up to scaling, to the quadratic form of the corresponding  $(n \times n)$  block of the SOS-symmetric matrix representation  $M_f$  of  $f$ . That is, for any  $I, J \in [n]^{d/2-1}$ , s.t.  $\alpha(I) + \alpha(J) = \beta$ ,

$$\bar{h}_\beta(x) = \frac{x^T M_f[I, J]x}{|\mathcal{O}(\beta)|}.$$

Let  $s$  be the  $(\bar{q}, 2q/d)$ -folded polynomial given by  $h^{q/d}$  where  $\bar{q} := (d-2)q/d$ . Consider any  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  and  $\gamma \in \{0, 1\}_{\bar{q}-2|\alpha|}^n$ , and let  $S_{2\alpha}$  be the multilinear component of  $s$  as defined in Lemma 6.1. Below the fold (no pun intended), we will show

$$\|f\|_2^{q/d} \geq \frac{2^{-O(q)} \cdot \|\overline{(S_{2\alpha})}_\gamma\|_{sp}}{(\bar{q} - 2|\alpha|)^{\bar{q}/2-|\alpha|} \cdot |\mathcal{O}(\alpha)|} \quad (6.2)$$

which would complete the proof after applying Lemma 6.2 to  $s$  and composing the result with Lemma 6.5. To elaborate, we apply Lemma 6.2 to  $s$  with  $d_1 = \bar{q}, d_2 = 2q/d$  and then for every  $\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n$  we apply Lemma 6.5 with  $d_1 = \bar{q} - 2|\alpha|, d_2 = 2q/d$ , to get

$$\Lambda(f^{q/d}) = \Lambda(s) \leq 2^{O(q)} \cdot \max_{\alpha \in \mathbb{N}_{\leq \bar{q}/2}^n} \max_{\gamma \in \{0, 1\}_{\bar{q}-2|\alpha|}^n} \frac{\|\overline{(S_{2\alpha})}_\gamma\|_{sp}}{(\bar{q} - 2|\alpha|)^{\bar{q}/2-|\alpha|} \cdot |\mathcal{O}(\alpha)|}$$

which on combining with Eq. (6.2) yields the claim.

Fix any  $\alpha, \gamma$  satisfying the above conditions. Let  $k := \bar{q} - 2\alpha$ . Let  $t := |\alpha|$ , and let

$$z := \Xi \cdot \frac{1}{\sqrt{k}} \cdot \gamma \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{t}}$$

$\Xi$  is an independent and uniformly randomly chosen  $(k+1)$ -th root of unity, and for any  $i \in [n]$ ,  $\zeta_i$  is an independent and uniformly randomly chosen  $(2\alpha_i + 1)$ -th root of unity, and for any  $i \in [n]$ ,  $b_i$  is an independent Bernoulli( $p$ ) random variable ( $p$  is a parameter that will be set later). By Lemma 5.7 and definition of  $h$ , we see that for any  $y$ ,  $\|f\|_2^c \geq \|(h(y))(x)\|_2^c$ . Thus we have,

$$\|f\|_2^{q/d} = \|f^{q/d}\|_2$$

$$\begin{aligned}
&\geq \|f^{q/d}\|_2^c \cdot 2^{-O(q)} && \text{(by Lemma 5.8)} \\
&\geq \max_{p \in [0,1]} \mathbb{E} \left[ \|h(z)^{q/d}(x)\|_2 \right] \cdot 2^{-O(q)} && \text{(by Lemma 5.7)} \\
&= \max_{p \in [0,1]} \mathbb{E} \left[ \|h(z)^{q/d}(x)\|_{sp} \right] \cdot 2^{-O(q)} && \text{(SOS exact on powered quadratics)} \\
&= \max_{p \in [0,1]} \mathbb{E} \left[ \|h(z)^{q/d}(x) \cdot \mathbb{E} \cdot \prod_{i \in [n]} \zeta_i\|_{sp} \right] \cdot 2^{-O(q)} \\
&\geq \max_{p \in [0,1]} \left\| \mathbb{E} \left[ h(z)^{q/d}(x) \cdot \mathbb{E} \cdot \prod_{i \in [n]} \zeta_i \right] \right\|_{sp} \cdot 2^{-O(q)} \\
&= \max_{p \in [0,1]} \left\| \mathbb{E} \left[ (s(z))(x) \cdot \mathbb{E} \cdot \prod_{i \in [n]} \zeta_i \right] \right\|_{sp} \cdot 2^{-O(q)} \\
&= \max_{p \in [0,1]} \left\| p^k \cdot \frac{(S_{2\alpha}(\gamma/\sqrt{k}))(x)}{|\mathcal{O}(\alpha)|} + r(p) \right\|_{sp} \cdot 2^{-O(q)} && \text{(by Lemma 6.4, } \deg(r) < k) \\
&= \max_{p \in [0,1]} \left\| p^k \cdot \frac{\overline{(S_{2\alpha})}_\gamma(x)}{k^{k/2} \cdot |\mathcal{O}(\alpha)|} + r(p) \right\|_{sp} \cdot 2^{-O(q)} \\
&\geq \frac{\left\| \overline{(S_{2\alpha})}_\gamma \right\|_{sp}}{k^{k/2} \cdot |\mathcal{O}(\alpha)|} \cdot 2^{-O(q+k)} && \text{(Chebyshev Inequality - Lemma 5.5)}
\end{aligned}$$

where the last inequality follows by the following argument: one would like to show that there always exists  $p \in [0, 1]$  such that  $\|p^k \cdot h_k(x) + \dots p^0 \cdot h_0(x)\|_{sp} \geq \|h_k(x)\|_{sp} \cdot 2^{-O(k)}$ . So let  $p$  be such that  $|p^k \cdot u^T M_k v + \dots p^0 \cdot u^T M_0 v| \geq |u^T M_k v| \cdot 2^{-O(k)}$  (such a  $p$  exists by Chebyshev inequality) where  $M_k$  is the matrix representation of  $h_k(x)$  realizing  $\|h_k\|_{sp}$  and  $u, v$  are the maximum singular vectors of  $M_k$ .  $M_{k-1}, \dots, M_0$  are arbitrary matrix representations of  $h_{k-1}, \dots, h_0$  respectively. But  $p^k \cdot M_k + \dots p^0 \cdot M_0$  is a matrix representation of  $p^k \cdot h_k + \dots p^0 \cdot h_0$ . Thus  $\|p^k \cdot h_k + \dots p^0 \cdot h_0\|_{sp} \geq |u^T M_k v| / 2^{-O(k)} = \|h_k\|_{sp} \cdot 2^{-O(q)}$ .

This completes the proof as we've established Eq. (6.2).  $\blacksquare$

## 6.7. Algorithms

It is straightforward to extract algorithms from the proofs of Theorem 6.7 and Theorem 6.8.

### 6.7.1. Non-negative coefficient polynomials

Let  $f$  be a degree- $d$  polynomial with non-negative coefficients and let  $h$  be a  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$ . Consider any  $q$  divisible by  $2d$  and let  $\bar{q} := (d-2)q/d$ . Pick and return the best vector from the set

$$\left\{ \frac{\mathbf{1}}{\sqrt{n}} + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} + \frac{\gamma}{\sqrt{|\gamma|}} + \arg \max \left\| h \left( \frac{\mathbf{1}}{\sqrt{n}} + \frac{\sqrt{\alpha}}{\sqrt{|\alpha|}} + \frac{\gamma}{\sqrt{|\gamma|}} \right) (x) \right\|_2 \mid \alpha \in \mathbb{N}_{\leq \bar{q}/2}^n, \gamma \in \mathbb{N}_{\bar{q}/2 - |\alpha|}^n \right\}$$

## 6.7.2. General Polynomials

Let  $f$  be a degree- $d$  polynomial and let  $h$  be the unique  $(d-2, 2)$ -folded polynomial whose unfolding yields  $f$  and such that for any  $\beta \in \mathbb{N}_{d-2}^n$ , the fold  $\bar{h}_\beta$  of  $h$  is equal up to scaling, to the quadratic form of the corresponding  $(n \times n)$  block of the SOS-symmetric matrix representation  $M_f$  of  $f$ . That is, for any  $I, J \in [n]^{d/2-1}$ , s.t.  $\alpha(I) + \alpha(J) = \beta$ ,

$$\bar{h}_\beta(x) = \frac{x^T M_f[I, J] x}{|\mathcal{O}(\beta)|}.$$

Consider any  $q$  divisible by  $2d$  and let  $\bar{q} := (d-2)q/d$ . Let the set  $S$  be defined by,

$$S := \left\{ \Xi \cdot \frac{1}{\sqrt{|\gamma|}} \cdot \gamma \circ \frac{1}{2\alpha + \mathbb{1}} \circ b + \frac{\sqrt{\alpha} \circ \zeta}{\sqrt{|\alpha|}} \mid \begin{array}{l} \Xi \in \Omega_{k+1}, \zeta_i \in \Omega_{2\alpha_i+1}, b \in \{0, 1\}^n, \\ \alpha \in \mathbb{N}_{\leq \bar{q}/2}^n, \gamma \in \{0, 1\}_{\bar{q}-2|\alpha|}^n \end{array} \right\}$$

where  $\Omega_p$  denotes the set of  $p$ -th roots of unity. Pick and return the best vector from the set

$$\left\{ c_1 \cdot y + c_2 \cdot \arg \max \|(h(y))(x)\|_2 \mid y \in S, c_1 \in [-(d-2), (d-2)], c_2 \in [-2, 2] \right\}$$

Note that one need only search through all roots of unity vectors  $\zeta$  supported on  $S(\gamma)$  and all  $\{0, 1\}$ -vectors  $b$  supported on  $S(\alpha)$ . Lemma 5.7 can trivially be made constructive in time  $2^{O(q)}$ . Lastly, to go from complexes to reals, Lemma 5.8 can trivially be made constructive using  $2^{O(d)}$  time. Thus the algorithm runs in time  $n^{O(q)}$ .

## 7. Constant Level Lower Bounds for Polynomials with Non-negative Coefficients

Let  $G = (V, E)$  be a random graph drawn from the distribution  $G_{n,p}$  for  $p \geq n^{-1/3}$ . Let  $\mathcal{C} \subseteq \binom{V}{4}$  be the set of 4-cliques in  $G$ . The polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in \mathcal{C}} x_{i_1} x_{i_2} x_{i_3} x_{i_4}.$$

Clearly,  $f$  is multilinear and every coefficient of  $f$  is nonnegative. In this section, we prove the following two lemmas that establish a polynomial gap between  $\|f\|_2$  and  $\Lambda(f)$ .

**Lemma 7.1 (Soundness).** *With probability at least  $1 - \frac{1}{n}$  over the choice of the graph  $G$ , we have  $\|f\|_2 \leq n^2 p^6 \cdot (\log n)^{O(1)}$ .*

**Lemma 7.2 (Completeness).** *With probability at least  $1 - \frac{1}{n}$  over the choice of the graph  $G$ , we have*

$$\Lambda(f) \geq \Omega\left(\frac{n^{1/2} \cdot p}{\log^2 n}\right)$$

when  $p \in [n^{-1/3}, n^{-1/4}]$ .

Note that the gap between the two quantities is  $\tilde{\Omega}(n^{1/6})$  when  $p = n^{-1/3}$ , which is the choice we make.

## 7.1. Upper Bound on $\|f\|_2$

### 7.1.1. Reduction to counting shattered cliques

We say that an ordered 4-clique  $(i_1, \dots, i_4)$  is *shattered* by 4 disjoint sets  $Z_1, \dots, Z_4$  if for each  $k \in [4]$ ,  $i_k \in Z_k$ . Let  $Y_{j_1}, \dots, Y_{j_4}$  be the sets containing the coordinates  $i_1, \dots, i_4$ . Let  $\mathcal{C}_G$  denote the set of (ordered) 4-cliques in  $G$ , and let  $\mathcal{C}_G(Z_1, Z_2, Z_3, Z_4)$  denote the set of cliques shattered by  $Z_1, \dots, Z_4$ .

We reduce the problem of bounding  $\|f\|_2$ , to counting shattered 4-cliques.

**Claim 7.3.** *There exist disjoint sets  $Z_1, \dots, Z_4 \subseteq [n]$  such that*

$$|\mathcal{C}_G(Z_1, Z_2, Z_3, Z_4)| \geq \left( \prod_{k=1}^4 |Z_k| \right)^{1/2} \cdot O\left( \frac{\|f\|_2}{(\log n)^4} \right).$$

*Proof.* Let  $x^* \in \mathbb{S}^{n-1}$  be the vector that maximizes  $f$ . Without loss of generality, assume that every coordinate of  $x^*$  is nonnegative. Let  $y^*$  be another unit vector defined as

$$y^* := \frac{(x^* + \mathbf{1}/\sqrt{n})}{\|x^* + \mathbf{1}/\sqrt{n}\|_2}.$$

Since both  $x^*$  and  $\frac{\mathbf{1}}{\sqrt{n}}$  are unit vectors, the denominator is at most 2. This implies that  $f(y^*) \geq \frac{f(x^*)}{2^4}$ , and each coordinate of  $y^*$  is at least  $\frac{1}{2\sqrt{n}}$ . For  $1 \leq j \leq \log_2 n$ , let  $Y_j$  be the set

$$Y_j := \left\{ i \in [n] \mid 2^{-j} < y_i^* \leq 2^{-(j-1)} \right\}.$$

The sets  $Y_1, \dots, Y_{\log_2 n}$  partition  $[n]$ . Since  $1 = \sum_{i \in [n]} y_i^2 > |Y_j| \cdot 2^{-2j}$ , we have for each  $j$ ,  $|Y_j| \leq 2^{2j}$ . Let  $Z_1, Z_2, Z_3$ , and  $Z_4$  be pairwise disjoint random subsets of  $[n]$  chosen as follows:

- Randomly partition each  $Y_j$  to  $Y_{j,1}, \dots, Y_{j,4}$  where each element of  $Y_j$  is put into exactly one of  $Y_{j,1}, \dots, Y_{j,4}$  uniformly and independently.
- Sample  $r_1, \dots, r_4$  independently and randomly from  $\{1, \dots, \log_2 n\}$ .
- For  $k = 1, \dots, 4$ , take  $Z_k := Y_{r_k, k}$

We use  $\mathcal{P}$  to denote random partitions  $\{(Y_{j,1}, \dots, Y_{j,4})\}_{j \in [\log n]}$  and  $r$  to denote the random choices  $r_1, \dots, r_4$ . Note that the events  $i_k \in Z_k$  are independent for different  $k$ , and that  $Z_1, \dots, Z_4$  are independent given  $\mathcal{P}$ . Thus, we have

$$\begin{aligned} \mathbb{E}_{\mathcal{P}, r} \left[ \frac{\mathbb{1}[(i_1, i_2, i_3, i_4) \text{ is shattered}]}{\sqrt{|Z_1| |Z_2| |Z_3| |Z_4|}} \right] &= \mathbb{E}_{\mathcal{P}} \left[ \prod_{k=1}^4 \mathbb{E}_{r_k} \left[ \frac{\mathbb{1}[i_k \in Z_k]}{\sqrt{|Z_k|}} \right] \right] \\ &= \mathbb{E}_{\mathcal{P}} \left[ \prod_{k=1}^4 \mathbb{E}_{r_k} \left[ \frac{\mathbb{1}[r_k = j_k] \cdot \mathbb{1}[i_k \in Y_{j_k, k}]}{\sqrt{|Y_{j_k, k}|}} \right] \right] \\ &\geq \mathbb{E}_{\mathcal{P}} \left[ \prod_{k=1}^4 \mathbb{E}_{r_k} \left[ \frac{\mathbb{1}[r_k = j_k] \cdot \mathbb{1}[i_k \in Y_{j_k, k}]}{\sqrt{|Y_{j_k}|}} \right] \right] \end{aligned}$$



$$\begin{aligned}
&= \mathbb{E}_{\mathcal{P}} \left[ \prod_{k=1}^4 \left( \frac{1}{\log n} \cdot \frac{\mathbb{1}[i_k \in Y_{j_k, k}]}{\sqrt{|Y_{j_k}|}} \right) \right] \\
&= \mathbb{E}_{\mathcal{P}} \left[ \prod_{k=1}^4 \left( \frac{1}{\log n} \cdot \frac{\mathbb{1}[i_k \in Y_{j_k, k}]}{\sqrt{|Y_{j_k}|}} \right) \right] \\
&= \frac{1}{(4 \log n)^4} \cdot \frac{1}{\sqrt{|Y_{j_1}| |Y_{j_2}| |Y_{j_3}| |Y_{j_4}|}} \\
&\geq \frac{1}{(4 \log n)^4} \cdot 2^{j_1 + j_2 + j_3 + j_4} \\
&\geq \frac{1}{(8 \log n)^4} \cdot y_{i_1}^* y_{i_2}^* y_{i_3}^* y_{i_4}^*.
\end{aligned}$$

Then, by linearity of expectation,

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}, r} \left[ \frac{|\mathcal{C}_G(Z_1, Z_2, Z_3, Z_4)|}{\sqrt{|Z_1| |Z_2| |Z_3| |Z_4|}} \right] &\geq \frac{1}{(8 \log n)^4} \cdot \sum_{(i_1, \dots, i_4) \in \mathcal{C}_G} y_{i_1}^* y_{i_2}^* y_{i_3}^* y_{i_4}^* \\
&= \frac{4!}{(8 \log n)^4} \cdot f(y^*) \\
&\geq \frac{4!}{(16 \log n)^4} \cdot f(x^*) = \frac{4!}{(16 \log n)^4} \cdot \|f\|_2,
\end{aligned}$$

which proves the claim.  $\blacksquare$

We will show that with high probability,  $G$  satisfies the property that every four disjoint sets  $Z_1, \dots, Z_4 \subseteq V$  shatter at most  $O\left(\sqrt{|Z_1| |Z_2| |Z_3| |Z_4|} \cdot n^2 p^6 \cdot (\log n)^{O(1)}\right)$  cliques, proving Lemma 7.1.

### 7.1.2. Counting edges and triangles

For a vertex  $i \in [n]$ , we use  $N(i)$  to denote the set of vertices in the graph  $G$ . For ease of notation, we use  $a \lesssim b$  to denote  $a \leq b \cdot (\log n)^{O(1)}$ . We first collect some simple consequences of Chernoff bounds.

**Claim 7.4.** *Let  $G \sim G_{n,p}$  with  $p \geq n^{-1/3}$ . Then, with probability  $1 - \frac{1}{n}$ , we have*

- For all distinct  $i_1, i_2 \in [n]$ ,  $|N(i_1) \cap N(i_2)| \lesssim np^2$ .
- For all distinct  $i_1, i_2, i_3 \in [n]$ ,  $|N(i_1) \cap N(i_2) \cap N(i_3)| \lesssim np^3$ .
- For all sets  $S_1, S_2 \subseteq [n]$ ,  $|E(S_1, S_2)| \lesssim \max\{|S_1| |S_2| p, |S_1| + |S_2|\}$ .

We also need the following bound on the number of triangles shattered by three disjoint sets  $S_1, S_2$  and  $S_3$ , denoted by  $\Delta_G(S_1, S_2, S_3)$ . As for 4-cliques, a triangle is said to be shattered if it has exactly one vertex in each the sets.

**Claim 7.5.** Let  $G \sim G_{n,p}$  with  $p \geq n^{-1/3}$ . Then, with probability  $1 - \frac{1}{n}$ , for all disjoint sets  $S_1, S_2, S_3 \subseteq [n]$

$$|\Delta_G(S_1, S_2, S_3)| \lesssim |S_3| + |E(S_1, S_2)| \cdot (np^3 \cdot |S_3|)^{1/2}.$$

*Proof.* With probability at least  $1 - \frac{1}{n}$ ,  $G$  satisfies the conclusion of Claim 7.4. Fix such a  $G$ , and consider arbitrary subsets  $S_1, S_2, S_3 \subseteq V$ . Consider the bipartite graph  $H$  where the left side vertices correspond to edges in  $E(S_1, S_2)$ , the right side vertices correspond to vertices in  $S_3$ , and there is an edge from  $(i_1, i_2) \in E(S_1, S_2)$  to  $i_3 \in S_3$  when both  $(i_1, i_3), (i_2, i_3) \in E$ . Clearly,  $|\Delta_G(S_1, S_2, S_3)|$  is equal to the number of edges in  $H$ .

Consider two different edges  $(i_1, i_2), (i'_1, i'_2) \in E(S_1, S_2)$ . These two edges are incident on at least 3 distinct vertices, say  $\{i_1, i_2, i'_1\}$ . Hence, the number of vertices  $i_3 \in [n]$  that are adjacent to all  $\{i_1, i_2, i'_1, i'_2\}$  in  $G$  is at most  $|\mathcal{N}(i_1) \cap \mathcal{N}(i_2) \cap \mathcal{N}(i'_1)| \lesssim np^3$ . This gives that the number of pairs triangles sharing a common vertex in  $S_3$  is at most  $|E(S_1, S_2)|^2 \cdot np^3 (\log n)^{O(1)}$ .

Let  $d_H(i_3)$  denote the degree of a vertex  $i_3$  in  $H$ , and let  $\Delta$  denote the number of shattered triangles. Counting the above pairs of triangles using the degrees gives

$$\sum_{i_3 \in S_3} \binom{d_H(i_3)}{2} \lesssim |E(S_1, S_2)|^2 \cdot np^3.$$

An application of Cauchy-Schwarz gives

$$\Delta^2 - \Delta \cdot |S_3| \lesssim |S_3| \cdot |E(S_1, S_2)|^2 \cdot np^3,$$

which proves the claim. ■

### 7.1.3. Bounding 4-clique Density

Let  $G \sim G_{n,p}$  be a graph satisfying the conclusions of Claims 7.4 and 7.5. Let  $S_1, \dots, S_4 \subseteq [n]$  be disjoint sets with sizes  $n_1 \leq n_2 \leq n_3 \leq n_4$ . We consider two cases:

- **Case 1:**  $|E(S_1, S_2)| \lesssim n_1 n_2 p$

Note that each edge  $(i_1, i_2)$  can only participate in at most  $|\mathcal{N}(i_1) \cap \mathcal{N}(i_2)|$  triangles, and each triangle  $(i_1, i_2, i_3)$  can only be extended to at most  $|\mathcal{N}(i_1) \cap \mathcal{N}(i_2) \cap \mathcal{N}(i_3)|$  4-cliques. Thus, Claim 7.4 gives

$$|\mathcal{C}_G(S_1, S_2, S_3, S_4)| \lesssim n_1 n_2 p \cdot np^2 \cdot np^3 \lesssim (n_1 n_2 n_3 n_4)^{1/2} \cdot n^2 p^6.$$

- **Case 2:**  $|E(S_1, S_2)| \lesssim n_1 + n_2$

Claim 7.5 gives

$$|\Delta_G(S_1, S_2, S_3)| \lesssim n_3 + (n_1 + n_2) \cdot (n_3 \cdot np^3)^{1/2},$$

which together with Claim 7.4 implies

$$|\mathcal{C}_G(S_1, S_2, S_3, S_4)| \lesssim n_3 \cdot np^3 + (n_1 + n_2) \cdot n_3^{1/2} \cdot (np^3)^{3/2}.$$

Considering the first term, we note that

$$n_3 \cdot np^3 \leq (n_3 n_4)^{1/2} \cdot n^2 p^6 \leq (n_1 n_2 n_3 n_4)^{1/2} \cdot n^2 p^6,$$

since  $n_3 \leq n_4$  and  $np^3 \geq 1$ . Similarly, for the second term, we have

$$(n_1 + n_2) \cdot n_3^{1/2} \cdot (np^3)^{3/2} \leq 2(n_2 n_3 n_4)^{1/2} \cdot (np^3)^{3/2} \leq 2 \cdot (n_1 n_2 n_3 n_4)^{1/2} \cdot n^2 p^6.$$

Combined with Claim 7.3, this completes the proof of Lemma 7.1.

## 7.2. Lower Bound on $\Lambda(f)$

Recall that given a random graph  $G = ([n], E)$  drawn from the distribution  $G_{n,p}$ , the polynomial  $f$  is defined as

$$f(x_1, \dots, x_n) := \sum_{\{i_1, i_2, i_3, i_4\} \in \mathcal{C}} x_{i_1} x_{i_2} x_{i_3} x_{i_4},$$

where  $\mathcal{C} \subseteq \binom{[n]}{4}$  is the set of 4-cliques in  $G$ . Let  $A \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the natural matrix representation of  $24f$  (corresponding to ordered copies of cliques) with

$$A[(i_1, i_2), (i_3, i_4)] = \begin{cases} 1 & \text{if } \{i_1, \dots, i_4\} \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases}$$

Let  $E' \subseteq [n]^2$  be the set of ordered edges i.e.,  $(i_1, i_2) \in E'$  if and only if  $\{i_1, i_2\} \in E$ . Note that  $|E'| = 2m$  where  $m$  is the number of edges in  $G$ . All nonzero entries of  $A$  are contained in the principal submatrix  $A_{E'}$ , formed by the rows and columns indexed by  $E'$ .

### 7.2.1. A simple lower bound on $\|f\|_{sp}$

We first give a simple proof that  $\|f\|_{sp} \geq \sqrt{n^2 p^5}$  with high probability.

**Lemma 7.6.**  $\|f\|_{sp} \geq \Omega(\sqrt{n^2 p^5}) = \Omega(n^{1/6})$  with high probability.

*Proof.* Consider any matrix representation  $M$  of  $24f$  and its principal submatrix  $M_{E'}$ . It is easy to observe that the Frobenius norm of  $M_{E'}$  satisfies  $\|M_{E'}\|_F^2 \geq 24|\mathcal{C}|$ , minimized when  $M = A$ . Since  $\|M_{E'}\|_F^2 \leq |E'| \cdot \|A_{E'}\|_2^2$ , we have that with high probability,

$$\|A\|_2 \geq \|A_{E'}\|_2 \geq \sqrt{\frac{24|\mathcal{C}|}{2|E'|}} = \Omega\left(\frac{\sqrt{n^4 p^6}}{\sqrt{n^2 p}}\right) = \Omega\left(\sqrt{n^2 p^5}\right). \quad \blacksquare$$

### 7.2.2. Lower bound for the stronger relaxation computing $\Lambda(f)$

We now prove Lemma 7.2, which says that  $\Lambda(f) \geq \frac{n^{1/6}}{\log^2 n}$  with high probability. In order to show a lower bound, we look at the dual SDP for computing  $\Lambda(f)$ , which is a maximization problem over positive semidefinite, SoS-symmetric matrices  $M$  with  $\text{Tr}(M) = 1$ . We exhibit such a matrix  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  for which the value of the objective  $\langle A, M \rangle$  is large.

For large  $\langle A, M \rangle$ , one natural attempt is to take  $M$  to be  $A$  and modify it to satisfy other conditions. Note that  $A$  is already SoS-symmetric. However,  $\text{Tr}(A) = 0$ , which implies that the minimum eigenvalue is negative.

Let  $\lambda_{\min}$  be the minimum eigenvalue of  $A$ , which is also the minimum eigenvalue of  $A_{E'}$ . Let  $I_{E'} \in \mathbb{R}^{[n]^2 \times [n]^2}$  be such that  $I[(i_1, i_2), (i_1, i_2)] = 1$  if  $(i_1, i_2) \in E'$  and all other entries are 0. Note that  $I_{E'}$  is a diagonal matrix with  $\text{Tr}(I_{E'}) = 2m$ . Adding  $-\lambda_{\min} \cdot I_{E'}$  to  $A$  makes it positive semidefinite, so setting

$$M = \frac{A - \lambda_{\min} I_{E'}}{\text{Tr}(A - \lambda_{\min} I_{E'})} = \frac{A - \lambda_{\min} I_{E'}}{-2m\lambda_{\min}} = \frac{A + |\lambda_{\min}| \cdot I_{E'}}{2m \cdot |\lambda_{\min}|} \quad (7.1)$$

makes sure that  $M$  is positive semidefinite,  $\text{Tr}(M) = 1$ , and  $\langle A, M \rangle = \frac{12|\mathcal{C}|}{m \cdot |\lambda_{\min}|}$  (each 4-clique in  $\mathcal{C}$  contributes 24). Since  $|\mathcal{C}| = \Theta(n^4 p^6)$  and  $m = \Theta(n^2 p)$  with high probability, if  $|\lambda_{\min}| = O(np^{5/2})$ ,  $\langle A, M \rangle = \Theta(n^2 p^{5/2})$ , which is  $\Omega(n^{1/6})$  when  $p = \Omega(n^{-1/3})$ .

The  $M$  defined in Eq. (7.1) does not directly work since it is not SoS-symmetric. However, the following claim proves that this issue can be fixed by losing a factor 2 in  $\langle A, M \rangle$ .

**Claim 7.7.** *There exists  $M$  such that it is SoS-symmetric, positive semidefinite with  $\text{Tr}(M) = 1$ , and  $\langle A, M \rangle \geq \frac{6|\mathcal{C}|}{m \cdot |\lambda_{\min}|}$ .*

*Proof.* Let  $Q_{E'} \in \mathbb{R}^{[n]^2 \times [n]^2}$  be the matrix such that

- For  $(i_1, i_2) \in E'$ ,  $Q_{E'}[(i_1, i_1), (i_2, i_2)] = Q_{E'}[(i_2, i_2), (i_1, i_1)] = 1$ .
- For  $i \in [n]$ ,  $Q_{E'}[(i, i), (i, i)] = \deg_G(i)$ , where  $\deg_G(i)$  denotes the degree of  $i$  in  $G$ .
- All other entries are 0.

We claim that  $I_{E'} + Q_{E'}$  is SoS-symmetric:  $(I_{E'} + Q_{E'})[(i_1, i_2), (i_3, i_4)]$  has a nonzero entry if and only if  $i_1 = i_2 = i_3 = i_4$  or two different numbers  $j_1, j_2$  appear exactly twice and  $(j_1, j_2) \in E'$  (in this case  $(I_{E'} + Q_{E'})[(i_1, i_2), (i_3, i_4)] = 1$ ). Since  $A$  is SoS-symmetric, so  $A + |\lambda_{\min}| \cdot (I_{E'} + Q_{E'})$  is also SoS-symmetric.

It is easy to see that  $Q_{E'}$  is diagonally dominant, and hence positive semidefinite. Since we already argued that  $A + |\lambda_{\min}| \cdot I_{E'}$  is positive semidefinite,  $A + |\lambda_{\min}| \cdot (I_{E'} + Q_{E'})$  is also positive semidefinite. Also,  $\text{Tr}(Q_{E'}) = \sum_{i \in [n]} \deg_G(i) = 2m$ . Thus, we take

$$M = \frac{A + |\lambda_{\min}| \cdot (I_{E'} + Q_{E'})}{\text{Tr}(A + |\lambda_{\min}| \cdot (I_{E'} + Q_{E'}))} = \frac{A + |\lambda_{\min}| \cdot I_{E'}}{4m \cdot |\lambda_{\min}|}.$$

By the above arguments, we have that  $M$  that is PSD, SoS-symmetric with  $\text{Tr}(M) = 1$ , and

$$\langle A, M \rangle = \frac{6|\mathcal{C}|}{m \cdot |\lambda_{\min}|}$$

as desired. ■

It only remains to bound  $\lambda_{\min}$ , which is the minimum eigenvalue of  $A$  and  $A_{E'}$ . For  $p$  in the range  $[n^{-1/3}, n^{-1/4}]$ , we will show a bound of  $\tilde{O}(n^{3/2} p^4)$  below, which when combined with the above claim, completes the proof of Lemma 7.2.

### 7.2.3. Bounding the smallest eigenvalue via the trace method

Our estimate  $|\lambda_{\min}| = O(np^{5/2})$  is based on the following observation:  $A_{E'}$  is a  $2m \times 2m$  random matrix where each row and column is expected to have  $\Theta(n^2p^5)$  ones (the expected number of 4-cliques an edge participates in). An adjacency matrix of a random graph with average degree  $d$  has a minimum eigenvalue  $-\Theta(\sqrt{d})$ , hence the estimate  $|\lambda_{\min}| = O(np^{5/2})$ . Even though  $A_{E'}$  is not sampled from a typical random graph model (and even  $E'$  is a random variable), we will be able to prove the following weaker estimate, which suffices for our purposes.

**Lemma 7.8.** *With high probability over the choice of the graph  $G$ , we have*

$$|\lambda_{\min}| = \begin{cases} \tilde{O}(n^{3/2} \cdot p^4) & \text{for } p \in [n^{-1/3}, n^{-1/4}] \\ \tilde{O}(n^{5/3} \cdot p^{14/3}) & \text{for } p \in [n^{-1/4}, 1/2] \end{cases}$$

*Proof.* Instead of  $A_{E'}$ , we directly study  $A$  to bound  $\lambda_{\min}$ . For simplicity, we consider the following matrix  $\hat{A}$ , where each row and column is indexed by an unordered pair  $\{i, j\} \in \binom{[n]}{2}$ , and  $\hat{A}[\{i_1, i_2\}, \{i_3, i_4\}] = 1$  if and only if  $i_1, i_2, i_3, i_4$  form a 4-clique.  $A$  has only zero entries in the rows or columns indexed by  $(i, i)$  for all  $i \in [n]$ , and for two pairs  $i_1 \neq i_2$  and  $i_3 \neq i_4$ , we have

$$\begin{aligned} \hat{A}[\{i_1, i_2\}, \{i_3, i_4\}] &:= \frac{1}{4} \cdot \{A[(i_1, i_2), (i_3, i_4)] + A[(i_1, i_2), (i_4, i_3)]\} \\ &\quad + \frac{1}{4} \cdot \{A[(i_2, i_1), (i_3, i_4)] + A[(i_2, i_1), (i_4, i_3)]\}. \end{aligned}$$

Therefore,  $|\lambda_{\min}(A)| \leq 4 \cdot |\lambda_{\min}(\hat{A})|$  and it suffices to bound the minimum eigenvalue of  $\hat{A}$ . We consider the matrix  $\hat{N}_E := \hat{A} - p^4 \cdot \hat{J}_E$ , where  $\hat{J}_E \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$  is such that

$$\hat{J}_E[\{i_1, i_2\}, \{i_3, i_4\}] = \begin{cases} 1 & \text{if } \{i_1, i_2\}, \{i_3, i_4\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

Since  $\hat{J}_E$  is a rank-1 matrix with a positive eigenvalue, the minimum eigenvalues of  $\hat{A}$  and  $\hat{N}_E$  are the same. In summary,  $\hat{N}_E$  is the following matrix.

$$\hat{N}_E[\{i_1, i_2\}, \{i_3, i_4\}] = \begin{cases} 1 - p^4 & \text{if } \{i_1, i_2, i_3, i_4\} \in \mathcal{C} \\ -p^4 & \text{if } \{i_1, i_2, i_3, i_4\} \notin \mathcal{C} \text{ but } \{i_1, i_2\}, \{i_3, i_4\} \in E \\ 0 & \text{otherwise} \end{cases}$$

We use the trace method to bound  $\|\hat{N}_E\|_2$ , based on the observation that for every even  $r \in \mathbb{N}$ ,  $\|\hat{N}_E\|_2 \leq \left(\text{Tr}((\hat{N}_E)^r)\right)^{1/r}$ . Fix an even  $r \in \mathbb{N}$ . The expected value of the trace can be represented as

$$\mathbb{E} \left[ \text{Tr}((\hat{N}_E)^r) \right] = \mathbb{E} \left[ \sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \prod_{k=1}^r \hat{N}_E[I^k, I^{k+1}] \right] = \sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \mathbb{E} \left[ \prod_{k=1}^r \hat{N}_E[I^k, I^{k+1}] \right]$$

where each  $I^j = \{i_1^j, i_2^j\} \in \binom{[n]}{2}$  is an edge of the complete graph on  $n$  vertices (call it a *potential edge*) and  $I^{r+1} := I^1$ .

Fix  $r$  potential edges  $I^1, \dots, I^r$ , let  $t := \prod_{k=1}^r \widehat{\mathbb{N}}_E[I^k, I^{k+1}]$ , and consider  $\mathbb{E}[t]$ . Let  $E_0 := \{I^1, \dots, I^r\}$  be the set of distinct edges represented by  $I^1, \dots, I^r$ . Note that the expected value is 0 if one of  $I^j$  does not become an edge. Therefore,  $\mathbb{E}[t] = p^{|E_0|} \cdot \mathbb{E}[t \mid E_0 \subseteq E]$ .

Let  $D \subseteq [r]$  be the set of  $j \in [r]$  such that all four vertices in  $I^j$  and  $I^{j+1}$  are distinct i.e.,

$$D := \left\{ j \in [r] \mid \left| \{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\} \right| = 4 \right\}.$$

For  $j \in [r] \setminus D$ ,  $\{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\}$  cannot form a 4-clique, so given that  $I^j, I^{j+1} \in E$ , we have  $\widehat{\mathbb{N}}_E[I^j, I^{j+1}] = -p^4$ . For  $j \in D$ , let  $E_j := \left\{ \{i_1^j, i_1^{j+1}\}, \{i_1^j, i_2^{j+1}\}, \{i_2^j, i_1^{j+1}\}, \{i_2^j, i_2^{j+1}\} \right\} \setminus E_0$  be the set of edges in the 4-clique created by  $\{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\}$  except ones in  $E_0$ . Then

$$\mathbb{E}[t] = p^{|E_0|} \cdot \mathbb{E}[t \mid E_0 \subseteq E] = p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \mathbb{E} \left[ \prod_{k \in D} \widehat{\mathbb{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E \right].$$

Suppose there exists  $j \in D$  such that  $|E_j| = 4$  and  $E_j \cap (\cup_{j' \in D \setminus \{j\}} E_{j'}) = \emptyset$ . Then, given that  $E_0 \subseteq E$ ,  $\widehat{\mathbb{N}}_E[I^j, I^{j+1}]$  is independent of all  $\left\{ \widehat{\mathbb{N}}_E[I^k, I^{k+1}] \right\}_{k \in D \setminus \{j\}}$ , and

$$\mathbb{E} \left[ \widehat{\mathbb{N}}_E[I^j, I^{j+1}] \mid E_0 \subseteq E \right] = p^4(1 - p^4) + (1 - p^4)(-p^4) = 0.$$

Therefore,  $\mathbb{E}[t] = 0$  unless for all  $j \in D$ , either  $|E_j| \leq 3$  or there exists  $j' \in D \setminus \{j\}$  with  $E_j \cap E_{j'} \neq \emptyset$ .

Let  $E_D := \cup_{j \in D} E_j$ . Note that  $E_0$  and  $E_D$  completely determines  $t$ .  $\mathbb{E}[t]$  can be written as

$$\begin{aligned} & p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \mathbb{E} \left[ \prod_{k \in D} \widehat{\mathbb{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E \right] \\ &= p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \sum_{F \subseteq E_D} \left( p^{|F|} (1-p)^{|E_D|-|F|} \cdot \mathbb{E} \left[ \prod_{k \in D} \widehat{\mathbb{N}}_E[I^k, I^{k+1}] \mid E_0 \subseteq E, E_D \cap E = F \right] \right) \\ &= p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \sum_{F \subseteq E_D} \left( p^{|F|} (1-p)^{|E_D|-|F|} \cdot (1-p^4)^{|D|-a(F)} (-p^4)^{a(F)} \right), \end{aligned}$$

where  $a(F)$  denotes the number of  $j \in D$  with  $E_j \not\subseteq F$ . Since  $E_D \subseteq F \cup (\cup_{j: E_j \not\subseteq F} E_j)$  and  $4a(F) + |F| \geq |E_D|$ , we have

$$\begin{aligned} \mathbb{E}[t] &= p^{|E_0|} \cdot (-p^4)^{r-|D|} \cdot \sum_{F \subseteq E_D} \left( p^{|F|} (1-p)^{|E_D|-|F|} \cdot (1-p^4)^{|D|-a(F)} (-p^4)^{a(F)} \right) \\ &\leq p^{|E_0|} \cdot (p^4)^{r-|D|} \cdot 2^{|E_D|} \cdot p^{|E_D|} \\ &\leq 2^{4r} \cdot p^{4(r-D)+|E_0|+|E_D|}. \end{aligned}$$

We now count the number of terms which contribute to the sum. Fix a graph  $H$  with  $r$  labelled edges  $I^1, \dots, I^r$  (possibly repeated) and  $q := q(H)$  vertices, without any isolated

vertex (so  $q \leq 2r$ ). There are at most  $\binom{q}{2}^r \leq (2r)^{2r}$  such graphs. Then  $I^1, \dots, I^r$ , as edges in  $\binom{[n]}{2}$ , are determined by a map  $V_H \rightarrow [n]$ . There are at most  $n^q$  such mappings.

Let  $E_0 := E_0(H), D := D(H), E_j := E_j(H), E' := E'(H)$  be defined as before. Note that  $E_0$  is set the edges of  $H$ . As observed before, the contribution from  $H$  is 0 if there exists  $j \in D$  such that  $|E_j| = 4$  and  $E_j$  is disjoint from  $\{E_{j'}\}_{j' \in D \setminus \{j\}}$ . Let  $\mathcal{H}$  be the set of  $H$  that has nonzero contribution. Then,

$$\begin{aligned} \mathbb{E} \left[ \text{Tr} \left( \left( \widehat{\mathbf{N}}_E \right)^r \right) \right] &= \sum_{I^1, \dots, I^r \in \binom{[n]}{2}} \mathbb{E} \left[ \prod_{k=1}^r \widehat{\mathbf{N}}_E [I^k, I^{k+1}] \right] \\ &\leq \sum_{H \in \mathcal{H}} n^{q(H)} \cdot 2^{4r} \cdot p^{4(r-D(H))+|E_0(H)|+|E_D(H)|} \\ &\leq (2r)^{2r} \cdot \max_{H \in \mathcal{H}} \left( n^{q(H)} 2^{4r} \cdot p^{4(r-D(H))+|E_0(H)|+|E_D(H)|} \right) \\ &\leq (8r)^{2r} \cdot \max_{H \in \mathcal{H}} \left( n^{q(H)} p^{4(r-D(H))+|E_0(H)|+|E_D(H)|} \right) \end{aligned}$$

We will prove the following bound on the maximum contribution of any  $H \in \mathcal{H}$ .

**Claim 7.9.** *Let  $\mathcal{H}$  be defined as above. Then, for all  $H \in \mathcal{H}$ , we have*

$$n^{q(H)} p^{4(r-D(H))+|E_0(H)|+|E_D(H)|} \leq n^2 \cdot B_p^r,$$

where

$$B_p = \begin{cases} n^{3/2} \cdot p^4 & \text{for } p \in [n^{-1/3}, n^{-1/4}] \\ n^{5/3} \cdot p^{14/3} & \text{for } p \in [n^{-1/4}, 1/2] \end{cases}.$$

Using the above claim, we can bound  $\mathbb{E} \left[ \text{Tr} \left( \left( \widehat{\mathbf{N}}_E \right)^r \right) \right]$  as

$$\begin{aligned} \mathbb{E} \left[ \text{Tr} \left( \left( \widehat{\mathbf{N}}_E \right)^r \right) \right] &\leq (8r)^{2r} \cdot \max_{H \in \mathcal{H}} \left( n^{q(H)} p^{4(r-D(H))+|E_0(H)|+|E_D(H)|} \right) \\ &\leq (8r)^{2r} \cdot n^2 \cdot B_p^r, \end{aligned}$$

where  $B_p$  is given by Claim 7.9 for different ranges of  $p$ . By Markov's inequality, we get that with probability  $1 - \frac{1}{n}$ , we have  $\text{Tr} \left( \left( \widehat{\mathbf{N}}_E \right)^r \right) \leq (8r)^{2r} \cdot n^3 \cdot B_p^r$ , which gives

$$\left\| \widehat{\mathbf{N}}_E \right\|_2 \leq (8r)^2 \cdot B_p \cdot n^{3/r}.$$

Choosing  $r = \Theta(\log n)$  then proves the lemma. ■

It remains to prove Claim 7.9.

#### 7.2.4. Analyzing contributing subgraphs

Recall that graphs  $H \in \mathcal{H}$  were constructed from edges  $\{I^1, \dots, I^r\}$ , with edge  $I^j$  consisting of vertices  $\{i_1^j, i_2^j\}$ . Also, we define  $q(H) = |V(H)|$ . Moreover, we defined the following sets for graph  $H$

$$E_0(H) := \{I^1, \dots, I^r\} \quad (\text{counting only distinct edges})$$

$$\begin{aligned}
D(H) &:= \left\{ j \in [r] \mid \left| \{i_1^j, i_2^j, i_1^{j+1}, i_2^{j+1}\} \right| = 4 \right\} \\
E_j(H) &:= \left\{ \{i_1^j, i_1^{j+1}\}, \{i_1^j, i_2^{j+1}\}, \{i_2^j, i_1^{j+1}\}, \{i_2^j, i_2^{j+1}\} \right\} \setminus E_0(H) \\
E_D(H) &:= \bigcup_{j \in D} E_j(H)
\end{aligned}$$

Moreover, the graph  $H$  is in  $\mathcal{H}$  only if for every  $j \in D$ , either  $|E_j(H)| \leq 3$  or there exists  $j' \in D \setminus \{j\}$  such that  $E_j(H) \cap E_{j'}(H) \neq \emptyset$ . Claim 7.9 then follows from the following combinatorial claim (taking  $b = \log(1/p) / \log n$ ).

**Claim 7.10.** *Any graph  $H \in \mathcal{H}$  satisfies, for all  $b \in [0, 1/3]$*

$$q(H) \leq 2 + b \cdot (4(r - |D(H)|) + |E_0(H)| + |E_D(H)|) + c \cdot r,$$

where  $c = 5/3 - 14b/3$  for  $b \in [0, 1/4]$  and  $c = 3/2 - 4b$  for  $b \in [1/4, 1/3]$ .

*Proof.* Fix a graph  $H \in \mathcal{H}$ . Let  $j = 1, \dots, r$ , let  $V_j := \{i_1^j, i_2^j\}_{1 \leq j \leq r}$  (i.e., the set of vertices covered by  $I^1, \dots, I^j$ ). For each  $j = 2, \dots, r$ , let  $v_j := |V_j| - |V_{j-1}|$  and classify the index  $j$  to one of the following types.

- Type  $-1$ :  $I^j \cap I^{j-1} \neq \emptyset$  (equivalently,  $j-1 \notin D$ ).
- Type  $k$  ( $0 \leq k \leq 2$ ):  $I^j$  and  $I^{j-1}$  are disjoint, and  $v_j = k$  (i.e., adding  $I^j$  introduces  $k$  new vertices).

Let  $T_k$  ( $-1 \leq k \leq 2$ ) be the set of indices of Type  $k$ , and let  $t_k := |T_k|$ . The number of vertices  $q$  is bounded by

$$q \leq 2 + 1 \cdot t_{-1} + 0 \cdot t_0 + 1 \cdot t_1 + 2 \cdot t_2 = 2 + t_{-1} + t_1 + 2t_2.$$

Let  $H_j$  be the graph with  $V_j$  as vertices and edges

$$E(H_j) = \{I^1, \dots, I^j\} \cup \left( \bigcup_{k \in D \cap [j-1]} E_k \right).$$

For  $j = 2, \dots, r$ , let  $e_j = |E(H_j)| - |E(H_{j-1})|$ . For an index  $j \in T_2$ , adding two vertices  $i_1^j, i_2^j$  introduces at least 5 edges in  $H_j$  compared to  $H_{j-1}$  (i.e., six edges in the 4-clique on  $\{i_1^{j-1}, i_2^{j-1}, i_1^j, i_2^j\}$  except  $I^{j-1}$ ), so  $e_j \geq 5$ . Similarly, we get  $e_j \geq 3$  for each  $j \in T_1$ .

The lemma is proved via the following charging argument. For each index  $j = 2, \dots, r$ , we get value  $b$  for each edge in  $H_j \setminus H_{j-1}$  and get value  $c$  for the new index. If  $j \in T_{-1}$ , we get an additional value of  $4b$ . We give this value to vertices in  $V_j \setminus V_{j-1}$ . If we do not give more value than we get and each vertex in  $V(H) \setminus V_1$  gets more than 1, this means

$$q - 2 \leq b \cdot (|E_0| + |E_D| + 4(r - |D(H)|)) + c \cdot r,$$

proving the claim. For example, if  $j$  is an index of Type 1, it gets a value at least  $3b + c$  and needs to give value 1, such a charging can be done if  $3b + c \geq 1$ . Similarly, a type 0 vertex does not need to give any value and has a surplus. We will choose parameters so that each



$j$  of types  $-1, 1$  or  $0$  can distribute the value to vertices added in  $V_j \setminus V_{j-1}$ . However, if  $j$  is an index of Type 2, it needs to distribute the value it gets ( $5b + c$ ) to two vertices, and we will allow it to be “compensated” by vertices of other types, which may have a surplus.

Consider an index  $j \in T_2$ . The fact that  $j \in T_2$  guarantees that earlier edges  $I^1, \dots, I^{j-1}$  are all vertex disjoint from  $I^j$ . If later edges  $I^{j+1}, \dots, I^r$  are all vertex disjoint from  $I^j$ , then  $|E_{j-1}| = 4$  and  $E_{j-1}$  is disjoint from  $\{E_{j'}\}_{j' \in D \setminus \{j-1\}}$ , and this means that  $H \notin \mathcal{H}$ . Thus, there exists  $j' > j$  such that  $I^{j'}$  and  $I^j$  share a vertex. Take the smallest  $j' > j$ , and say that  $j'$  compensates  $j$ . Note that  $j' \notin T_2$ .

We will allow a type 1 index to compensate at most one type 2 index, and a type -1 or 0 index to compensate at most two type 2 indices. We consider below the constraints implied by each kind of compensation.

1. *One Type 1 index  $j'$  compensates one Type 2 index  $j$*   
 $v_{j'} + v_j = 3$  and  $e_{j'} + e_j \geq 8$  (5 from  $e_j$  and 3 from  $e_{j'}$ ). This is possible if  $8b + 2c \geq 3$ .
2. *One Type 0 index  $j'$  compensates one Type 2 index  $j$*   
 $v_{j'} + v_j = 2$  and  $e_{j'} + e_j \geq 5$  (5 from  $e_j$ ). This is possible if  $5b + 2c \geq 2$ .
3. *One Type 0 index  $j'$  compensates two Type 2 indices  $j_1$  and  $j_2$  (say  $j_1 < j_2$ ).*

There are two cases.

- (a)  $e_{j'} + e_{j_1} + e_{j_2} \geq 11$ :  $v_{j'} + v_{j_1} + v_{j_2} = 4$ . This is possible if  $11b + 3c \geq 4$ .
- (b)  $e_{j'} + e_{j_1} + e_{j_2} = 10$ : since  $e_{j_1}, e_{j_2} \geq 5$ , this means that  $e_{j'} = 0$ .

First, we note that since  $j_1$  is a type 2 index and  $j'$  is the smallest index  $j$  such that  $I^{j_1} \cap I_j \neq \emptyset$ , in the graph  $H_{j'-1}$ , vertices in  $I^{j_1}$  only have edges to vertices in  $I^{j_1-1}$  and  $I^{j_1+1}$ . Similarly, vertices in  $I^{j_2}$  only have edges to vertices in  $I^{j_2-1}$  and  $I^{j_2+1}$ .

Since  $I^{j'}$  shares one vertex each with  $I^{j_1}$  and  $I^{j_2}$ , let  $I^{j'} = \{i_1^{j'}, i_2^{j'}\}$  with  $i_1^{j'} \in I^{j_1}$  and  $i_2^{j'} \in I^{j_2}$ . Since  $e_{j'} = 0$  means that  $I^{j'} = \{i_1^{j'}, i_2^{j'}\}$  was in  $H_{j'-1}$ . However, this is an edge between vertices in  $I^{j_1}$  and  $I^{j_2}$ . By the above argument, this is only possible if  $j_2 = j_1 + 1$ . Also, since  $j'$  is type 0 and  $I^{j'}$  shares a vertex with  $I^{j_2}$ , we must have  $j' > j_2 + 1$  (otherwise  $j'$  would be type -1).

Consider  $I^{j'-1}$ , which are vertex disjoint from both  $I^{j_1}$  and  $I^{j_2}$ . If  $I^{j'-1} \neq I^{j_1-1}$ , at least one edge between  $I^{j'-1}$  and  $I^{j'}$  was not in  $H_{j'-1}$ , contradicting the assumption  $e_{j'} = 0$ . Therefore,  $I^{j'-1} = I^{j_1-1}$ . For the same reason,  $I^{j'-1} = I^{j_2+1}$ . Thus, in particular, we have  $I^{j_2+1} = I^{j_1-1}$ . Thus,  $j_2 + 1$  is also a type 0 index. Moreover, it cannot compensate any previous index, since any such index would already be compensated by  $j_1 - 1$ .

In this case we consider that  $I^{j_2+1}$  and  $I^{j'}$  jointly compensate  $j_1$  and  $j_2$ .  $v_{j'} + v_{j_2+1} + v_{j_1} + v_{j_2} = 4$  and  $e_{j_2+1} + e_{j'} + e_{j_1} + e_{j_2} \geq 10$ . Compensation is possible if  $10b + 4c \geq 4$ .

4. *One Type  $-1$  index  $j'$  compensates one Type 2 index  $j$ .*  
 $v_{j'} + v_j \leq 3$  and  $e_{j'} + e_j \geq 5$  (5 from  $e_j$ ). Compensation is possible if  $5b + 4b + 2c \geq 2$ .
5. *One Type  $-1$  index  $j'$  compensates two Type 2 indices  $j_1$  and  $j_2$*

We have  $v_{j'} + v_{j_1} + v_{j_2} \leq 5$  and  $e_{j'} + e_{j_1} + e_{j_2} \geq 10$ . Compensation is possible if  $10b + 4b + 3c \geq 5$ .

Each index  $j$  of Type 2 is compensated by exactly one other index  $j'$ . We also require indices of types 1 and  $-1$  which do not compensate any other index, to have value at least 1 (to account for the one vertex added). This is true if  $3b + c \geq 1$  and  $4b + c \geq 1$ .

Aggregating the above conditions (and discarding the redundant ones), we take

$$c = \max \left\{ \frac{3}{2} - 4b, 1 - \frac{5b}{2}, \frac{4}{3} - \frac{11b}{3}, \frac{5}{3} - \frac{14b}{3} \right\}$$

It is easy to check that the maximum is attained by  $c = 5/3 - 14b/3$  when  $b \in [0, 1/4]$  and  $c = 3/2 - 4b$  when  $b \in [1/4, 1/3]$ .  $\blacksquare$

## 8. Lifting $\|\cdot\|_{sp}$ lower bounds to higher levels

For a matrix  $B \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ , let  $B^S$  denote the matrix obtained by symmetrizing  $B$ , i.e. for any  $I, J \in [n]^{q/2}$ ,

$$B^S[I, J] := \frac{1}{|\mathcal{O}(\alpha(I) + \alpha(J))|} \cdot \sum_{\substack{\alpha(I') + \alpha(J') \\ = \alpha(I) + \alpha(J)}} B[I', J']$$

Equivalently,  $B^S$  can be defined as follows:

$$B^S = \frac{1}{q!} \cdot \sum_{\pi \in \mathcal{S}_q} B^\pi$$

where for any  $K \in [n]^q$ ,  $B^\pi[K] := B[\pi(K)]$ .

For a matrix  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  let  $T \in \mathbb{R}^{[n]^4}$  denote the tensor given by,  $T[i_1, i_2, i_3, i_4] = M[(i_1, i_2), (i_3, i_4)]$ . Also for any non-negative integers  $x, y$  satisfying  $x + y = 4$ , let  $M_{x,y} \in \mathbb{R}^{[n]^x \times [n]^y}$  denote the matrix given by,  $M[(i_1, \dots, i_x), (j_1, \dots, j_y)] = T[i_1, \dots, i_x, j_1, \dots, j_y]$ . We will use the following result that we prove in Section 8.3.

**Theorem 8.1 (Lifting ‘‘Stable’’  $\|\cdot\|_{sp}$  Lower Bounds).** *Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4-SOS-symmetric matrix satisfying*

$$\|M\|_{S_1}, \|M_{3,1}\|_{S_1} \leq 1.$$

*Then for any  $q$  divisible by 4,*

$$\left\| \left( M^{\otimes q/4} \right)^S \right\|_{S_1} = 2^{O(q)}$$

### 8.1. Gap between $\|\cdot\|_{sp}$ and $\|\cdot\|_2$ for Non-Neg. Coefficient Polynomials

**Lemma 8.2.** *Consider any homogeneous polynomial  $g$  of even degree- $t$  and let  $M_g \in \mathbb{R}^{[n]^{t/2} \times [n]^{t/2}}$  be its SoS-symmetric matrix representation. Then  $\|g\|_{sp} \geq \|M_g\|_F^2 / \|M_g\|_{S_1}$ .*

*Proof.* We know by strong duality, that

$$\|g\|_{sp} = \max \left\{ \langle X, M_g \rangle \mid \|X\|_{S_1} = 1, X \text{ is SoS-Symmetric}, X \in \mathbb{R}^{[n]^{t/2} \times [n]^{t/2}} \right\}.$$

The claim follows by substituting  $X := M_g / \|M_g\|_{S_1}$ . ■

**Theorem 8.3.** *For any  $q$  divisible by 4 and  $f$  as defined in Section 7, we have that w.h.p.*

$$\frac{\|f^{q/4}\|_{sp}}{\|f^{q/4}\|_2} \geq \frac{n^{q/24}}{(q \log n)^{O(q)}}.$$

*Proof.* Let  $f$  be the degree-4 homogeneous polynomial as defined in Section 7 and let  $M = M_f$  be its SoS-symmetric matrix representation. Let  $g := f^{q/4}$  and let  $M_g$  be its SoS-symmetric matrix representation. Thus  $M_g = (M^{\otimes q/4})^S$  and it is easily verified that w.h.p.,  $\|M\|_F^2 \geq \tilde{\Omega}(n^4 p^6) = \tilde{O}(n^2)$  and also  $\|M_g\|_F^2 \geq \tilde{\Omega}((n^4 p^6)^{q/4} / q^{O(q)}) = \tilde{\Omega}(n^{q/2} / q^{O(q)})$ .

It remains to estimate  $\|M_g\|_{S_1}$  so that we may apply Lemma 8.2. Implicit, in the proof of Lemma 7.8, is that w.h.p.  $M$  has one eigenvalue of magnitude  $O(n^2 p) = O(n^{5/3})$  and at most  $O(n^2 p) = O(n^{5/3})$  eigenvalues of magnitude  $\tilde{O}(n^{3/2} p^4) = \tilde{O}(n^{1/6})$ . Thus  $\|M\|_{S_1} = \tilde{O}(n^{11/6})$  w.h.p. Now we have that  $\|M_{1,3}\|_{S_1} \leq \sqrt{n} \cdot \|M_{1,3}\|_F = \sqrt{n} \cdot \|M\|_F = \tilde{O}(n^{3/2})$  w.h.p. Thus on applying Theorem 8.1 to  $M/\tilde{O}(n^{11/6})$ , we get that  $\|M_g\|_{S_1} / \tilde{O}(n^{11q/24}) \leq 2^{O(q)}$  w.h.p.

Thus, applying Lemma 8.2 yields the claim. ■

## 8.2. Tetris Theorem

Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4 SoS-Symmetric matrix, let  $M_A := M_{3,1} \otimes M_{0,4} \otimes M_{3,1}$ , let  $M_B := M_{3,1} \otimes M_{1,3}$ , let  $M_C := M$  and let  $M_D := \text{Vec}(M) \text{Vec}(M)^T = M_{0,4} \otimes M_{4,0}$ . For any permutation  $\pi \in \mathfrak{S}_{q/2}$  let  $\bar{\pi} \in \mathfrak{S}_{n^{q/2}}$  denote the permutation that maps any  $i \in [n]^{q/2}$  to  $\pi(i)$ . Also let  $P_\pi \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  denote the row-permutation matrix induced by the permutation  $\bar{\pi}$ . Let  $P := \sum_{\pi \in \mathfrak{S}_{q/2}} P_\pi$ . Let  $\mathcal{R}(a, b, c, d) := (c!^2 2!^{2c}) (b! (2a+b)! 3!^{2a+2b}) (d! (a+d)! 4!^{a+2d})$ . Define

$$\begin{aligned} \mathfrak{M}(a, b, c, d) &:= \frac{M_A^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d}}{\mathcal{R}(a, b, c, d)} \\ \bar{\mathfrak{M}}(a, b, c, d) &:= \frac{(M_A^T)^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d}}{\mathcal{R}(a, b, c, d)} \\ S^{\mathfrak{M}} &:= \left\{ P \cdot \mathfrak{M}(a, b, c, d) \cdot P^T \mid 12a + 8b + 4c + 8d = q \right\} \cup \\ &\quad \left\{ P \cdot \bar{\mathfrak{M}}(a, b, c, d) \cdot P^T \mid 12a + 8b + 4c + 8d = q \right\}. \end{aligned}$$

**Theorem 8.4.** *Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4-SOS-symmetric matrix. Then*

$$(q/4)! \cdot 4!^{q/4} \cdot \sum_{\mathfrak{M} \in S^{\mathfrak{M}}} \mathfrak{M} = \sum_{\pi \in \mathfrak{S}_q} \left( M^{\otimes q/4} \right)^\pi = q! \cdot \left( M^{\otimes q/4} \right)^S \quad (8.1)$$

We shall prove this claim in Section 8.4 after first exploring its consequences.

### 8.3. Lifting Stable Degree-4 Lower Bounds

**Theorem 8.5 (Lifting “Stable”  $\|\cdot\|_{sp}$  Lower Bounds: Restatement of Theorem 8.1).**

Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4-SOS-symmetric matrix satisfying

$$\|M\|_{S_1}, \|M_{3,1}\|_{S_1} \leq 1.$$

Then for any  $q$  divisible by 4,

$$\left\| \left( M^{\otimes q/4} \right)^S \right\|_{S_1} = 2^{O(q)}$$

*Proof.* Implicit in the proof of Theorem 8.4 is the following:

$$\begin{aligned} & q! \cdot \left( M^{\otimes q/4} \right)^S \\ = & \sum_{12a+8b+4c+8d=q} \frac{(q/4)! \cdot 4!^{q/4}}{\mathcal{R}(a,b,c,d)} \sum_{\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}} \left( P_{\sigma_1}^T \left( M_A^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d} \right) P_{\sigma_2} \right) \\ + & \sum_{12a+8b+4c+8d=q} \frac{(q/4)! \cdot 4!^{q/4}}{\mathcal{R}(a,b,c,d)} \sum_{\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}} \left( P_{\sigma_1}^T \left( (M_A^T)^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d} \right) P_{\sigma_2} \right) \end{aligned} \quad (8.2)$$

First note that  $(q/4)! \cdot 4!^{q/4} / \mathcal{R}(a,b,c,d) \leq 2^{O(q)}$  since for any integers  $i, j, k, l$

$$(i+j+k+l)! / (i! \cdot j! \cdot k! \cdot l!) \leq 4^{i+j+k+l}.$$

Next note that  $\|M_{0,4}\|_{S_1} = \|M\|_F \leq \|M\|_{S_1} \leq 1$ . Combining this with the fact that  $\|M_{1,3}\|_{S_1}, \|M\|_{S_1} \leq 1$ , we get that  $\|M_A\|_{S_1}, \|M_B\|_{S_1}, \|M_C\|_{S_1}, \|M_D\|_{S_1} \leq 1$ , since  $\|X \otimes Y\|_{S_1} = \|X\|_{S_1} \cdot \|Y\|_{S_1}$  for any (possibly rectangular) matrices  $X$  and  $Y$ . Further note that Schatten 1-norm is invariant to multiplication by a permutation matrix. Thus the claim follows by applying triangle inequality to the  $O_q(q^q)$  terms in Eq. (8.2). ■

**Corollary 8.6 (lifting “stable”  $\Lambda(\cdot)$  lower bounds).** Let  $M \in \mathbb{R}^{[n]^2 \times [n]^2}$  be a degree-4-SOS-symmetric matrix satisfying

$$M \succeq 0, \quad M_A := M_{3,1} \otimes M_{0,4} \otimes M_{3,1} \succeq 0, \quad \text{and} \quad M_B := M_{3,1} \otimes M_{1,3} \succeq 0.$$

Then for any  $q$  divisible by 4,

$$\left( M^{\otimes q/4} \right)^S \succeq 0$$

(i.e.  $(M^{\otimes q/4})^S$  is a degree- $q$  SOS moment matrix).

*Proof.* Observe that Theorem 8.4 implies the claim since  $M_A, M_B, M_C$ , and  $M_D$  are PSD and the set of PSD matrices is closed under transpose, Kronecker product, scaling, conjugation, and addition. ■

### 8.4. Proof of Tetris Theorem

We start with defining a *hypergraphical matrix* which will allow a more intuitive paraphrasing of Theorem 8.4. By now, this is an important formalism in the context of SoS, and closely-related objects have been defined in several works, including [DM15], [RRS16], [BHK<sup>+</sup>16].

### 8.4.1. Hypergraphical Matrix

**Definition 8.7.** For symbolic sets  $L = \{\ell_1, \dots, \ell_{q_1}\}, R = \{r_1, \dots, r_{q_2}\}$ , a  $d$ -uniform **template-hypergraph** represented by  $(L, R, E)$ , is a  $d$ -uniform hypergraph on vertex set  $L \uplus R$  with  $E$  being the set of hyperedges.

For  $I = (i_1, \dots, i_{q_1}) \in [n]^{q_1}, J = (j_1, \dots, j_{q_2}) \in [n]^{q_2}$ , we also define a related object called **edge-set instantiation** (and denoted by  $E(I, J)$ ) as the set of size- $d$  multisets induced by  $E$  on substituting  $\ell_t = i_t$  and  $r_t = j_t$ .

**Remark.** There is a subtle distinction between  $E$  and  $E(I, J)$  above, in that  $E$  is a set of  $d$ -sets and  $E(I, J)$  is a set of size- $d$  multisets (i.e.  $e \in E(I, J)$  can have repeated elements).

**Definition 8.8.** Given an SoS-symmetric order- $d$  tensor  $\mathbb{T}$  and a  $d$ -uniform template-hypergraph  $H = (L, R, E)$  with  $|L| = q_1, |R| = q_2$ , we define the  $d$ -uniform degree- $(q_1, q_2)$  hypergraphical matrix  $M_{hyp}^{\mathbb{T}}(H)$  as

$$M_{hyp}^{\mathbb{T}}(H)[I, J] = \prod_{e \in E(I, J)} \mathbb{T}[e]$$

for any  $I \in [n]^{q_1}, J \in [n]^{q_2}$ .

In order to represent Theorem 8.4 in the language of hypergraphical matrices, we first show how to represent  $M^{\otimes q/4}$  and  $M_A^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d}$  in this language.

### 8.4.2. Kronecker Products of Hypergraphical Matrices

We begin with the observation that the kronecker product of hypergraphical matrices yields another hypergraphical matrix (corresponding to the "disjoint-union" of the template-hypergraphs).

**Definition 8.9.** let  $H = (L, R, E), H' = (L', R', E')$  be template-hypergraphs with  $|L| = q_1, |R| = q_2, |L'| = q_3, |R'| = q_4$ . Let  $\bar{H} = (\bar{L}, \bar{R}, \bar{E})$  be a template-hypergraph with  $|\bar{L}| = q_1 + q_3, |\bar{R}| = q_2 + q_4$ , where  $\bar{\ell}_t = \ell_t$  if  $t \in [q_1], \bar{\ell}_t = \ell'_t$  if  $t \in [q_1 + 1, q_1 + q_3], \bar{r}_t = r_t$  if  $t \in [q_2], \bar{r}_t = r'_t$  if  $t \in [q_2 + 1, q_2 + q_4]$ , and  $\bar{E} = E \uplus E'$ . We call  $\bar{H}$  the **disjoint-union** of  $H$  and  $H'$ , which we denote by  $H \uplus H'$ .

**Observation 8.10.** Let  $\mathbb{T}$  be an SOS-symmetric order- $d$  tensor and let  $H = (L, R, E), H' = (L', R', E')$  be template-hypergraphs. Then,

$$M_{hyp}^{\mathbb{T}}(H) \otimes M_{hyp}^{\mathbb{T}}(H') = M_{hyp}^{\mathbb{T}}(H \uplus H')$$

**Remark.** Note that the disjoint-union operation on template-hypergraphs does not commute, i.e.  $M_{hyp}^{\mathbb{T}}(H \uplus H') \neq M_{hyp}^{\mathbb{T}}(H' \uplus H)$  (since kronecker-product does not commute).

Now consider a degree-4 SoS-symmetric matrix  $M$  (as in the statement of Theorem 8.4) and let  $\mathbb{T}$  be the SoS-symmetric tensor corresponding to  $M$ . Then for any  $x + y = 4$  we have that  $M_{x,y} = M_{hyp}^{\mathbb{T}}(H_{x,y})$ , where  $H_{x,y} = (L, R, E)$  is the template-hypergraph satisfying  $L = \{\ell_1, \dots, \ell_x\}, R = \{r_1, \dots, r_y\}$  and  $E = \{\{\ell_1, \dots, \ell_x, r_1, \dots, r_y\}\}$ . Combining this observation with Observation 8.10 yields that  $M_A = M_{hyp}^{\mathbb{T}}(H_A), M_B = M_{hyp}^{\mathbb{T}}(H_B), M_C =$

$M_{\text{hyp}}^T(H_C)$ ,  $M_D = M_{\text{hyp}}^T(H_D)$ , where  $H_A := H_{3,1} \uplus H_{0,4} \uplus H_{3,1}$ ,  $H_B := H_{3,1} \uplus H_{1,3}$ ,  $H_C := H_{2,2}$ , and  $H_D := H_{4,0} \uplus H_{0,4}$ . Lastly, another application of Observation 8.10 to the above, yields

**Observation 8.11.** *For a template-hypergraph  $H$ , let  $H^{\uplus t}$  denote  $\biguplus_{g \in [t]} H$ . Then,*

$$(1) M^{\otimes q/4} = M_{\text{hyp}}^T(H_{2,2}^{\uplus q/4}).$$

$$(2) M_A^{\otimes a} \otimes M_B^{\otimes b} \otimes M_C^{\otimes c} \otimes M_D^{\otimes d} = M_{\text{hyp}}^T(H(a, b, c, d)) \text{ where}$$

$$H(a, b, c, d) := H_A^{\uplus a} \uplus H_B^{\uplus b} \uplus H_C^{\uplus c} \uplus H_D^{\uplus d}$$

For technical reasons we also define the following related template-hypergraph:

$$\bar{H}(a, b, c, d) := \bar{H}_A^{\uplus a} \uplus \bar{H}_B^{\uplus b} \uplus \bar{H}_C^{\uplus c} \uplus \bar{H}_D^{\uplus d},$$

where  $\bar{H}_A$  is the template-hypergraph whose corresponding hypergraphical matrix is  $M_A^T$ .

To finish paraphrasing Theorem 8.4, we are left with studying the effect of permutations on hypergraphical matrices - which is the content of the following section.

### 8.4.3. Hypergraphical Matrices under Permutation

Recall that for any matrix  $B \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  and  $\pi \in \mathbb{S}_q$ ,  $B^\pi$  is the matrix satisfying  $B^\pi[K] := B[\pi(K)]$  where  $K \in [n]^{q/2} \times [n]^{q/2}$ .

Also recall that for any permutation  $\sigma \in \mathbb{S}_{q/2}$ ,  $\bar{\sigma} \in \mathbb{S}_{n^{q/2}}$  denotes the permutation that maps any  $i \in [n]^{q/2}$  to  $\sigma(i)$  and also that  $P_\sigma \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$  denotes the  $[n]^{q/2} \times [n]^{q/2}$  row-permutation matrix induced by the permutation  $\bar{\sigma}$ .

We next define a permuted template hypergraph in order to capture how permuting a hypergraphical matrix (in the senses above) can be seen as permutations of the vertex set of the hypergraph.

**Definition 8.12 (Permuted Template-Hypergraph).** *For any  $\pi \in \mathbb{S}_q$  (even  $q$ ), and a  $d$ -uniform template-hypergraph  $H = (L, R, E)$  with  $|L| = |R| = q/2$ , let  $H^\pi = (L', R', E)$  denote the template-hypergraph obtained by setting  $\ell'_t := k_t$  and  $r'_t = k_{t+q/2}$  for  $t \in [q/2]$ , where  $K = (k_1, \dots, k_q) = \pi(L \oplus R)$ .*

*Similarly for any  $\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}$ , let  $H^{\sigma_1, \sigma_2} = (L', R', E)$  denote the template-hypergraph obtained by setting  $\ell'_t := \sigma_1(\ell_t)$  and  $r'_t = \sigma_2(r_t)$ .*

We then straightforwardly obtain

**Observation 8.13.** *For any  $\pi \in \mathbb{S}_q$ ,  $\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}$ , SoS-symmetric order- $d$  tensor  $\mathbb{T}$  and any  $d$ -uniform template-hypergraph  $H$ ,*

$$(1) \left( M_{\text{hyp}}^T(H) \right)^\pi = M_{\text{hyp}}^T(H^\pi)$$

$$(2) P_{\sigma_1} \cdot M_{\text{hyp}}^T(H) \cdot P_{\sigma_2}^T = M_{\text{hyp}}^T(H^{\sigma_1, \sigma_2})$$

Thus to prove Theorem 8.4, it remains to establish

$$\begin{aligned}
& \frac{1}{4!^{q/4} \cdot (q/4)!} \cdot \sum_{\pi \in \mathbb{S}_q} \mathbf{M}_{\text{hyp}}^{\top}((H_{2,2}^{\uplus q/4})^{\pi}) \\
= & \sum_{12a+8b+4c+8d=q} \frac{1}{\mathcal{R}(a,b,c,d)} \cdot \sum_{\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}} \mathbf{M}_{\text{hyp}}^{\top}(H(a,b,c,d)^{\sigma_1, \sigma_2}) + \\
& \sum_{12a+8b+4c+8d=q} \frac{1}{\mathcal{R}(a,b,c,d)} \cdot \sum_{\sigma_1, \sigma_2 \in \mathbb{S}_{q/2}} \mathbf{M}_{\text{hyp}}^{\top}(\bar{H}(a,b,c,d)^{\sigma_1, \sigma_2}) \tag{8.3}
\end{aligned}$$

We will establish this in the next section by comparing the template-hypergraphs generated (with multiplicities) in the LHS with those generated in the RHS.

#### 8.4.4. Proof of Eq. (8.3)

We start with some definitions to track the template-hypergraphs generated in the LHS and RHS of Eq. (8.3). For any  $12a + 8b + 4c + 8d = q$ , let

$$\begin{aligned}
\mathcal{F}(a,b,c,d) &:= \left\{ H(a,b,c,d)^{\sigma_1, \sigma_2} \mid \sigma_1, \sigma_2 \in \mathbb{S}_{q/2} \right\} \\
\bar{\mathcal{F}}(a,b,c,d) &:= \left\{ \bar{H}(a,b,c,d)^{\sigma_1, \sigma_2} \mid \sigma_1, \sigma_2 \in \mathbb{S}_{q/2} \right\} \\
\mathcal{F} &:= \left\{ (H_{2,2}^{\uplus q/4})^{\pi} \mid \pi \in \mathbb{S}_q \right\}
\end{aligned}$$

Firstly, it is easily verified that whenever  $(a,b,c,d) \neq (a',b',c',d')$ ,  $\mathcal{F}(a,b,c,d) \cap \mathcal{F}(a',b',c',d') = \emptyset$ , and that  $\mathcal{F}(a,b,c,d) \cap \bar{\mathcal{F}}(a,b,c,d) = \emptyset$ . It is also easily verified that for any  $12a + 8b + 4c + 8d = q$ , and any  $H \in \mathcal{F}(a,b,c,d)$ ,

$$\mathcal{R}(a,b,c,d) = \left| \left\{ (\sigma_1, \sigma_2) \in \mathbb{S}_{q/2}^2 \mid H(a,b,c,d)^{\sigma_1, \sigma_2} = H \right\} \right|$$

and for any  $H \in \mathcal{F}$ ,

$$4!^{q/4} \cdot (q/4)! = \left| \left\{ \pi \in \mathbb{S}_q \mid (H_{2,2}^{\uplus q/4})^{\pi} = H \right\} \right|.$$

Thus in order to prove Eq. (8.3), it is sufficient to establish that

$$\mathcal{F} = \bigsqcup_{12a+8b+4c+8d=q} (\mathcal{F}(a,b,c,d) \uplus \bar{\mathcal{F}}(a,b,c,d)) \tag{8.4}$$

It is sufficient to establish that

$$\mathcal{F} \subseteq \bigsqcup_{12a+8b+4c+8d=q} (\mathcal{F}(a,b,c,d) \uplus \bar{\mathcal{F}}(a,b,c,d)) \tag{8.5}$$

since the other direction is straightforward. To this end, consider any  $H = (L, R, E) \in \mathcal{F}$ , and for any  $x + y = 4$ , define

$$s_{x,y} := \left| \left\{ e \in E \mid |e \cap L| = x, |e \cap R| = y \right\} \right|.$$

Now clearly  $H \in \mathcal{F}(a, b, c, d)$  iff

$$s_{0,4} = a + d, s_{3,1} = 2a + b, s_{1,3} = b, s_{2,2} = c, s_{4,0} = d. \quad (8.6)$$

and  $H \in \overline{\mathcal{F}}(a, b, c, d)$  iff

$$s_{4,0} = a + d, s_{1,3} = 2a + b, s_{3,1} = b, s_{2,2} = c, s_{0,4} = d. \quad (8.7)$$

Thus we need only find  $12a' + 8b' + 4c' + 8d' = q$ , such that Eq. (8.6) or Eq. (8.7) is satisfied.

We will assume w.l.o.g. that  $s_{0,4} \geq s_{4,0}$  and show that one can satisfy Eq. (8.6), since if  $s_{(0,4)} < s_{(4,0)}$ , an identical argument allows one to show that Eq. (8.7) is satisfiable. So let  $d' = s_{4,0}$ ,  $c' = s_{2,2}$ ,  $b' = s_{1,3}$  and  $a' = (s_{3,1} - s_{1,3})/2$ . Since  $H \in \mathcal{F}$ , it must be true that  $4s_{4,0} + 3s_{3,1} + 2s_{2,2} + s_{(1,3)} = q/2$ . Thus,  $12a' + 8b' + 4c' + 8d' = 8s_{4,0} + 6s_{3,1} + 4s_{2,2} + 2s_{1,3} = q$  as desired. We will next see that  $(a', b', c', d')$  and  $s_{x,y}$  satisfy Eq. (8.6). We have by construction that  $s_{4,0} = d'$ ,  $s_{2,2} = c'$ ,  $s_{1,3} = b'$  and  $s_{3,1} = 2a' + b'$ . It remains to show that  $s_{0,4} = a' + d'$ . Now we know that  $4s_{4,0} + 3s_{3,1} + 2s_{2,2} + s_{1,3} = q/2$  and  $4s_{0,4} + 3s_{1,3} + 2s_{2,2} + s_{3,1} = q/2$ . Subtracting the two equations yields  $s_{0,4} - s_{4,0} = (s_{3,1} - s_{1,3})/2$ . This implies  $a' + d' = s_{4,0} + (s_{3,1} - s_{1,3})/2 = s_{0,4}$ , and furthermore it implies that  $a'$  is non-negative since we assumed  $s_{0,4} \geq s_{4,0}$ . So Eq. (8.6) is satisfied. Thus we have established Eq. (8.5), which completes the proof of Theorem 8.4.

## 9. Open problems

Our work makes progress on polynomial optimization based on new spectral techniques for dealing with higher order matrix representations of polynomials. Several interesting questions in the subject remain open, and below we collect some of the salient ones brought to the fore by our work.

1. What is the largest possible ratio between  $\Lambda(f)$  and  $\|f\|_2$  for arbitrary homogeneous polynomials of degree  $d$ ? Recall that we have an upper bound of  $O_d(n^{d/2-1})$  and a lower bound of  $\Omega_d(n^{d/4-1/2})$ , and closing this quadratic gap between these bounds is an interesting challenge. Even a lower bound for  $\|\cdot\|_{sp}$  that improves upon the current  $\Omega_d(n^{d/4-1/2})$  bound by polynomial factors would be very interesting.
2. A similar goal to pursue would be closing the gap between upper and lower bounds for polynomials with non-negative coefficients.
3. We discussed two relaxations of  $\|h\|_2 - \Lambda(h)$  which minimizes the maximum eigenvalue  $\lambda_{\max}(M_h)$  over matrix representations  $M_h$  of  $h$ , and  $\|h\|_{sp}$  which minimizes the spectral norm  $\|M_h\|_2$ . How far apart, if at all, can these quantities be for arbitrary polynomials  $h$ ?
4. Are there low-degree SoS lower bounds that satisfy the stability conditions of Corollary 8.6 (for the arbitrary polynomial case or other special classes)?
5. We studied three classes of polynomials: arbitrary, those with non-negative coefficients, and sparse. Are there other natural classes of polynomials for which we can give improved SoS-based (or other) approximation algorithms? Can our techniques be used in sub-exponential algorithms for special classes?



6. Despite being such a natural problem for which known algorithms give weak polynomially large approximation factors, the known NP-hardness results for polynomial optimization over the unit sphere only rule out an FPTAS. Can one obtain NP-hardness results for bigger approximation factors?

## References

- [BBH<sup>+</sup>12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 1, 2
- [BGK<sup>+</sup>01] Andreas Brieden, Peter Gritzmann, Ravindran Kannan, Victor Klee, László Lovász, and Miklós Simonovits. Deterministic and randomized polynomial-time approximation of radii. *Mathematika*, 48(1-2):63–105, 2001. 6, 57, 58
- [BGL16] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Certifying random polynomials over the unit sphere via sum of squares hierarchy. *arXiv preprint arXiv:1605.00903*, 2016. 2, 5, 7, 8, 15
- [BH13] Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 861–870. ACM, 2013. 1, 2
- [BH15] Fernando GSL Brandao and Aram W Harrow. Estimating operator norms using covering nets. *arXiv preprint arXiv:1509.05065*, 2015. 7, 8
- [BHK<sup>+</sup>16] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Proceedings of the 57th annual IEEE symposium on Foundations of Computer Science, FOCS '16*, pages 428–437, 2016. 6, 17, 49
- [BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 31–40. ACM, 2014. 1, 2, 7, 18
- [BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 143–151. ACM, 2015. 1
- [BKS17] Boaz Barak, Pravesh Kothari, and David Steurer. Quantum entanglement, sum of squares, and the log rank conjecture. *arXiv preprint arXiv:1701.06321*, 2017. 2
- [BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014. 18

- [BV09] S Charles Brubaker and Santosh S Vempala. Random tensors and planted cliques. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 406–419. Springer, 2009. 1
- [Cha12] Sourav Chatterjee. The missing log in large deviations for triangle counts. *Random Structures & Algorithms*, 40(4):437–451, 2012. 16
- [DK08] Etienne De Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central European Journal of Operations Research*, 16(2):111–125, 2008. 1, 5
- [DK12a] Bobby DeMarco and Jeff Kahn. Upper tails for triangles. *Random Structures & Algorithms*, 40(4):452–459, 2012. 16
- [DK12b] Robert DeMarco and Jeff Kahn. Tight upper tail bounds for cliques. *Random Structures & Algorithms*, 41(4):469–487, 2012. 16
- [dKLP06] Etienne de Klerk, Monique Laurent, and Pablo A Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, 361(2):210–225, 2006. 1, 11
- [dKLS14] Etienne de Klerk, Monique Laurent, and Zhao Sun. Convergence analysis for Lasserre’s measure-based hierarchy of upper bounds for polynomial optimization. *arXiv preprint arXiv:1411.6867*, 2014. 2, 5
- [dKLS15] Etienne de Klerk, Monique Laurent, and Zhao Sun. An alternative proof of a PTAS for fixed-degree polynomial optimization over the simplex. *Mathematical Programming*, 151(2):433–457, 2015. 11
- [DIPG12] Victor De la Pena and Evarist Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012. 4
- [DM15] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *COLT*, pages 523–562, 2015. 49
- [DW12] Andrew C Doherty and Stephanie Wehner. Convergence of sdp hierarchies for polynomial optimization on the hypersphere. *arXiv preprint arXiv:1210.5048*, 2012. 2, 5, 6
- [FK08] Alan Frieze and Ravi Kannan. A new approach to the planted clique problem. In *LIPICs-Leibniz International Proceedings in Informatics*, volume 2. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008. 1
- [FV<sup>+</sup>62] David G Feingold, Richard S Varga, et al. Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem. *Pacific J. Math*, 12(4):1241–1250, 1962. 31
- [GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 829, 2015. 1

- [Hås96] Johan Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 627–636. IEEE, 1996. [1](#)
- [HKP<sup>+</sup>17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. *Personal communication*, 2017. [5](#), [17](#)
- [HLZ10] Simai He, Zhening Li, and Shuzhong Zhang. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Mathematical Programming*, 125(2):353–383, 2010. [1](#), [2](#), [6](#), [11](#), [23](#), [58](#)
- [HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Proceedings of The 28th Conference on Learning Theory*, pages 956–1006, 2015. [1](#)
- [HV04] Johan Håstad and Srinivasan Venkatesh. On the advantage over a random assignment. *Random Structures & Algorithms*, 25(2):117–149, 2004. [1](#)
- [JOR04] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142(1):61–92, 2004. [16](#)
- [KMOW17] Pravesh K Kothari, Ryuhei Mori, Ryan O’Donnell, and David Witmer. Sum of squares lower bounds for refuting any csp. *arXiv preprint arXiv:1701.04521*, 2017. [17](#)
- [KN08] Subhash Khot and Assaf Naor. Linear equations modulo 2 and the  $L_1$  diameter of convex bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008. [2](#), [6](#), [7](#), [11](#), [57](#), [58](#)
- [KV04] Jeong Han Kim and Van H Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures & Algorithms*, 24(2):166–174, 2004. [16](#)
- [Las09] Jean Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2009. [1](#), [2](#), [6](#)
- [Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009. [1](#), [5](#), [6](#), [18](#)
- [LZ16] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 2016. [16](#)
- [MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014. [1](#)
- [Nes03] Yurii Nesterov. Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, UCL, 2003. [1](#)
- [OZ13] Ryan O’Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. Society for Industrial and Applied Mathematics, 2013. [18](#)

- [RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. *arXiv preprint arXiv:1605.00058*, 2016. [2](#), [8](#), [49](#)
- [So11] Anthony Man-Cho So. Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems. *Mathematical programming*, 129(2):357–382, 2011. [1](#), [2](#), [6](#), [7](#), [11](#), [57](#), [58](#)
- [Vu01] Van H Vu. A large deviation result on the number of small subgraphs of a random graph. *Combinatorics, Probability & Computing*, 10(1):79–94, 2001. [16](#)

## A. Oracle Lower Bound

Khot and Naor [KN08] observed that the problem of maximizing a polynomial over unit sphere can be reduced to computing diameter of centrally symmetric convex body. This observation was also used by So [So11] later. We recall the reduction here: For a convex set  $K$ , let  $K^\circ$  denote the polar of  $K$ , i.e.,  $K^\circ = \{y : \forall x \in K \langle x, y \rangle \leq 1\}$ . For a degree-3 polynomial  $P(x, y, z)$  on  $3n$  variables, let  $\|x\|_P = \|P(x, \cdot, \cdot)\|_{sp}$  where  $P(x, \cdot, \cdot)$  is a degree-2 restriction of  $P$  with  $x$  variables set. Let  $\mathbb{B}_P = \{x : \|x\|_P \leq 1\}$ . From the definition of polar and  $\|\cdot\|_{sp}$ , we have:

$$\begin{aligned} \max_{\|x\|_2, \|y\|_2, \|z\|_2 \leq 1} P(x, y, z) &= \max_{x \in \mathbb{B}_2} \|x\|_P \\ &= \max_{x \in \mathbb{B}_P^\circ} \|x\|_2 \end{aligned}$$

For general convex bodies, a lower bound for number of queries with “weak separation oracle” for approximating the diameter of the convex body was proved by Brieden et al. [BGK<sup>+</sup>01] and later improved by Khot and Naor [KN08]. We recall the definition:

**Definition A.1.** For a given a convex body  $P$ , a weak separation oracle  $A$  is an algorithm which on input  $(x, \varepsilon)$  behaves as following:

- If  $x \in A + \varepsilon \mathbb{B}_2$ ,  $A$  accepts it.
- Else  $A$  outputs a vector  $c \in \mathbb{Q}^n$  with  $\|c\|_\infty = 1$  such that for all  $y$  such that  $y + \varepsilon \mathbb{B}_2 \subset P$  we have  $c^T x + \varepsilon \geq c^T y$ .

Let  $K_{s,v}$  be the convex set  $K_{s,v}^{(n)} = \text{conv}(\mathbb{B}_2^n \cup \{sv, -sv\})$ , for unit vector  $v$ . Brieden et al. [BGK<sup>+</sup>01] proved the following theorem:

**Theorem A.2.** Let  $A$  be a randomized algorithm, for every convex set  $P$ , with access to a weak separation oracle for  $P$ . Let  $\mathcal{K}(n, s) = \{K_{s,u}^{(n)}\}_{u \in \mathbb{S}_2^{n-1}} \cup \{\mathbb{B}_2^n\}$ . If for every  $K \in \mathcal{K}(n, s)$  and  $s = \frac{\sqrt{n}}{\lambda}$ , we have:

$$\Pr \left[ A(K) \leq \text{diam}(K) \leq \frac{\sqrt{n}}{\lambda} A(K) \right] \geq \frac{3}{4}$$

where  $\text{diam}(K)$  is the diameter of  $K$ , then  $A$  must use at least  $\mathcal{O}(\lambda 2^{\lambda^2/2})$  oracle queries for  $\lambda \in [\sqrt{2}, \sqrt{n/2}]$ .

Using  $\lambda = \log n$ , we get that to get  $s = \frac{\sqrt{n}}{\log n}$  approximation to diameter,  $A$  must use super-polynomial number of queries to the weak separation oracle. We note that this was later improved to give analogous lower bound on the number of queries for an approximation factor  $\sqrt{\frac{n}{\log n}}$  by Khot and Naor [KN08].

Below, we show that the family of hard convex bodies considered by Brieden et al. [BGK<sup>+</sup>01] can be realized as  $\{\mathbb{B}_P^\circ\}_{P \in \mathcal{P}}$  by a family of polynomials  $\mathcal{P}$  – which, in turn, establishes a lower bound of  $\Omega\left(\frac{\sqrt{n}}{\log n}\right)$  on the approximation for polynomial optimization, achievable using this approach, for the case of  $d = 3$ . For an unit vector  $u \in \mathbb{S}_2^{n-1}$ , let  $P_u$  be the polynomial defined as:

$$P_u(x, y, z) = \sum_{i=1}^n x_i y_i z_i + s \cdot (u^T x) y_n z_n.$$

A matrix representation of  $P_u(x, \cdot, \cdot)$ , with rows indexed by  $y$  and columns indexed by  $z$  variables is as follows:

$$A_u = \begin{pmatrix} x_1 & 0 & \dots & 0 & 0 \\ x_2 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n-1} & 0 & \dots & 0 & 0 \\ x_n & 0 & \dots & 0 & s \cdot (u^T x) \end{pmatrix} \text{ and so, } A_u^T A_u = \begin{pmatrix} \|x\|_2^2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & s^2 \cdot |u^T x|^2 \end{pmatrix}.$$

This proves:  $\|x\|_{P_u} = \|P_u(x, \cdot, \cdot)\|_{sp} = \|A_u\|_{sp} = \max\{\|x\|_2, s |u^T x|\}$ .

Let  $B = \{x : \|x\|_2 \leq 1\}$  and  $C_u = \{x : s \cdot |u^T x| \leq 1\}$ . We note that,  $B^\circ = \{y \in \mathbb{R}^n : \|y\|_2 \leq 1\}$  and,  $C_u^\circ = \{\lambda \cdot u : \lambda \in [-s, s]\} = \text{conv}(\{-s \cdot u, s \cdot u\})$ .

Next, we observe:  $\mathbb{B}_{P_u} = B \cap C_u$ . It follows from De Morgan's law of polars that:  $\mathbb{B}_{P_u}^\circ = (B \cap C_u)^\circ = \text{conv}(B^\circ \cup C_u^\circ) = \text{conv}(\mathbb{B}_2^n \cup \{-s \cdot u, s \cdot u\}) = K_{s,u}^{(n)}$ . Finally, we observe that for the polynomial  $P_0 = \sum_{i=1}^n x_i y_i z_i$ , we have:  $\mathbb{B}_{P_0} = \mathbb{B}_2^n$ .

Hence for polynomial  $Q \in \mathcal{P} = \{P_u\}_{u \in \mathbb{S}_2^{n-1}} \cup \{P_0\}$ , no randomized polynomial can approximate  $\text{diam } \mathbb{B}_Q$  within factor  $\frac{\sqrt{n}}{q}$  without using more than  $2^{\Omega(q)}$  number of queries. Since the algorithm of Khot and Naor [KN08] reduces the problem of optimizing polynomial  $Q$  to computing  $\text{diam}(\mathbb{B}_Q)$ ,  $\mathcal{P}$  shows that their analysis is almost tight.

## B. Maximizing $|f(x)|$ vs. $f(x)$

Let  $f_{\max}$  denote  $\sup_{\|x\|=1} f(x)$ . Note that for polynomials with odd-degree, we have  $\|f\|_2 = f_{\max}$ . When the degree is even, a multiplicative approximation for  $f_{\max}$  is not possible since  $f_{\max}$  may be 0 or even negative. Moreover, even when  $f_{\max}$  is positive, any constructive multiplicative approximation of  $f_{\max}$  with a factor (say)  $B$ , can be turned into a  $1 + \varepsilon$  approximation by considering  $f' = f - C \cdot \|x\|_2^d$ , for  $C = (1 - \varepsilon) \cdot f_{\max}$  (one can use binary search on the values of  $C$  and use the solution give by the constructive algorithm to check).

An alternate notion considered in the literature [HLZ10, So11] is that of relative approximation where one bounds the ratio  $(\Lambda - f_{\min}) / (f_{\max} - f_{\min})$  (known as a relative approximation), where  $\Lambda$  is the estimate by the algorithm, and  $f_{\min}$  is defined analogously

to  $f_{\max}$ . While this notion is stronger than approximating  $\|f\|_2$  in some cases, one can use a shift of  $f$  as in the example above (by  $C \cdot f_{\min}$ ) to obtain a relative approximation unless  $|f_{\max} - f_{\min}| / |f_{\min}| = n^{-\omega(1)}$ .