Sum-of-Squares Lower Bounds for Sherrington-Kirkpatrick via Planted Affine Planes

Mrinalkanti Ghosh
Toyota Technological Institute
Chicago, USA
mkghosh@ttic.edu

Fernando Granha Jeronimo
Computer Science Department
University of Chicago
Chicago, USA
granha@uchicago.edu

Chris Jones
Computer Science Department
University of Chicago
Chicago, USA
csj@uchicago.edu

Aaron Potechin
Computer Science Department
University of Chicago
Chicago, USA
potechin@uchicago.edu

Goutham Rajendran
Computer Science Department
University of Chicago
Chicago, USA
goutham@uchicago.edu

***Abstract*—The Sum-of-Squares (SoS) hierarchy is a semi-definite programming meta-algorithm that captures state-of-the-art polynomial time guarantees for many optimization problems such as Max-k-CSPs and Tensor PCA. On the flip side, a SoS lower bound provides evidence of hardness, which is particularly relevant to average-case problems for which NP-hardness may not be available.**

In this paper, we consider the following average case problem, which we call the *Planted Affine Planes* (PAP) problem: Given m random vectors $d_1, \ldots, d_m$ in $\mathbb{R}^n$, can we prove that there is no vector $v \in \mathbb{R}^n$ such that for all $u \in [m]$, $\langle v, d_u \rangle^2 = 1$? In other words, can we prove that m random vectors are not all contained in two parallel hyperplanes at equal distance from the origin? We prove that for $m \leq n^{3/2-\varepsilon}$, with high probability, degree-$n^{\Omega(\varepsilon)}$ SoS fails to refute the existence of such a vector v.

When the vectors $d_1, \ldots, d_m$ are chosen from the multivariate normal distribution, the PAP problem is equivalent to the problem of proving that a random n-dimensional subspace of $\mathbb{R}^m$ does not contain a boolean vector. As shown by Mohanty–Raghavendra–Xu [STOC 2020], a lower bound for this problem implies a lower bound for the problem of certifying energy upper bounds on the Sherrington-Kirkpatrick Hamiltonian, and so our lower bound implies a degree-$n^{\Omega(\varepsilon)}$ SoS lower bound for the certification version of the Sherrington-Kirkpatrick problem.

The full version of the paper is available at http://arxiv.org/abs/2009.01874.

I. Introduction

The Sum-of-Squares (SoS) hierarchy is a semi-definite programming (SDP) hierarchy which provides a meta-algorithm for polynomial optimization [Las15]. Given a polynomial objective function and a system of polynomial equalities and inequalities as constraints, the SoS framework specifies a family of increasingly "larger" SDP programs, where each program provides a convex relaxation to the polynomial optimization problem. The family is indexed by a size parameter D called the SoS degree. Roughly speaking, the larger the SoS degree D, the tighter the relaxation, but also, the greater the computational time required to solve the convex program, with $D = O(1)$ corresponding to polynomial time and $D = n$ able to exactly solve an optimization problem on n boolean variables. Due to the versatility of polynomials in modeling computational problems, the SoS hierarchy can be applied to a vast range of optimization problems. It has been shown to be quite successful in this regard, as it captures state-of-the-art approximation guarantees for many problems such as Sparsest Cut [ARV04], MaxCut [GW95], Tensor PCA [HSS15] and all Max-k-CSPs [Rag08].

The success of SoS for optimization confers on it an important role as an algorithmic tool. For this reason, on the flip side, understanding the degree range for which SoS *fails* to provide a desired guarantee to a computational problem can be useful to the algorithm designer in two ways. Firstly and more concretely, since SoS is a proof system capturing a broad class of algorithmic reasoning [FKP19], an SoS lower bound can inform the algorithm designer not only of the minimum degree required within the SoS hierarchy, but also to avoid methods of proof that are captured by low-degree SoS reasoning. Secondly, an SoS lower

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00093

bound can serve as strong evidence for computational hardness [HKP+17], [Hop18], even though it is not a formal guarantee against all types of algorithms. This hardness evidence is particularly relevant to average-case problems for which we do not have NP-hardness (see, e.g., the SoS lower bound on the Planted Clique problem [BHK+16]).

Our main results concern the performance of SoS on the following basic optimization problem

$$\mathsf{OPT}(W) := \max_{x\in\{\pm 1\}^n} x^\intercal W x, \tag{1}$$

where W is a symmetric matrix in $\mathbb{R}^{n\times n}$. This problem arises in the fields of computer science and statistical physics, though the terminology can sometimes differ. Computer scientists might regard $x \in \{\pm 1\}^n$ as encoding a bipartition of $[n] = \{1, 2, \ldots, n\}$. Note that by taking W to be a graph Laplacian [HLW06, Section 4] the problem is equivalent to the MaxCut problem, a well-known NP-hard problem in the worst case [Kar72]. A statistical physicist might regard x as encoding spin values in a spin-glass model. The matrix $-W$ is regarded as the *Hamiltonian* of the underlying physical system, where entry $-W_{i,j}$ models the interaction between spin x_i and x_j (with $-W_{i,j} \geq 0$ being ferromagnetic and $-W_{i,j} < 0$ being antiferromagnetic). Then, the optimized x corresponds to the minimum-energy, or ground, state of the system.

Instead of considering $\mathsf{OPT}(W)$ for a worst-case W, one can consider the average-case problem in which W is sampled according to some distribution. One of the simplest models of random matrices is the Gaussian Orthogonal Ensemble, denoted $\mathrm{GOE}(n)$ for n-by-n matrices and defined as follows.

Definition I.1. *The Gaussian Orthogonal Ensemble, denoted* $\mathrm{GOE}(n)$*, is the distribution of* $\frac{1}{\sqrt{2}}(A + A^\intercal)$ *where* A *is a random* $n \times n$ *matrix with i.i.d. standard Gaussian entries.*

Taking $W \sim \mathrm{GOE}(n)$ for the optimization problem $\mathsf{OPT}(W)$ of Eq. (1) gives rise to the so-called Sherrington–Kirkpatrick (SK) Hamiltonian [SK75]. Note that $\mathrm{GOE}(n)$ is a particular kind of Wigner matrix ensemble, thereby satisfying the semicircle law, which in this case establishes that the largest eigenvalue of W is $(2 + \mathrm{o}_n(1)) \cdot \sqrt{n}$ with probability $1 - \mathrm{o}_n(1)$. Thus, a trivial spectral bound establishes $\mathsf{OPT}(W) \leq (2 + \mathrm{o}_n(1)) \cdot n^{3/2}$ with probability $1 - \mathrm{o}_n(1)$. However, in a foundational work based on a variational argument [Par79], Parisi conjectured that

$$\mathop{\mathbb{E}}_{W\sim\mathrm{GOE}(n)} [\mathsf{OPT}(W)] \approx 2 \cdot P^* \cdot n^{3/2},$$

where $P^* \approx 0.7632$ is now referred to as the Parisi constant. In a breakthrough result, Talagrand [Tal06] gave a rigorous proof of Parisi's conjecture [1]. The question then became, "is there a polynomial-time algorithm that given $W \sim \mathrm{GOE}(n)$ computes an x achieving close to $\mathsf{OPT}(W)$?" As it turns out, the answer was essentially shown to be yes by Montanari [Mon19]!

The natural question we study is that of certification: "is there an efficient algorithm to certify an upper bound on $\mathsf{OPT}(W)$ for any input W, that improves upon the trivial spectral bound?" In particular, we can ask how well SoS does as a certification algorithm. The natural upper bound of $(2 + \mathrm{o}_n(1)) \cdot n^{3/2}$ obtained via the spectral norm of W is also the value of the degree-2 SoS relaxation [MS16]. Two independent recent works of Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] show that degree-4 SoS does not perform much better, and a heuristic argument from [BKW19] suggests that even degree-$(n/\log n)$ SoS cannot certify anything stronger than the trivial spectral bound. Thus we ask,

Can higher-degree SoS certify better upper bounds for the Sherrington–Kirkpatrick problem, hopefully closer to the true bound $2 \cdot P^* \cdot n^{3/2}$*?*

Our Results. We answer the question above negatively by showing that even at degree as large as n^δ, SoS cannot improve upon the basic spectral algorithm. More precisely, we have the following theorem which is our first main result and our most important contribution.

Theorem I.2. *[Main I] There exists a constant* $\delta > 0$ *such that, w.h.p. for* $W \sim \mathrm{GOE}(n)$*, there is a degree-*n^δ *SoS solution for the Sherrington–Kirkpatrick problem with value at least* $(2 - \mathrm{o}_n(1)) \cdot n^{3/2}$*.*

In light of the result of Montanari [Mon19], the situation is intriguing. Montanari showed that for all $\varepsilon > 0$, there is a $\mathrm{O}_\varepsilon(n^2)$ time randomized algorithm that given a random W drawn from the Gaussian Orthogonal Ensemble, outputs an x such that $x^T W x \geq (1 - \varepsilon)\mathsf{OPT}(W)$. The correctness of the algorithm assumes a widely-believed con-

[1] The results of Talagrand [Tal06] were for the Sherrington–Kirkpatrick and mixed p-spin systems with p even. In [Pan14], Panchenko generalized these results to arbitrary mixed p-spin systems (also including odd p).

jecture from statistical physics known as the full replica symmetry breaking assumption. However, we show an integrality gap for SoS.

Based on this, it is an interesting question whether SoS, together with an appropriate rounding scheme, is optimal for the Sherrington-Kirkpatrick problem. On the one hand, the situation could be similar to the Feige-Schechtman integrality gap instance for MaxCut [FS02]. For the Feige-Schechtman integrality gap instance, SoS fails to certify the value of the optimal solution. However, applying hyperplane rounding to the SoS solution gives an almost-optimal solution for these instances. It could be the case that there is a rounding scheme which takes an SoS solution for the Sherrington-Kirkpatrick problem on a random W and returns an almost optimal solution x. On the other hand, we currently don't know what this rounding scheme would be.

In order to prove Theorem I.2, we first introduce a new average-case problem we call Planted Affine Planes (PAP) for which we directly prove a SoS lower bound. We then use the PAP lower bound to prove a lower bound on the Sherrington–Kirkpatrick problem. The PAP problem can be informally described as follows (see Definition II.1 for the formal definition).

Definition I.3 (Informal statement of PAP)**.** *Given m random vectors $d_1, \dots, d_m$ in $\mathbb{R}^n$, can we prove that there is no vector $v \in \mathbb{R}^n$ such that for all $u \in [m]$, $\langle v, d_u \rangle^2 = 1$? In other words, can we prove that m random vectors are not all contained in two parallel hyperplanes at equal distance from the origin?*

This problem, when we restrict v to a Boolean vector in $\{\pm\frac{1}{\sqrt{n}}\}^n$, can be encoded as the feasibility of the polynomial system

$$\exists v \in \mathbb{R}^n \text{ s.t.} \qquad \forall i \in [n],\ v_i^2 = \frac{1}{n}, \qquad \forall u \in [m],\ \langle v, d_u \rangle^2 = 1.$$

Hence it is a ripe candidate for SoS. However, we show that SoS fails to refute a random instance. The Boolean restriction on v actually makes the lower bound result stronger since SoS cannot refute even a smaller subset of vectors in $\mathbb{R}^n$. In this work, we will consider two different random distributions, namely when $d_1, \dots, d_m$ are independent samples from the multivariate normal distribution and when they are independent samples from the uniform distribution on the boolean hypercube.

Theorem I.4 (Main II)**.** *For both the Gaussian and Boolean settings, there exists a constant $c > 0$ such that for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $m \le n^{3/2-\varepsilon}$, w.h.p. there is a feasible degree-n^δ SoS solution for Planted Affine Planes.*

It turns out that the Planted Affine Plane problem introduced above is closely related to the following "Boolean vector in a random subspace" problem, which we call the Planted Boolean Vector problem, introduced by [MRX19] in the context of studying the performance of SoS on computing the Sherrington–Kirkpatrick Hamiltonian.

The Planted Boolean Vector problem is to certify that a random subspace of $\mathbb{R}^n$ is far from containing a boolean vector. Specifically, we want to certify an upper bound for

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b,$$

where V is a uniformly random p-dimensional subspace[2] of $\mathbb{R}^n$, and Π_V is the projector onto V. In brief, the relationship to the Planted Affine Plane problem is that the PAP vector v represents the coefficients on a linear combination for the vector b in the span of a basis of V.

An argument of [MRX19] shows that, when $p \ll n$, w.h.p., $\mathsf{OPT}(V) \approx \frac{2}{\pi}$, whereas they also show that w.h.p. assuming $p \ge n^{0.99}$, there is a degree-4 SoS solution with value $1 - o_n(1)$. They ask whether or not there is a polynomial time algorithm that can certify a tighter bound; we rule out SoS-based algorithms for a larger regime both in terms of SoS degree and the dimension p of the random subspace.

Theorem I.5. *[Main III] There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $p \ge n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value 1.*

Our Approach. We now provide a brief high-level description of our approach (see Section III for a more detailed overview). The bulk of our technical contribution lies in the SoS lower bound for the Planted Affine Planes problem, Theorem I.4. We then show that Planted Affine Planes in the Gaussian setting is equivalent to the Planted Boolean Vector problem. The reduction from Sherrington-Kirkpatrick to the Planted Boolean Vector problem is due to Mohanty–Raghavendra–Xu [MRX19].

[2] V can be specified by a basis, which consists of p i.i.d. samples from $\mathcal{N}(0, I)$.

As a starting point to the PAP lower bound, we employ the general techniques introduced by Barak et al. [BHK+16] for SoS lower bounds. We use their pseudocalibration machinery to produce a good candidate SoS solution $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ unfortunately does not exactly satisfy the PAP constraints "$\langle v, d_u\rangle^2 = 1$", it only satisfies them up to a tiny error. We use an interesting and rather generic approach to round $\widetilde{\mathbb{E}}$ to a nearby pseudoexpectation operator $\widetilde{\mathbb{E}}'$ which does exactly satisfy the constraints.

For degree D, the candidate SoS solution can be viewed as a (pseudo) moment matrix $\mathcal{M}$ with rows and columns indexed by subsets $I, J \subset [n]$ with size bounded by $D/2$ and with entries

$$\mathcal{M}[I, J] := \widetilde{\mathbb{E}}[v^I v^J].$$

The matrix $\mathcal{M}$ is a random function of the inputs $d_1, \ldots, d_m$, and the most challenging part of the analysis consists of showing that $\mathcal{M}$ is positive semi-definite (PSD) with high probability.

Similarly to [BHK+16], we decompose $\mathcal{M}$ as a linear combination of graph matrices, i.e., $\mathcal{M} = \sum_\alpha \lambda_\alpha \cdot M_\alpha$, where M_α is the graph matrix associated with shape α. In brief, each graph matrix aggregates all terms with shape α in the Fourier expansions of the entries of $\mathcal{M}$ – the shape α is informally a graph with labeled edges with size bounded by $\mathrm{poly}(D)$. A graph matrix decomposition of $\mathcal{M}$ is particularly handy in the PSD analysis since the operator norm of individual graph matrices M_α is (with high probability) determined by simple combinatorial properties of the graph α. One technical difference from [BHK+16] is that our graph matrices have two types of vertices □ and ◯; these graph matrices fall into the general framework developed by Ahn et al. in [AMP20].

To show that the matrix $\mathcal{M}$ is PSD, we need to study the graph matrices that appear with nonzero coefficients in the decomposition. The matrix $\mathcal{M}$ can be split into blocks and each diagonal block contains in the decomposition a (scaled) identity matrix. From the graph matrix perspective, this means that certain "trivial" shapes appear in the decomposition, with appropriate coefficients. If we could bound the norms of all other graph matrices that appear against these trivial shapes and show that, together, they have negligible norm compared to the sum of these scaled identity blocks, then we would be in good shape.

Unfortunately, this approach will not work. The kernel of the matrix $\mathcal{M}$ is nontrivial, as a consequence of satisfying the PAP constraints "$\langle v, d_u\rangle^2 = 1$", and hence there is no hope of showing that the contribution of all nontrivial shapes in the decomposition of $\mathcal{M}$ has small norm. Indeed, certain shapes α appearing in the decomposition of $\mathcal{M}$ are such that $\|\lambda_\alpha \cdot M_\alpha\|$ is large. As it turns out, all such shapes have a simple graphical substructure, and so we call these shapes *spiders*.

To get around the null space issue, we restrict ourselves to $\mathrm{Null}(\mathcal{M})^\perp$, which is the complement of the nullspace of $\mathcal{M}$. We show that the substructure present in a spider implies that the spider is close to the zero matrix in $\mathrm{Null}(\mathcal{M})^\perp$. Because of this, we can almost freely add and subtract M_α for spiders α while preserving the action of $\mathcal{M}$ on $\mathrm{Null}(\mathcal{M})^\perp$. Our strategy is to "kill" the spiders by subtracting off $\lambda_\alpha \cdot M_\alpha$ for each spider α. But because M_α is only approximately in $\mathrm{Null}(\mathcal{M})^\perp$, this strategy could potentially introduce new graph matrix terms, and in particular it could introduce new spiders. To handle this, we recursively kill them while carefully analyzing how the coefficients of all the graph matrices change. After all spiders are killed, the resulting moment matrix becomes

$$\sum_{0 \le k \le D/2} \frac{1}{n^k} \cdot I_k + \sum_{\gamma\colon \text{non-spiders}} \lambda'_\gamma \cdot M_\gamma,$$

for some new coefficients λ'_γ. Here, I_k is the matrix which has an identity in the kth block and the remaining entries 0. Using a novel charging argument, we finally show that the latter term is negligible compared to the former term, thus establishing $\mathcal{M} \succeq 0$.

Summary of Related Work and Our Contributions. We now summarize the existing work on these problems and our contributions. Degree-4 SoS lower bounds on the Sherrington-Kirkpatrick Hamiltonian problem were proved independently by Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] whereas we prove an improved degree-n^δ SoS lower bound for some constant $\delta > 0$. Our result is obtained by reducing the Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" problem which is equivalent to our new Planted Affine Planes problem on the normal distribution. The reduction from Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" is due to Mohanty–Raghavendra–Xu [MRX19]. The results of Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] build on a degree-2 SoS lower bounds of Montanari and Sen [MS16]. Regarding upper bounds, Montanari [Mon19] gave

an efficient randomized message passing algorithm to estimate $\mathsf{OPT}(W)$ in the SK problem within a $(1-\varepsilon)$ factor under the full replica symmetry breaking assumption.

Degree-4 SoS lower bounds on the "Boolean Vector in a Random Subspace" problem for $p \geq n^{0.99}$ were proved by Mohanty–Raghavendra–Xu in [MRX19] where this problem was introduced. We improve the dependence on p to $p \geq n^{2/3+\varepsilon}$ for any $\varepsilon > 0$ and obtain a stronger degree-$n^{c\varepsilon}$ SoS lower bound for some absolute constant $c > 0$.

II. Technical Preliminaries

In this section we record problem statements, then define and discuss the main objects in our SoS lower bound: pseudoexpectation operators, the moment matrix, and graph matrices.

For a vector or variable $v \in \mathbb{R}^n$, and $I \subseteq [n]$, we use the notation $v^I := \prod_{i \in I} v_i$. When a statement holds with high probability (w.h.p.), it means it holds with probability $1 - o_n(1)$. In particular, there is no requirement for small n.

A. Problem statements

We introduce the Planted Affine Planes problem over a distribution $\mathcal{D}$.

Definition II.1 (Planted Affine Planes (PAP) problem)**.** *Given $d_1, \ldots, d_m \sim \mathcal{D}$ where each d_u is a vector in $\mathbb{R}^n$, determine whether there exists $v \in \{\pm\frac{1}{\sqrt{n}}\}^n$ such that*

$$\langle v, d_u \rangle^2 = 1,$$

for every $u \in [m]$.

Our results hold for the Gaussian setting $\mathcal{D} = \mathcal{N}(0, I)$ and the boolean setting where $\mathcal{D}$ is uniformly sampled from $\{\pm 1\}^n$, though we conjecture (see the full version of the paper) that similar SoS bounds hold under more general conditions on $\mathcal{D}$.

Observe that in both settings the solution vector v is restricted to be Boolean (in the sense that the entries are either $\frac{1}{\sqrt{n}}$ or $\frac{-1}{\sqrt{n}}$) and an SoS lower bound for this restricted version of the problem is stronger than when v can be an arbitrary vector from $\mathbb{R}^n$.

The Sherrington–Kirkpatrick (SK) problem comes from the spin-glass model in statistical physics [SK75].

Definition II.2 (Sherrington-Kirkpatrick problem)**.** *Given $W \sim \mathrm{GOE}(n)$, compute*

$$\mathsf{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x.$$

The Planted Boolean Vector problem was introduced by Mohanty–Raghavendra–Xu [MRX19], where it was called the "Boolean Vector in a Random Subspace".

Definition II.3 (Planted Boolean Vector problem)**.** *Given a uniformly random p-dimensional subspace V of $\mathbb{R}^n$ in the form of a projector Π_V onto V, compute*

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b.$$

B. Sum-of-Squares solutions

We will work with two equivalent definitions of a degree-D SoS solution: a pseudoexpectation operator and a moment matrix. We tailor these definitions to our setting of feasibility of systems of polynomial equality constraints given by the common zero set of a collection of polynomials $\mathcal{P}$ on $\pm\frac{1}{\sqrt{n}}$ Boolean variables $v_1, \ldots, v_n$. For a degree-D solution to be well defined, we need D to be at least the maximum degree of a polynomial in $\mathcal{P}$. Let $\mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ be the subset of polynomials of degree at most D from the polynomial ring $\mathbb{R}(v_1, \ldots, v_n)$. We denote the degree of a polynomial $f \in \mathbb{R}(v_1, \ldots, v_n)$ by $\deg(f)$.

1) Pseudoexpectation operator: We formally define the pseudoexpectation operators used in our setting.

Definition II.4 (Pseudoexpectation)**.** *Given a finite collection of "constraint" polynomials $\mathcal{P}$ of degree at most D on $\pm\frac{1}{\sqrt{n}}$ Boolean variables $v_1, \ldots, v_n$, a degree-D pseudoexpectation operator $\widetilde{\mathbb{E}}$ is an operator $\widetilde{\mathbb{E}} \colon \mathbb{R}^{\leq D}(v_1, \ldots, v_n) \to \mathbb{R}$ satisfying:*

1) $\widetilde{\mathbb{E}}[1] = 1$,
2) *$\widetilde{\mathbb{E}}$ is an $\mathbb{R}$-linear operator, i.e., $\widetilde{\mathbb{E}}[f + g] = \widetilde{\mathbb{E}}[f] + \widetilde{\mathbb{E}}[g]$ for every $f, g \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$,*
3) *$\widetilde{\mathbb{E}}[f^2] \geq 0$ for every $f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f^2) \leq D$.*
4) *$\widetilde{\mathbb{E}}[(v_i^2 - \frac{1}{n}) \cdot f] = 0$ for all $i \in [n]$ and for every $f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f) \leq D - 2$, and*
5) *$\widetilde{\mathbb{E}}[g \cdot f] = 0$ for every $g \in \mathcal{P}, f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f \cdot g) \leq D$.*

Note that $\widetilde{\mathbb{E}}$ behaves similarly to an expectation operator restricted to $\mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with

the caveat that $\widetilde{\mathbb{E}}$ is only guaranteed to be non-negative on sum-of-squares polynomials.

The degree-D SoS algorithm checks feasibility of a polynomial system by checking whether or not a degree-D pseudoexpectation operator exists. To show an SoS lower bound, one must construct a pseudoexpectation operator.

2) Moment matrix: We define the moment matrix associated with a degree-D pseudoexpectation $\widetilde{\mathbb{E}}$.

Definition II.5 (Moment Matrix of $\widetilde{\mathbb{E}}$)**.** *The moment matrix $\mathcal{M} = \mathcal{M}(\widetilde{\mathbb{E}})$ associated to a pseudoexpectation $\widetilde{\mathbb{E}}$ is a $\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}$ matrix with rows and columns indexed by subsets of $I, J \subseteq [n]$ of size at most $D/2$ and defined as*

$$\mathcal{M}[I, J] := \widetilde{\mathbb{E}}\left[v^I \cdot v^J\right].$$

To show that a candidate pseudoexpectation satisfies Item 3 in Definition II.4, we will rely on the following standard fact.

Fact II.6. *In the definition of pseudoexpectation, Definition II.4, the condition in Item 3 is equivalent to $\mathcal{M} \succeq 0$.*

C. Graph matrices

To study $\mathcal{M}$, we decompose it using the framework of *graph matrices*. Originally developed in the context of the planted clique problem, graph matrices are random matrices whose entries are symmetric functions of an underlying random object – in our case, the set of vectors $d_1, \ldots, d_m$. We take the general presentation and results from [AMP20]. For our purposes, the following definitions are sufficient.

The graphs that we study have two types of vertices, circles ◯ and squares □. We let $\mathcal{C}_m$ be a set of m circles labeled 1 through m, which we denote by $\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$, and let $\mathcal{S}_n$ be a set of n squares labeled 1 through n, which we denote by $\boxed{1}, \boxed{2}, \ldots, \boxed{n}$. We will work with bipartite graphs with edges between circles and squares, which have positive integer labels on the edges. When there are no multiedges (the graph is simple), such graphs are in one-to-one correspondence with Fourier characters on the vectors d_u. An edge between $\textcircled{u}$ and $\boxed{i}$ with label l represents $h_l(d_{u,i})$ where $\{h_k\}$ is the Fourier basis (e.g. Hermite polynomials).

$$\text{simple graph with labeled edges} \iff \prod_{\substack{\textcircled{u} \in \mathcal{C}_m, \\ \boxed{i} \in \mathcal{S}_n}} h_{l(\textcircled{u},\boxed{i})}(d_{u,i})$$

An example of a Fourier polynomial as a graph with labeled edges is given in Fig. 1. Unlabeled edges are implicitly labeled 1.

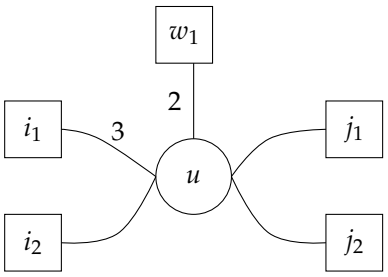

Figure 1. The Fourier polynomial $h_3(d_{u,i_1})h_1(d_{u,i_2})h_2(d_{u,w_1})h_1(d_{u,j_1})h_1(d_{u,j_2})$ represented as a graph.

Define the degree of a vertex v, denoted $\deg(v)$, to be the sum of the labels incident to v, and $|E|$ to be the sum of all labels. For intuition it is mostly enough to work with simple graphs, in which case these quantities make sense as the edge multiplicities in an implicit multigraph.

Definition II.7 (Proper)**.** *We say an edge-labeled graph is proper if it has no multiedges.*

The definitions allow for "improper" edge-labeled multigraphs which simplify multiplying graph matrices.

Definition II.8 (Matrix indices)**.** *A matrix index is a set A of elements from $\mathcal{C}_m \cup \mathcal{S}_n$.*

We let $A(\boxed{i})$ or $A(\textcircled{u})$ be 0 or 1 to indicate if the vertex is in A.

Definition II.9 (Ribbons)**.** *A ribbon is an undirected, edge-labeled graph $R = (V(R), E(R), A_R, B_R)$, where $V(R) \subseteq \mathcal{C}_m \cup \mathcal{S}_n$ and A_R, B_R are two matrix indices (possibly not disjoint) with $A_R, B_R \subseteq V(R)$, representing two distinguished sets of vertices. Furthermore, all edges in $E(R)$ go between squares and circles.*

We think of A_R and B_R as being the "left" and "right" sides of R, respectively. We also define the set of "middle vertices" $C_R := V(R) \setminus (A_R \cup B_R)$. If $e \notin E(R)$, then we define its label $l(e) = 0$. We also abuse notation and write $l(\boxed{i}, \textcircled{u})$ instead of $l(\{\boxed{i}, \textcircled{u}\})$.

Akin to the picture above, each ribbon corresponds to a Fourier polynomial. This Fourier polynomial lives inside a single entry of the matrix M_R. In the definition below, the $h_k(x)$ are the Fourier basis corresponding to the respective setting. In the Gaussian case, they are the (unnormalized)

Hermite polynomials, and in the boolean case, they are just the parity function, represented by

$$h_0(x) = 1, \qquad h_1(x) = x, \qquad h_k(x) = 0 \ \ (k \geq 2)$$

Definition II.10 (Matrix for a ribbon)**.** *The matrix M_R has rows and columns indexed by subsets of $\mathcal{C}_m \cup \mathcal{S}_n$, with a single nonzero entry defined by*

$$M_R[I,J] = \begin{cases} \prod_{\substack{e \in E(R), \\ e = \{\boxed{i}, \textcircled{u}\}}} h_{l(e)}(d_{u,i}) & I = A_R, J = B_R \\ 0 & \text{Otherwise} \end{cases}$$

Next we describe the shape of a ribbon, which is essentially the ribbon when we have forgotten all the vertex labels and retained only the graph structure and the distinguished sets of vertices.

Definition II.11 (Index shapes)**.** *An index shape is a set U of formal variables. Furthermore, each variable is labeled as either a "circle" or a "square".*

We let $U(\boxed{i})$ and $U(\textcircled{u})$ be either 0 or 1 for whether $\boxed{i}$ or $\textcircled{u}$, respectively, is in U.

Definition II.12 (Shapes)**.** *A shape is an undirected, edge-labeled graph $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ where $V(\alpha)$ is a set of formal variables, each of which is labeled as either a "circle" or a "square". U_α and V_α are index shapes (possibly with variables in common) such that $U_\alpha, V_\alpha \subseteq V(\alpha)$. The edge set $E(\alpha)$ must only contain edges between the circle variables and the square variables.*

We'll also use $W_\alpha := V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ to denote the "middle vertices" of the shape.

Remark II.13. *We will abuse notation and use $\boxed{i}, \boxed{j}, \textcircled{u}, \textcircled{v}, \ldots$ for both the vertices of ribbons and the vertices of shapes. If they are ribbon vertices, then the vertices are elements of $\mathcal{C}_m \cup \mathcal{S}_n$ and if they are shape vertices, then they correspond to formal variables with the appropriate type.*

Definition II.14 (Trivial shape)**.** *Define a shape α to be trivial if $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ and $E(\alpha) = \emptyset$.*

Definition II.15 (Transpose of a shape)**.** *The transpose of a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is defined to be the shape $\alpha^\intercal = (V(\alpha), E(\alpha), V_\alpha, U_\alpha)$.*

For a shape α and an injective map $\sigma : V(\alpha) \to \mathcal{C}_m \cup \mathcal{S}_n$, we define the realization $\sigma(\alpha)$ as a ribbon in the natural way, by labeling all the variables using the map σ. We also require σ to be type-preserving i.e. it takes square variables to $\mathcal{S}_n$ and circle variables to $\mathcal{C}_m$. The ribbons that result are referred to as *ribbons of shape α*; notice that this partitions the set of all ribbons according to their shape[3][4].

Finally, given a shape α, the graph matrix M_α consists of all Fourier characters for ribbons of shape α.

Definition II.16 (Graph matrices)**.** *Given a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$, the graph matrix M_α is*

$$M_\alpha = \sum_{R \text{ is a ribbon of shape } \alpha} M_R$$

The moment matrix for PAP will turn out to be defined using graph matrices M_α whose left and right sides only have square vertices, and no circles. However, in the course of the analysis we will factor and multiply graph matrices with circle vertices in the left or right.

D. Norm bounds

The spectral norm of a graph matrix is determined, up to logarithmic factors, by relatively simple combinatorial properties of the graph. For a subset $S \subseteq \mathcal{C}_m \cup \mathcal{S}_n$, we define the weight $w(S) := (\#\text{ circles in } S) \cdot \log_n(m) + (\#\text{ squares in } S)$. Observe that $n^{w(S)} = m^{\#\text{ circles in } S} \cdot n^{\#\text{ squares in } S}$.

Definition II.17 (Minimum vertex separator)**.** *For a shape α, a set $S_{\min}$ is a minimum vertex separator if all paths from U_α to V_α pass through $S_{\min}$ and $w(S_{\min})$ is minimized over all such separating sets.*

Let W_{iso} denote the set of isolated vertices in W_α. Then essentially the following norm bound holds for all shapes α with high probability (a formal statement can be found in the full version of the paper):

$$\|M_\alpha\| \leq \widetilde{O}\left(n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}\right)$$

In fact, the only probabilistic property required of the inputs $d_1, \ldots, d_m$ by our proof is that the above norm bounds hold for all shapes that arise in the analysis. We henceforth assume that the norm bounds for the Gaussian case and for the boolean case hold.

[3]Partitions up to equality of shapes, where two shapes are equal if there is a type-preserving bijection between their variables that converts one shape to the other. When we operate on sets of shapes below, we implicitly use each distinct shape only once.

[4]Note that in our definition two realizations of a shape may give the same ribbon.

III. Proof Strategy

Here we explain in more detail the ideas for the Planted Affine Planes lower bound. Towards the proof of Theorem I.4, fix a constant $\varepsilon > 0$ and a random instance $d_1, \ldots, d_m$ with $n \le m \le n^{3/2-\varepsilon}$. We will construct a pseudoexpectation operator and show that it is PSD up to degree $D = 2 \cdot n^{\delta}$ with high probability.

We start by pseudocalibrating to obtain a pseudoexpectation operator $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ will exactly satisfy the "booleanity" constraints "$v_i^2 = \frac{1}{n}$" though it may not exactly satisfy the constraints "$\langle v, d_u \rangle^2 = 1$" due to truncation error in the pseudocalibration. Taking the truncation parameter n^{τ} to be larger than the degree D of the SoS solution, i.e., $\delta \ll \tau$, the truncation error is small enough that we can round $\widetilde{\mathbb{E}}$ to a nearby $\widetilde{\mathbb{E}}'$ that exactly satisfies the constraints. This is formally accomplished by viewing $\widetilde{\mathbb{E}} \in \mathbb{R}^{\binom{[n]}{\le D}}$ as a vector and expressing the constraints as a matrix Q such that $\widetilde{\mathbb{E}}$ satisfies the constraints iff it lies in the null space of Q. The choice of $\widetilde{\mathbb{E}}'$ is then the projection of $\widetilde{\mathbb{E}}$ to $\mathrm{Null}(Q)$. The end result is that we construct a moment matrix $\mathcal{M}_{fix} = \mathcal{M} + \mathcal{E}$ that exactly satisfies the constraints such that $\|\mathcal{E}\|$ is tiny.

After performing pseudocalibration, in both settings, we will have essentially the graph matrix decomposition

$$\mathcal{M} = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha = \sum_{\substack{\text{shapes } \alpha: \\ \deg(\boxed{i})+U(\boxed{i})+V(\boxed{i}) \text{ even}, \\ \deg(\textcircled{u}) \text{ even}}} \frac{1}{n^{\frac{|U_\alpha|+|V_\alpha|}{2}}} \cdot \left(\prod_{\textcircled{u} \in V(\alpha)} h_{\deg(\textcircled{u})}(1) \right) \cdot \frac{M_\alpha}{n^{|E(\alpha)|/2}}$$

Here $h_k(1)$ is in both settings the k-th Hermite polynomial, evaluated on 1.

In this decomposition of $\mathcal{M}$, the trivial shapes will be the dominant terms which we will use to bound the other terms. Recall that a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is trivial if $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ and $E(\alpha) = \emptyset$. These shapes contribute scaled identity matrices on different blocks of the main diagonal of $\mathcal{M}$, with trivial shape α contributing an identity matrix with coefficient $n^{-|U_\alpha|}$. Two trivial shapes are illustrated in Fig. 2.

Let $\mathcal{M}_{\text{triv}}$ be this diagonal matrix of trivial shapes in the above decomposition of $\mathcal{M}$. To prove that $\mathcal{M} \succeq 0$, we attempt the simple strategy of showing that the norm of all other terms can be "charged" against this diagonal matrix $\mathcal{M}_{\text{triv}}$. For several shapes this strategy is indeed viable. To illustrate, let's consider one such shape α depicted in Fig. 3.

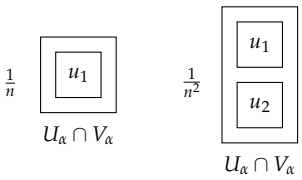

Figure 2. Two examples of trivial shapes.

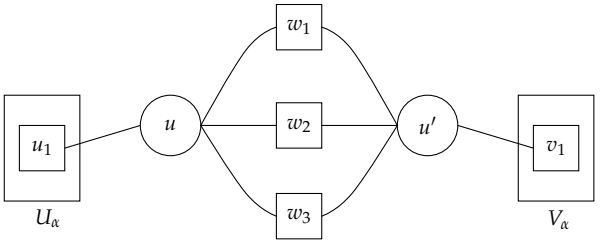

Figure 3. Picture of basic non-spider shape α.

This graph matrix has $|\lambda_\alpha| = \Theta(\frac{1}{n^5})$. Using the graph matrix norm bounds, with high probability the norm of this graph matrix is $\tilde{O}(n^2 m)$: there are four square vertices and two circle vertices which are not in the minimum vertex separator. Thus, for this shape α, with high probability $|\lambda_\alpha| \|M_\alpha\|$ is $\tilde{O}\left(\frac{m}{n^3}\right)$ and thus $\lambda_\alpha M_\alpha \preceq \frac{1}{n} Id$ (which is the multiple of the identity appearing in the corresponding block of $\mathcal{M}_{\text{triv}}$).

Unfortunately, as pointed out in the introduction, some shapes α that appear in the decomposition have $\|\lambda_\alpha M_\alpha\|$ too large to be charged against $\mathcal{M}_{\text{triv}}$. These are shapes with a certain substructure (actually the same structure that appears in the matrix Q used to project the pseudoexpectation operator!) whose norms cannot be handled by the preceding argument, and which we denote *spiders*. The following graph depicts one such *spider* shape (and also motivates this terminology):

The norm $\|\lambda_\alpha M_\alpha\|$ of this graph is $\widetilde{\Omega}(\frac{1}{n^2})$, as can be easily estimated through the norm bounds (the coefficient is $\lambda_\alpha = \frac{-2}{n^4}$, the minimum vertex separator is $\textcircled{u}$, and there are no isolated vertices). This is too large to bound against $\frac{1}{n^2} Id$, which is the coefficient of M_{triv} on this spider's block.

To skirt this and other spiders, we restrict ourselves to vectors $x \perp \mathrm{Null}(M)$, and observe that this spider α satisfies $x^\intercal M_\alpha \approx 0$. To be more precise, consider the following argument. Consider the two shapes in Fig. 5, β_1 and β_2 (take note of

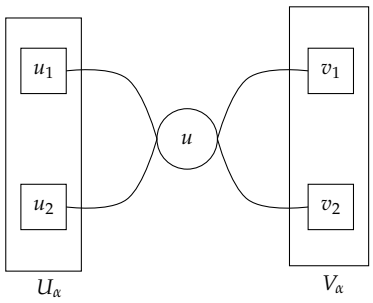

Figure 4. Picture of basic spider shape α.

the label 2 on the edge in β_2).

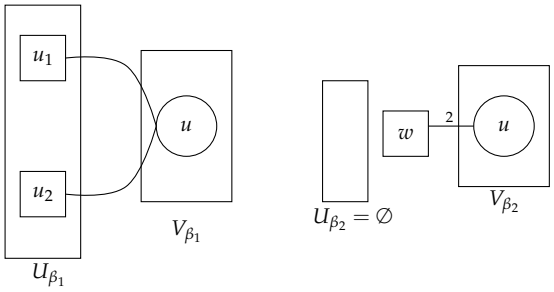

Figure 5. Picture of shapes β_1 and β_2.

We claim that every column of the matrix $2M_{\beta_1} + \frac{1}{n}M_{\beta_2}$ is in the null space of $\mathcal{M}$. There are m nonzero columns indexed by assignments to V, which can be a single circle $\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$. The nonzero rows are $\emptyset$ in β_2 and $\{\boxed{i}, \boxed{j}\}$ for $i \neq j$ in β_1. Fixing $I \subseteq [n]$, entry $(I, \textcircled{u})$ of the product matrix $\mathcal{M}(2M_{\beta_1} + \frac{1}{n}M_{\beta_2})$ is

$$
\begin{aligned}
&2\sum_{i<j} \widetilde{\mathbb{E}}[v^I v_i v_j] \cdot d_{ui} d_{uj} + \frac{1}{n}\widetilde{\mathbb{E}}[v^I] \cdot \sum_i (d_{ui}^2 - 1) \\
&= 2\sum_{i<j} \widetilde{\mathbb{E}}[v^I v_i v_j] \cdot d_{ui} d_{uj} + \\
&\quad \widetilde{\mathbb{E}}[v^I v_i^2] \cdot \sum_i d_{ui}^2 - \widetilde{\mathbb{E}}[v^I] \qquad (\widetilde{\mathbb{E}} v_i^2 = \frac{1}{n}) \\
&= \sum_{i,j} \widetilde{\mathbb{E}}[v^I v_i v_j] d_{ui} d_{uj} - \widetilde{\mathbb{E}}[v^I] \\
&= \widetilde{\mathbb{E}}[v^I (\langle v, d_u \rangle^2 - 1)] \\
&= 0 \qquad (\widetilde{\mathbb{E}} \langle v, d_u \rangle^2 = 1)
\end{aligned}
$$

In words, the constraint "$\langle v, d_u \rangle^2 = 1$" creates a shape $2\beta_1 + \frac{1}{n}\beta_2$ that lies in the null space of the moment matrix. On the other hand, we can approximately factor the spider α across its central vertex, and when we do so, the shape β_1 appears on the left side.

Therefore $M_\alpha \approx M_{\beta_1} M_{\beta_1}^\intercal \approx (M_{\beta_1} + \frac{1}{2n}M_{\beta_2})M_{\beta_1}^\intercal$. The columns of the matrix $M_{\beta_1} + \frac{1}{2n}M_{\beta_2}$ are in the null space of $\mathcal{M}$, so for $x \perp \text{Null}(\mathcal{M})$ we have $x^\intercal M_\alpha \approx 0$.

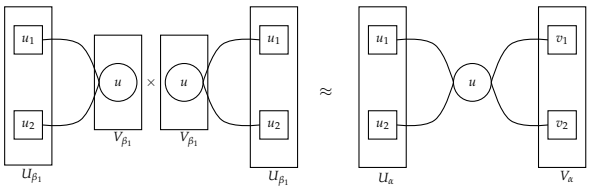

Figure 6. Approximation $\beta_1 \times \beta_1^\intercal \approx \alpha$.

More formally, we are able to find coefficients c_β so that all columns of the matrix

$$A = M_\alpha + \sum_\beta c_\beta M_\beta$$

are in $\text{Null}(\mathcal{M})$. We then observe the following fact:

Fact III.1. *If $x \perp \text{Null}(\mathcal{M})$ and $\mathcal{M}A = 0$, then $x^\intercal(AB + \mathcal{M})x = x^\intercal(B^\intercal A^\intercal + \mathcal{M})x = x^\intercal \mathcal{M} x$.*

Using the fact, we can freely add multiples of A to $\mathcal{M}$ without changing the action of $\mathcal{M}$ on $\text{Null}(\mathcal{M})^\perp$. A judicious choice is to subtract $\lambda_\alpha A$ which will "kill" the spider from $\mathcal{M}$. Doing this for all spiders, we produce a matrix whose action is equivalent on $\text{Null}(\mathcal{M})^\perp$, and which has high minimum eigenvalue by virtue of the fact that it has no spiders, showing that $\mathcal{M}$ is PSD.

The catch is two-fold: first, the coefficients c_β may contribute to the coefficients on the non-spiders; second, the further intersection terms M_β may themselves be spiders (though they will always have fewer square vertices than α). Thus we must recursively kill these spiders, until there are no spiders remaining in the decomposition of $\mathcal{M}$. The resulting matrix has some new coefficients on the non-spiders

$$\mathcal{M}' = \sum_{\text{non-spiders } \beta} \lambda'_\beta M_\beta.$$

We must bound the accumulation on the coefficients λ'_β. We do this by considering the *web* of spiders and non-spiders created by each spider and using bounds on the c_β and λ_α to argue that the contributions do not blow up, via an interesting charging scheme that exploits the structure of these graphs.

IV. Sherrington-Kirkpatrick Lower Bounds

Here, we show the reductions to Planted Affine Planes in Theorem I.5 and Theorem I.2. Recall that in the Planted Boolean Vector problem,

we wish to optimize

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b,$$

where V is a uniformly random p-dimensional subspace of $\mathbb{R}^n$.

Theorem I.5. *[Main III] There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \leq c\varepsilon$, for $p \geq n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value 1.*

Proof: We wish to produce an SoS solution $\widetilde{\mathbb{E}}$ on boolean variables $b_1, \ldots, b_n$ such that $\widetilde{\mathbb{E}}[b^\intercal \Pi_V b] = n$. Instead of sampling a uniformly random p-dimensional subspace V of $\mathbb{R}^n$, we first sample $d_1, \ldots, d_n$ i.i.d. p-dimensional Gaussian vectors from $\mathcal{N}(0, I)$, then form an n-by-p matrix A with rows $d_1, \ldots, d_n$, and finally take V to be the span of the columns of A. Since the columns of A are isotropic i.i.d. random Gaussian vectors, we have that V is a uniform p-dimension subspace[5] of $\mathbb{R}^n$.

We will consider V as the input for the Planted Boolean Vector problem while the vectors $d_1, \ldots, d_n$ will be used to construct a pseudoexpectation operator for the Planted Affine Planes problem[6]. Since $n \leq p^{3/2-\Omega(\varepsilon)}$, by Theorem I.4, for all $\delta \leq c\varepsilon$ for a constant $c > 0$, w.h.p., there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}'$ on formal variables $v = (v_1, \ldots, v_p)$ such that $\widetilde{\mathbb{E}}'[\langle v, d_u \rangle^2] = 1$ for every $u \in [n]$.

Define $\widetilde{\mathbb{E}}$ by $\widetilde{\mathbb{E}}[b_u] := \widetilde{\mathbb{E}}'[\langle v, d_u \rangle]$ for all $u \in [n]$ and extending it to all polynomials on $\{b_u\}$ by multilinearity. This is well defined because $\widetilde{\mathbb{E}}'[\langle v, d_u \rangle^2] = 1$. Note that $\widetilde{\mathbb{E}}$ is a valid pseudoexpectation operator of the same degree as $\widetilde{\mathbb{E}}'$. Finally, observe that

$$\frac{1}{n}\widetilde{\mathbb{E}}[b^\intercal \Pi_V b] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^\intercal A^\intercal \Pi_V A v] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^\intercal A^\intercal A v] = 1.$$

■

Now we prove lower bounds for the Sherrington-Kirkpatrick problem, using a reduction and proof due to [MRX19]. We include it here for completeness. Recall that the SK problem is to compute

$$\mathsf{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x,$$

where W is sampled from $\mathrm{GOE}(n)$.

[5]Except for a zero measure event.

[6]Note that the vectors d_u are not "given" in the Planted Boolean Vector problem, though the construction of $\widetilde{\mathbb{E}}$ is not required to be algorithmic in any sense anyway.

Theorem I.2. *[Main I] There exists a constant $\delta > 0$ such that, w.h.p. for $W \sim \mathrm{GOE}(n)$, there is a degree-n^δ SoS solution for the Sherrington–Kirkpatrick problem with value at least $(2 - o_n(1)) \cdot n^{3/2}$.*

We will use the following standard results from random matrix theory of $\mathrm{GOE}(n)$.

Fact IV.1. *Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of $W \sim \mathrm{GOE}(n)$ with corresponding normalized eigenvectors $w_1, \ldots, w_n$. Then,*

1) *For every $p \in [n]$, the span of $w_1, \ldots, w_p$ is a uniformly random p-dimensional subspace of $\mathbb{R}^n$ (see e.g. [OVW16, Section 2]).*
2) *W.h.p., $\lambda_{n^{0.67}} \geq (2 - o(1))\sqrt{n}$ (Corollary of Wigner's semicircle law [Wig93])*

Proof of Theorem I.2: Let $p = n^{0.67}$ and $W \sim \mathrm{GOE}(n)$. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of W with corresponding orthonormal set of eigenvectors $w_1, \ldots, w_n$. By Fact IV.1, we have that $\lambda_p \geq (2 - o(1))\sqrt{n}$ and that $w_1, \ldots, w_p$ span a uniformly random p-dimensional subspace V of $\mathbb{R}^n$.

We consider V as the input of the Boolean Planted Vector problem and by Theorem I.5, for some constant $\delta > 0$, w.h.p. there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ such that $\widetilde{\mathbb{E}}[x_i^2] = 1$ and $\widetilde{\mathbb{E}}[\sum_{i=1}^p \langle x, w_i \rangle^2] = \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] = n$. Now,

$$\begin{aligned}
\widetilde{\mathbb{E}}[x^\intercal W x] = \widetilde{\mathbb{E}}[\sum_{i=1}^n \lambda_i \langle x, w_i \rangle^2] &\geq \lambda_p \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] - |\lambda_n| \, \widetilde{\mathbb{E}}[\sum_{i=p+1}^n \langle x, w_i \rangle^2] \\
&\geq (2 - o(1)) n^{3/2} - |\lambda_n| \, \widetilde{\mathbb{E}}[\langle x, x \rangle - \sum_{i=1}^p \langle x, w_i \rangle^2] \\
&= (2 - o(1)) n^{3/2}.
\end{aligned}$$

■

Remark IV.2. *Using the same proof as above, we can obtain Theorem I.2 even if we were only able to prove SoS lower bounds for Planted Affine Planes for some $m = \omega(n)$. So, pushing the value of m up to $n^{3/2-\varepsilon}$, which is Theorem I.4, offers only a modest improvement.*

V. Open Problems

We conjecture that for the Planted Affine Planes problem, the problem remains difficult even with the number of vectors increased to $m = n^{2-\varepsilon}$.

Conjecture V.1. *Theorem I.4 holds with the bound on the number of sampled vectors m loosened to $m \leq n^{2-\varepsilon}$.*

The justification for the conjecture is that this is the regime in which $\widetilde{\mathbb{E}}[1] = 1 + o(1)$ in pseudocalibration. Analyzing $\widetilde{\mathbb{E}}[1]$ is an established way to hypothesize about the power of SoS in hypothesis testing problems (see [HKP+17], [Hop18]).

Dual to the Planted Affine Planes problem, we conjecture a similar bound for Planted Boolean Vector problem whenever $d \geq n^{1/2+\varepsilon}$.

Conjecture V.2. *Theorem I.5 holds with the bound on the dimension p of a random subspace loosened to $p \geq n^{1/2+\varepsilon}$.*

We conjecture that the Planted Boolean Vector problem/Planted Affine Planes problem is still hard for SoS if the input is no longer i.i.d. Gaussian or boolean entries, but is drawn from a "random enough" distribution. For example, if in the random instance of PAP the vectors d_u are i.i.d. samples from S^n, or a random orthonormal system, degree n^δ SoS should still believe the instance is satisfiable (after appropriate normalization of v). Or, taking the view of Planted Boolean Vector, if the subspace is the eigenspace of the bottom eigenvectors of a random adjacency matrix, the instance should still be difficult. This last setting arises in MaxCut, for which we conjecture the following.

Conjecture V.3. *Let $d \geq 3$, and let G be a random d-regular graph on n vertices. For some $\delta > 0$, w.h.p. there is a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ on boolean variables x_i with MaxCut value at least*

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}(1 - o_{d,n}(1))$$

The above expression is w.h.p. the value of the spectral relaxation for MaxCut, therefore qualitatively this conjecture expresses that degree n^δ SoS cannot significantly beat the basic spectral relaxation.

We should remark that, with respect to the goal of showing that SoS cannot significantly outperform the Goemans-Williamson relaxation, random instances are not integrality gap instances. The main difficulty in comparing (even degree-4) SoS to the Goemans-Williamson algorithm seems to be the lack of a candidate hard input distribution.

Evidence for this conjecture comes from the fact that the only property required of the random inputs $d_1, \ldots, d_m$ was that norm bounds hold for the graph matrices with Hermite polynomial entries. When the variables $\{d_{u,i}\}$ are sampled i.i.d from some other distribution, if we use graph matrices for the orthonormal polynomials under that distribution and assume suitable bounds on the moments of the distribution, similar norm bounds hold [AMP20]. When $d_u \in_R S^n$ or another distribution for which the coordinates are not i.i.d, it seems likely that if we use e.g. the basis of spherical harmonics, then similar norm bounds hold.

Acknowledgements

We thank Madhur Tulsiani and Pravesh K. Kothari for several enlightening discussions in the initial phases of this work. G.R. thanks Sidhanth Mohanty for useful discussions regarding the Sherrington-Kirkpatrick problem. We also thank the anonymous reviewers for their useful suggestions on improving the text. Mrinalkanti, Fernando and Goutham are supported in part by NSF grant CCF-1816372. Chris and Aaron are supported in part by NSF grant CCF-2008920.

References

[AMP20] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. abs/1604.03423, 2020. URL: https://arxiv.org/abs/1604.03423, arXiv:1604.03423. 4, 6, 11

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004. 1

[BHK+16] B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science*, pages 428–437, 2016. 2, 4

[BKW19] Afonso S. Bandeira, Dmitriy Kunisky, and Alexander S. Wein. Computational hardness of certifying bounds on constrained pca problems, 2019. arXiv:1902.07324. 2

[FKP19] N. Fleming, P. Kothari, and T. Pitassi. *Semialgebraic Proofs and Efficient Algorithm Design*. 2019. 1

[FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Structures & Algorithms*, 20(3):403–440, 2002. 3

[GW95] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. Preliminary version in *Proc. of STOC'94*. 1

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 720–731. IEEE, 2017. 2, 10

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006. 2

[Hop18] Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018. 2, 10

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-squares proofs. In *Conference on Learning Theory*, pages 956–1006, 2015. 1

[Kar72] R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. 2

[KB19] Dmitriy Kunisky and Afonso S. Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. abs/1907.11686, 2019. URL: https://arxiv.org/abs/1907.11686, arXiv:1907.11686. 2, 4

[Las15] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2015. doi:10.1017/CBO9781107447226. 1

[Mon19] A. Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science*, pages 1417–1433, 2019. 2, 4

[MRX19] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: Degree-2 to degree-4. abs/1911.01411, 2019. URL: https://arxiv.org/abs/1911.01411, arXiv:1911.01411. 2, 3, 4, 5, 10

[MS16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the 48th ACM Symposium on Theory of Computing*, pages 814–827, 2016. 2, 4

[OVW16] Sean O'Rourke, Van Vu, and Ke Wang. Eigenvectors of random matrices. *J. Comb. Theory Ser. A*, 144(C):361–442, November 2016. 10

[Pan14] Dmitry Panchenko. The parisi formula for mixed p-spin models. *Ann. Probab.*, 42(3):946–958, 05 2014. 2

[Par79] G. Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754–1756, Dec 1979. 2

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 245–254, 2008. 1

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796, Dec 1975. 2, 5

[Tal06] Michel Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006. 2

[Wig93] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions i. In *The Collected Works of Eugene Paul Wigner*, pages 524–540. Springer, 1993. 10