Lifting Sum-of-Squares Lower Bounds: Degree-2 to Degree-4

Sidhanth Mohanty
Prasad Raghavendra
Jeff Xu
sidhanthm@cs.berkeley.edu
prasad@cs.berkeley.edu
jeffxusichao@berkeley.edu
EECS Department, University of California at Berkeley
Berkeley, CA, USA

ABSTRACT

The degree-4 Sum-of-Squares (SoS) SDP relaxation is a powerful algorithm that captures the best known polynomial time algorithms for a broad range of problems including MaxCut, Sparsest Cut, all MaxCSPs and tensor PCA. Despite being an explicit algorithm with relatively low computational complexity, the limits of degree-4 SoS SDP are not well understood. For example, existing integrality gaps do not rule out a $(2-\varepsilon)$-algorithm for Vertex Cover or a $(0.878+\varepsilon)$-algorithm for MaxCut via degree-4 SoS SDPs, each of which would refute the notorious Unique Games Conjecture.

We exhibit an explicit mapping from solutions for degree-2 Sum-of-Squares SDP (Goemans-Williamson SDP) to solutions for the degree-4 Sum-of-Squares SDP relaxation on boolean variables. By virtue of this mapping, one can lift lower bounds for degree-2 SoS SDP relaxation to corresponding lower bounds for degree-4 SoS SDPs. We use this approach to obtain degree-4 SoS SDP lower bounds for MaxCut on random d-regular graphs, Sherington-Kirkpatrick model from statistical physics and PSD Grothendieck problem.

Our constructions use the idea of pseudocalibration towards candidate SDP vectors, while it was previously only used to produce the candidate matrix which one would show is PSD using much technical work. In addition, we develop a different technique to bound the spectral norms of *graphical* matrices that arise in the context of SoS SDPs. The technique is much simpler and yields better bounds in many cases than the *trace method* – which was the sole technique for this purpose.

CCS CONCEPTS

• **Mathematics of computing → Random graphs**; • **Theory of computation → Semidefinite programming**.

KEYWORDS

Sum-of-Squares lower bounds, max cut on random d-regular graph, random matrices, Sherrington-Kirkpatrick Model

STOC '20, June 22–26, 2020, Chicago, IL, USA

ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384319

ACM Reference Format:
Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. 2020. Lifting Sum-of-Squares Lower Bounds: Degree-2 to Degree-4. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3357713.3384319

1 INTRODUCTION

Sum-of-Squares (SoS) semidefinite programming hierarchy is one of the most powerful frameworks for algorithm design. Its foundations lie in the so-called "Positivestellensatz" whose history dates back to more than a century to the work of Hilbert and others. The algorithmic insight of finding Sum-of-Squares proofs via the technique of semi-definite programming was only codified at the turn of the century by Parrillo [33] and Lasserre [23] (also see [36]).

Given a system of polynomial equations/inequalities $\mathcal{P}$, the SoS SDP hierarchy yields a sequence of semi-definite programming relaxations to reason about the feasibility of $\mathcal{P}$. The d-th relaxation in the sequence is referred to as the *degree-d SoS SDP relaxation*. Successive relaxations get increasingly accurate in reasoning about $\mathcal{P}$ at the expense of computational complexity that grows exponentially with the degree.

SoS SDP hierarchy is an incredibly powerful algorithmic technique. The best known approximation algorithms for a variety of combinatorial optimization problems including Maximum Cut, all Max-CSPs and Sparsest Cut are all subsumed by the first two levels (degree-4) of the hierarchy. More recently, there has been a flurry of work that uses SoS SDP hierarchy on problems in unsupervised learning such as dictionary learning, estimating parameters of mixtures of Gaussians, tensor PCA and linear regression.

The limits of SoS SDP hierarchy remain largely a mystery even at degree four. The degree four SoS SDP relaxation could possibly yield a $(2-\varepsilon)$-approximation for Minimum Vertex Cover or a $(0.878+\varepsilon)$-approximation for Maximum Cut and thereby refute the notorious Unique Games Conjecture. Despite the immense consequences, the integrality gap of degree-4 SoS SDP relaxations of Maximum Cut and Vertex Cover remain unresolved.

Understanding the precise limits of SoS SDP hierarchy has compelling implications even in the context of average case problems. Specifically, the SoS SDP hierarchy can serve as a *lens* to understand the terrain of average case complexity. For example, consider the problem of refuting a random 3-SAT formula. Here the input consists of a random 3-SAT formula Φ with $m = pn$ clauses chosen uniformly at random on n variables. For all densities p that are

larger than some fixed constant, the formula Φ is unsatisfiable with high probability. The goal of a refutation algorithm is to certify that Φ is unsatisfiable. Formally, a refutation algorithm outputs 1 only on instances that are unsatisfiable and it does so on a non-negligible fraction of random 3-SAT formulae. Although the computational complexity of refuting random 3-SAT formulae conceivably varies with the density p of clauses, it seems difficult to glean this structure using *reductions* – the central tool in worst-case computational complexity. In particular, it is quite difficult to devise reductions that produce random instances from simple probability distributions such as random 3-SAT, though this has been sometimes achieved [4, 6]. In such a setting, the *smallest degree of SoS SDP hierarchy* that can solve the refutation problem (henceforth referred to as just "SoS degree") can serve as a proxy for computational complexity. While SoS SDP hierarchy doesn't capture all efficient algorithms in every context, it unifies and subsumes many of the state-of-the-art algorithms for basic combinatorial optimization problems.

This paradigm has been fruitful for random 3-SAT. Nearly matching upper and lower bounds on SoS degree of refutation [15, 34, 35] have been established, thereby painting a precise picture of how the complexity of the problem changes with density of clauses. Specifically, for all $\omega(1) < p < n^{3/2}$, the Sum-of-Squares degree is $\tilde{\Theta}(n/p^2)$, yielding a complexity of $2^{\tilde{\Theta}(n/p^2)}$.

There is a rich landscape of average case problems with many having sharper computational thresholds than random 3-SAT. For example, the random regular NAESAT promises to exhibit an abrupt change in computational complexity as soon as the degree exceeds 13.5 [10]. Chromatic number of random d-regular graphs and community detection on stochastic block models are two other prominent examples with very sharp but conjectural computational thresholds. Much is known about structural characterestics and phase transitions in the solution space as one varies the underlying parameters in these models. Heuristically, certain phase transitions in the solution space are conjectured to be associated with abrupt changes in the computational complexity. The Sum-of-Squares SDP can be harnessed towards quantitatively demonstrating these phenomena.

1.1 Our Results

Our main result is an explicit mapping from solutions to degree-2 SoS SDP to solutions to degree-4 SoS SDP for boolean optimization. To formally state the theorem, let us begin by setting up some notation.

First, the degree-d SoS SDP relaxation can be succinctly described in terms of *pseudodistributions*. Intuitively, a *pseudodistribution* corresponds to a function that looks like an actual distribution over solutions, to low degree polynomial squares. The definition is succinct and simple enough that we reproduce the formal definition here.

Definition 1.1. Fix a natural number $d \in \mathbb{N}$. A *degree d pseudodistribution* μ is a function $\mu : \{\pm 1\}^n \to \mathbb{R}$ satisfying

(1) (Normalization)
$$\mathbf{E}_{x \in \{-1,1\}^n}[\mu(x)] = 1$$
(2) (Positivity) For all $p \in \mathbb{R}[x_1, \ldots, x_n]$, $\deg(p) \leqslant d/2$,
$$\mathbf{E}_{x \in \{-1,1\}^n}[p^2(x) \cdot \mu(x)] \geqslant 0$$

While the above description of degree-d SoS SDP is accurate, we will now describe the associated semidefinite programs for degree two and four in detail. By the degree-2 SoS SDP for boolean optimization, we refer to the Goemans-Williamson SDP relaxation, first introduced in the context of the MaxCut problem. Specifically, a feasible solution to the degree-2 SoS SDP solution is given by a p.s.d matrix $X \succeq 0$ whose diagonal entries are identically 1. Formally, the set of degree-2 SoS SDP solutions denoted by SoS_2 is given by,

$$\mathrm{SoS}_2 = \{X \in \mathbb{R}^{n \times n} | X \succeq 0 \text{ and } X_{ii} = 1 \text{ for all } i \in [n]\}$$

The solution to a degree-4 SoS SDP for boolean optimization consists of a matrix $\mathcal{M}$ of dimension $\binom{n}{\leqslant 2} = 1 + \binom{n}{1} + \binom{n}{2}$. The matrix $\mathcal{M}$ is indexed by subsets of $[n] = \{1, \ldots, n\}$ of size at most 2. The set SoS_4 is specified by the following SDP:

$$\mathcal{M}[S, T] = \mathcal{M}[S', T'] \tag{1}$$
$$\text{for all } S, T, S', T' \in \binom{[n]}{\leqslant 2} \text{ such that } S \Delta T = S' \Delta T'$$
$$\mathcal{M}[\emptyset, \emptyset] = 1 \tag{2}$$
$$\mathcal{M} \succeq 0 \tag{3}$$

The above semidefinite programs are equivalent to the definition of SoS relaxations in terms of pseudodistributions. Specifically, the entries of the matrix $\mathcal{M}$ are *pseudomoments* upto degree four of the pseudodistribution μ. Formally, the entry $\mathcal{M}[S, T]$ corresponds to the following moment:

$$\mathcal{M}[S, T] = \mathbf{E}_{x \in \{-1,1\}^n} \left[\mu(x) \prod_{i \in S} x_i \prod_{j \in T} x_j \right]$$

We are now ready to state the main theorem of this work.

Theorem 1.2 (Main theorem). *There is an explicit map $\Phi : \mathrm{SOS}_2 \to \mathrm{SOS}_4$ such that $\Phi(X)[i, j]$ [1] is given by*

$$\Phi(X)[i, j] = \frac{X_{ij} + X_{ij}^3}{1 + C\alpha_{\mathrm{mag}} \cdot (1 + \alpha_{\mathrm{row}}^4) \cdot (1 + \alpha_{\mathrm{spec}}^2)} \tag{4}$$

where $\alpha_{\mathrm{mag}}, \alpha_{\mathrm{row}}$ and α_{spec} are the maximum off-diagonal entry, maximum row norm and spectral norm respectively of the degree two SDP solution X, and C is an absolute constant. Moreover for every pair of subsets $S, T \in \binom{[n]}{\leqslant 2}$, $\Phi(X)[S, T]$ is an explicit function of $\{X_{ij} | i, j \in S \cup T\}$.

All the entries of $\Phi(X)$ are explicit constant degree polynomials in X. We refer the reader to Section 2 for the definition of Φ and the proof of Theorem 1.2. Let us suppose we have an objective value given by $\langle A, X \rangle = \sum_{i,j} A_{ij} X_{ij}$ for a Hermitian matrix A. The corresponding objective value of degree-4 SoS SDP is given by $\langle A, \mathcal{M} \rangle = \sum_{i,j} A_{ij} \mathcal{M}[i, j]$. We show the following bound on change in objective value (see Lemma 2.4 in Section 2):

Theorem 1.3. *Let $\alpha := C\alpha_{\mathrm{mag}} \cdot (1 + \alpha_{\mathrm{row}}^4) \cdot (1 + \alpha_{\mathrm{spec}}^2)$ where $\alpha_{\mathrm{mag}}, \alpha_{\mathrm{row}}$ and α_{spec} are as defined in Theorem 1.2, then for any Hermitian matrix $A \in \mathbb{R}^{n \times n}$, let $\Phi^{(2)}(X)$ be the restriction of $\Phi(X)$ to the degree-2 part,*

$$\langle A, \Phi^{(2)}(X) \rangle \geqslant \frac{1}{1 + \alpha} \langle A, X \rangle - \frac{\alpha}{1 + \alpha} \cdot \left(\sqrt{n} \|A\|_F - \mathrm{Trace}\,(A) \right)$$

[1] We are using $\Phi(X)[i, j]$ to denote $\Phi(X)[\{i\}, \{j\}]$.

The existence of a non-trivial and useful mapping from degree-2 SoS SDP solutions to degree-4 SoS SDP solutions comes as a surprise to the authors. Consider the following immediate consequence of such a mapping. Given the degree-2 SoS SDP on an instance of MaxCut, the above theorem yields an easily computable lower bound on the degree-4 SoS SDP value on the same instance. For example, this yields an efficiently verifiable sufficient condition (checkable in time $O(n^2)$) under which the degree-4 SoS SDP yields no better bound than the degree-2 SoS.

We use the lifting theorem to recover lower bounds for degree-4 SoS SDP relaxations for a few average case problems – which was the original motivation behind this work. The problems and the corresponding lower bounds are described below.

Sherrington–Kirkpatrick Model. Let $\boldsymbol{W}$ be a random $n \times n$ matrix with independent Gaussian entries, let $\boldsymbol{G} := \frac{1}{\sqrt{2}}\left(\boldsymbol{W} + \boldsymbol{W}^\dagger\right)$; we say that $\boldsymbol{G}$ is sampled from GOE(n), a distribution known as the Gaussian Orthogonal Ensemble. A fundamental model in the study of spin glasses from statistical physics is the *Sherrington–Kirkpatrick (SK) model* where the energy of a system of n particles in a state $x \in \{-1,+1\}^n$ states is given by $-x^\dagger \boldsymbol{G} x$. The Sherrington-Kirkpatrick (SK) model has been extensively studied in various areas including the study of spin glasses, random satisfiability problems, and learning theory [11, 24–27, 29].

For the SK model, a quantity of particular interest is the minimum possible energy, i.e.,

$$\mathsf{OPT}(\boldsymbol{G}) = \max_{x \in \{-1,1\}^n} x^\dagger \boldsymbol{G} x \,.$$

In a highly influential work, Parisi predicted in [31, 32] that $\mathsf{OPT}(\boldsymbol{G})$ concentrates around $2 \cdot \mathsf{P}^* n^{3/2}$, where P^* is an explicit constant now referred to as the Parisi constant. The value of P^* is roughly 0.763166. This prediction was eventually rigorously proven twenty five years later in a celebrated work of Talagrand [37], thereby confirming that $\mathsf{OPT}(\boldsymbol{G}) \approx (1.52633\ldots) \cdot n^{3/2}$.

This brings us to our natural average case refutation problem, that of certifying an upper bound on $x^\dagger \boldsymbol{G} x$ for $x \in \{-1,1\}^n$. A natural refutation algorithm is the *spectral refutation*. Indeed

$$\mathsf{OPT}(\boldsymbol{G}) = \max_{x \in \{\pm 1\}^n} x^\dagger \boldsymbol{G} x \leqslant n \cdot \max_{\|x\|=1} x^\dagger \boldsymbol{G} x = n \cdot \lambda_{\max}(\boldsymbol{G}),$$

the algorithm which outputs $\lambda_{\max}(\boldsymbol{G})$ given $\boldsymbol{G}$ as input is an efficient refutation algorithm. Since $\lambda_{\max}(\boldsymbol{G})$ concentrates around $2\sqrt{n}$, it certifies an upper bound $OPT(\mathcal{G}) \leqslant 2n^{3/2}$ which is larger than the true value of the optimum $OPT(\mathcal{G}) = 2\mathsf{P}^* \cdot n^{3/2} = 1.52 \cdot n^{3/2}$.

This raises the question whether efficient algorithms can certify an upper bound stronger than the simple spectral bound? In this work, we show that the degree-4 SoS SDP fails to certify a bound better than the spectral bound. To this end, we start with a feasible solution to the degree-2 SoS SDP relaxation for the SK model and apply our lifting theorem Theorem 1.2 to construct a degree-4 SoS SDP solution.

Theorem 1.4 (Degree-4 SoS lower bound for Sherrington–Kirkpatrick). *Let $\boldsymbol{G} \sim \mathrm{GOE}(n)$. With probability $1 - o_n(1)$, there exists a degree-4 SoS SDP solution with value at least $(2 - o_n(1)) \cdot n^{3/2}$*

In an independent and concurrent work, Kunisky and Bandeira [22] also obtained a degree-4 SoS integrality gap for the Sherrington–Kirkpatrick refutation problem.

MaxCut *in random d-regular graphs.* Akin to the Sherrington–Kirkpatrick model, it is known from the work of Dembo et al. [8] that the fraction of edges cut by the max-cut in a random d-regular graph $\boldsymbol{G}$ on n vertices is concentrated around

$$\frac{1}{2} + \frac{\mathsf{P}^*}{\sqrt{d}} + o_d\left(\frac{1}{\sqrt{d}}\right) + o_n(1).$$

On the other hand, it was proved in [5, 13] that the spectral refutation algorithm, which outputs the maximum eigenvalue of $\frac{L_G}{4m}$, certifies an upper bound of

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d} + o_n(1).$$

Once again the question remains whether more sophisticated refutation algorithms can beat the spectral bound. Through our lifting theorem, we show that degree 4 SoS SDP is no better than spectral algorithm asymptotically as $d \to \infty$[2].

Theorem 1.5 (Degree-4 SoS lower bound for MaxCut in random d-regular graphs). *Let $\boldsymbol{G}$ be a random d-regular graph. For every constant $\varepsilon > 0$ with probability $1 - o_n(1)$, there is a degree-4 SoS SDP solution with MaxCut value at least*

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}\left(1 - \varepsilon - \frac{\gamma(\varepsilon)}{d^{1/2}}\right)$$

for some constant γ that depends only on ε.

The degree-2 SoS SDP solution for the SK model on which we apply our lifting theorem is presented in Theorem 4.8. Analogously, Theorem 5.1 describes the degree 2 SoS SDP solution we use for the MaxCut problem.

"Boolean Vector in Random Subspace" Problem. The refutation problem for the SK model is closely tied to the following problem: given a random subspace $\boldsymbol{V}$ of dimension d in $\mathbb{R}^n$, can we certify that there is no hypercube vector $\{\pm 1\}^n$ 'close' to $\boldsymbol{V}$ in polynomial-time? Formally, if $\Pi_{\boldsymbol{V}}$ denotes the projection operator onto a random subspace, then let $\mathsf{OPT}(\boldsymbol{V})$ denote the maximum correlation of a boolean vector with $\boldsymbol{V}$, i.e.,

$$\mathsf{OPT}(\boldsymbol{V}) = \frac{1}{n} \max_{x \in \{-1,1\}^n} x^\dagger \Pi_{\boldsymbol{V}} x \,.$$

Using a simple ε-net argument, one can show that with high probability $\mathsf{OPT}(\boldsymbol{V}) \sim \frac{2}{\pi} + \gamma(d/n)$ for some function $\gamma : [0,1] \to \mathbb{R}^+$ such that $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$[3]. In other words, for a low dimensional subspace with $d \ll n$, $\mathsf{OPT}(\boldsymbol{V})$ is close to $2/\pi$ with high probability over choice of $\boldsymbol{V}$.

The spectral algorithm can only certify $\mathsf{OPT}(\boldsymbol{V}) \leqslant \|\Pi_{\boldsymbol{V}}\| = 1$ which is a trivial bound. A natural question is whether one can

[2] We believe that Theorem 1.5 is not tight and conjecture that there should exist pseudoexpectations with objective value $\frac{1}{2} + (1 - o_n(1))\frac{\sqrt{d-1}}{d}$ for all values of d.

[3] $\mathsf{OPT}(\boldsymbol{V}) = \|A_{\boldsymbol{V}}\|^2_{2\to 1}$, where columns of $A_{\boldsymbol{V}}$ are an orthogonal basis for $\boldsymbol{V}$. So for fixed unit $x \in \mathbb{R}^d$, $\|A_{\boldsymbol{V}} x\|_1$ concentrates around $\sqrt{2/\pi}$ with a subgaussian tail. A union bound over an ε-net of $\mathbb{R}^d$ completes the calculation.

efficiently certify a stronger upper bound. We show that the degree-4 SoS SDP fails to improve on the spectral bound by a non-negligible amount.

Theorem 1.6 (Boolean Vector in Random Subspace). *If $\boldsymbol{V}$ is a random d-dimensional subspace where $d \geqslant n^{.99}$, then with probability $1 - o_n(1)$ there exists a degree-4 SoS SDP solution with value at least $1 - o_n(1)$.*

1.2 Related Work

Early work on lower bounds for Sum-of-Squares SDPs arose out of the literature on proof complexity. In particular, these included lower bounds on Sum-of-Squares refutations of Knapsack [14], Parity principle (non-existence of a perfect matching in a complete graph on odd number of vertices) [15] and 3XOR/3SAT [15]. For 3SAT/3XOR, it was proven by Grigoriev [15] and later independently by Schoenbeck [35] that the polynomial time regime of Sum-of-Squares fails to refute random instances whenever the density of clauses is $o(\sqrt{n})$. This lower bound for 3SAT is the starting point of lower bounds for a host of other problems. Specifically, the use of polynomial time reductions to convert integrality gaps for one problem into another, first pioneered in [20], was shown to be applicable to the SoS SDP hierarchy [38]. By harnessing the known reductions, Tulsiani [38] recovers exponential lower bounds for a variety of constraint satisfaction problems (CSP) starting from 3SAT.

More recently, Kothari et al. [21] obtained lower bounds for all CSPs corresponding to predicates whose satisfying assignments support a pairwise independent distribution. This class of CSPs is well beyond the reach of current web of NP-hardness reductions. 2-CSPs such as MaxCut are not pairwise independent, and are thus not within the realm of known lower bounds for SoS SDPs.

The problem of certifying the size of maximum clique on Erdos-Renyi random graphs (closely related to the planted clique problem) has received much attention lately. Following a series of works [9, 16] that obtained the tight lower bounds for degree four, the breakthrough tour-de-force of Barak et al. [2] obtained lower bounds for upto degree $O(\log n)$. In this work, Barak et al. [2] introduced a heuristic technique for constructing candidate solutions to Sum-of-Squares SDPs called *pseudocalibration*. Subsequently, the pseudocalibration technique was used in [17] to show SoS lower bounds for Tensor PCA and Sparse PCA. Building on ideas from pseudocalibration, Hopkins and Steurer [18] recovered conjectured computational thresholds in community detection, while [19] use it towards showing LP extended formulation lower bounds for Random 3SAT.

In an independent work, Kunisky and Bandeira [22] also obtained a degree-4 SoS integrality gap for the Sherrington–Kirkpatrick refutation problem.

1.3 Technical Overview

The mapping Φ alluded to in Theorem 1.2 is quite intricate and we are unable to motivate the construction of the mapping in a canonical fashion. Instead, we focus on how the map Φ was first constructed in the context of the Boolean Vector in Random Subspace problem.

Fix a randomly chosen subspace $\boldsymbol{V}$ of dimension d in $\mathbb{R}^n$. With high probability, no boolean vector $x \in \{-1,1\}^n$ is close to $\boldsymbol{V}$ (every boolean vector x has correlation less than $\frac{2}{\pi} + o_n(1)$ with $\boldsymbol{V}$). To prove that the degree 4 SoS SDP cannot refute the existence of a boolean vector in $\boldsymbol{V}$, we need to construct a degree 4 pseudodistribution μ such that,

$$\mathbf{E}_{x \in \{-1,1\}^n} [\mu(x) x^\dagger \Pi_{\boldsymbol{V}} x] \approx n \ .$$

In words, the pseudodistribution μ is seemingly supported on vectors x in the subspace $\boldsymbol{V}$.

Pseudocalibration. We will now use the pseudocalibration recipe of Barak [3] to arrive at the pseudodistribution μ.

The idea is to construct a planted distribution Θ over pairs $(x, \boldsymbol{V})$ where $x \in \{-1,1\}^n$, $x \in \boldsymbol{V}$ and the subspace $\boldsymbol{V}$ is a *seemingly* random subspace. For example, a natural planted distribution Θ would be given by the following sampling procedure:

- Sample $x \in \{-1,1\}^n$ uniformly at random.
- Sample a uniformly random subspace W of dimension $d-1$ and set $\boldsymbol{V} = \mathrm{Span}(W \cup \{x\})$.

It is clear that the pair $(x, \boldsymbol{V})$ satisfies all the desired properties of the planted distribution.

Let $\mathrm{Gr}(n,d)$ denote the space of all d-dimensional subspaces of $\mathbb{R}^n$. Let Θ denote the density associated with the planted distribution, i.e., Θ is a function over $\mathrm{Gr}(n,d) \times \{-1,1\}^n$. [4]

For any specific $\boldsymbol{V} \in \mathrm{Gr}(n,d)$ that contains a boolean vector in the subspace, notice that the restriction $\Theta_{\boldsymbol{V}}(x) = \Theta(x, \boldsymbol{V})$ is up to a factor normalization, a valid probability distribution over $\{-1,1\}^n$. Therefore, $\Theta_{\boldsymbol{V}}$ is a solution to the degree d SoS SDP relaxation for all d, upto the normalization factor. Ignoring the issue of the normalization factor for now, the candidate degree 4 moment matrix would be given by,

$$\mathcal{M}^*_{\boldsymbol{V}}[S,T] = \mathbf{E}_{x \in \{-1,1\}^n} \left[\left(\prod_{i \in S} x_i \right) \left(\prod_{j \in T} x_j \right) \cdot \Theta(x, \boldsymbol{V}) \right] \tag{5}$$

The matrix $\mathcal{M}^*$ is clearly positive semidefinite for each $\boldsymbol{V}$. To formally construct the Cholesky factorization of $\mathcal{M}^*$, one defines the vectors $\{V_S : \{-1,1\}^n \to \mathbb{R}\}$ to be the functions $V^*_S(x) = \Pi_{i \in S} x_i \cdot (\Theta(x, \boldsymbol{V}))^{1/2}$. The inner product between the vectors f, g is given by

$$\langle f(x), g(x) \rangle = \mathbf{E}_{x \in \{-1,1\}^n} [f(x) g(x)] \, .$$

With these definitions, we will have

$$\mathcal{M}^*[S,T] = \langle V^*_S, V^*_T \rangle \tag{6}$$

as desired. While the above ideal SDP solution and vectors satisfies most of the constraints, it fails the normalization. In fact, the normalization factor $\Gamma_{\boldsymbol{V}} = \mathbf{E}_{x \in \{-1,1\}^n} [\Theta_{\boldsymbol{V}}(x)]$ is very spiky, it is zero on almost all instances $\boldsymbol{V}$ except being very large on subspaces $\boldsymbol{V}$ containing a boolean vector.

The key insight of pseudocalibration is to project the planted density Θ to low degree functions in Θ, or equivalently truncate away the part of Θ that is high degree in the instance $\boldsymbol{V}$. Let $\Theta^{\leqslant D}$ denote the low degree truncation of the planted density Θ. For any $\boldsymbol{V} \in \mathrm{Gr}(n,d)$, the pseudo-calibrated pseudodensity

[4] Technically, the density Θ needs to be represented by a distribution

$\Theta^{\leqslant D}[V] : \{-1,1\}^n \to \mathbb{R}$ is given by $\Theta^{\leqslant D}[V](x) = \Theta^{\leqslant D}(V,x)$. More concretely, the candidate SDP solution specified by pseudo-calibration is

$$\mathcal{M}_V[S,T] = \mathbf{E}_{x\in\{-1,1\}^n}\left[\left(\prod_{i\in S} x_i\right)\left(\prod_{j\in T} x_j\right)\cdot \Theta^{\leqslant D}[V](x)\right] \quad (7)$$

for all S, T. The feasibility of $\mathcal{M}_V$ needs to be established, which often requires considerable technical work, especially the proof of positive semidefiniteness of $\mathcal{M}_V$.

A natural approach to prove psdness of $\mathcal{M}_V$ is to construct the corresponding SDP vectors (Cholesky factorization) by using a low degree truncation of the ideal SDP vectors V_S^* defined above. Since $\mathcal{M}_V$ is obtained by truncating an ideal solution $\mathcal{M}^*$ to low degree polynomials, it would be conceivable that the low degree truncation of the ideal SDP vectors yield Cholesky factorization of $\mathcal{M}_V$. Unfortunately, this hope does not come to fruition and to our knowledge does not hold for any problem.

Representations. Executing the above strategy over $\mathrm{Gr}(n,d)$ is technically challenging since low degree polynomials over $\mathrm{Gr}(n,d)$ are complicated. To cope with the technical difficulty, it is better to work with an explicit representation of the subspace V. Specifically, V can be represented by a $n\times\kappa$ matrix M_κ in that $V = \text{Col-Span}(M_\kappa)$. Any choice of $\kappa \geqslant d$ would suffice to represent a d-dimensional subspace V, and in our construction we will set $\kappa \to \infty$.

With this representation, a candidate planted distribution (x, M_κ) is sampled as follows:

- Sample $x \in \{-1,1\}^n$ uniformly at random.
- Sample $d-1$ vectors $w_1,\ldots,w_{d-1} \in \mathbb{R}^n$ from the standard normal distribution $N(0,1)^n$. Let M be the $n\times d$ matrix whose columns are x and $w_1,\ldots,w_{d-1}$.
- Let $U_\kappa \in \mathbb{R}^{\kappa\times\kappa}$ be a random unitary matrix, and let $U_\kappa^{\leqslant n} \in \mathbb{R}^{n\times\kappa}$ matrix denote the first n rows of U_κ. Set $M_\kappa = M\cdot U_\kappa^{\leqslant n}$

First, notice that $x \in \text{Col-Span}(M_\kappa)$ as needed. However, the representations are not unique in that each subspace V has infinitely many different representations. Further, the original SoS optimization problem depends solely on the subspace V, and is independent of the matrix M_κ representing V.

At first, these redundant representations or inherent symmetries of the planted density, seem to be an issue to be dealt with. It turns out that these redundancy in representations is actually useful in constructing the SDP vectors!

Planted Distribution. Before proceeding, we will first simplify our planted distribution even further. Since computations over random unitary matrices are technically difficult, we will select a much simpler finite subgroup of the unitary group to work with. In particular, the planted distribution Θ over pairs (x, M) is sampled as follows:

- Sample $x \in \{-1,1\}^n$ uniformly at random.
- Sample $d-1$ vectors $w_1,\ldots,w_{d-1} \in \mathbb{R}^n$ from the standard normal distribution $N(0,1)^n$. Let M be the $n\times d$ matrix whose columns are x and $w_1,\ldots,w_{d-1}$.
- Let $H_\kappa^{\leqslant n}$ denote the $n\times\kappa$ matrix obtained by taking the first n rows of the Hadamard matrix H_κ. Let $Z \in \mathbb{R}^{\kappa\times\kappa}$ denote a diagonal matrix with random $\{\pm 1\}$ entries. Set $M_\kappa = MH_\kappa^{\leqslant n}Z$

The above construction uses $H_\kappa Z$ instead of a unitary random matrix U_κ. In particular, the continous unitary group is replaced with a finite set of 2^κ transformations indexed by the familiar $\{-1,1\}^\kappa$, making the calculations tractable.

Exploiting multiple representations. Applying the pseudo-calibration heuristic to the planted density (x, M_κ) defined above, we get a candidate *ideal* SDP solution $\mathcal{M}_{M_\kappa}$

$$\mathcal{M}^*_{M_\kappa}[S,T] = \mathbf{E}_{x\in\{-1,1\}^n}\left[\left(\prod_{i\in S} x_i\right)\left(\prod_{j\in T} x_j\right)\cdot \Theta(M_\kappa, x)\right] \quad (8)$$

This ideal SDP solution needs to be truncated to low degree with Θ to be replaced by $\Theta^{\leqslant D}$. The specifics of the low degree projection used to define $\Theta^{\leqslant D}$ are intentionally left vague at this time.

The construction thus far is essentially the pseudocalibration heuristic albeit on a somewhat complicated planted distribution. It is at this time that we will exploit the symmetries of the planted density. Recall that the underlying subspace V depends only on $\text{Col-Span}(M_\kappa) = \text{Col-Span}(M)$, and so does the underlying SoS SDP relaxation. Therefore, it is natural to average out the above pseudocalibrated solution over the various representations of V, i.e., define the solution $\mathcal{M}_V$ as,

$$\mathcal{M}^*_V[S,T] = \mathbf{E}_{Z,x\in\{\pm1\}^n}\left[\left(\prod_{i\in S} x_i\right)\left(\prod_{j\in T} x_j\right)\cdot \Theta[MH_\kappa^{\leqslant n}Z](x)\right] \quad (9)$$

Analogous to the ideal SDP vectors (6), one can define SDP vectors V_S^* here, but this time as functions over both x and Z. That is if we let $V_S^*(x,Z) = (\Pi_{i\in S}x_i)\cdot\sqrt{\Theta[MH_\kappa^{\leqslant n}Z](x)}$ then,

$$\mathcal{M}^*_V[S,T] = \langle V_S^*(x,Z), V_T^*(x,Z)\rangle$$

where $\langle f(x,Z), g(x,Z)\rangle = \mathbf{E}_Z \mathbf{E}_{x\in\{-1,1\}^n}[f(x,Z)g(x,Z)]$.

The above construction looks similar to (7) and (6) with one important difference. The quantities are a function of the matrix M defining the subspace and a set of redundancies in representation given by Z. In particular, *low degree truncation* $\Theta^{\leqslant D}$ can include truncation in the degree over M and over Z separately.

Somewhat mysteriously, it turns out that by choosing a low degree truncation (in **both** M **and** Z) of both the ideal SDP solution $\mathcal{M}^*$ and the ideal vectors V_S^*, we can recover SDP solution along with an approximate Cholesky factorization (analogous to (6)). While the above discussion describes how we arrive at the definition of the mapping. The proof that the mapping works amounts to showing that the truncated vectors yield an approximate Cholesky factorization of the pseudo-calibrated matrix, which forms the technical heart of the paper. We defer the details of the construction to Section 2.

Bounding Spectral Norm. We exhibit a candidate SoS SDP solution $\mathcal{M}^{(1)}$, and show that there exists a positive semi-definite matrix $\mathcal{M}^{(2)}$ that is close in spectral norm to $\mathcal{M}^{(2)}$. The difference $\mathcal{M}^{(1)} - \mathcal{M}^{(2)}$ is matrix with entries that are low degree polynomials

in the input M, and our goal is to upper bound the spectral norm $\|\mathcal{M}^{(2)} - \mathcal{M}^{(1)}\|$.

As is typical, this involves obtaining spectral norm bounds on matrices whose entries are low degree polynomials. Earlier works on Planted Clique [2, 9] and others have developed technical machinery based on the trace method towards bounding spectral norms. We present a simpler factorization based technique to obtain bounds on spectral norms here. Owing to its simplicity, it is broadly applicable to more complicated ensembles of random matrices such as those arising in sparse d-regular random graphs. Furthermore, in some cases, the technique yields tighter bounds than trace method. For example, consider the following random matrix. Let $A \in \mathbb{R}^{n\times n}$ be a random symmetric matrix with $A_{ii} = 0$ for all i and A_{ij} being independent $\{\pm 1\}$ entry otherwise. Consider the random matrix $B \in \mathbb{R}^{[n]^2\times[n]^2}$ defined as,

$$B[(i_1, i_2), (j_1, j_2)] = A_{i_1 j_1} \cdot A_{i_2 j_1} \cdot A_{i_2 j_2} .$$

The best known bounds for $\|B\|$ using the trace method imply that $\|B\| \leqslant n \cdot (\log n)^c$ for some constant c [9]. On the other hand, the factorization technique outlined in Section 3 can be easily used to obtain a $\Theta(n)$ upper bound (specifically, an upper bound of $4n$).

All our spectral norm bounds are obtained via the factorization method, starting from bounds on the norm of the original matrix A.

2 LIFTS OF A DEGREE-2 PSEUDOEXPECTATION

In this section, we describe how to obtain a degree-4 pseudoexpectation $\widetilde{\mathbb{E}}_4$ from a degree-2 pseudoexpectation $\widetilde{\mathbb{E}}_2$. We specify $\widetilde{\mathbb{E}}_4$ via its pseudomoment matrix $\mathcal{M}$ whose rows and columns are indexed by sets of size at most 2, with $\mathcal{M}[S, T] = \widetilde{\mathbb{E}}_4 \left[x^{S\Delta T}\right]$. Let $\mathcal{M}'$ be the following $n \times n$ submatrix of the degree-2 pseudomoment matrix:

$$\mathcal{M}'[\{i\}, \{j\}] \coloneqq \widetilde{\mathbb{E}}_2[x_i x_j] \qquad i, j \in [n].$$

Since $\mathcal{M}'$ is positive semidefinite, we can write $\mathcal{M}'$ in its Cholesky decomposition $MM^\dagger$ where M is some $n \times n$ matrix.

For each $\kappa \geqslant n$ that is a power of 2, let $H_\kappa^{\leqslant n}$ denote the $n \times \kappa$ matrix obtained by taking the first n rows of the Hadamard matrix H_κ. We first define a $n \times \kappa$ matrix $M_\kappa \coloneqq MH_\kappa^{\leqslant n}$. A key property of M we use is:

Fact 1. *$\langle M[i], M[j]\rangle = \langle M_\kappa[i], M_\kappa[j]\rangle$ where $M[t]$ denotes the t-th row of M since the rows of $H_\kappa^{\leqslant n}$ are orthogonal unit vectors.*

Fix a set of indeterminates $z_1, \ldots, z_\kappa$ obeying $z_i^2 = 1$. For each $i \in [n]$, we define "seed polynomials"

$$q_{i,\kappa}(z) \coloneqq \sum_{j\in[\kappa]} M_\kappa[i, j]z_j - 2 \sum_{\{j_1,j_2,j_3\}\subseteq[\kappa]} M_\kappa[i, j_1]M_\kappa[i, j_2]M_\kappa[i, j_3]z_{j_1}z_{j_2}z_{j_3}$$

and for each subset $S \subseteq [n]$ define "set polynomials"

$$q_{S,\kappa}(z) \coloneqq \prod_{i\in S} q_{i,\kappa}(z).$$

We now define matrix $\mathcal{M}^{(1)}$ as follows:

$$\mathcal{M}^{(1)}[S, T] \coloneqq \lim_{\kappa\to\infty} \mathbf{E}_{z\sim\{\pm1\}^\kappa}[q_{S\Delta T,\kappa}(z)] \tag{10}$$

We pick our pseudomoment matrix $\mathcal{M}$ as a mild adjustment to $\mathcal{M}^{(1)}$. Specifically, we define

$$\mathcal{M} \coloneqq (1 - \eta)\mathcal{M}^{(1)} + \eta \cdot \mathrm{Id}.$$

where we choose η later.

It is clear that $\mathcal{M}$ satisfies the "Booleanness" and "symmetry" constraints. It remains to prove that $\mathcal{M}$ is positive semidefinite for appropriate choice of η.

Towards doing so, we define a new matrix $\mathcal{M}^{(2)}$. Define "truncated polynomials"

$$p_{S,\kappa}(z) \coloneqq q_{S,\kappa}(z)^{\leqslant|S|}$$

where $q_{S,\kappa}(z)^{\leqslant\tau}$ denotes the projection of $q_{S,\kappa}$ onto the space of polynomials spanned by χ_T where $|T| \leqslant \tau$. And define $\mathcal{M}^{(2)}$ as:

$$\mathcal{M}^{(2)}[S, T] \coloneqq \lim_{\kappa\to\infty} \mathbf{E}_{z\sim\{\pm1\}^\kappa}[p_{S,\kappa}(z)p_{T,\kappa}(z)] \tag{11}$$

Once again, we defer the proof that the limit on the right-hand side exists to the full paper. $\mathcal{M}^{(2)}$ is PSD as it is the limit of second moment matrices, each of which is PSD.

To show $\mathcal{M}$ is PSD, we first bound the spectral norm of $\mathcal{M}^{(1)} - \mathcal{M}^{(2)}$.

Lemma 2.1. *Let $\alpha_{\mathrm{spec}} \coloneqq \|\mathcal{M}'\|_2$, $\alpha_{\mathrm{mag}} \coloneqq \max_{i,j:i\neq j} \mathcal{M}'[i, j]$, $\alpha_{\mathrm{row}} \coloneqq \max_{i\in[n]} \sqrt{\Sigma_{j\neq i}\mathcal{M}'[i, j]^2}$. There is an absolute constant $C > 0$ such that $\alpha \coloneqq C\alpha_{\mathrm{mag}} \cdot (1+\alpha_{\mathrm{row}}^4) \cdot (1+\alpha_{\mathrm{spec}}^2)$ and $\|\mathcal{M}^{(1)} - \mathcal{M}^{(2)}\|_2 \leqslant \alpha$.*

Lemma 2.1 is an immediate consequence of Lemma 3.1, which Section 3 is dedicated to proving.

Corollary 2.2. *Let α be as in the statement of Lemma 2.1. Then $\lambda_{\min}(\mathcal{M}^{(1)}) \geqslant -\alpha$.*

Set $\eta \coloneqq \frac{\alpha}{1+\alpha}$. The PSDness of $\mathcal{M}$ follows from Corollary 2.2 and the fact that adding $\eta \cdot \mathrm{Id}$ to any matrix increases all its eigenvalues by η.

Theorem 2.3. *$\mathcal{M} \succeq 0$.*

Lemma 2.4. *Let α be as in the statement of Lemma 2.1. For any Hermitian matrix $A \in \mathbb{R}^{n\times n}$,*

$$\widetilde{\mathbb{E}}_4[x^\dagger Ax] \geqslant \left(1 - \frac{\alpha}{1+\alpha}\right)\left(\widetilde{\mathbb{E}}_2[x^\dagger Ax] - \alpha\sqrt{n}\|A\|_F\right) + \frac{\alpha}{1+\alpha}\mathrm{Trace}\,(A).$$

3 SPECTRAL NORM BOUNDS

This section is dedicated to proving Lemma 2.1. We first make some structural observations about $\mathcal{E} \coloneqq \mathcal{M}^{(1)} - \mathcal{M}^{(2)}$.

Observation 1. *Suppose $|S\Delta T|$ is odd. Then $\mathcal{E}[S, T] = 0$.*

Proof. Since $q_{i,\kappa}(z)$ is a sum over odd degree terms in z, so is $q_{S\Delta T,\kappa}(z)$ when $|S\Delta T|$ is odd, and so the expected value of each term over the choice of random z is 0. Thus, $\mathcal{M}_\kappa^{(1)}[S, T] = 0$, and by extension $\mathcal{M}^{(1)}[S, T] = 0$. Note that for any set S all terms in $p_{S,\kappa}$ have the same parity as $|S|$, and thus all terms in $p_{S,\kappa}p_{T,\kappa}$ have the same parity as $|S| + |T|$, whose parity is the same as $|S\Delta T|$. Thus, $\mathcal{M}_\kappa^{(2)}[S, T] = 0$ and consequently $M^{(2)}[S, T] = 0$. □

Observation 2. *Suppose $S = \emptyset$ or $T = \emptyset$. Then $\mathcal{E}[S,T] = 0$.*

Thus, we can split $\mathcal{E}$ into four parts.

$$\mathcal{E}^{(1)}[S,T] := \begin{cases} \mathcal{E}[S,T] & S = T \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{E}^{(2)}[S,T] := \begin{cases} \mathcal{E}[S,T] & \text{if } |S| = |T| = 1, |S \cap T| = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{E}^{(3)}[S,T] := \begin{cases} \mathcal{E}[S,T] & \text{if } |S| = |T| = 2, |S \cap T| = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{E}^{(4)}[S,T] := \begin{cases} \mathcal{E}[S,T] & \text{if } |S| = |T| = 2, |S \cap T| = 0 \\ 0 & \text{otherwise} \end{cases}$$

Since $\mathcal{E} = \mathcal{E}^{(1)} + \mathcal{E}^{(2)} + \mathcal{E}^{(3)} + \mathcal{E}^{(4)}$, proving a spectral norm bound on each individual piece also gives a bound of the spectral norm of $\mathcal{E}$ via the triangle inequality. In later parts of the section, the following are proved.

Lemma 3.1. *The following spectral norm bounds hold:*

$$\|\mathcal{E}^{(1)}\| \leqslant O(\alpha_{\text{mag}})$$
$$\|\mathcal{E}^{(2)}\| \leqslant O(\alpha_{\text{row}}^2 \cdot \alpha_{\text{mag}})$$
$$\|\mathcal{E}^{(3)}\| \leqslant O(\alpha_{\text{mag}} \cdot (1 + \alpha_{\text{spec}} + \alpha_{\text{row}}^2))$$
$$\|\mathcal{E}^{(4)}\| \leqslant O(\alpha_{\text{mag}} \cdot (1 + \alpha_{\text{row}}^4) \cdot (1 + \alpha_{\text{spec}}^2)).$$

In particular, this implies $\|\mathcal{E}\| \leqslant O(\alpha_{\text{mag}} \cdot (1 + \alpha_{\text{row}}^4) \cdot (1 + \alpha_{\text{spec}}^2))$.

The full for this lemma is highly mechanical and technical, and we defer the interested readers to the full version of this paper while we give an outline of the graphical matrices, an important ingredient in our analysis.

3.1 Graphical Polynomials and Graphical Matrices

Akin to [2], we give a way to associate matrices with constant sized graphs. To motivate studying graphical matrices, we start with some simple examples. Let H be some graph with vertex set $[n]$. Now, consider the graph $\mathcal{G}$ in the figure below.

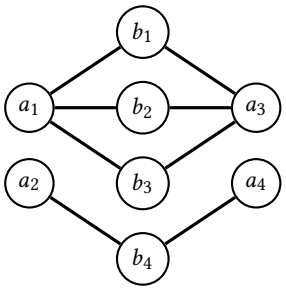


Figure 1: Graph $\mathcal{G}$

Next, define an $n^2 \times n^2$ matrix $Q_{\mathcal{G}}$, which is the "graphical matrix" of $\mathcal{G}$ with rows and columns indexed by size-2 subsets of $[n]$ where

$$Q_{\mathcal{G}}[\{i,j\},\{k,\ell\}] := \#\{\text{subgraphs of } H \text{ isomorphic to } \mathcal{G} \text{ so that } a_1, a_2, a_3, a_4 \text{ map to } i, j, k, \ell\}.$$

Our reason for considering matrices that encode 'constant-sized graph statistics' such as the above, which we call *graphical matrices*, is that we are able to naturally view $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ as a sum of simple graphical matrices[5]. Thus, a natural way to obtain a handle on the spectral norm of $\mathcal{M}^{(1)} - \mathcal{M}^{(2)}$ is understanding the spectral behavior of the graphical matrices that constitute it.

3.1.1 Sketch of Graphical Matrices.

We dig into the specific graphical matrices that arise in this section. We view the matrix M_κ as a weighted bipartite graph with left vertex set $[n]$ and right vertex set $[\kappa]$, where the weight of the edge between $i \in [n]$ and $j \in [\kappa]$ is $M_\kappa[i,j]$ — we call this Bipartite(M_κ). Now, let $\mathcal{G}$ be a bipartite graph on constant number of vertices where each left vertex of $\mathcal{G}$ is one or two of two colors, `row` or `column`, and each right vertex is uncolored. The *graphical matrix* associated with $\mathcal{G}$ is the $n^{|\texttt{row}(\mathcal{G})|} \times n^{|\texttt{column}(\mathcal{G})|}$ matrix $Q_{\mathcal{G}}$ with rows and columns indexed by subsets of $[n]$ of size $|\texttt{row}(\mathcal{G})|$ and $|\texttt{column}(\mathcal{G})|$ respectively where we obtain the S,T entry in the following way.

> Enumerate over all subgraphs of Bipartite(M_κ) that are isomorphic to $\mathcal{G}$, and vertices colored `row` map into S and the vertices colored `column` map into T, take the product of edge weights of each subgraph, and then take the sum over all subgraphs enumerated over.

Symbolically,

$$Q_{\mathcal{G},\kappa}[S,T] := \sum_{\substack{\mathcal{H} \text{ subgraph of Bipartite}(M_\kappa) \\ \mathcal{H} \text{ isomorphic to } \mathcal{G} \\ \texttt{row}(\mathcal{G}) \text{ maps into } S \\ \texttt{column}(\mathcal{G}) \text{ maps into } T}} \prod_{\{i,j\} \in \mathcal{H}} M_\kappa[i,j].$$

3.1.2 Definitions.

Definition 3.2 (Half-Glyph). A half-glyph $\mathcal{HG}$ is a bipartite (multi-)graph with a left vertex set $L(\mathcal{HG}) := \{\ell_1, \ldots, \ell_{|L(\mathcal{HG})|}\}$, a middle vertex set $M(\mathcal{HG}) := \{m_1, \ldots, m_{|M(\mathcal{HG})|}\}$ and edges $E(\mathcal{HG})$. We use $\mathcal{HG}_{a,b}$ to represent the number of edges between ℓ_a and m_b.

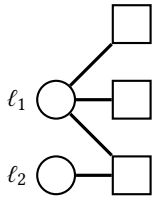
Definition 3.3 (Half-Glyph Labeling). For a half-glyph $\mathcal{HG}$, we call $S : L(\mathcal{HG}) \to [n]$ a valid labeling if

(1) It is a injective map from $L(\mathcal{HG})$ to $[n]$.
(2) $S(\ell_i) < S(\ell_j)$ if and only if $i < j$.[7]

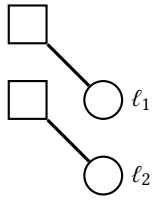
Remark 1. *For simplicity, we represent each valid labeling as a size-$|L(\mathcal{HG})|$ subset of $[n]$.*

Definition 3.4 (Cluster of $M(\mathcal{HG})$). For a half-glyph $\mathcal{HG}$, we call a set of vertices $\{v_1, \ldots, v_{|B|}\}$ in cluster B if they have the same neighborhood on $L(\mathcal{HG})$, i.e., $\forall i,j \in B$, $\mathcal{HG}(\ell,i) = \mathcal{HG}(\ell,j)$ for any $\ell \in L(\mathcal{HG})$. We let $\mathcal{B}(\mathcal{HG}) = (B_1, \ldots, B_k)$ be the set of clusters in $\mathcal{HG}$ where $k \leqslant \kappa$ is the number of clusters.

[5] Where H is replaced with a complete (n, L)-bipartite graph, and the edges are equipped with weights from the matrix M_κ.
[6] We will use circles to represent vertices in $L(\mathcal{HG})$ (and later $L(\mathcal{G})$ and $R(\mathcal{G})$) that should be thought as vertices in $[n]$ and square to represent vertices in $M(\mathcal{G})$ that should be thought as indeterminates z.
[7] This "order-preserving" requirement is an artifact of our proof.

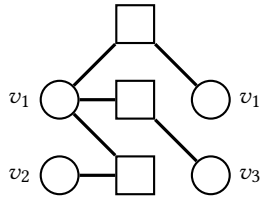


(a) Half-Glyph $\mathcal{HG}_1$

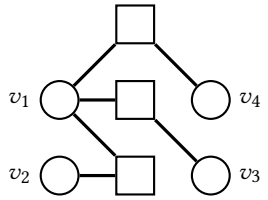


(b) Half-Glyph $\mathcal{HG}_2$

Figure 2: Half-Glyphs [6]



(a) Glyph $\mathcal{G}_1$



(b) Glyph $\mathcal{G}_2$

Figure 3: Glyphs

Definition 3.5 (z-labeling of half-glyph). We say $\pi : M(\mathcal{HG}) \to [\kappa]$ is a z-labeling if it is injective, and if for each cluster $B_i \in \mathcal{B}(\mathcal{HG})$ and $m_a, m_b \in B_i$, $\pi(m_a) < \pi(m_b)$ if and only if $a < b$. We denote the set of z-labelings by $\Pi(\mathcal{HG})$.

Definition 3.6 (κ-Graphical Polynomial of a Half-Glyph). For any κ, every half-glyph $\mathcal{HG}$ with a valid labeling S is associated with a polynomial over indeterminates $z = (z_1, \ldots, z_\kappa)$ given by

$$\beta_{\mathcal{HG},\kappa,S}(z) \coloneqq \sum_{\pi \in \Pi(\mathcal{HG})} \Pi_{i \in L(\mathcal{HG})} \prod_{j \in M(\mathcal{HG})} (M_\kappa[S(i), \pi(j)] \cdot z_{\pi(j)})^{\mathcal{HG}_{i,j}}$$

Definition 3.7 (Glyph). A glyph $\mathcal{G}$ is a multi-graph on the vertex set $V(\mathcal{G}) = L(\mathcal{G}) \cup M(\mathcal{G}) \cup R(\mathcal{G})$ and edge set $E(\mathcal{G})$, where $L(\mathcal{G}) \cup R(\mathcal{G}) = \{v_1, v_2, \ldots, v_{|L(\mathcal{G})\cup R(\mathcal{G})|}\}$ and $M(\mathcal{G}) = \{m_1, m_2, \ldots, m_{|M(\mathcal{G})|}\}$. We use $\mathcal{G}_{a,b}$ to represent the number of edges between v_a and m_b.

Remark 2. *Our definition of cluster and z-labeling for half-glyph extends naturally to glyph.*

We will refer to $L(\mathcal{G})$ as *left* vertices, $M(\mathcal{G})$ as *middle* vertices, and $R(\mathcal{G})$ as *right* vertices of the glyph. We emphasize that $L(\mathcal{G})$ and $R(\mathcal{G})$ need not be disjoint; in particular some vertices can be both *left* and *right* vertices. In the following figure, $\mathcal{G}_1$ and $\mathcal{G}_2$ are different glyphs because L and R intersect in $\mathcal{G}_1$ but not in $\mathcal{G}_2$.

Observe that any glyph can be seen as being "composed" of two half-glyphs: the *left half-glyph* $\mathcal{L}(\mathcal{G})$ which is the induced subgraph on $L(\mathcal{G}) \cup M(\mathcal{G})$, and the *right half-glyph* $\mathcal{R}(\mathcal{G})$ which is the induced subgraph on $R(\mathcal{G}) \cup M(\mathcal{G})$. We now extend the definition of labeling and graphical polynomial to glyphs.

Definition 3.8 (Glyph Labeling). For any glyph $\mathcal{G}$, let S be a valid labeling for $\mathcal{L}(\mathcal{G})$, and T be a valid labeling for $\mathcal{R}(\mathcal{G})$, S and T are *$\mathcal{G}$-compatible* if they agree on $L(\mathcal{G}) \cap R(\mathcal{G})$, i.e. $S|_{L(\mathcal{G})\cap R(\mathcal{G})} = T|_{L(\mathcal{G})\cap R(\mathcal{G})}$ and are disjoint on their symmetric difference, i.e. $S(L(\mathcal{G}) \setminus R(\mathcal{G})) \cap T(R(\mathcal{G}) \setminus L(\mathcal{G})) = \emptyset$. For two $\mathcal{G}$-compatible labelings S and T, let $S \circ T : L(\mathcal{G}) \cup R(\mathcal{G}) \to [n]$ denote the joint labelling induced by both.

Definition 3.9 (κ-Graphical Polynomial of a Glyph). For any κ, for a glyph $\mathcal{G}$ with half-glyphs $\mathcal{L}(\mathcal{G})$ and $\mathcal{R}(G)$ and a pair of compatible labelings S, T, we associate it with a polynomial over indeterminates $z = (z_1, \ldots, z_\kappa)$ given by

$$\beta_{\mathcal{G},\kappa,S\circ T}(z) \coloneqq \sum_{\pi \in \Pi(\mathcal{G})} \prod_{i \in L(\mathcal{G})\cup R(\mathcal{G})} \prod_{j \in M(\mathcal{G})} (M_\kappa[S \circ T(i), \pi(j)] \cdot z_{\pi(j)})^{\mathcal{G}_{i,j}}$$

Definition 3.10. A glyph $\mathcal{G}$ is called *well-glued* if every middle vertex has even degree.

Remark 3. *The κ-graphical polynomial of a well-glued glyph does not depend on z. Specifically,*

$$\beta_{\mathcal{G},\kappa,S\circ T} = \sum_{\pi \in \Pi(\mathcal{G})} \prod_{i \in L(\mathcal{G})\cup R(\mathcal{G})} \prod_{j \in M(\mathcal{G})} M_\kappa[S \circ T(i), \pi(j)]^{\mathcal{G}_{i,j}}$$

Definition 3.11 (κ-Graphical Matrix of a Well-Glued Glyph). For each well-glued glyph $\mathcal{G}$, we associate a matrix indexed by $\binom{[n]}{L(\mathcal{G})} \times \binom{[n]}{R(\mathcal{G})}$ defined as

$$Q_{\mathcal{G},\kappa}[S, T] \coloneqq 1[S, T \text{ are } \mathcal{G}\text{-compatible}] \cdot \beta_{\mathcal{G},\kappa,S\circ T}$$

which we call the κ-graphical matrix of $\mathcal{G}$.

Claim 1. *Let $\mathcal{G}$ be a well-glued (A, B)-glyph. The limit $\lim_{\kappa\to\infty} Q_{\mathcal{G},\kappa}$ exists.*

We defer the proof of the claim to Appendix **??**.

Definition 3.12 (Graphical matrix of a well-glued glyph). For a well-glued glyph $\mathcal{G}$, we call the matrix

$$Q_{\mathcal{G}} \coloneqq \lim_{\kappa\to\infty} Q_{\mathcal{G},\kappa}$$

the *graphical matrix* of $\mathcal{G}$.

Definition 3.13. Given a well-glued glyph $\mathcal{G}$ and a length-2 walk that starts at $u \in L(\mathcal{G}) \cup R(\mathcal{G})$, takes an edge to middle vertex $m \in M(\mathcal{G})$, and takes a different edge from m to $u' \in L(\mathcal{G}) \cup R(\mathcal{G})$. We call the length-2 walk a *cyclic walk* if $u = u'$; otherwise, we call it an *acyclic walk*.

We also give an explicit expression for the entries of $Q_{\mathcal{G}}$.

Lemma 3.14. *Let $\mathcal{G}$ be a well-glued glyph. Suppose any middle vertex of $\mathcal{G}$ has degree $\geqslant 4$, $Q_{\mathcal{G}} = 0$. Suppose all middle vertices of $\mathcal{G}$ have degree 2 and $S \circ T$ is a valid labeling of $\mathcal{G}$ and for $i, j \in L(\mathcal{G}) \cup R(\mathcal{G})$ let $\mathcal{P}_{i,j}$ be the collection of length-2 walks from i to j. Then:*

$$Q_{\mathcal{G}}[S,T] = \Pi_{i \leqslant j \in L(\mathcal{G}) \cup R(\mathcal{G})} \frac{\langle M[S \circ T(i)], M[S \circ T(j)] \rangle^{|\mathcal{P}_{i,j}|}}{|\mathcal{P}_{i,j}|!}.$$

3.2 Glyph Factorization and Spectral Norm Bound

A useful ingredient towards our goal is a generic way to bound the spectral norm of a graphical matrix. In Lemma 3.14, we show that the entries of the graphical matrix of a well-glued graph can be written as a product of inner products. We use this insight to factor the graphical matrices we need to deal with into simpler matrices. We start with a few basic definitions of types of simple matrices we encounter.

Definition 3.15 (Growth and shrinkage matrices). We call a matrix a *growth* matrix if it is block-diagonal and each block is a subrow of $MM^\dagger$. We define a *shrink* matrix as one that can be written as the transpose of a growth matrix.

Definition 3.16 (Swap matrices). We call a matrix a *swap* matrix if it is block diagonal and each block can be written as either (a) $W - \mathrm{Id}$ where W is a principal submatrix of $MM^\dagger$, or (b) W where W is a (not necessarily principal) submatrix of $MM^\dagger$.

Definition 3.17 (Residue matrices). We call a matrix a *residue* matrix if it is a diagonal matrix and each entry is an off-diagonal entry of $MM^\dagger$.

Lemma 3.18. *If $\mathcal{L}$ is a growth/shrinkage matrix, its spectral norm is bounded by α_{row}; if it is a swap matrix, its spectral norm is bounded by α_{spec}; and if it is a residue matrix, its spectral norm is bounded by α_{mag}.*

Before jumping into the full proof, we illustrate the efficacy of our method on the following toy example that will appear in our analysis of $\mathcal{E}^{(4)}$. Consider the following glyph $\mathcal{G}$ with entries:

$$Q_{\mathcal{G}}[\{i,j\},\{k,\ell\}] = \frac{1}{3!}\langle M[i], M[k] \rangle^3 \langle M[j], M[\ell] \rangle$$

for $i, j, k, \ell \in [n]$ distinct and $i < j$, $k < \ell$.

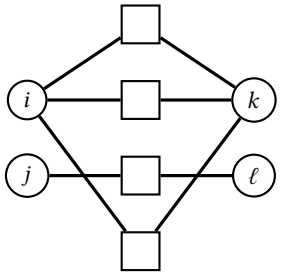


Figure 4: Glyph $\mathcal{G}$

$Q_{\mathcal{G}}$ can be written as a product of simpler matrices — define matrices $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4$ as follows. For all i, j, k, ℓ distinct in $[n]$ with $i < j$ and $k < \ell$,

$$\begin{aligned}
\mathcal{L}_1[\{i,j\},\{i,j,k\}] &:= \langle M[i], M[k] \rangle \\
\mathcal{L}_2[\{i,j,k\},\{i,j,k\}] &:= \langle M[i], M[k] \rangle \\
\mathcal{L}_3[\{i,j,k\},\{j,k\}] &:= \langle M[i], M[k] \rangle \\
\mathcal{L}_4[\{j,k\},\{k,\ell\}] &:= \langle M[j], M[\ell] \rangle
\end{aligned}$$

The above matrices are set to 0 wherever they are undefined. It can be verified that

$$Q_{\mathcal{G}} = \mathcal{L}_1 \cdot \mathcal{L}_2 \cdot \mathcal{L}_3 \cdot \mathcal{L}_4$$

A major advantage of glyph factorization is that it offers a unified framework to bound the spectral norm of graphical matrices of the complex glyphs in terms of spectral norms of simpler matrices. In our example, we have

$$\left\|Q_{\mathcal{G}}\right\| \leqslant \|\mathcal{L}_1\| \cdot \|\mathcal{L}_2\| \cdot \|\mathcal{L}_3\| \cdot \|\mathcal{L}_4\|.$$

We wrap up by giving spectral norm bounds on $\mathcal{L}_i$, and we will generalize from them all the basic glyphs that we will use throughout this section.

Bounding $\|\mathcal{L}_1\|$ and $\|\mathcal{L}_3\|$. $\mathcal{L}_1$ and $\mathcal{L}_3$ are growth and shrinkage matrices respectively and hence their spectral norms are bounded by α_{row}.

Bounding $\|\mathcal{L}_2\|$. $\mathcal{L}_2$ is a residue matrix and hence its spectral norm is at most α_{mag}.

Bounding $\|\mathcal{L}_4\|$. $\mathcal{L}_4$ is a swap matrix and hence its spectral norm is at most α_{spec}.

Combining the above gives $\|Q_{\mathcal{G}}\| \leqslant \alpha_{\mathrm{row}}^2 \cdot \alpha_{\mathrm{mag}} \cdot \alpha_{\mathrm{spec}}$. More generally:

Lemma 3.19. *Let $\mathcal{G}$ be a well-glued glyph whose graphical matrix factorizes as $Q_{\mathcal{G}} = \mathcal{L}_1 \cdot \ldots \cdot \mathcal{L}_k$ where each $\mathcal{L}_i$ is either a growth/shrinkage/swap residue matrix. Let the number of growth (or shrinkage) matrices be t_1, the number of residue matrices be t_2, and the number of swap matrices be t_3, then*

$$\|Q_{\mathcal{G}}\| \leqslant \alpha_{\mathrm{row}}^{t_1} \cdot \alpha_{\mathrm{mag}}^{t_2} \cdot \alpha_{\mathrm{spec}}^{t_3}.$$

4 DEGREE-4 SOS LOWER BOUND FOR THE SHERRINGTON–KIRKPATRICK HAMILTONIAN

4.1 Gaussian Concentration

In this section, we give a brief review of standard concentration results related to Gaussian random variables, vectors, and matrices.

As in previous sections, let $\boldsymbol{M}$ be a $n \times d$ matrix where each entry is independently sampled from $\mathcal{N}\left(0, \frac{1}{d}\right)$ and assume $d < n$.

Lemma 4.1 (Concentration of singular values of Gaussian matrices, [39, Corollary 5.35]). *Except with probability $2\exp\left(-\frac{t^2}{2}\right)$,*

$$\frac{\sqrt{n} - \sqrt{d} - t}{\sqrt{d}} \leqslant s_{\min}(\boldsymbol{M}) \leqslant s_{\max}(\boldsymbol{M}) \leqslant \frac{\sqrt{n} + \sqrt{d} + t}{\sqrt{d}}.$$

Corollary 4.2. *Except with probability* $2\exp\left(\frac{-t^2}{2}\right)$,

$$\|\boldsymbol{M}\boldsymbol{M}^\dagger\| \leqslant \|\boldsymbol{M}\|^2 \leqslant \frac{n+d+2\sqrt{dn}+t^2+2(\sqrt{d}+\sqrt{n})t}{d}$$

Corollary 4.3. *Except with probability* n^{-100}, *for all* i,

$$\langle \boldsymbol{M}_i, \boldsymbol{M}_i\rangle \in \left[1-100\sqrt{\frac{\log n}{d}}, 1+100\sqrt{\frac{\log n}{d}}\right]$$

Lemma 4.4. *Except with probability at least* n^{-100}, *for all pairs of distinct* i, j,

$$\langle \boldsymbol{M}_i, \boldsymbol{M}_j\rangle \in \left[-100\sqrt{\frac{\log n}{d}}, 100\sqrt{\frac{\log n}{d}}\right].$$

Lemma 4.5 ($\frac{d}{n}\boldsymbol{M}\boldsymbol{M}^\dagger$ approximates a projection matrix). *With probability at least* $1-2e^{-d/2}$, *for all* $x \in \mathbb{R}^n$,

$$x^\dagger \boldsymbol{M}\left(\boldsymbol{M}^\dagger\boldsymbol{M}\right)^{-1}\boldsymbol{M}^\dagger x = \left(1 \pm O\left(\sqrt{\frac{d}{n}}\right)\right)\frac{d}{n}x^\dagger \boldsymbol{M}\boldsymbol{M}^\dagger x.$$

Note: $\boldsymbol{M}\left(\boldsymbol{M}\boldsymbol{M}^\dagger\right)^{-1}\boldsymbol{M}^\dagger x$ *is the projection matrix onto the column space of* $\boldsymbol{M}$.

Lemma 4.6. *With probability at least* $1-2e^{-t^2/2}$,

$$\left\|\boldsymbol{M}\boldsymbol{M}^\dagger\right\|_F^2 \geqslant \left(1-4\frac{\sqrt{d}+t}{\sqrt{n}}\right)\frac{n^2}{d}$$

4.2 Degree-2 Pseudoexpectation for SubspaceBooleanVector

We call the following problem SubspaceBooleanVector. Given a $n\times d$ matrix $\boldsymbol{M}$ where each entry is independently sampled from $\mathcal{N}\left(0, \frac{1}{d}\right)$, certify an upper bound on $x^\dagger \boldsymbol{M}\boldsymbol{M}^\dagger x$ over the boolean hypercube. Let $\boldsymbol{M}$ be a $n\times d$ matrix where each entry is independently sampled from $\mathcal{N}\left(0, \frac{1}{d}\right)$. The degree-2 Sum-of-Squares relaxation is as follows:

$$\max_{\widetilde{\mathbf{E}} \text{ degree-2}} \widetilde{\mathbf{E}}[x^\dagger \boldsymbol{M}\boldsymbol{M}^\dagger x] \qquad \text{s.t. } \widetilde{\mathbf{E}}[x_i^2] = 1.$$

Lemma 4.7. *Except with probability* n^{-90}, *there is a degree-2 pseudoexpectation* $\widetilde{\mathbf{E}}$ *with pseudomoment matrix* $\mathcal{M}$ *such that its maximum magnitude off-diagonal entry is at most* $100\sqrt{\frac{\log n}{d}}$, *the* ℓ_2 *norms of its rows are bounded by* $\sqrt{\frac{n\log n}{d}}$, *its spectral norm is bounded by* $1.2\frac{n}{d}$, *and*

$$\frac{d}{n}\widetilde{\mathbf{E}}[x^\dagger \boldsymbol{M}\boldsymbol{M}^\dagger x] \geqslant \left(1-O\left(\sqrt{\frac{\log n}{d}}\right)-O\left(\sqrt{\frac{d}{n}}\right)\right)n.$$

Proof. A degree-2 pseudoexpectation $\widetilde{\mathbf{E}}$ (that is due to [28]) can be constructed in the following way. Let $\gamma := 100\sqrt{\frac{\log n}{d}}$.

$$\widetilde{\mathbf{E}}[x^S] = \begin{cases} 1 & \text{when } |S| = 0 \\ 0 & \text{when } |S| = 1 \\ (1-\gamma)\left(\boldsymbol{M}\boldsymbol{M}^\dagger\right)[i,j] & \text{when } S = \{i,j\} \end{cases}$$

The pseudomoment matrix $\mathcal{M}$ of $\widetilde{\mathbf{E}}$ can thus be written as

$$\begin{bmatrix} 1 & 0 \\ 0 & (1-\gamma)\boldsymbol{M}\boldsymbol{M}^\dagger + \boldsymbol{D}\end{bmatrix}$$

where $\boldsymbol{D}$ is some diagonal matrix.

It remains to prove that $\widetilde{\mathbf{E}}$ is a valid Boolean pseudoexpectation. It is clear that $\widetilde{\mathbf{E}}$ satisfies the Booleanness and symmetry constraints. It remains to prove that $\mathcal{M}$ is PSD. And to do so, it suffices to show that $(1-\gamma)\boldsymbol{M}\boldsymbol{M}^\dagger + \boldsymbol{D}$ is PSD. $\boldsymbol{D}[i,i] = 1-(1-\gamma)\boldsymbol{M}\boldsymbol{M}[i,i]$. From Corollary 4.3 along with a union bound over all diagonal entries of $\boldsymbol{D}$ we can conclude that for all $i \in [n]$, $1 \geqslant \boldsymbol{D}[i,i] \geqslant 0$ with probability at least $1-n^{-99}$ which means $\boldsymbol{D}$ is PSD. $(1-\gamma)\boldsymbol{M}\boldsymbol{M}^\dagger$ is clearly PSD, which means $\mathcal{M}$ is PSD.

Next, we determine the objective value attained by $\widetilde{\mathbf{E}}[\cdot]$.

$$\begin{aligned}\frac{d}{n}\widetilde{\mathbf{E}}[x^\dagger \boldsymbol{M}\boldsymbol{M}^\dagger x] &= \frac{d}{n}\langle \boldsymbol{M}\boldsymbol{M}^\dagger, (1-\gamma)\boldsymbol{M}\boldsymbol{M}^\dagger + \boldsymbol{D}\rangle \\ &= \frac{d}{n}\left((1-\gamma)\langle \boldsymbol{M}\boldsymbol{M}^\dagger, \boldsymbol{M}\boldsymbol{M}^\dagger\rangle + \langle \boldsymbol{M}\boldsymbol{M}^\dagger, \boldsymbol{D}\rangle\right) \\ &\geqslant \frac{d}{n}(1-\gamma)\|\boldsymbol{M}\boldsymbol{M}^\dagger\|_F^2.\end{aligned}$$

From Lemma 4.6, the above is at least $(1-\gamma)\left(1-O\left(\sqrt{\frac{d}{n}}\right)\right)n$ except with probability at most n^{-100}.

Finally, we establish bounds on the maximum absolute off-diagonal entry, the row norm, and the spectral norm of $\mathcal{M}$.

From Corollary 4.4 except with probability n^{-100} all off-diagonal entries of $\mathcal{M}$ are bounded in magnitude by $100\sqrt{\frac{\log n}{d}}$; combined with the fact that the diagonal entries are equal to 1, we see that the ℓ_2 norm of each row is bounded by $\sqrt{\frac{n\log n}{d}}$. The spectral norm of $\|\boldsymbol{M}\boldsymbol{M}^\dagger\|$ is bounded by $1.1\frac{n}{d}$ and each $\boldsymbol{D}[i,i]$ is between 0 and 1 except with with probability at most n^{-100}. Thus, the spectral norm of $\mathcal{M}$ is bounded by $1.2\frac{n}{d}$ except with probability at most n^{-100}. □

4.3 Degree-2 Pseudoexpectation for the Sherrington–Kirkpatrick Hamiltonian

Recall that $G \sim \mathrm{GOE}(n)$ and $\boldsymbol{M}$ is a $n\times d$ matrix where each entry is independently sampled from $\mathcal{N}\left(0, \frac{1}{d}\right)$.

Theorem 4.8. *With probability* $1-o_n(1)$, *there is a degree-2 Boolean pseudoexpectation* $\widetilde{\mathbf{E}}$ *such that*

$$\frac{1}{n^{3/2}}\widetilde{\mathbf{E}}[x^\dagger G x] \geqslant 2-o_n(1).$$

The pseudomoment matrix $\mathcal{M}$ *satisfies the following:*

1. *The off-diagonal entries of* $\mathcal{M}$ *are bounded in magnitude by* $100\sqrt{\frac{\log n}{n^{.99}}}$.
2. *The* ℓ_2 *norms of rows of* $\mathcal{M}$ *are bounded by* $\sqrt{n^{.01}\log n}$.
3. *The spectral norm of* $\mathcal{M}$ *is at most* $1.2n^{.01}$.

Towards proving Theorem 4.8 we first recall the following facts from random matrix theory.

Fact 2 ([12, Sec. 1.14]). *The empirical distribution of eigenvalues of any* $G \sim \mathrm{GOE}(n)$ *follows a universal pattern, namely the Wigner*

Semicircle Law. For any real numbers $a \leqslant b$,

$$\frac{1}{n}\#\{i : \lambda_i \in [a,b]\} = (1 \pm o_n(1)) \int_a^b \rho_{sc}(x)dx$$

with probability $1 - o_n(1)$, where $\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{\max(4-x^2, 0)}$.

Corollary 4.9. *For every $\varepsilon > 0$, there is $\delta > 0$ such that $\lambda_{\delta n}(G) \geqslant (2-\varepsilon)\sqrt{n}$ with probability $1 - o_n(1)$. In particular $\lambda_{n^{.99}} \geqslant (2 - o_n(1))\sqrt{n}$.*

Lemma 4.10. *The distribution of the column space of $\boldsymbol{M}$ is that of a d-dimensional uniformly random subspace in $\mathbb{R}^n$.*

Lemma 4.11 ([30]). *Let $G \sim \mathrm{GOE}(n)$. Its sequence of normalized eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ has the same distribution as choosing a uniformly random orthonormal basis of $\mathbb{R}^n$, i.e., the distribution of first choosing unit $\boldsymbol{v}_1$ uniformly at random on $\mathbb{S}^{n-1}$, then choosing unit $\boldsymbol{v}_2$ uniformly at random orthogonal to $\boldsymbol{v}_1$, then choosing unit $\boldsymbol{v}_3$ uniformly at random orthogonal to $\mathrm{span}\{v_1, v_2\}$ and so on.*

Lemma 4.12. *Let V be a uniformly random subspace of $\mathbb{R}^n$ of dimension d, and let Π_V be the projection matrix onto V. With probability $1 - o_n(1)$ there is a degree-2 pseudoexpectation operator $\widetilde{\mathbf{E}}_V[\cdot]$ over polynomials in x on the hypercube $\{\pm 1\}^n$ such that*

$$\widetilde{\mathbf{E}}_V\left[x^\dagger \Pi_V x\right] \geqslant (1 - o_n(1))n.$$

Additionally, the pseudomoment matrix of $\widetilde{\mathbf{E}}$ satisfies identical bounds on its off-diagonal entries, its row norms and its spectral norm as $\mathcal{M}$ from the statement of Lemma 4.7.

Proof of Theorem 4.8. Let $\{\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_{n^{.99}}\}$ be the top δn eigenvalues of G, let V be the subspace spanned by the top $n^{.99}$ eigenvectors of G, and let Π_V be the projection matrix onto V. By Lemma 4.11, V is a uniformly random $n^{.99}$-dimensional subspace of $\mathbb{R}^n$. Let $\widetilde{\mathbf{E}}_V$ be the promised pseudoexpectation from Lemma 4.12.

$$\begin{aligned}
\frac{1}{n^{3/2}}\widetilde{\mathbf{E}}_V[x^\dagger G x] &\geqslant \widetilde{\mathbf{E}}_V\left[\frac{\boldsymbol{\lambda}_{n^{.99}}}{n^{3/2}}\langle \Pi_V, xx^\dagger\rangle\right] \\
&\quad + \widetilde{\mathbf{E}}_V\left[\frac{\lambda_{\min}(G)}{n^{3/2}}\langle \Pi_{V^\perp}, xx^\dagger\rangle\right] \\
&\geqslant (1 - o_n(1))\frac{\boldsymbol{\lambda}_{n^{.99}}}{n^{3/2}}\widetilde{\mathbf{E}}_V\left[x^\dagger \Pi_V x\right] - o_n(1) \\
&\geqslant (1 - o_n(1))\frac{\boldsymbol{\lambda}_{n^{.99}}}{\sqrt{n}} - o_n(1) \text{(by Lemma 4.12)} \\
&\geqslant 2 - o_n(1). \text{(by Corollary 4.9)}
\end{aligned}$$

The bounds on off-diagonal entries, row norms and spectral norm of the pseudomoment matrix of $\widetilde{\mathbf{E}}_V$ follow by plugging in $d = n^{.99}$ into the bounds from Lemma 4.12. □

4.4 Wrap-up

The degree-4 Sum-of-Squares lower bound is then an immediate consequence of Theorem 4.8 and our lifting theorem Theorem 1.2/Theorem 1.3

Theorem 4.13 (Restatement of Theorem 1.4). *Let $G \sim \mathrm{GOE}(n)$. With probability $1 - o_n(1)$, there exists a degree-4 SoS SDP solution with value at least $(2 - o_n(1)) \cdot n^{3/2}$.*

5 DEGREE-4 SOS LOWER BOUND FOR MaxCut IN RANDOM d-REGULAR GRAPHS

In this section, we first give a degree-2 pseudoexpectation for MaxCut in random d-regular graphs, which is used as a "seed" to derive a degree-4 pseudoexpectation from Theorem 1.2 and Theorem 1.3.

This degree-2 pseudoexpectation is only a slight variant of the known construction of [7, 28].

Theorem 5.1. *Let G be a random d-regular graph. For every constant $\varepsilon > 0$ with probability $1 - o_n(1)$ there is a degree-2 Boolean pseudoexpectation $\widetilde{\mathbf{E}}$ such that:*

$$\widetilde{\mathbf{E}}[x^\dagger(-A_G)x] \geqslant (1 - 2\varepsilon - o_n(1))2\sqrt{d-1}n.$$

Additionally, the pseudomoment matrix $\mathcal{M}$ of $\widetilde{\mathbf{E}}$ satisfies the following:

1. *Its row norms are bounded by a constant $\gamma(\varepsilon)$ which only depends on ε.*
2. *Its spectral norm is bounded by constant $\gamma'(\varepsilon)$ which only depends on ε.*
3. *Its off-diagonal entries are bounded in magnitude by $\frac{\gamma''(\varepsilon)}{\sqrt{d}}$ where $\gamma''(\varepsilon)$ is some constant that only depends on ε.*

We first develop some tools and then prove Theorem 5.1 in Section 5.6.

5.1 The [7, 28] Construction

We first revisit the degree-2 pseudoexpectation for Max Cut due to [7, 28]. Given a random d-regular graph G on n vertices, we state the moment matrix of a degree-2 pseudoexpectation. We call a vertex C-*good* if its radius-$(2C+1)$ neighborhood is a tree, and C-*bad* otherwise.

First, we define vector x_v corresponding to vertex v. Let ρ, C, α be constants that we'll set later. If v is C-bad, then we let

$$\boldsymbol{x}_v[u] := \begin{cases} 1 & \text{if } u = v \\ 0 & \text{if } u \neq v, \end{cases}$$

otherwise, we let

$$\boldsymbol{x}_v[u] := \begin{cases} \alpha \cdot \rho^{d_G(u,v)} & \text{if } d_G(u,v) \leqslant C \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we also define a vector $\boldsymbol{x}_\emptyset$ which is orthogonal to all $\{\boldsymbol{x}_v\}_{v \in G}$.

Once ρ, C are chosen, we pick α so that the vectors $\boldsymbol{x}_v$ for C-good v have unit norm. The degree-2 pseudomoment matrix $\mathcal{M}$ is indexed by pairs of sets S, T such that $|S|, |T| \leqslant 1$ and is defined as follows:

$$\mathcal{M}[S,T] := \langle \boldsymbol{x}_S, \boldsymbol{x}_T \rangle.$$

A nice feature of this solution is that one can derive a closed form for $\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle$ when $\{v, w\}$ is an edge between two C-good vertices.

Lemma 5.2. *Let $\{v, w\}$ be an edge in G. If v, w are both C-good, then*

$$\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle = 2 \cdot \left(\frac{d-1}{d}\right) \cdot \rho \cdot \left(1 - \alpha^2 \rho^{2C} d(d-1)^{C-1}\right)$$

otherwise, $\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle = 0$.

Remark 4. *For any $0 < \varepsilon \leqslant 1$, if we choose $\rho = -\frac{1-\varepsilon}{\sqrt{d-1}}$, then for an edge between C-good vertices $\{v, w\}$ we would have*

$$\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle = -\frac{2\sqrt{d-1}(1-\varepsilon)}{d} \cdot \left(1 - \alpha^2 \cdot \left(\frac{d}{d-1}\right) \cdot (1-\varepsilon)^C)\right).$$

One can make $(1-\varepsilon)^C$ arbitrarily small by increasing C, and additionally, increasing C only makes α smaller. Further, since $\frac{d}{d-1} \leqslant \frac{3}{2}$ for $d \geqslant 3$, there exists a choice for C depending only on ε such that

$$\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle \leqslant -(1-2\varepsilon)\frac{2\sqrt{d-1}}{d}.$$

For the purposes of our proof, we will also need bounds on $|\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle|$ when v and w are within distance C of each other. A similar calculation to that in the proof of Lemma 5.2 lets us show:

Lemma 5.3. *Let v and w be any two vertices. We have*

$$|\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle| \leqslant \begin{cases} |\rho|^{d_G(v,w)}(d_G(v,w)+1) & d_G(v,w) \leqslant C \\ 0 & \text{otherwise} \end{cases}$$

Proof. If v or w are C-bad, then $\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle = 0$, in which case the bound holds. Thus, for the rest of the proof we will assume v and w are both C-good. Let a be a C-good vertex and b be a vertex with distance at most C from a. We use P_{ab} denote the unique path of length at most C between vertices a and b.

$$\begin{aligned}
\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle &= \sum_{u \in V(G)} \boldsymbol{x}_v[u] \cdot \boldsymbol{x}_w[u] \\
&= \alpha^2 \sum_{s \in P_{vw}} \sum_{\substack{u \in V(G) \\ d_G(u,v),\ d_G(u,w) \leqslant C \\ s \in P_{vu},\ s \in P_{wu}}} \rho^{d_G(v,w)} \rho^{2d_G(s,u)} \\
&\leqslant \sum_{s \in P_{vw}} |\rho|^{d_G(v,w)} \sum_{\ell=0}^{C} d(d-1)^{\ell-1}\rho^{2\ell} \\
&= \sum_{s \in P_{vw}} |\rho|^{d_G(v,w)} \\
&= |\rho|^{d_G(v,w)} \cdot (d_G(v,w)+1)
\end{aligned}$$

□

5.2 Nonbacktracking Polynomials

We define a sequence of polynomials $g_0, g_1, \ldots$ which we call *nonbacktracking polynomials* below (see, for example, [1]):

Definition 5.4. Let the *nonbacktracking polynomials* be the following sequence of polynomials defined recursively below.

$$\begin{aligned}
g_0(x) &= 1 \\
g_1(x) &= x \\
g_2(x) &= x^2 - d \\
g_t(x) &= x g_{t-1}(x) - (d-1) g_{t-2}(x) \qquad \text{for } t \geqslant 3.
\end{aligned}$$

An elementary fact about nonbacktracking polynomials, which earns them their name is:

Fact 3. *For any d-regular graph G,*

$$g_i(A_G)_{uv} = \#\text{ of nonbacktracking walks from } u \text{ to } v. \tag{12}$$

We will be interested in $g_i(\lambda)$ for eigenvalues λ of A_G. The following can be extracted from [1, Proof of Lemma 2.3]:

Lemma 5.5. *When $x \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$, $|g_i(x)| \leqslant 2(i+1)\sqrt{(d-1)^i}$.*

By a simple continuity argument, this implies:

Corollary 5.6. *For any $\varepsilon > 0$, there exists $\delta > 0$ such that $|g_i(x)| \leqslant 2(i+1)\sqrt{(d-1)^i} + \varepsilon$ when $x \in [-2\sqrt{d-1}-\delta, 2\sqrt{d-1}+\delta]$.*

5.3 Random Graphs

We need the following two facts about random regular graphs.

Lemma 5.7 (Easy consequence of [40, Theorem 2.5]). *Let $d \geqslant 3$ be a fixed constant, let $\boldsymbol{G}$ be a random d-regular graph on n vertices, and let C be any constant. Then w.h.p. the number of C-bad vertices in $\boldsymbol{G}$ is $O(\log n)$.*

Theorem 5.8 (Friedman's theorem [5, 13]). *Let $d \geqslant 3$ be a fixed constant, and let $\boldsymbol{G}$ be a random d-regular graph on n-vertices. Then with probability $1 - o_n(1)$:*

$$\max\{\lambda_2(\boldsymbol{G}), |\lambda_n(\boldsymbol{G})|\} \leqslant 2\sqrt{d-1} + o_n(1).$$

5.4 Construction

Stage 1. First choose constant $\varepsilon > 0$, and let ρ, C, α be chosen according to Remark 4 so that each $\boldsymbol{x}_v$ is a unit vector, and $\langle \boldsymbol{x}_v, \boldsymbol{x}_w \rangle \leqslant -(1-2\varepsilon)\frac{2\sqrt{d-1}}{d}$ for every edge $\{v, w\}$ between two C-good vertices v and w. Next, define polynomial g as follows:

$$g(x) := \alpha \sum_{i=0}^{C} \rho^i g_i(x).$$

Stage 2. Let $\mathcal{W} := g(A_G)^2 - g(d)^2 \cdot \left(\frac{\vec{1}\vec{1}^\dagger}{n}\right)$.

Claim 2. *$\mathcal{W} \succeq 0$. (Proof deferred to the full version of our paper.)*

Stage 3. Let S_G be the collection of C-bad vertices in $\boldsymbol{G}$. Let $\mathcal{W}'$ be the matrix obtained by zeroing out all rows and columns in S_G and then setting $\mathcal{W}'[v, v]$ to 1 for all $v \in V(\boldsymbol{G})$. Symbolically,

$$\mathcal{W}'[v,w] := \begin{cases} 1 & \text{if } v = w \\ \mathcal{W}[v,w] & \text{if } v \neq w \text{ and } v, w \notin S_G \\ 0 & \text{otherwise} \end{cases}$$

Remark 5. *$\mathcal{W}'$ is a PSD matrix since it is a 2×2 block diagonal matrix where each block is PSD. In particular one block, $\mathcal{W}'[S_G, S_G]$, is an identity matrix and is thus PSD. The other block can be seen to satisfy:*

$$\mathcal{W}'[V(G) \setminus S_G, V(G) \setminus S_G] \succeq \mathcal{W}[V(G) \setminus S_G, V(G) \setminus S_G].$$

Thus, the other block is also PSD since it PSD-dominates a principal submatrix of the PSD matrix $\mathcal{W}$.

Remark 6. *Note that while the vectors $\{\boldsymbol{x}_u\}_{u \in V(G)}$ didn't play an explicit role in the construction, they have a role in the analysis.*

5.5 Various Norm Bounds

In this section, we give bounds on the ℓ_2 norm of a subset of indices of rows/columns of $\mathcal{W}'$ and the spectral norm of $\mathcal{W}'$.

Observation 3. *For any pair of vertices v, w,*

$$\left|\langle \boldsymbol{x}_v, \boldsymbol{x}_w\rangle - \mathcal{W}'[v,w]\right| \leqslant \frac{\kappa(\varepsilon,d)}{n}$$

where the $\kappa(\varepsilon,d)$ is a constant depending on ε and d.

Lemma 5.9. *Let $\mathcal{W}'[u]$ be the u-th row of $\mathcal{W}'$. Then when n, the number of vertices in the graph is large enough,*

$$\|\mathcal{W}'[u, V(G)\setminus\{u\}]\|_2 \leqslant \gamma(\varepsilon)$$

where $\gamma(\varepsilon) > 0$ is a constant that depends only on ε chosen in Stage 1 of the construction in Section 5.4.

Next, we upper bound the spectral norm of $\mathcal{W}'$.

Lemma 5.10. *When n, the number of vertices in $V(G)$ is large enough, $\|\mathcal{W}'\| \leqslant \gamma'(\varepsilon)$ where $\gamma'(\varepsilon)$ is a constant that depends only on ε chosen in Stage 1 of the construction in Section 5.4.*

Proof. First, recall the notation S_G to denote the set of C-bad vertices in G and that up to permutation of rows and columns, $\mathcal{W}'$ has the following block diagonal structure:

$$\mathcal{W}' = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

where $A = \mathcal{W}[V(G)\setminus S_G, V(G)\setminus S_G] + \frac{g(d)^2}{n}\cdot \mathrm{Id}$ and B is an identity matrix. Thus, $\|\mathcal{W}'\| \leqslant \max\{\|A\|, \|B\|\}$. We already know that $\|B\| \leqslant 1$, and thus it remains to obtain a bound on $\|A\|$.

$$\begin{aligned}\|A\| &= \|\mathcal{W}[V(G)\setminus S_G, V(G)\setminus S_G]\| + \frac{g(d)^2}{n}\\ &\leqslant \|\mathcal{W}\| + o_n(1)\\ &= \left\|\sum_{i=2}^{n} g(\lambda_i)^2 v_i v_i^\dagger\right\| + o_n(1)\\ &\leqslant \max_{i\in\{2,\ldots,n\}} g(\lambda_i(G))^2 + o_n(1).\end{aligned}$$

Now, recall Friedman's theorem Theorem 5.8, according to which whp $\lambda_2(G),\ldots,\lambda_n(G)$ are all in $[-2\sqrt{d-1}-o_n(1), 2\sqrt{d-1}+o_n(1)]$. Thus it suffices to bound $|g(x)|$ on the specified interval. For the below calculation, assume $x \in [-2\sqrt{d-1}-o_n(1), 2\sqrt{d-1}+o_n(1)]$.

$$\begin{aligned}|g(x)| &\leqslant \alpha\sum_{i=0}^{C}\left(\frac{1-\varepsilon}{\sqrt{d-1}}\right)^i |g_i(x)|\\ &\leqslant \alpha\sum_{i=0}^{C} 2(i+1)\left(\frac{1-\varepsilon}{\sqrt{d-1}}\right)^i \sqrt{(d-1)^i} + o_n(1)\\ &\leqslant 2\alpha\sum_{i=0}^{C}(i+1)(1-\varepsilon)^i + o_n(1)\end{aligned}$$

which bounds $\|A\|$ by a constant $\gamma'(\varepsilon)$ only depending on ε (as C also depends only on ε) when n is large enough. □

5.6 MaxCut Wrap-Up

We are now finally ready to prove Theorem 5.1 and Theorem 1.5.

Proof of Theorem 5.1. Define $\widetilde{\mathbf{E}}$ in the following way:

$$\widetilde{\mathbf{E}}[x^S] = \begin{cases} 1 & \text{if } |S| = 0\\ 0 & \text{if } |S| = 1\\ \mathcal{W}'[u,v] & \text{if } S = \{u,v\}.\end{cases}$$

Its pseudomoment matrix is then

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{W}' \end{bmatrix}$$

and hence is PSD. The bounds on the row norms and spectral norm on $\mathcal{M}$ follow from Lemma 5.9 and Lemma 5.10 respectively and the bound on the magnitude of off-diagonal entries follows from Lemma 5.3 and Observation 3. Finally, we show that the objective value is indeed at least $(1-2\varepsilon-o_n(1))2\sqrt{d-1}n$. Our choice of parameters combined with Observation 3 tells us that $\widetilde{\mathbf{E}}[x_u x_v] \leqslant -(1-2\varepsilon-o_n(1))\frac{2\sqrt{d-1}}{d}$ for edges $\{u,v\}$ between C-good vertices. Since we additionally know that the number of C-bad vertices is $O(\log n)$, the fraction of edges that are between C-good vertices is $1-o_n(1)$. Consequently, it follows that

$$\widetilde{\mathbf{E}}[x^\dagger(-A_G)x] \geqslant (1-2\varepsilon-o_n(1))2\sqrt{d-1}n.$$

□

Theorem 5.11 (Restatement of Theorem 1.5). *Let G be a random d-regular graph. For every constant $\varepsilon > 0$ with probability $1-o_n(1)$, there is a degree-4 SoS SDP solution with* MaxCut *value at least*

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}\left(1 - 2\varepsilon - \frac{\gamma(\varepsilon)}{d^{1/2}}\right)$$

for some constant γ that depends only on ε.

Proof. By applying our lifting theorem Theorem 1.2 to the degree-2 pseudoexpectation $\widetilde{\mathbf{E}}_2$ from Theorem 5.1, we obtain a degree-4 pseudoexpectation $\widetilde{\mathbf{E}}_4$ such that

$$\widetilde{\mathbf{E}}_4[x^\dagger(-A_G)x] \geqslant (1-2\varepsilon-\frac{\gamma(\varepsilon)}{d^{1/2}})2\sqrt{d-1}n \tag{13}$$

where $\gamma(\varepsilon)$ is a constant that depends only on ε. As a result:

$$\begin{aligned}\frac{1}{4|E(G)|}\widetilde{\mathbf{E}}_4[x^\dagger(D_G - A_G)x] &= \frac{dn}{4|E(G)|} + \widetilde{\mathbf{E}}_4[x^\dagger(-A_G)x]\\ &\geqslant \frac{1}{2} + \frac{\sqrt{d-1}}{d}\left(1-\varepsilon-\frac{\gamma(\varepsilon)}{d^{1/2}}\right)\end{aligned}$$

□

ACKNOWLEDGMENTS

S.M. and P.R. are supported by NSF grant CCF-1718695. We would like to thank Vijay Bhattiprolu, Siu On Chan, Sam Hopkins, Ray Li, Ryan O'Donnell, Goutham Rajendran, Tselil Schramm, Nikhil Srivastava, and anonymous reviewers.

REFERENCES

[1] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. 2007. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics* 9, 04 (2007), 585–603.

[2] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. 2019. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. *SIAM J. Comput.* 48, 2 (2019), 687–735.

[3] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. 2016. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. *CoRR* abs/1604.03084 (2016). arXiv:1604.03084 http://arxiv.org/abs/1604.03084

[4] Quentin Berthet and Philippe Rigollet. 2013. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*. 1046–1066.

[5] Charles Bordenave. 2019. *A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts.* Technical Report 1502.04482v4. arXiv. To appear in Annales scientifiques de l'École normale supérieure.

[6] Matthew Brennan, Guy Bresler, and Wasim Huleihel. 2018. Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure. In *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*. 48–166. http://proceedings.mlr.press/v75/brennan18a.html

[7] Endre Csóka, Balázs Gerencsér, Viktor Harangi, and Bálint Virág. 2015. Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures & Algorithms* 47, 2 (2015), 284–303.

[8] Amir Dembo, Andrea Montanari, Subhabrata Sen, et al. 2017. Extremal cuts of sparse random graphs. *The Annals of Probability* 45, 2 (2017), 1190–1217.

[9] Yash Deshpande and Andrea Montanari. 2015. Improved Sum-of-Squares Lower Bounds for Hidden Clique and Hidden Submatrix Problems. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*. 523–562. http://proceedings.mlr.press/v40/Deshpande15.html

[10] Yash Deshpande, Andrea Montanari, Ryan O'Donnell, Tselil Schramm, and Subhabrata Sen. 2019. The threshold for SDP-refutation of random regular NAE-3SAT. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2305–2321.

[11] Andreas Engel and Christian P. L. Van den Broeck. 2001. *Statistical Mechanics of Learning*. Cambridge University Press, New York, NY, USA.

[12] Laszlo Erdős. 2011. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys* 66, 3 (jun 2011), 507–626. https://doi.org/10.1070/rm2011v066n03abeh004749

[13] Joel Friedman. 2003. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM, 720–724.

[14] Dima Grigoriev. 2001. Complexity of Positivstellensatz proofs for the knapsack. *computational complexity* 10, 2 (2001), 139–154.

[15] Dima Grigoriev. 2001. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science* 259, 1 (2001), 613 – 622. https://doi.org/10.1016/S0304-3975(00)00157-2

[16] Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. 2018. On the Integrality Gap of Degree-4 Sum of Squares for Planted Clique. *ACM Trans. Algorithms* 14, 3 (2018), 28:1–28:31. https://doi.org/10.1145/3178538

[17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. 2017. The Power of Sum-of-Squares for Detecting Hidden Structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 720–731.

[18] Samuel B. Hopkins and David Steurer. 2017. Efficient Bayesian Estimation from Few Samples: Community Detection and Related Problems. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*. 379–390. https://doi.org/10.1109/FOCS.2017.42

[19] Jonah Issac-Brown Cohen and Prasad Raghavendra. 2020. Extended Formulation Lower Bounds for Refuting Random CSPs. *ACM Symposium on Discrete Algorithms (SODA)* (2020).

[20] Subhash A Khot and Nisheeth K Vishnoi. 2015. The Unique Games Conjecture, Integrality Gap for Cut Problems and Embeddability of Negative-Type Metrics into âĎŞ 1. *Journal of the ACM (JACM)* 62, 1 (2015), 8.

[21] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. 2017. Sum of Squares Lower Bounds for Refuting Any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* (Montreal, Canada) *(STOC 2017)*. ACM, New York, NY, USA, 132–145. https://doi.org/10.1145/3055399.3055485

[22] Dmitriy Kunisky and Afonso S Bandeira. 2019. A Tight Degree 4 Sum-of-Squares Lower Bound for the Sherrington-Kirkpatrick Hamiltonian. *arXiv preprint arXiv:1907.11686* (2019).

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

[24] Marc Mezard and Andrea Montanari. 2009. *Information, Physics, and Computation*. Oxford University Press, Inc., New York, NY, USA.

[25] M. Mezard, G. Parisi, and M. Virasoro. 1987. Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications *(World Scientific Lecture Notes in Physics)*. World Scientific. https://books.google.com/books?id=DwY8DQAAQBAJ

[26] M. Mézard, G. Parisi, and R. Zecchina. 2002. Analytic and Algorithmic Solution of Random Satisfiability Problems. *Science* 297, 5582 (2002), 812–815. https://doi.org/10.1126/science.1073287 arXiv:https://science.sciencemag.org/content/297/5582/812.full.pdf

[27] Andrea Montanari. 2018. Optimization of the Sherrington-Kirkpatrick Hamiltonian.

[28] Andrea Montanari and Subhabrata Sen. 2016. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM, 814–827.

[29] Hidetoshi Nishimori. 2001. *Statistical Physics of Spin Glasses and Information Processing: an Introduction*. Oxford University Press, Oxford; New York.

[30] Sean O'Rourke, Van Vu, and Ke Wang. 2016. Eigenvectors of Random Matrices. *J. Comb. Theory Ser. A* 144, C (Nov. 2016), 361–442. https://doi.org/10.1016/j.jcta.2016.06.008

[31] G. Parisi. 1979. Infinite Number of Order Parameters for Spin-Glasses. *Phys. Rev. Lett.* 43 (Dec 1979), 1754–1756. Issue 23. https://doi.org/10.1103/PhysRevLett.43.1754

[32] Giovanni P. Parisi. 1980. A sequence of approximated solutions to the S-K model for spin glasses.

[33] Pablo A Parrilo. 2000. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. Ph.D. Dissertation. California Institute of Technology.

[34] Prasad Raghavendra, Satish Rao, and Tselil Schramm. 2017. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*. 121–131. https://doi.org/10.1145/3055399.3055417

[35] Grant Schoenebeck. 2008. Linear level Lasserre lower bounds for certain k-CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 593–602.

[36] N. Z. Shor. 1987. Class of global minimum bounds of polynomial functions. *Cybernetics* 23, 6 (01 Nov 1987), 731–734. https://doi.org/10.1007/BF01070233

[37] Michel Talagrand. 2006. The Parisi formula. *Annals of mathematics* (2006), 221–263.

[38] Madhur Tulsiani. 2009. CSP Gaps and Reductions in the Lasserre Hierarchy. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing* (Bethesda, MD, USA) *(STOC '09)*. ACM, New York, NY, USA, 303–312. https://doi.org/10.1145/1536414.1536457

[39] Roman Vershynin. 2010. Introduction to the non-asymptotic analysis of random matrices.

[40] Nicholas C Wormald. 1999. Models of random regular graphs. (1999).