58th Annual IEEE Symposium on Foundations of Computer Science

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

Vijay Bhattiprolu*, Venkatesan Guruswami*, Euiwoong Lee†
*Computer Science Department*
*Carnegie Mellon University*
*Pittsburgh, USA*
*vpb@cs.cmu.edu, guruswami@cmu.edu, euiwoonl@cs.cmu.edu*
* *Supported by NSF CCF-1422045 and CCF-1526092.*
† *Supported by a Samsung Fellowship, Simons award, and NSF CCF-1526092.*

Mrinalkanti Ghosh ‡, Madhur Tulsiani‡
*Toyota Technological Institute*
*Chicago, USA*
*mkghosh@ttic.edu, madhurt@ttic.edu*
‡ *Supported by NSF CCF-1254044.*

***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 R^n that maximizes abs(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(n) / q)^(d/2-1) for arbitrary polynomials, (O(n) / q)^(d/4-1/2) for polynomials with non-negative coefficients, and (m / q)^(1/2) 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(n)^(d/4-1/2). For polynomials with non-negative coefficients, we prove an Omega(n^(1/6) / polylogs) 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).**

## I. 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

$$\|f\|_2 \;:=\; \sup_{\|x\|=1} |f(x)| \qquad \text{(I.1)}$$

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 [1], [2], [3].

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 [4], [5], 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 $2 \to 4$ 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) [6], [5]. 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 [7], [8], [9], [10]. Frieze and Kannan [11] (see also [12]) also established a connection between the problem of approximating the spectral norm of a tensor (or equivalently, computing $\|f\|_2$

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.97

for a polynomial $f$), and finding planted cliques in random graphs.

The problem of polynomial optimization has been studied [1] over various compact sets [3], [14], and is natural to ask how well polynomial time algorithms can *approximate* the optimum value over a given compact set (see [14] for a survey). While the maximum of a degree-$d$ polynomial over the simplex admits a PTAS for every fixed $d$ [15], the problem of optimizing even a degree 3 polynomial over the hypercube does not admit any approximation better than $2^{(\log n)^{1-\varepsilon}}$ (for arbitrary $\varepsilon > 0$) assuming NP cannot be solved in time $2^{(\log n)^{O(1)}}$ [16].

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 [1], [17]. On the hardness side, Nesterov [18] gave a reduction from Maximum Independent Set to optimizing a homogeneous cubic polynomial over $\mathbb{S}^{n-1}$. 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 [19], this rules out an FPTAS for optimization over the unit sphere. Assuming the Exponential Time Hypothesis, Barak et al. [6] proved that computing $2 \to 4$ 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 $\epsilon > 0$.

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 [3] for various applications). We refer the reader to the recent work of Doherty and Wehner [20] and de Klerk, Laurent, and Sun [21] 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 [1], [17], approximation guarantees have been proved for several special cases including $2 \to q$ norms [6], polynomials with non-negative coefficients [5], some polynomials that arise in quantum information theory [22], [4], and random polynomials [23], [24]. 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 [17] 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 [25] (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 I-B, but first turn to stating our approximation guarantees.

### A. 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 $(x^{\otimes q/2})^T \cdot M_h \cdot x^{\otimes q/2} = h(x) \;\; \forall x \in \mathbb{R}^n$. Next we define the quantity,

$$\Lambda(h) \;:=\; \inf \left\{ \sup_{\|z\|_2=1} z^T M_h\, z \;\middle|\; M \text{ is a representation of } h \right\}. \tag{I.2}$$

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 $(h^2)_{\max} = \left\|h^2\right\|_2$. More generally, for a degree-$d$ homogeneous polynomial $f$ and an integer $q$ divisible by $2d$, we have the upper estimate

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

The following result shows that $\Lambda\left(f^{q/d}\right)^{d/q}$ approximates $\|f\|_2$ within polynomial factors, and also gives an algorithm to approximate $\|f\|_2$ with respect to the upper bound

[1] 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 the full version [13].

$\Lambda\left(f^{q/d}\right)^{d/q}$. 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 I.1.** *Let $f$ be an $n$-variate homogeneous polynomial of degree-$d$, and let $q \leq n$ be an integer divisible by $2d$. Then,*

*Arbitrary $f$:*

$$\left(\Lambda\left(f^{q/d}\right)\right)^{d/q} \leq O_d\left((n/q)^{d/2-1}\right) \cdot \|f\|_2$$

*$f$ with Non-neg. Coefficients:*

$$\left(\Lambda_C\left(f^{q/d}\right)\right)^{d/q} \leq O_d\left((n/q)^{d/4-1/2}\right) \cdot \|f\|_2$$

*$f$ with Sparsity $m$:*

$$\left(\Lambda\left(f^{q/d}\right)\right)^{d/q} \leq O_d\left(\sqrt{m/q}\right) \cdot \|f\|_2 .$$

*(where $\Lambda_C(\cdot)$ is a a related efficiently computable quantity)*

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

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

*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 I.2.** *Interestingly, our deterministic algorithms only involve computing the maximum eigenvectors of $n^{O(q)}$ different matrices in $\mathbb{R}^{n\times n}$, and actually don't require computing $\Lambda\left(f^{q/d}\right)^{d/q}$ (even though this quantity can also be computed in $n^{O(q)}$ time by the sum-of-squares SDP; see Section I-B). The quantity $\Lambda\left(f^{q/d}\right)^{d/q}$ is only used in the analysis.*

**Remark I.3.** *If $m = n^{\rho\cdot d}$ 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 I.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 I.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 $\Lambda\left(f^{q/d}\right)$, 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 I.5 (Informal).** *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*

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

*and $m = q^{O(q)}$.*

By combining Theorem I.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, which also have variety of applications in functional analysis and probability [26]. However, such methods increase the number of variables to $nq$, which would completely nullify any advantage obtained from the increased degrees of freedom. This is because the approximation obtained would be of the form $(\#vars/q)^{d/2-1} = n^{d/2-1}$.

Our proof of Theorem I.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 III-A4 and [13]). 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.

### *B. 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 $\mathsf{Tr}\big(M_h^T X\big)$ over all positive semidefinite matrices $X \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ with $\mathsf{Tr}(X) = 1$. (This is a relaxation because $X = x^{\otimes q/2}(x^{\otimes q/2})^T$ is psd with $\mathsf{Tr}(X) = \|x\|_2^q$.)

It is well known (see for instance [2]) that the quantity $\Lambda(h)$ from (I.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 the full version [13] for more detailed SoS preliminaries.) In light of this, our results from Theorem I.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 [20], [21] (see Section I-C for a comparison of results).

**SoS Lower Bounds.** While the approximation factors in our upper bounds of Theorem I.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 [24] 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. [27] 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 $\Lambda\big(f^{q/d}\big)^{d/q}$ and $\|f\|_2$ for the case of *arbitrary* polynomials. Note that this lower bound is roughly square root of our upper bound from Theorem I.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 out knowledge, the only previous lower bound for this problem, was known through Nesterov's reduction [14], 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 I.6.** *There exists an $n$ variate degree-4 homogeneous polynomial $f$ with non-negative coefficients such that*

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

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 (I.2) by the *spectral norm* $\|M_h\|_2$. (See [13] 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 I.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 I.7.** *Let $g := f^{q/4}$ where $f$ is the degree-4 polynomial as in Theorem I.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 $X$ (see full version [13]) — this inspired us to name our tool "Tetris theorem." Recently, the insightful pseudo-calibration approach [28] 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.

### C. 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 [2], [3] for a survey.

**Arbitrary Polynomials.** For general homogeneous polynomials of degree-$d$, a polytime $O_d\left(n^{d/2-1}\right)$-approximation was given by He et al. [1], which was improved to $O_d\left((n/\log n)^{d/2-1}\right)$ by So [17]. The convergence of SDP hierarchies for polynomial optimization was analyzed by Doherty and Wehner [20]. However, their result only applies to relaxations given by $\Omega(n)$ levels of the SoS hierarchy (Theorem 7.1 in [20]). Thus, our results can be seen an giving an interpolation between the polynomial time algorithms obtained by [1], [17] and the exponential time algorithms given by $\Omega(n)$ levels of SoS, although the bounds obtained by [20] 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. [29] used by So [17]. However, to the best of our knowledge, this is not published. In particular, So uses the techniques of Khot and Naor [25] 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 [29], 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 $O_d\big((n/q)^{d/2-1}\big)$ approximation in time $2^q$. We note here that the method of Khot and Naor [25] cannot be improved further (up to polylog) for the case $d = 3$ (see full version [13]). 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 [25], [17] 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 [5] 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 $\|f\|_{sp}$.[2] 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 $\widetilde{\Omega}_d((n/d)^{d/4-1/2})$ (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 [24] 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 [5] that for any matrix $A$, $q$ levels of the SoS hierarchy approximates $\|A\|_{2\to4}^4 = \big\|\|Ax\|_4^4\big\|_2$ (i.e. the fourth power of the 2-to-4-norm) within a factor of

$$1 + \left(\frac{\log n}{q}\right)^{1/2} \cdot \frac{\|A\|_{2\to2}^2\|A\|_{2\to\infty}^2}{\|A\|_{2\to4}^4}.$$

Brandao and Harrow [30] 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 $\|A\|_{2\to2}^2\|A\|_{2\to\infty}^2$ and $\|A\|_{2\to4}^4$ 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, $\|A\|_{2\to2}^2 = \Theta(m+n)$, $\|A\|_{2\to\infty}^2 = \Theta(n)$, and $\|A\|_{2\to4}^2 = \Theta(m+n^2)$ 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 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 [30] already implies a $\sqrt{m/q}$-approximation for all $m$, and our result implies a $\sqrt{m}/q$-approximation when $m = \Omega(n^2)$).

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

**Random Polynomials.** For the case when $f$ is a degree-$d$ homogeneous polynomial with i.i.d. random $\pm 1$ coefficients [24], [23] showed that degree-$q$ SoS certifies an upper bound on $\|f\|_2$ that is with high probability at most $\widetilde{O}((n/q)^{d/4-1/2}) \cdot \|f\|_2$. 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 [23] motivated by applications to refuting random CSPs.

## II. 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 $x_1, \ldots, x_n$) 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

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

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 $(\alpha_1^r, \ldots, \alpha_n^r)$ 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 II.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*

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

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

- *We refer to the polynomials $\overline{f}_\alpha$ 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 $\mathsf{U}(f)$.*
- *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 $I = (i_1, \ldots, i_k)$ 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 $\left(x^{\otimes k}\right)\left(x^{\otimes k}\right)^T$ 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 II.2 (SoS-Symmetry).** *A matrix $\mathsf{M}$ which satisfies $\mathsf{M}[I, J] = \mathsf{M}[K, L]$ 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.

## III. 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.

### A. 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.

*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 III.1). Thus, we have,

$$\begin{aligned}\|f\|_2 \; &= \; \Omega(\|\mathcal{F}\|_2) \; \geq \; \Omega\left(|\mathcal{F}(e_{i_1}, e_{i_2}, e_{i_3}, e_{i_4})|\right) \\ &= \; \Omega(B) \; = \; \Omega\left(\Lambda(f)/n^2\right) .\end{aligned}$$

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

*2) $(n^2/q^2)$-Approximation Assuming Theorem I.5:* We will next show that $\Lambda\left(f^{q/4}\right)^{4/q} \leq O(n^2/q^2) \cdot \|f\|_2$ (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 [13] 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 I.5 which on combining with the aforementioned multi-linear 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 I.5 and it's 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 I.5.

*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 [25], [1], [17] (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 III.1 ([1]).** *Let $\mathcal{A}$ be a SoS-symmetric $d$-tensor and let $h(x) := \langle \mathcal{A}, x^{\otimes d}\rangle$. Then $\|h\|_2 \geq 2^{-O(d)} \cdot \max_{\|x^i\|=1} \langle \mathcal{A}, x^1 \otimes \cdots \otimes x^d\rangle$.*

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

$$\begin{aligned} g(x) \; &= \; \sum_{|\alpha| \leq q/2} x^{2\alpha} \cdot G_{2\alpha}(x) \\ &= \; \sum_{r=0}^{q/2} \sum_{|\alpha|=r} x^{2\alpha} \cdot G_{2\alpha}(x) \\ &=: \; \sum_{r=0}^{q/2} g_r(x) \, , \end{aligned} \qquad \text{(III.1)}$$

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

$$\frac{\Lambda(g)}{\|g\|_2} \; \leq \; \max_{\alpha \in \mathbb{N}^n_{\leq q/2}} \frac{\Lambda(G_{2\alpha})}{\|G_{2\alpha}\|_2} \cdot 2^{O(q)} \qquad \text{(III.2)}$$

As a simple and immediate example of its applicability, (III.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 [15], [31]). 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 (III.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:

$$\Lambda(g) \;\; \leq \;\; (1+q/2) \cdot \max_{\alpha \in \mathbb{N}^n_{\leq q/2}} \Lambda(G_{2\alpha}) \, . \qquad \text{(III.3)}$$

We can further strengthen (III.3) by averaging the "best" representation of $G_{2\alpha}$ over $|\mathcal{O}(\alpha)|$ diagonal-blocks which all correspond to $x^{2\alpha}$. We show (see [13])

$$\Lambda(g) \;\; \leq \;\; (1+q/2) \cdot \max_{\alpha \in \mathbb{N}^n_{\leq q/2}} \frac{\Lambda(G_{2\alpha})}{|\mathcal{O}(\alpha)|} \, . \qquad \text{(III.4)}$$

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. (III.1) not only gives a block-diagonal decomposition for matrix representations of $g$, but can in fact be thought of as a "decomposition" of the *tensor* corresponding to $g$ (with regards to computing $\|g\|_2$). We show that

$$\|g\|_2 \;\geq\; 2^{-O(q)} \cdot \max_{\alpha \in \mathbb{N}^n_{\leq q/2}} \frac{\|G_{2\alpha}\|_2}{|\mathcal{O}(\alpha)|}\,. \qquad \text{(III.5)}$$

The above inequality together with (III.4), implies (III.2).

*4) Bounding $\|g\|_2$ via a new weak decoupling lemma:* Recall that the expansion of $g(x)$ in Eq. (III.1), contains the term $x^{2\alpha} \cdot G_{2\alpha}(x)$. The key part of proving the bound in (III.5) is to show the following "weak decoupling" result for $x^{2\alpha}$ and $G_{2\alpha}$.

$$\begin{aligned} \forall\alpha \qquad \|g\|_2 \;&\geq\; \max_{\|y\|=\|x\|=1} y^{2\alpha} \cdot G_{2\alpha}(x) \cdot 2^{-O(q)} \\ &=\; \max_{\|y\|=1} y^{2\alpha} \cdot \|G_{2\alpha}\|_2 \cdot 2^{-O(q)}. \end{aligned}$$

The proof of (III.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 $\mathbb{E}\left[|g(Z)|\right]$.

**The extremal case of $\alpha = 0^n$.** We first consider the extremal case of $\alpha = 0^n$, 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, $\mathbb{E}\left[h(Z)\right]$ 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 then gives that there exists a value of the probability $p$ such that

$$\begin{aligned} \|h\|_2 \;&\geq\; \mathbb{E}\left[|h(Z)|\right] \;\geq\; \left|\mathbb{E}\left[h(Z)\right]\right| \;\geq\; 2^{-O(t)} \cdot |h_m(x^*)| \\ &=\; 2^{-O(t)} \cdot \|h_m\|_2\,. \end{aligned}$$

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 III.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 extremal case of $g = g_r$.** In this case, the problem reduces to showing that for all $\alpha \in \mathbb{N}^n_r$ 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 $\alpha \in \mathbb{N}^n_r$, 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 $|\alpha'| = r, \quad |\gamma| = q - 2r, \quad \gamma \leq \mathbb{1}$,

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

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

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

**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.

### *B. 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.

*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,

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

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

$$\begin{aligned} \Lambda(f) \ &\leq\ \|f\|_{sp} \ \leq\ n \cdot \max_{|\beta|=2,\, \beta\leq \mathbb{1}} \left\|\bar{h}_\beta\right\|_{sp} \\ &=\ n \cdot \max_{|\beta|=2,\, \beta\leq \mathbb{1}} \|\bar{h}_\beta\|_2 , \end{aligned}$$

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.

*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 III-B1 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 $f := \mathsf{U}(h)$ 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 our work (see [13]), is the following theorem we believe to be of independent interest:

**Theorem III.2.** *Let $h, f$ be as above, and let*

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

*Then for any $q$ divisible by $2d$,*

$$\Lambda\Big(f^{q/d}\Big)^{d/q} \ \leq\ \ O_d\Big(\Gamma \cdot (n/q)^{d_1/2}\Big) \cdot \|f\|_2 .$$

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 focused on general polynomials for the overview, let us remark that in the case of polynomials with non-negative coefficients, the approximation factor in Theorem III.2 becomes $O_d\big(\delta \cdot (n/q)^{d_1/4}\big)$.

### *C. Lower Bounds for Polynomials with Non-negative Coefficients*

*1) Degree-4 Lower Bound for Polynomials with Non-Negative Coefficients:* We discuss some of the important ideas from the proof of Theorem I.6. The lower bound proved by a subset of the authors in [24] 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 $\mathsf{A} \in \mathbb{R}^{[n]^2 \times [n]^2}$ be the natural matrix representation of $f$ (i.e., each coefficient $f_\alpha$ is distributed uniformly among the corresponding entries of $\mathsf{A}$), the Perron-Frobenius theorem shows that $\|\mathsf{A}\|_2$ is less than the maximum row sum $\max(\tilde{O}(n^2p), 1)$ of $\mathsf{M}$, which is also an upper bound on $\Lambda(f)$. However, we can match this bound by (within constant factors) choosing $x = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ 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

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

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 $\mathsf{A}$ 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 $\|f\|_{sp} \geq \Omega\left(\sqrt{n^2p^5}\right) = \Omega(n^{1/6})$ by considering the Frobenius norm of the $n^2p \times n^2p$ principal submatrix, over any matrix representation (indeed, $\mathsf{A}$ is the minimizer). To prove $\Lambda(f) \geq \tilde{\Omega}(n^{1/6})$, we construct a moment matrix $\mathsf{M}$ that is SoS-symmetric, positive semidefinite, and has a large $\langle \mathsf{M}, \mathsf{A}\rangle$ (see the dual form of $\Lambda(f)$ in [13]). It turns out that the $n^2p \times n^2p$ submatrix of $\mathsf{A}$ shares spectral properties of the adjacency matrix of a random graph $G_{n^2p,p^4}$, and taking $\mathsf{M} := c_1\mathsf{A} + c_2\mathsf{I}$ for some identity-like matrix $\mathsf{I}$ 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 $\{v_1,\ldots,v_4\}$ of $G$ with $v_i \in S_i$ for $i = 1,\ldots,4$. The number of copies of a fixed subgraph $H$ in $G_{n,p}$, including its tail bound, has been actively studied in probabilistic combinatorics [32], [33], [34], [35], [36], [37], [38], 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 $O(n^2p^6 \cdot \log^{O(1)} n)$ when $|S_1| = \cdots = |S_4|$, 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 $O(n^2p^6 \cdot \log^{O(1)} n)$.

*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 $f$, $\left\|f^{q/4}\right\|_{sp}^{4/q}$ 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 $\mathsf{M}_f$ the canonical matrix representation of $f$, and $\|\mathsf{X}\|_{S_1}$ is the Schatten 1-norm (nuclear norm) of $X$, which is the sum of it's singular values.

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

Now let $\mathsf{X}$ be a solution realizing a gap of $\delta$ between $\|f\|_{sp}$ and $\|f\|_2$. We shall next see how assuming reasonable conditions on $\mathsf{X}$ 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 $\left\|f^{q/4}\right\|_{sp}$, a natural choice for a solution is the symmetrized version of the matrix $\mathsf{X}^{\otimes q/4}$ normalized by its Schatten-1 norm i.e., for $Y = \mathsf{X}^{\otimes q/4}$, we take

$$\begin{aligned} \mathsf{Z} &:= Y^S / \left\|Y^S\right\|_{S_1} \\ &\text{where } \; Y^S[K] = \mathop{\mathbb{E}}_{\pi \in \mathbb{S}_q}[Y[\pi(K)]] \quad \forall K \in [n]^q\,. \end{aligned}$$

To get a good gap, we are now left with showing that $\left\|Y^S\right\|_{S_1}$ 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 $\left\|Y^S\right\|_{S_1}$. 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, $M[(i_1,\ldots,i_x),(j_1,\ldots j_y)] = T[i_1,\ldots,i_x,j_1,\ldots j_y]$. 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 $M_D := \mathsf{Vec}(M)\,\mathsf{Vec}(M)^T = M_{0,4} \otimes M_{4,0}$.

We show (see [13]) 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 $M := \mathsf{X}$, we are able to establish that if $\mathsf{X}$ satisfies the additional condition of $\|\mathsf{X}_{1,3}\|_{S_1} \leq 1$ (note that we already know $\|X\|_{S_1} \leq 1$), then

$\left\|Y^S\right\|_{S_1} = 2^{O(q)}$. Thus Z realizes a $\langle M_f^{\otimes q/4}, Y^S\rangle / 2^{O(q)}$ gap for $\left\|f^{q/4}\right\|_{sp}$. On composing this result with the degree-4 gap from the previous section, we obtain an $\|\cdot\|_{sp}$ gap of $n^{q/24}/\left(q \cdot \log n\right)^{O(q)}$ 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 [13]). 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 [28], [39], [27]. 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.

## IV. Conclusion

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) 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?
5) 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?


## Acknowledgment

The authors would like to thank the anonymous referees of FOCS 2017 for particularly detailed and informative reviews.


## References


[1] S. He, Z. Li, and S. Zhang, "Approximation algorithms for homogeneous polynomial optimization with quadratic constraints," *Mathematical Programming*, vol. 125, no. 2, pp. 353–383, 2010.

[2] M. Laurent, "Sums of squares, moment matrices and optimization over polynomials," in *Emerging applications of algebraic geometry*. Springer, 2009, pp. 157–270.

[3] J. B. Lasserre, *Moments, positive polynomials and their applications*. World Scientific, 2009, vol. 1.

[4] F. G. Brandao and A. W. Harrow, "Quantum de finetti theorems under local measurements with applications," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 861–870.

[5] B. Barak, J. A. Kelner, and D. Steurer, "Rounding sum-of-squares relaxations," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, 2014, pp. 31–40.

[6] B. Barak, F. G. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou, "Hypercontractivity, sum-of-squares proofs, and their applications," in *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 2012, pp. 307–326.

[7] B. Barak, J. A. Kelner, and D. 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*. ACM, 2015, pp. 143–151.

[8] R. Ge and T. Ma, "Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms," *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, p. 829, 2015.

[9] A. Montanari and E. Richard, "A statistical model for tensor PCA," in *Advances in Neural Information Processing Systems*, 2014, pp. 2897–2905.

[10] S. B. Hopkins, J. Shi, and D. Steurer, "Tensor principal component analysis via sum-of-square proofs," in *Proceedings of The 28th Conference on Learning Theory*, 2015, pp. 956–1006.

[11] A. Frieze and R. Kannan, "A new approach to the planted clique problem," in *LIPIcs-Leibniz International Proceedings in Informatics*, vol. 2. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008.

[12] S. C. Brubaker and S. S. Vempala, "Random tensors and planted cliques," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2009, pp. 406–419.

[13] V. Bhattiprolu, M. Ghosh, V. Guruswami, E. Lee, and M. Tulsiani, "Multiplicative approximations for polynomial optimization over the unit sphere," *CoRR*, vol. abs/1611.05998, 2016. [Online]. Available: http://arxiv.org/abs/1611.05998

[14] E. De Klerk, "The complexity of optimizing over a simplex, hypercube or sphere: a short survey," *Central European Journal of Operations Research*, vol. 16, no. 2, pp. 111–125, 2008.

[15] E. de Klerk, M. Laurent, and P. A. Parrilo, "A PTAS for the minimization of polynomials of fixed degree over the simplex," *Theoretical Computer Science*, vol. 361, no. 2, pp. 210–225, 2006.

[16] J. Håstad and S. Venkatesh, "On the advantage over a random assignment," *Random Structures & Algorithms*, vol. 25, no. 2, pp. 117–149, 2004.

[17] A. M.-C. So, "Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems," *Mathematical programming*, vol. 129, no. 2, pp. 357–382, 2011.

[18] Y. Nesterov, "Random walk in a simplex and quadratic optimization over convex polytopes," UCL, Tech. Rep., 2003.

[19] J. Håstad, "Clique is hard to approximate within n 1-&epsiv," in *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*. IEEE, 1996, pp. 627–636.

[20] A. C. Doherty and S. Wehner, "Convergence of sdp hierarchies for polynomial optimization on the hypersphere," *arXiv preprint arXiv:1210.5048*, 2012.

[21] E. de Klerk, M. Laurent, and Z. Sun, "Convergence analysis for Lasserre's measure–based hierarchy of upper bounds for polynomial optimization," *arXiv preprint arXiv:1411.6867*, 2014.

[22] B. Barak, P. Kothari, and D. Steurer, "Quantum entanglement, sum of squares, and the log rank conjecture," *arXiv preprint arXiv:1701.06321*, 2017.

[23] P. Raghavendra, S. Rao, and T. Schramm, "Strongly refuting random csps below the spectral threshold," *arXiv preprint arXiv:1605.00058*, 2016.

[24] V. Bhattiprolu, V. Guruswami, and E. Lee, "Certifying random polynomials over the unit sphere via sum of squares hierarchy," *arXiv preprint arXiv:1605.00903*, 2016.

[25] S. Khot and A. Naor, "Linear equations modulo 2 and the l_1 diameter of convex bodies," *SIAM Journal on Computing*, vol. 38, no. 4, pp. 1448–1463, 2008.

[26] V. De la Pena and E. Giné, *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

[27] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, "The power of sum-of-squares for detecting hidden structures," *Personal communication*, 2017.

[28] B. Barak, S. B. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. 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*, ser. FOCS '16, 2016, pp. 428–437.

[29] A. Brieden, P. Gritzmann, R. Kannan, V. Klee, L. Lovász, and M. Simonovits, "Deterministic and randomized polynomial-time approximation of radii," *Mathematika*, vol. 48, no. 1-2, pp. 63–105, 2001.

[30] F. G. Brandao and A. W. Harrow, "Estimating operator norms using covering nets," *arXiv preprint arXiv:1509.05065*, 2015.

[31] E. de Klerk, M. Laurent, and Z. Sun, "An alternative proof of a PTAS for fixed-degree polynomial optimization over the simplex," *Mathematical Programming*, vol. 151, no. 2, pp. 433–457, 2015.

[32] V. H. Vu, "A large deviation result on the number of small subgraphs of a random graph," *Combinatorics, Probability & Computing*, vol. 10, no. 1, pp. 79–94, 2001.

[33] J. H. Kim and V. H. Vu, "Divide and conquer martingales and the number of triangles in a random graph," *Random Structures & Algorithms*, vol. 24, no. 2, pp. 166–174, 2004.

[34] S. Janson, K. Oleszkiewicz, and A. Ruciński, "Upper tails for subgraph counts in random graphs," *Israel Journal of Mathematics*, vol. 142, no. 1, pp. 61–92, 2004.

[35] S. Chatterjee, "The missing log in large deviations for triangle counts," *Random Structures & Algorithms*, vol. 40, no. 4, pp. 437–451, 2012.

[36] B. DeMarco and J. Kahn, "Upper tails for triangles," *Random Structures & Algorithms*, vol. 40, no. 4, pp. 452–459, 2012.

[37] R. DeMarco and J. Kahn, "Tight upper tail bounds for cliques," *Random Structures & Algorithms*, vol. 41, no. 4, pp. 469–487, 2012.

[38] E. Lubetzky and Y. Zhao, "On the variational problem for upper tails in sparse random graphs," *Random Structures & Algorithms*, 2016.

[39] P. K. Kothari, R. Mori, R. O'Donnell, and D. Witmer, "Sum of squares lower bounds for refuting any csp," *arXiv preprint arXiv:1701.04521*, 2017.