Sum-of-Squares Lower Bounds for Sparse Independent Set

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

Madhur Tulsiani
Toyota Technological Institute
Chicago, USA
madhurt@ttic.edu

Jeff Xu
Computer Science Department
Carnegie Mellon University
Pittsburgh, USA
jeffxusichao@cmu.edu

***Abstract*—The Sum-of-Squares (SoS) hierarchy of semidefinite programs is a powerful algorithmic paradigm which captures state-of-the-art algorithmic guarantees for a wide array of problems. In the average case setting, SoS lower bounds provide strong evidence of algorithmic hardness or information-computation gaps. Prior to this work, SoS lower bounds have been obtained for problems in the "dense" input regime, where the input is a collection of independent Rademacher or Gaussian random variables, while the sparse regime has remained out of reach. We make the first progress in this direction by obtaining strong SoS lower bounds for the problem of Independent Set on sparse random graphs. We prove that with high probability over an Erdős-Rényi random graph $G \sim G_{n,\frac{d}{n}}$ with average degree $d > \log^2 n$, degree-D_{SoS} SoS fails to refute the existence of an independent set of size $k = \Omega\left(\frac{n}{\sqrt{d}(\log n)(D_{\mathrm{SoS}})^{c_0}}\right)$ in G (where c_0 is an absolute constant), whereas the true size of the largest independent set in G is $O\left(\frac{n \log d}{d}\right)$.**

Our proof involves several significant extensions of the techniques used for proving SoS lower bounds in the dense setting. Previous lower bounds are based on the pseudo-calibration heuristic of Barak et al. [FOCS 2016] which produces a candidate SoS solution using a planted distribution indistinguishable from the input distribution via low-degree tests. In the sparse case the natural planted distribution *does* admit low-degree distinguishers, and we show how to adapt the pseudo-calibration heuristic to overcome this.

Another notorious technical challenge for the sparse regime is the quest for matrix norm bounds. In this paper, we obtain new norm bounds for *graph matrices* in the sparse setting. While in the dense setting the norms of graph matrices are characterized by the size of the minimum vertex separator of the corresponding graph, this turns not to be the case for sparse graph matrices. Another contribution of our work is developing a new combinatorial understanding of structures needed to understand the norms of sparse graph matrices.

***Keywords*-Sum-of-Squares; Independent Set; Sparse Erdős-Rényi graphs;**

I. Introduction

The Sum-of-Squares (SoS) hierarchy is a powerful convex programming technique that has led to successful approximation and recovery algorithms for various problems in the past decade. SoS captures the best-known approximation algorithms for several classical combinatorial optimization problems. Some of the additional successes of SoS also include Tensor PCA [HSS15], [MSS16] and Constraint Satisfaction Problems with additional structure [BRS11], [GS11]. SoS is a family of convex relaxations parameterized by *degree*; by taking larger degree, one gets a better approximation to the true optimum at the expense of a larger SDP instance. Thus we are interested in the tradeoff between degree and approximation quality. For an introduction to Sum-of-Squares algorithms, see [BS16], [FKP19].

The success of SoS on the upper bound side has also conferred on it an important role for the investigation of algorithmic hardness. Lower bounds for the SoS hierarchy provide strong unconditional hardness results for several optimization problems and are of particular interest when NP-hardness results are unavailable. An important such setting is the study of average case complexity of optimization problems, where relatively few techniques exist for establishing NP-hardness results [ABB19]. In this setting, a study of the SoS hierarchy not only provides a powerful benchmark for average-case complexity, but also helps in understanding the structural properties of the problems: what makes them algorithmically challenging? Important examples of such results include an improved understanding of sufficient conditions for average-case hardness of CSPs [KMOW17] and lower bounds for the planted clique problem [BHK+16].

An important aspect of previous lower bounds for the SoS hierarchy is that they apply for the so-called *dense setting*, which corresponds to cases when the input distribution can be specified by a collection of independent Rademacher or Gaussian variables. In the case of planted clique, this corresponds to the case when the input is a random graph distributed according to $G_{n,\frac{1}{2}}$ i.e. specified by a collection of $\binom{n}{2}$ independent Rademacher variables. In the case of CSPs, one fixes the structure of the lower bound instance and only considers an instance to be specified by the signs of the literals, which can again be taken as uniformly random $\{-1,1\}$ variables. Similarly, recent results by a subset of the authors [PR20] for tensor PCA apply when the input tensor has independent Rademacher or Gaussian entries. The techniques used to establish these lower bounds have proved difficult to extend to the case when the input distribution naturally corresponds to a *sparse graph* (or more generally, when it is specified by a collection of independent sub-gaussian variables, with Orlicz norm $\omega(1)$ instead of $O(1)$).

In this paper we are interested in extending lower bound technology for SoS to the *sparse setting*, where the input is a graph with average degree $d \leq n/2$. We use as a case study the fundamental combinatorial optimization problem of independent set. For the dense case $d = n/2$, finding an independent set is equivalent to finding a clique and the paper [BHK+16] shows an average-case lower bound against the Sum-of-Squares algorithm. We extend the techniques introduced there, namely pseudocalibration, graph matrices, and the approximate decomposition into positive semidefinite matrices, in order to show the first average-case lower bound for the sparse setting. We hope that the techniques developed in this paper offer a gateway for the analysis of SoS on other sparse problems. Section IV lists several such problems that are likely to benefit from an improved understanding of the sparse setting.

Sample $G \sim G_{n,\frac{d}{n}}$ as an Erdős-Rényi random graph[1] with average degree d, where we think of $d \ll n$. Specializing to the problem of independent set, a maximum independent set in G has size:

Fact I.1 ([COE15], [DM11], [DSS16])**.** *W.h.p. the max independent set in G has size $(1 + o_d(1)) \cdot \frac{2 \ln d}{d} \cdot n$.*

The value of the degree-2 SoS relaxation for independent set equals the Lovász ϑ function, which is an upper bound on the independence number $\alpha(G)$, by virtue of being a relaxation. For random graphs $G \sim G_{n,d/n}$ this value is larger by a factor of about $\sqrt{d}$ than the true value of $\alpha(G)$ with high probability.

Fact I.2 ([CO05])**.** *W.h.p. $\vartheta(G) = \Theta(\frac{n}{\sqrt{d}})$.*

We will prove that the value of higher-degree SoS is also on the order of $n/\sqrt{d}$, rather than n/d, and thereby demonstrate that the information-computation gap against basic SDP/spectral algorithms persists against higher-degree SoS.

A. Our main results

The solution to the convex relaxations obtained via the SoS hierarchy can be specified by the so-called "pseudoexpectation operator".

Definition I.3 (Pseudoexpectation)**.** *A degree-D pseudoexpectation operator $\widetilde{\mathbb{E}}$ is a linear functional on polynomials of degree at most D (in n variables) such that $\widetilde{\mathbb{E}}[1] = 1$ and $\widetilde{\mathbb{E}}[f^2] \geq 0$ for every polynomial f with degree at most $D/2$. A pseudoexpectation is said to satisfy a polynomial constraint $g = 0$ if $\widetilde{\mathbb{E}}[f \cdot g] = 0$ for all polynomials f when $\deg(f \cdot g) \leq D$.*

In considering relaxations for independent set of a graph $G = (V, E)$, with variables x_v being the $0/1$ indicators of the independent set, the SoS relaxation searches for pseudoexpectation operators satisfying the polynomial constraints

$$\forall v \in V, x_v^2 = x_v \quad \text{and} \quad \forall (u,v) \in E, x_u x_v = 0$$

The objective value of the convex relaxation is given by the quantity $\widetilde{\mathbb{E}}[\sum_{v \in V} x_v] = \sum_{v \in V} \widetilde{\mathbb{E}}[x_v]$. For the results below, we say that an event occurs with high probability (w.h.p.) when it occurs with probability at least $1 - O(1/n^c)$ for some $c > 0$. The following theorem states our main result.

Theorem I.4. *There is an absolute constant $c_0 \in \mathbb{N}$ such that for sufficiently large $n \in \mathbb{N}$ and*

[1] Unfortunately our techniques do not work for a random d-regular graph. See the open problems (Section IV).

$d \in [(\log n)^2, n^{0.5}]$, and parameters k, D_{SoS} satisfying

$$k \;\le\; \frac{n}{D_{SoS}^{c_0} \cdot \log n \cdot d^{1/2}}\,,$$

it holds w.h.p. for $G = (V, E) \sim G_{n,\, d/n}$ that there exists a degree-D_{SoS} pseudoexpectation satisfying

$$\forall v \in V, x_v^2 = x_v \quad \textit{and} \quad \forall (u,v) \in E, x_u x_v = 0\,,$$

and objective value $\widetilde{\mathbb{E}}[\sum_{v \in V} x_v] \;\ge\; (1 - o(1))k$.

Remark I.5. *This is a non-trivial lower bound whenever $D_{SoS} \;\le\; \left(\frac{d^{1/2}}{\log n}\right)^{1/c_0}$.*

Remark I.6. *It suffices to set $c_0 = 20$ for our current proof. We did not optimize the tradeoff in D_{SoS} with k, but we did optimize the log factor (with the hope of eventually removing it).*

Remark I.7. *Using the same technique, we can prove an $n^{\Omega(\varepsilon)}$ SoS-degree lower bound for all $d \in [\sqrt{n}, n^{1-\varepsilon}]$.*

For $n^{\varepsilon} \le d \le n^{0.5}$, the theorem gives a polynomial n^{δ} SoS-degree lower bound. For smaller d, the bound is still strong against low-degree SoS, but it becomes trivial as D_{SoS} approaches $(d^{1/2}/\log n)^{1/c_0}$ or d approaches $(\log n)^2$ since k matches the size of the maximum independent set in G, hence there is an actual distribution over independent sets of this size (the expectation operator for which is trivially is also a pseudoexpectation operator).

The above bound says nothing about the "almost dense" regime $d \in [n^{1-\varepsilon}, n/2]$. To handle this regime, we observe that our techniques, along with the ideas from the $\Omega(\log n)$-degree SoS bound from [BHK+16] for the dense case, prove a lower bound for any degree $d \ge n^{\varepsilon}$.

Theorem I.8. *For any $\varepsilon_1, \varepsilon_2 > 0$ there is $\delta > 0$, such that for $d \in [n^{\varepsilon_1}, n/2]$ and $k \le \frac{n}{d^{1/2+\varepsilon_2}}$, it holds w.h.p. for $G = (V, E) \sim G_{n,\, d/n}$ that there exists a degree-$(\delta \log d)$ pseudoexpectation satisfying*

$$\forall v \in V, x_v^2 = x_v \quad \textit{and} \quad \forall (u,v) \in E, x_u x_v = 0\,,$$

and objective value $\widetilde{\mathbb{E}}[\sum_{v \in V} x_v] \;\ge\; (1 - o(1))k$.

In particular, these theorems rule out polynomial-time certification (i.e. constant degree SoS) for any $d \ge \mathrm{polylog}(n)$.

B. Our approach

Proving lower bounds for the case of sparse graphs requires extending the previous techniques for SoS lower bounds in multiple ways. The work closest to ours is the planted clique lower bound of [BHK+16]. The idea there is to view a random graph $G \sim G_{n,\,1/2}$ as a random input in $\{-1,1\}^{\binom{n}{2}}$, and develop a canonical method called "pseudo-calibration" for obtaining the pseudoexpectation $\widetilde{\mathbb{E}}$ as a function of G. The pseudocalibration method takes the low-degree Fourier coefficients of μ based on a *different* distribution on inputs G (with large planted cliques), and takes higher degree coefficients to be zero. This is based on the heuristic that distribution $G_{n,\,1/2}$ and the planted distribution are indistinguishable by low-degree tests. The pseudoexpectation obtained via this heuristic is then proved to be PSD (i.e., to satisfy $\widetilde{\mathbb{E}}[f^2] \ge 0$) by carefully decomposing its representation as a (moment) matrix Λ. One then needs to estimate the norms of various terms in this decomposition, known as "graph matrices", which are random matrices with entries as low-degree polynomials (in $\{-1,1\}^{\binom{n}{2}}$), and carefully group terms together to form PSD matrices.

Each of the above components require a significant generalization in the sparse case. To begin with, there is no good planted distribution to work with, as the natural planted distribution (with a large planted independent set) *is distinguishable from $G_{n,\,d/n}$ via low-degree tests*! While we still use the natural planted distribution to compute *some* pseudocalibrated Fourier coefficients, we also truncate (set to zero) several low-degree Fourier coefficients, in addition to the high-degree coefficients as in [BHK+16]. In particular, when the Fourier coefficients correspond to subgraphs where certain vertex sets are disconnected (viewed as subsets of $\binom{n}{2}$), we set them to zero. This is perhaps the most conceptually interesting part of the proof, and we hope that the same "connected truncation" will be useful for other integrality gap constructions.

The technical machinery for understanding norm bounds, and obtaining PSD terms, also requires a significant update in the sparse case. Previously, norm bounds for graph matrices were understood in terms of minimum vertex separators for the corresponding graphs, and arguments for obtaining PSD terms required working with the combinatorics of vertex separators [AMP20]. However, the number of vertices in a vertex separator turns out to be insufficient to control the relevant norm bounds in the sparse case. This is because of the fact that unlike random ± 1 variables, their p-biased analogs no longer have Orlicz norm $O(1)$ but instead $O(\frac{1}{\sqrt{p}})$, which results in both the vertices as well as edges in the graph playing a role in the norm bounds. To handle this, we characterize the norms of the relevant random matrices in terms of vertex separators, where the cost of a separator depends on the number of vertices and also the number of induced edges. Another issue is that the estimates on spectral norms obtained via the trace power method can fluctuate significantly due to rare events (presence of some dense subgraphs), and we need to carefully condition on the absence of these events.

A more detailed overview of our approach is presented in Section III.

C. Related work

Several previous works prove SoS lower bounds in the dense setting, when the inputs can be viewed as independent Gaussian or Rademacher random variables. Examples include the planted clique lower bound of Barak et al. [BHK+16], CSP lower bounds of Kothari et al. [KMOW17], and the tensor PCA lower bounds [HKP+17], [PR20]. The technical component of decomposing the moment matrix in the dense case, as a sum of PSD matrices, is developed into a general "machinery" in a recent work by a subset of the authors [PR20]. A different approach than the ones based on pseudocalibration, which also applies in the dense regime, was developed by Kunisky [Kun20].

For the case of independent set in random sparse graphs, many works have considered the *search* problem of finding a large independent set in a random sparse graph. Graphs from $G_{n,d/n}$ are known to have independent sets of size $(1 + o_d(1)) \cdot \frac{2 \ln d}{d} \cdot n$ with high probability, and it is possible to find an independent set of size $(1 + o_d(1)) \cdot \frac{\ln d}{d} \cdot n$, either by greedily taking a maximal independent set in the dense case [GM75] or by using a local algorithm in the sparse case [Wor95]. This is conjectured to be a computational phase transition, with concrete lower bounds against search beyond $\frac{\ln d}{d} \cdot n$ for local algorithms [RV17] and low-degree polynomials [Wei20]. The game in the search problem is all about the constant 1 vs 2, whereas our work shows that the integrality gap of SoS is significantly worse, on the order of $\sqrt{d}$. Lower bounds against search work in the regime of constant d (though in principle they could be extended to at least some $d = \omega(1)$ with additional technical work), while our techniques require $d \geq \log(n)$. For search problems, the *overlap distribution* of two high-value solutions has emerged as a heuristic indicator of computational hardness, whereas for certification problems it is unclear how the overlap distribution plays a role.

Norm bounds for sparse graph matrices were also obtained using a different method of matrix deviation inequalities, by a subset of the authors [RT20].

The work [BBK+20] constructs a *computationally quiet* planted distribution that is a candidate for pseudocalibration. However, their distribution is not quite suitable for our purposes. [2] [3]

A recent paper by Pang [Pan21] fixes a technical shortcoming of [BHK+16] by constructing a pseudoexpectation operator that satisfies "$\sum_{v \in V} x_v = k$" as a polynomial constraint (whereas the shortcoming was $\widetilde{\mathbb{E}}[\sum_{v \in V} x_v] \geq (1 - o(1))k$ like we have here).

II. Technical Preliminaries

A. The Sum-of-Squares hierarchy

The Sum-of-Squares (SoS) hierarchy is a hierarchy of semidefinite programs parameterized by its degree D. We will work with two equivalent definitions of a degree-D SoS solution: a pseudoexpectation operator $\widetilde{\mathbb{E}}$ (Definition I.3) and a moment matrix. For a degree-D solution to be well defined, we need D to be at least the maximum degree of all constraint polynomials. The degree-D SoS algorithm checks feasibility of a polynomial system by checking whether or not a degree-D pseudoexpectation operator exists. This can be done in time $n^{O(D)}$ via semidefinite programming (ignoring some issues of bit complexity [RW17]). To show an SoS lower bound, one must construct a pseudoexpectation operator that exhibits the desired integrality gap.

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

Definition II.1 (Moment Matrix of $\widetilde{\mathbb{E}}$)**.** *The moment matrix $\Lambda = \Lambda(\widetilde{\mathbb{E}})$ associated to a pseudoexpectation $\widetilde{\mathbb{E}}$*

[2] [BBK+20] provide evidence that their distribution is hard to distinguish from $G_{n,d/n}$ with probability $1 - o(1)$ (it is not "strongly detectable"). However, their distribution *is* distinguishable with probability $\Omega(1)$, via a triangle count (it is "weakly detectable"). In SoS pseudocalibration, this manifests as $\widetilde{\mathbb{E}}[1] = \Theta_d(1)$. We would like the low-degree distinguishing probability to be $o(1)$ i.e. $\widetilde{\mathbb{E}}[1] = 1 + o_d(1)$ so that normalizing by $\widetilde{\mathbb{E}}[1]$ does not affect the objective value.

[3] Another issue is that their planted distribution introduces noise by adding a small number of edges inside the planted independent set.

is a $\binom{[n]}{\le D/2} \times \binom{[n]}{\le D/2}$ *matrix with rows and columns indexed by subsets of* $I, J \subseteq [n]$ *of size at most* $D/2$ *and defined as*

$$\Lambda[I, J] := \widetilde{\mathbb{E}}\left[x^I \cdot x^J\right].$$

To show that a candidate pseudoexpectation satisfies $\widetilde{\mathbb{E}}[f^2] \ge 0$ in Definition I.3, we will rely on the following standard fact.

Fact II.2. *In the definition of pseudoexpectation, Definition I.3, the condition* $\widetilde{\mathbb{E}}[f^2] \ge 0$ *for all* $\deg(f) \le D/2$ *is equivalent to* $\Lambda \succeq 0$.

B. p-biased Fourier analysis

Since we are interested in sparse Erdös-Rényi graphs in this work, we will resort to p-biased Fourier analysis [O'D14, Section 8.4]. Formally, we view the input graph $G \sim G_{n,p}$ as a vector in $\{0,1\}^{\binom{n}{2}}$ indexed by sets $\{i,j\}$ for $i, j \in [n], i \ne j$, where each entry is independently sampled from the p-biased Bernoulli distribution, $\text{Bernoulli}(p)$. Here, by convention $G_e = 1$ indicates the edge e is present, which happens with probability p. The Fourier basis we use for analysis on G is the set of p-biased Fourier characters (which are naturally indexed by graphs H on $[n]$).

Definition II.3. χ *denotes the p-biased Fourier character,*

$$\chi(0) = \sqrt{\frac{p}{1-p}}, \qquad \chi(1) = -\sqrt{\frac{1-p}{p}}.$$

For H *a subset or multi-subset of* $\binom{[n]}{2}$, *let* $\chi_H(G) := \prod_{e \in H} \chi(G_e)$.

We will also need the function $1 - G_e$ which indicates that an edge is not present.

Definition II.4. *For* $H \subseteq \binom{[n]}{2}$, *let* $1_{\overline{H}}(G) = \prod_{e \in H}(1 - G_e)$.

When H is a clique, this is the *independent set indicator* for the vertices in H.

Proposition II.5. *For* $e \in \{0,1\}$, $1 + \sqrt{\frac{p}{1-p}}\chi(e) = \frac{1}{1-p}(1-e)$. *Therefore, for any* $H \subseteq \binom{[n]}{2}$,

$$\sum_{T \subseteq H} \left(\frac{p}{1-p}\right)^{|T|/2} \chi_T(G) = \frac{1}{(1-p)^{|H|}} \cdot 1_{\overline{H}}(G).$$

C. Ribbons and graph matrices

A degree-D pseudoexpectation operator is a vector in $\mathbb{R}^{\binom{[n]}{\le D}}$. The matrices we consider will have rows and columns indexed by all subsets of $[n]$. We express the moment matrix Λ in terms of the Fourier basis on $G_{n,p}$. A particular Fourier character in a particular matrix entry is identified by a combinatorial structure called a *ribbon*.

Definition II.6 (Ribbon)**.** *A ribbon is a tuple* $R = (V(R), E(R), A_R, B_R)$, *where* $(V(R), E(R))$ *is an undirected multigraph without self-loops,* $V(R) \subseteq [n]$, *and* $A_R, B_R \subseteq V(R)$. *Let* $C_R := V(R) \setminus (A_R \cup B_R)$.

Definition II.7 (Matrix for a ribbon)**.** *For a ribbon* R, *the matrix* M_R *has rows and columns indexed by all subsets of* $[n]$ *and has a single nonzero entry,*

$$M_R[I, J] = \begin{cases} \chi_{E(R)}(G) & I = A_R, J = B_R \\ 0 & \text{Otherwise} \end{cases}$$

Definition II.8 (Ribbon isomorphism)**.** *Two ribbons* R, S *are isomorphic, or have the same shape, if there is a bijection between* $V(R)$ *and* $V(S)$ *which is a multigraph isomorphism between* $E(R), E(S)$ *and is a bijection from* A_R *to* A_S *and* B_R *to* B_S. *Equivalently, letting* S_n *permute the vertex labels of a ribbon, the two ribbons are in the same* S_n*-orbit.*

If we ignore the labels on the vertices of a ribbon, what remains is the *shape* of the ribbon.

Definition II.9 (Shape)**.** *A shape is an equivalence class of ribbons with the same shape. Each shape has associated with it a representative* $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$, *where* $U_\alpha, V_\alpha \subseteq V(\alpha)$. *Let* $W_\alpha := V(\alpha) \setminus (U_\alpha \cup V_\alpha)$.

Definition II.10 (Embedding)**.** *Given a shape* α *and an injective function* $\varphi : V(\alpha) \to [n]$, *we let* $\varphi(\alpha)$ *be the ribbon obtained by labeling* α *in the natural way.*

Definition II.11 (Graph matrix)**.** *For a shape* α, *the graph matrix* M_α *is*

$$M_\alpha = \sum_{\text{injective } \varphi: V(\alpha) \to [n]} M_{\varphi(\alpha)}.$$

Injectivity is an important property of graph matrices. On the one hand, we have a finer partition of ribbons than allowing all assignments, and this allows more control. On the other hand, injectivity introduces technically challenging "intersection terms" into graph matrix multiplication. A graph matrix is essentially a sum over all ribbons with shape α (this is not entirely accurate as each ribbon will be repeated $|\text{Aut}(\alpha)|$ times).

Definition II.12 (Automorphism)**.** *For a shape* α, $\text{Aut}(\alpha)$ *is the group of bijections from* $V(\alpha)$ *to itself such that all vertices in* $U_\alpha \cup V_\alpha$ *are fixed and the map is a multigraph automorphism on* $E(\alpha)$ *and a graph*

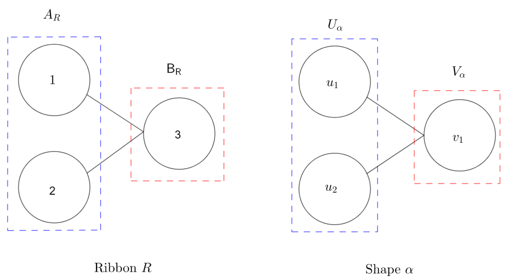

Figure 1. An example of a ribbon and a shape

automorphism on $F(\alpha)$. Equivalently, $\mathrm{Aut}(\alpha)$ is the stabilizer subgroup (of S_n) of any ribbon of shape α.

Fact II.13.

$$\begin{aligned} M_\alpha &= \sum_{\text{injective } \varphi: V(\alpha)\to[n]} M_{\varphi(\alpha)} \\ &= |\mathrm{Aut}(\alpha)| \sum_{R \text{ ribbon of shape } \alpha} M_R \end{aligned}$$

Example II.14 (Ribbon)**.** *As an example, consider the ribbon in Fig. 1. We have $A_R = \{1,2\}, B_R = \{3\}, V(R) = \{1,2,3\}, E(R) = \{\{1,2\},\{2,3\}\}$. The Fourier character is $\chi_{E_R} = \chi_{1,3}\chi_{2,3}$. And finally, M_R is a matrix with rows and columns indexed by subsets of $[n]$, with exactly one nonzero entry $M_R(\{1,2\},\{3\}) = \chi_{\{1,3\}}\chi_{\{2,3\}}$. Succinctly,*

$$M_R = \begin{array}{cc} & \text{column } \{3\} \\ & \downarrow \\ \text{row } \{1,2\} \rightarrow & \begin{pmatrix} 0 & \vdots & 0 \\ \cdots\cdots & \chi_{1,3}\chi_{2,3} & \cdots\cdots \\ 0 & \vdots & 0 \end{pmatrix} \end{array}$$

Example II.15 (Shape)**.** *In Fig. 1, consider the shape α as shown. We have $U_\alpha = \{u_1, u_2\}, V_\alpha = \{v_1\}, W_\alpha = \emptyset, V(\alpha) = \{u_1, u_2, v_1\}$, and $E(\alpha) = \{\{u_1, v_1\}, \{u_2, v_1\}\}$. M_α is a matrix with rows and columns indexed by subsets of $[n]$. The nonzero entries will have rows and columns indexed by $\{a_1, a_2\}$ and b_1 respectively for all distinct $a_1, a_2, b_1 \in [n]$, with the corresponding entry being $M_\alpha(\{a_1,a_2\},\{b_1\}) = \chi_{a_1,b_1}\chi_{a_2,b_1}$. Succinctly,*

$$M_\alpha = \begin{array}{cc} & \text{column } \{b_1\} \\ & \downarrow \\ \text{row } \{a_1,a_2\} \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots\cdots & \chi_{a_1,b_1}\chi_{a_2,b_1} & \cdots\cdots \\ & \vdots & \end{pmatrix} \end{array}$$

Definition II.16 (Proper)**.** *A ribbon or shape is proper if it has no multi-edges. Otherwise, it is improper. Let $\mathsf{mul}_\alpha(e)$ be the multiplicity of edge e in ribbon or shape α.*

An improper ribbon or shape with an edge e of multiplicity 2, e.g., has a squared Fourier character χ_e^2. Since this is a function on $\{0,1\}$, by expressing it in the Fourier basis an improper ribbon or shape can be decomposed in a unique way into a linear combination of proper ones, which we call *linearizations*.

Definition II.17 (Linearization)**.** *Given an improper ribbon or shape α, a linearization β is a proper ribbon or shape such that $\mathsf{mul}_\beta(e) \le \mathsf{mul}_\alpha(e)$ for all $e \in E(\alpha)$.*

Definition II.18 (Isolated vertex)**.** *For a shape α, an isolated vertex is a degree-0 vertex in W_α. Let I_α denote the set of isolated vertices in α. Similarly, for a ribbon R, the isolated vertices are denoted I_R.*

We stress that an isolated vertex never refers to degree-0 vertices inside $U_\alpha \cup V_\alpha$.

Definition II.19 (Trivial shape)**.** *A shape α is trivial if $V(\alpha) = U_\alpha = V_\alpha$ and $E(\alpha) = \emptyset$.*

M_α for a trivial α is the identity matrix restricted to the degree-$|U_\alpha|$ block.

Definition II.20 (Transpose)**.** *Given a ribbon R or shape α, we define its transpose by swapping A_R and B_R (resp. U_α and V_α). Observe that this transposes the matrix for the ribbon/shape.*

III. An overview of the proof techniques

Here, we will give a sketch of the proof techniques that we utilize in our SoS lower bound. Recall that we are given a graph $G \sim G_{n,p}$ where $d = pn$ is the average degree and our goal is to show that for any constant $\varepsilon > 0$, $D_{\mathrm{SoS}} \approx n^\delta$ for some $\delta > 0$, degree D_{SoS} SoS thinks there exists an independent set of size $k \approx \frac{n}{d^{1/2+\varepsilon}(D_{\mathrm{SoS}}\log n)^{c_0}}$ whereas the true independent set has size $\approx \frac{n \log d}{d}$ for some absolute constant c_0.

To prove the lower bound, we review the Planted Clique lower bound [BHK+16] and describe the obstacles that need to be overcome in the sparse setting.

A. Modified pseudocalibration

Since SoS is a convex program, the goal of an SoS lower bound is to construct a dual object: a set of *pseudomoments* $\widetilde{\mathbb{E}}[x^S]$ for each small

$S \subseteq V(G)$, which are summarized in the *moment matrix*. The moment matrix must (i) obey the problem constraints (ii) be SoS-symmetric, and (iii) be positive semidefinite (PSD). Following the recipe of pseudocalibration introduced by [BHK+16], we can produce a candidate moment matrix which is guaranteed to satisfy the first two conditions, while like all other SoS lower bounds, the hard work remains in verifying the PSDness of the moment matrix. Pseudocalibration has been successfully exploited in a multitude of SoS lower bound applications, e.g., [BHK+16], [KMOW17], [MRX20], [GJJ+20], [PR20]. So, this is a natural starting point for us.

Failure of pseudocalibration: The first obstacle we overcome is the lack of a planted distribution. Pseudocalibration requires a planted and random distribution which are hard to distinguish using the *low-degree, likelihood ratio test* (i.e. $\widetilde{\mathbb{E}}[1]$ is bounded whp) [HKP+17], [Hop18]. In the case of sparse independent set, we have the following natural hypothesis testing problem with which one may hope to pseudocalibrate.

- Null Hypothesis: Sample a graph $G \sim G_{n,p}$.
- Alternate Hypothesis: Sample a graph $G \sim G_{n,p}$. Then, sample a subset $S \subseteq [n]$ where each vertex is chosen with probability $\frac{k}{n}$. Then, plant an independent set in S, i.e. remove all the edges inside S.

In the case of sparse independent set, the naïve planted distribution *is* distinguishable from a random instance via a simple low-degree test – counting 4-cycles. In all uses of pseudocalibration that we are aware of, the two distributions being compared are conjecturally hard to distinguish by all polynomial-time algorithms. We are still searching for a suitable planted distribution for sparse independent set, and we believe this is an exciting question on its own.

Fixing pseudocalibration via connected truncation: To get around with this issue, we close our eyes and "pretend" the planted distribution is quiet, ignoring the obvious distinguisher, and make a "connected truncation" of the moment matrix to remove terms which correspond to counting subgraphs in G. What remains is that $\widetilde{\mathbb{E}}[x^S]$ is essentially independent of the global statistics of G. It should be pointed out here that this is inherently distinct from the local truncation for weaker hierarchies (e.g. Sherali-Adams) where the moment matrix is an entirely local function [CM18]. In contrast, our $\widetilde{\mathbb{E}}[x^S]$ may depend on parts of the graph that are far away from S, in fact, even up to radius n^δ, exceeding the diameter of the random graph!

At this point, the candidate moment matrix can be written as follows.

$$\Lambda := \sum_{\alpha \in \mathcal{S}} \left(\frac{k}{n}\right)^{|V(\alpha)|} \cdot \left(\frac{p}{1-p}\right)^{\frac{|E(\alpha)|}{2}} M_\alpha.$$

Here, $\mathcal{S}$ ranges over all proper shapes α of appropriately bounded size such that *all vertices of α are connected to $U_\alpha \cup V_\alpha$*. The latter property is the important distinction from standard pseudocalibration and will turn out to be quite essential for our analysis.

Using connected objects to take advantage of correlation decay is also a theme in the cluster expansion from statistical physics (see Chapter 5 of [FV18]). Although not formally connected with connected truncation, the two methods share some similar characteristics.

B. Approximate PSD decompositions, norm bounds and conditioning

Continuing, to show the moment matrix is PSD, Planted Clique [BHK+16] performs an approximate factorization of the moment matrix in terms of graph matrices. A crucial part of this approach is to identify "dominant" and "non-dominant" terms in the approximate PSD decomposition. Then, the dominant terms are shown to be PSD and the non-dominant terms are bounded against the dominant terms. In this approach, a crucial component in the latter step is to control the norms of graph matrices.

Tighter norm bounds for sparse graph matrices: Existing norm bounds in the literature [MP16], [AMP20] for graph matrices have focused exclusively on the dense setting $G_{n,1/2}$. Unfortunately, while these norm bounds apply for the sparse setting, they're too weak to be useful. Consider the case where we sample $G \sim G_{n,p}$ and try to bound the spectral norm of the centered adjacency matrix. Existing norm bounds give a bound of $\tilde{O}(\frac{\sqrt{n(1-p)}}{\sqrt{p}})$ whereas the true norm is $O(\sqrt{n})$ regardless of d. This is even more pronounced when we use shapes with more vertices. So, our first step is to tighten the existing norm bounds in the literature for sparse graph matrices.

For a shape α, a vertex separator S is a subset of vertices such that there are no paths from U_α to

V_α in $\alpha \setminus S$. It is known from previous works that in the dense case the spectral norm is controlled by the number of vertices in the minimum vertex separator between U_α and V_α. Assuming α does not have isolated vertices and $U_\alpha \cap V_\alpha = \emptyset$ for simplicity, the norm of M_α is given by the following expression, up to polylog factors and the leading coefficient of at most $|V(\alpha)|^{|V(\alpha)|}$,

$$\|M_\alpha\| \le \max_{\text{vertex separator } S} \sqrt{n}^{|V(\alpha)|-|V(S)|}$$

However, it turns out this is no longer the controlling quantity if the underlying input matrix is sparse, and tightly determining this quantity arises as a natural task for our problem, and for future attack on SoS lower bounds for other problems in the sparse regime. To motivate the difference, we want to point out this is essentially due to the following simple observation. For $k \ge 2$,

$$|\mathbb{E}[X^k]| = 1$$
$$|\mathbb{E}[Y^k]| \le \left(\sqrt{\frac{1-p}{p}}\right)^{k-2} \approx \left(\sqrt{\frac{n}{d}}\right)^{k-2}$$

for X a uniform ± 1 bit and Y a p-biased random variable $\mathrm{Ber}(p)$. This suggests that in the trace power method, there will be a preference among vertex separators of the same size if some contain more edges inside the separator (because vertices inside vertex separators are "fixed" in the dominant term in the trace calculation, and thus edges within the separator will contribute some large power of $\sqrt{\frac{n}{d}}$, creating a noticeable influence on the final trace). Finally, this leads us to the following characterization for sparse matrix norm bounds, up to polylog factors and the leading coefficient of at most $|V(\alpha)|^{|V(\alpha)|}$,

$$\|M_\alpha\| \le \max_{\text{vertex separator } S} \sqrt{n}^{|V(\alpha)|-|V(S)|} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|}$$

We prove this via an application of the trace method followed by a careful accounting of the large terms.

The key conceptual takeaway is that we need to redefine the weight of a vertex separator to also incorporate the edges within the separator, as opposed to only considering the number of vertices. We clearly distinguish these with the terms *Dense Minimum Vertex Separator* (DMVS) and *Sparse Minimum Vertex Separator* (SMVS). When $p = \frac{1}{2}$, these two bounds are the same up to lower order factors.

Approximate PSD decomposition: We then perform an approximate PSD decomposition of the graph matrices that make up Λ. The general factoring strategy is the same as [BHK+16], though in the sparse regime we must be very careful about what kind of combinatorial factors we allow. Each shape comes with a natural "vertex decay" coefficient arising from the fractional size of the independent set and an "edge decay" coefficient arising from the sparsity of the graph. The vertex decay coefficients can be analyzed in a method similar to Planted Clique (which only has vertex decay). For the edge decay factors, we use novel charging arguments. At this point, the techniques are strong enough to prove Theorem I.8, an SoS-degree $\Omega(\log n)$ lower bound for $d \ge n^\varepsilon$. The remaining techniques are needed to push the SoS degree up and the graph degree down.

Conditioning: In our analysis, it turns out that to obtain strong SoS lower bounds in the sparse regime, a norm bound from the vanilla trace method is not quite sufficient. Sparse random matrices' spectral norms are fragile with respect to the influence of an unlikely event, exhibiting deviations away from the expectation with polynomially small probability (rather than exponentially small probability, like what is obtained from a good concentration bound). These "bad events" are small dense subgraphs present in a graph sampled from $G_{n,p}$.

To get around this, we condition on the high probability event that G has no small dense subgraphs. For example, for $d = n^{1-\varepsilon}$ whp every small subgraph S has $O(|S|)$ edges (even up to size n^δ). For a **shape** which is dense (i.e. v vertices and more than $O(v)$ edges) we can show that its norm falls off extremely rapidly under this conditioning. This allows us to throw away dense shapes, which is critical for controlling combinatorial factors that would otherwise dominate the moment matrix.

This type of conditioning is well-known: a long line of work showing tight norm bounds for the simple adjacency matrix appeals to a similar conditioning argument within the trace method [BLM15], [Bor19], [FM17], [DMO+19]. We instantiate the conditioning in two ways. The first is through the following identity.

Observation III.1. *Given a set of edges $E \subseteq \binom{[n]}{2}$, if we know that not all of the edges of E are in $E(G)$ then*

$$\chi_E(G) = \sum_{E' \subseteq E : E' \ne E} \left(\sqrt{\frac{p}{1-p}}\right)^{|E|-|E'|} \chi_{E'}(G)$$

This simple observation can be applied recursively to replace a dense shape α by a sum of its sparse subshapes $\{\beta\}$. The second way we eliminate dense shapes is by using a bound on the Frobenius norm which improves on the trace calculation for dense shapes. After conditioning, we can restrict our attention to sparse shapes, which allows us to avoid several combinatorial factors which would otherwise overwhelm the "charging" argument.

Handling the subshapes $\{\beta\}$ requires some care. Destroying edges from a shape can cause its norm to either go up or down: the vertex separator gets smaller (increasing the norm), but if we remove edges from inside the SMVS, the norm goes down. An important observation is we do not necessarily have to apply Observation III.1 on the entire set of edges of a shape, but we can also just apply it on some of the edges. We will choose a set of edges $\mathsf{Res}(\alpha) \subseteq E(\alpha)$ that "protects the minimum vertex separator" and only apply conditioning on edges outside $\mathsf{Res}(\alpha)$. In this way the norm of subshapes β will be guaranteed to be less than α. The fact that it's possible to reserve such edges is shown separately for the different kind of shapes we encounter in our analysis

Finally, we are forced to include certain dense shapes that encode indicator functions of independent sets. These shapes must be factored out and tracked separately throughout the analysis.[4] After handling all of these items we have shown that Λ is PSD.

IV. Open Problems

Several other problems on sparse graphs are conjectured to be hard for SoS and it is our hope that the techniques here can help prove that these problems are hard for SoS. These problems include MaxCut and k-Coloring. For MaxCut in particular, since there are no constraints other than booleanity of the variables it may be possible to truncate away dense shapes, which we could not do here due to the presence of independent set indicator functions.

Another direction for further research is to handle random graphs which are not Erdős-Rényi. Since the techniques here depend on graph matrix norms, one would hope that they generalize to distributions such as d-regular graphs for which low-degree polynomials are still concentrated. However, in the non-iid setting, it is not clear what the analogue of graph matrices should be used due to the lack of a Fourier basis that is friendly to work with.

The polynomial constraint "$\sum_{v \in V} x_v = k$" is not satisfied exactly by our pseudoexpectation operator. It's possible that techniques from [Pan21] can be used to fix this.

The parameters in this paper can likely be improved. One direction is to remove the final factor of $\log n$ from our bound. This would allow us to prove an SoS lower bound for the "ultrasparse regime" $d = O(1)$ rather than $d \geq \log^2 n$. This setting is interesting as there is a nontrivial algorithm that finds an independent set of half optimal size [GS17], [RV17]. Furthermore, this algorithm is local in a sense that we don't define here. It would be extremely interesting if this algorithm could be converted into a rounding algorithm for constant-degree SoS.

Another direction is to improve the dependence on D_{SoS}. While our bound has a $\frac{1}{poly(D_{\mathrm{SoS}})}$ dependence on D_{SoS}, we conjecture that the dependence should actually be $(1-p)^{O(D_{\mathrm{SoS}})}$. If so, this would provide strong evidence for the prevailing wisdom in parameterized complexity and proof complexity that a maximum independent set of size k requires $n^{\Omega(k)}$ time to find/certify (corresponding to SoS degree $\Omega(k)$).

Acknowledgments

We thank anonymous reviewers for their useful suggestions in improving the manuscript. Chris and Aaron are supported in part by NSF grant CCF-2008920. Goutham and Madhur are supported in part by NSF grant CCF-1816372. Jeff is supported by NSF CAREER Award #2047933.

References

[ABB19] Enric Boix Adserà, Matthew Brennan, and Guy Bresler. The average-case complexity of counting cliques in Erdős-Rényi hypergraphs. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, pages 1256–1280. IEEE Comput. Soc. Press, Los Alamitos, CA, [2019] ©2019. 1

[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. 3, 7

[4] One may ask if we could definitionally get rid of dense shapes, like we did for disconnected shapes via the connected truncation, but these dense shapes are absolutely necessary.

[BBK+20] Afonso S. Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S. Wein. Spectral planting and hardness of refuting cuts, colorability, and communities in random graphs. *arXiv preprint arXiv:2008.12237*, 2020. 4

[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, 3, 4, 6, 7, 8

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: Community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357, 2015. doi:10.1109/FOCS.2015.86. 8

[Bor19] Charles Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. Technical Report 1502.04482v4, arXiv, 2019. To appear in Annales scientifiques de l'École normale supérieure. 8

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2011. 1

[BS16] Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares. *Course notes: http://www. sumofsquares. org/public/index. html*, 2016. 1

[CM18] Eden Chlamtác and Pasin Manurangsi. Sherali-adams integrality gaps matching the log-density threshold. *CoRR*, abs/1804.07842, 2018. URL: http://arxiv.org/abs/1804.07842, arXiv:1804.07842. 7

[CO05] Amin Coja-Oghlan. The Lovász number of random graphs. *Combinatorics, Probability and Computing*, 14(4):439–465, 2005. 2

[COE15] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015. 2

[DM11] Varsha Dani and Cristopher Moore. Independent sets in random graphs from the weighted second moment method. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 472–482. Springer, Heidelberg, 2011. URL: https://doi.org/10.1007/978-3-642-22935-0_40, doi:10.1007/978-3-642-22935-0_40. 2

[DMO+19] Yash Deshpande, Andrea Montanari, Ryan O'Donnell, Tselil Schramm, and Subhabrata Sen. The threshold for sdp-refutation of random regular nae-3sat. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2305–2321. SIAM, 2019. 8

[DSS16] Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *Acta Math.*, 217(2):263–340, 2016. doi:10.1007/s11511-017-0145-9. 2

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Foundations and TrendsÂ® in Theoretical Computer Science*, 14(1-2):1–221, 2019. URL: http://dx.doi.org/10.1561/0400000086, doi:10.1561/0400000086. 1

[FM17] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614. ACM, 2017. 8

[FV18] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems*. Cambridge University Press, Cambridge, 2018. 7

[GJJ+20] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science—FOCS 2020*, pages 954–965. IEEE Computer Soc., Los Alamitos, CA, [2020] ©2020. 7

[GM75] G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Math. Proc. Cambridge Philos. Soc.*, 77:313–324, 1975. doi:10.1017/S0305004100051124. 4

[GS11] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *FOCS*, pages 482–491, 2011. 1

[GS17] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Ann. Probab.*, 45(4):2353–2376, 2017. doi:10.1214/16-AOP1114. 9

[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. 4, 7

[Hop18] Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018. 7

[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

[KMOW17] Pravesh Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, 2017. 2, 4, 7

[Kun20] Dmitriy Kunisky. Positivity-preserving extensions of sum-of-squares pseudomoments over the hypercube. *arXiv preprint arXiv:2009.07269*, 2020. 4

[MP16] Dhruv Medarametla and Aaron Potechin. Bounds on the norms of uniform low degree graph matrices. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016. 7

[MRX20] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: degree-2 to degree-4. In *Proceedings of the 52nd ACM Symposium on Theory of Computing*, pages 840–853, 2020. 7

[MSS16] Tengyu Ma, Jonathan Shi, and David Steurer. Polynomial-time tensor decompositions with sum-of-squares. In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science*, pages 438–446. IEEE, 2016. 1

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014. doi:10.1017/CBO9781139814782. 5

[Pan21] Shuo Pang. SOS lower bound for exact planted clique. In *36th Computational Complexity Conference*, volume 200 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. 26, 63. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021. 4, 9

[PR20] Aaron Potechin and Goutham Rajendran. Machinery for proving sum-of-squares lower bounds on certification problems. abs/2011.04253, 2020. URL: https://arxiv.org/abs/2011.04253, arXiv:2011.04253. 2, 4, 7

[RT20] Goutham Rajendran and Madhur Tulsiani. Nonlinear concentration via matrix efronstein. Manuscript, 2020. 4

[RV17] Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. *Ann. Probab.*, 45(3):1543–1577, 2017. doi:10.1214/16-AOP1094. 4, 9

[RW17] Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In *Proceedings of the 44th International Colloquium on Automata, Languages and Programming*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 4

[Wei20] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *arXiv preprint arXiv:2010.06563*, 2020. 4

[Wor95] Nicholas C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, 5(4):1217–1235, 1995. URL: http://links.jstor.org/sici?sici=1050-5164(199511)5:4<1217:DEFRPA>2.0.CO;2-A&origin=MSN. 4