THEORY OF COMPUTING, Volume 17 (9), 2021, pp. 1–37
www.theoryofcomputing.org

SPECIAL ISSUE: CCC 2019

Sherali–Adams Strikes Back

Ryan O'Donnell* Tselil Schramm†

Received August 31, 2019; Revised October 29, 2020; Published September 28, 2021

Abstract. Let G be any n-vertex graph whose random walk matrix has its nontrivial eigenvalues bounded in magnitude by $1/\sqrt{\Delta}$ (for example, a random graph G of average degree $\Theta(\Delta)$ typically has this property). We show that the $\exp\left(c\frac{\log n}{\log \Delta}\right)$-round Sherali–Adams linear programming hierarchy certifies that the maximum cut in such a G is at most 50.1% (in fact, at most $\frac{1}{2}+2^{-\Omega(c)}$). For example, in random graphs with $n^{1.01}$ edges, $O(1)$ rounds suffice; in random graphs with $n \cdot \mathrm{polylog}(n)$ edges, $n^{O(1/\log\log n)} = n^{o(1)}$ rounds suffice.

Our results stand in contrast to the conventional beliefs that linear programming hierarchies perform poorly for MAX-CUT and other CSPs, and that eigenvalue/SDP methods are needed for effective refutation. Indeed, our results imply that constant-round Sherali–Adams can strongly refute random Boolean k-CSP instances with $n^{\lceil k/2 \rceil+\delta}$ constraints; previously this had only been done with spectral algorithms or the SOS SDP hierarchy. We also show similar results for other classical combinatorial optimization problems such as independent set, max clique, and vertex cover.

ACM Classification: G.1.6

AMS Classification: 03F20, 68W25

Key words and phrases: hierarchies, Sherali–Adams, combinatorial optimization

---

A conference version of this paper appeared in the Proceedings of the 34th Computational Complexity Conference (CCC'19) [37].

*Supported by NSF grants CCF-1618679, CCF-1717606. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

†This work was supported by NSF grants CCF 1565264 and CNS 1618026.

 DOI: 10.4086/toc.2021.v017a009

1 Introduction

Linear programming (LP) is a fundamental algorithmic primitive, and is the method of choice for a huge number of optimization and approximation problems. Still, there are some very basic tasks where it performs poorly. A classic example is the simplest of all constraint satisfaction problems (CSPs), the MAX-CUT problem: Given a graph $G=(V,E)$, partition V into two parts so as to maximize the fraction of "cut" (crossing) edges. The standard LP relaxation for this problem [5, 38] involves optimizing over the *metric polytope*. Using "± 1 notation," we have a variable Y_{uv} for each pair of vertices $\{u,v\}$ (with Y_{uv} supposed to be -1 if the edge is cut, $+1$ otherwise); the LP is:

$$\textsc{max-cut}(G) \le \max \frac{1}{2}-\frac{1}{2}\cdot\frac{1}{|E|}\sum_{uv\in E}\mathsf{Y}_{uv} \qquad \text{s.t.} \quad \begin{array}{rl} -1 \le \mathsf{Y}_{uv} \le 1 & (\text{for all } u,v\in V) \\ -\mathsf{Y}_{uv}-\mathsf{Y}_{vw}-\mathsf{Y}_{wu} \le 1 & (\text{for all } u,v,w\in V) \\ -\mathsf{Y}_{uv}+\mathsf{Y}_{vw}+\mathsf{Y}_{wu} \le 1 & (\text{for all } u,v,w\in V) \end{array}$$

While this LP gives the optimal bound for some graphs (precisely, all graphs not contractible to K_5 [5]), it can give a very poor bound in general. Indeed, although there are graphs with maximum cut arbitrarily close to $1/2$ (e. g., K_n), the above LP bound is at least $2/3$ for every graph, since $\mathsf{Y}_{uv}\equiv -1/3$ is always a valid solution. Worse, there are graphs G with MAX-CUT(G) arbitrarily close to $1/2$ but with LP value arbitrarily close to 1—i. e., graphs where the *integrality ratio* is $2-o(1)$. For example, this is true [39] of an Erdős–Rényi $\mathcal{G}(n,\Delta/n)$ random graph with high probability (whp) when $\Delta=\Delta(n)$ satisfies $\omega(1)<\Delta<n^{o(1)}$.

There have been two main strategies employed to overcome this deficiency: strengthened LPs, and eigenvalue methods.

Strengthened LPs. One way to try to improve the performance of LPs on MAX-CUT is to add more valid inequalities to the LP relaxation, beyond just the "triangle inequalities." Innumerable valid inequalities have been considered: $(2k+1)$-gonal, hypermetric, negative type, gap, clique-web, suspended tree, as well as inequalities from the Lovász–Schrijver hierarchy; see Deza and Laurent [18, Ch. 28–30] for a review.

It is now known that the most principled and general form of this strategy is the *Sherali–Adams LP hierarchy* [45], reviewed in Section 2.4. At a high level, the Sherali–Adams LP hierarchy gives a standardized way to tighten LP relaxations of Boolean integer programs, by adding variables and constraints. The number of new variables/constraints is parameterized by a positive integer R, called the number of "rounds." Given a Boolean optimization problem with n variables, the R-round Sherali–Adams LP has variables and constraints corresponding to monomials of degree up to R, and thus has size $O(n)^R$. A remarkable recent line of work [12, 33] has shown that for any CSP (such as MAX-CUT), the R-round Sherali–Adams LP relaxation achieves essentially[1] the tightest integrality ratio among all LPs of its size. Nevertheless, even this most powerful of LPs arguably struggles to certify good bounds for MAX-CUT. In a line of work [23, 44] concluding in a result of Charikar–Makarychev–Makarychev [13], it was demonstrated that for any constant $\varepsilon>0$, there are graphs (random Δ-regular ones, $\Delta=O(1)$) for which

[1] Provided that the LP is an extended formulation, which means that the feasible region is required to contain a representation of each solutions $\{\pm 1\}^n$, so only the objective function of the LP can depend on the specific max cut instance.

the $n^{\Omega(1)}$-round Sherali–Adams LP has a MAX-CUT integrality gap of $2-\varepsilon$. As a consequence, *every* MAX-CUT LP relaxation of size up to $2^{n^{\Omega(1)}}$ has such an integrality gap.

Eigenvalue and SDP methods. But for MAX-CUT, there is a simple, non-LP, algorithm that works very well to certify that random graphs have maximum cut close to $1/2$: the *eigenvalue* bounds. There are two slight variants here (that coincide in the case of regular graphs): Given graph $G=(V,E)$ with adjacency matrix A and diagonal degree matrix D, the eigenvalue bounds are

$$\text{MAX-CUT}(G) \leq \frac{|V|}{4|E|}\lambda_{\max}(D-A)\,, \tag{1.1}$$

$$\text{MAX-CUT}(G) \leq \frac{1}{2}+\frac{1}{2}\lambda_{\max}(-D^{-1}A)\,. \tag{1.2}$$

Here $D-A$ and $D^{-1}A$ are the *Laplacian* matrix and the *random walk* matrix, respectively. The use of eigenvalues to bound various cut values in graphs (problems like MAX-CUT, MIN-BISECTION, 2-XOR, expansion, etc.) has a long history dating back to Fieldler and Donath–Hoffman [24, 19] among others (Inequality (1.1) is specifically from Mohar–Poljak [36]). It was recognized early on that eigenvalue methods work particularly well for solving planted-random instances (e. g., of 2-XOR [31] and MIN-BISECTION [10]) and for certifying MAX-CUT values near $1/2$ for truly random instances. Indeed, as soon as one knows (as we now do [46, 22]) that $D^{-1}A$ has all nontrivial eigenvalues bounded in magnitude by $O(1/\sqrt{\Delta})$ (whp) for a random Δ-regular graph (or an Erdős–Rényi $\mathcal{G}(n,\Delta/n)$ graph with $\Delta \gtrsim \log n$), the eigenvalue bound (1.2) certifies that $\text{MAX-CUT}(G) \leq 1/2+O(1/\sqrt{\Delta})$. This implies an integrality ratio tending to 1; indeed, $\text{MAX-CUT}(G) = 1/2+\Theta(1/\sqrt{\Delta})$ in such random graphs (whp).

Furthermore, if one extends the eigenvalue bound (1.1) above to

$$\text{MAX-CUT}(G) \leq \min_{\substack{U \text{ diagonal} \\ \operatorname{tr}(U)=0}} \frac{|V|}{4|E|}\lambda_{\max}(D-A+U)$$

(as suggested by Delorme and Poljak [17], following [19, 10]), one obtains the polynomial-time computable *semidefinite programming (SDP)* bound. Goemans and Williamson [28] showed this bound has integrality ratio less than $1.14 \approx 1/.88$ for *worst-case* G, and it was subsequently shown [49, 21, 15] that the SDP bound is $1/2+o(1)$ whenever $\text{MAX-CUT}(G) \leq 1/2+o(1)$.

LPs cannot compete with eigenvalues/SDPs? This seemingly striking separation between the performance of LPs and SDPs in the context of random MAX-CUT instances is now taken as a matter of course. To quote, e. g., [47],

> [E]xcept for semidefinite programming, we know of no technique that can provide, for every graph of max cut optimum $\leq .501$, a certificate that its optimum is $\leq .99$. Indeed, the results of [23, 44, 13] show that large classes of Linear Programming relaxations of max cut are unable to distinguish such instances.

Specifically, the last statement here is true for Δ-regular random graphs when Δ is a certain large constant. The conventional wisdom is that for such graphs, linear programs cannot compete with semidefinite programs, and cannot certify even the eigenvalue bound.

Our main result challenges this conception.

1.1 Our results

We show that whenever the eigenvalue bound (1.2) certifies the bound $\text{MAX-CUT}(G) \le 1/2 + o(1)$, then $n^{o(1)}$-round Sherali–Adams can certify this as well.[2]

Theorem 1.1 (Simplified version of Theorem 3.6 and Theorem 3.7)**.** *Let G be a simple n-vertex graph and assume that $|\lambda| < \rho$ for all eigenvalues λ of G's random walk matrix $D^{-1}A$ (excluding the trivial eigenvalue of 1). Then for any $1 \le c \le \Omega(\log(1/\rho))$, Sherali–Adams with $n^{O(c/\log(1/\rho))}$ rounds certifies that $\text{MAX-CUT}(G) \le 1/2 + 2^{-c}$.*

For example, if G's random walk matrix has its nontrivial eigenvalues bounded in magnitude by $n^{-.001}$, as is the case (whp) for random graphs with about $n^{1.002}$ edges, then Sherali–Adams can certify $\text{MAX-CUT}(G) \le 50.1\%$ with *constantly* many rounds. We find this result surprising, and in defiance of the common belief that polynomial-sized LPs cannot take advantage of spectral properties of the underlying graph.[3]

(As an aside, in Theorem 4.2, we show that the R-round Sherali–Adams relaxation for MAX-CUT has value at least $1/2 + \Omega(1/R)$ for *every* graph G. This at least demonstrates that some dependence of the refutation strength on the number of Sherali–Adams rounds is always necessary.)

One might ask whether Theorem 1.1 even requires the assumption of small eigenvalues. That is, perhaps $n^{o(1)}$-round Sherali–Adams can certify $\text{MAX-CUT} \le 1/2 + o(1)$ whenever this is *true*. As far as we know, this may be possible; after all, the basic SDP relaxation has this property [49, 21, 15]. On the other hand, the eigenvalue bound itself does *not* have this property; there exist graphs with large (nontrivial) eigenvalues even though the maximum cut is close to $1/2$.[4]

1.1.1 Subexponential-sized LPs for MAX-CUT in sparse random graphs

One setting in which the spectral radius ρ is understood concretely is in random regular graphs. Building upon [26, 11, 16], the following was recently shown:

Theorem 1.2 ([46])**.** *There is a fixed constant C such that for all $3 \le \Delta \le n/2$ with Δn even, it holds that a uniformly random n-vertex Δ-regular simple graph $\boldsymbol{G}$ satisfies the following with high probability: all eigenvalues of $\boldsymbol{G}$'s normalized adjacency matrix, other than 1, are at most $C/\sqrt{\Delta}$ in magnitude.*

---

[2] Actually, there is a slight mismatch between our result and Inequality (1.2): in Theorem 1.1 we need the maximum eigenvalue *in magnitude* to be small; i. e., we need $\lambda_{\min}(-D^{-1}A)$ to be not too negative. This may well just be an artifact of our proof.

[3] We should emphasize that it is not concomitant $n^{O(1)}$ running time that is surprising, since the eigenvalue bound itself already achieves that.

[4] Consider, for example, a graph given by the union of a Δ-regular random graph on n vertices and a Δ-regular bipartite graph on $\sqrt{n}$ vertices. This will have MAX-CUT value close to $1/2$, but will also have large negative eigenvalues coming from the bipartite component.

Combining the above with Theorem 1.1, we have the following consequence for MAX-CUT on random regular graphs:

Corollary 1.3. *Let n, $3 \le \Delta \le n/2$, and $1 \le c \le \Omega(\log \Delta)$ be positive integers. Then if G is a random Δ-regular n-vertex graph, with high probability $n^{O(c/\log \Delta)}$-round Sherali–Adams can certify that* $\text{MAX-CUT}(G) \le \frac{1}{2} + 2^{-c}$.

For example, if $\Delta \ge C \cdot 10^6$ (for C the constant in the bound on $\lambda(G)$), then $n^{1/3}$-rounds of Sherali–Adams can certify $\text{MAX-CUT}(G) \le .51$. This result serves as a partial converse to [13]:

Theorem 1.4. ([13, Theorem 5.3]) *For every fixed integer $\Delta \ge 3$, with high probability over the choice of an n-vertex Δ-regular random graph G,*[5] *the $n^{\Theta(1/f(\Delta))}$-round Sherali–Adams relaxation for* MAX-CUT *has value at least* $\text{MAX-CUT}(G) \ge 0.99$*, where $f(\Delta)$ is a function that grows with Δ.*[6]

While [13] show that Δ-regular random graphs require Sherali–Adams (and by [33] any LP) relaxations of at least subexponential size, our result implies that subexponential LPs are sufficient. Further, though the function $f(\Delta)$ is not specified in [13], by tracing back through citations (e. g., [3, 4, 14]) to extract a dependence, it appears we may take $f(\Delta) = \log \Delta$. So our upper bound is *tight* as a function of Δ, up to constant factors.

Prior to our result, it was unclear whether even $(n/\text{polylog}\, n)$-round Sherali–Adams could certify that the MAX-CUT value was < 0.9 for sparse random regular graphs. Indeed, it was equally if not more conceivable that Charikar et al.'s result was not tight, and could be extended to $\tilde{\Omega}(n)$-rounds. In light of our result, we are left to wonder whether there are instances of MAX-CUT which have truly exponential extension complexity.

1.1.2 Refuting random CSPs with linear programs

With minor modifications, our argument extends as well to 2-XOR. Then following the framework in [2], we have the following consequence for certifying bounds on the value of random k-CSPs:

Theorem 1.5 (Simplified version of Theorem 5.2)**.** *Suppose that $P : \{\pm 1\}^k \to \{0,1\}$ is a k-ary Boolean predicate, and that $\delta, \varepsilon > 0$. Let $\mathbf{E}[P]$ be the probability that a random $x \in \{\pm 1\}^k$ satisfies P. Then for a random instance $\mathcal{I}$ of P on n variables with $m \ge n^{\lceil k/2 \rceil + \delta}$ expected clauses, with high probability Sherali–Adams can certify that $\text{OBJ}_{\mathcal{I}}(x) \le \mathbf{E}[P] + \varepsilon$ using $R = O_{\varepsilon,\delta,k}(1)$ rounds.*

This almost matches the comparable performance of sum-of-squares and spectral algorithms [2], which are known to require $m \ge n^{k/2}$ clauses to certify comparable bounds in polynomial time [30, 43, 34].[7] Prior to our work it was known that Sherali–Adams admits *weak* refutations (i. e., a certificate that

[5]In [13], the graph is actually a *pruned* random graph, in which $o(n)$ edges are removed; this does not affect compatibility with our results, as the LP value is Lipschitz and so the pruning changes the LP value by $o(1)$.

[6]Though $f(\Delta)$ is not specified in [13] (and in their proof $f(\Delta)$ is dictated by a combination of lemmas in prior work) it appears we can take $f(\Delta) = \log \Delta$.

[7]The expert may notice that we require the number of clauses $m \gg n^{\lceil k/2 \rceil}$, whereas the best sum-of-squares and spectral algorithms require only $m \gg n^{k/2}$. This is because we do not know how to relate the Sherali–Adams value of the objective function to its square (local versions of the Cauchy–Schwarz argument result in a loss). Such a relation would allow us to apply our techniques immediately to prove that Sherali Adams matches the SOS and spectral performance for odd as well as even k.

$\mathrm{OBJ} \le 1 - o(1)$) when $m \ge n^{k/2}$, but it was conceivable (and even conjectured) that $O(1)$-rounds could not certify $\mathrm{OBJ} \le 1 - \delta$ for constant δ at $m = o(n^k)$.

The result above also extends to t-wise independent predicates as in [2] (see Section 5). Also, one may extract the dependence on the parameters ε, δ to give nontrivial results when these parameters depend on n.[8]

1.1.3 Independent Set, Clique, and Vertex Cover

Our result also extends to constrained optimization problems such as Independent Set, Vertex Cover, and Clique in regular[9] graphs. Here again, our results complement strong LP lower bounds obtained by Charikar et al. and others (e. g., [3, 14]). We prove a theorem for Independent Set in regular graphs:

Theorem 1.6 (Simplified version of Theorem 3.11)**.** *Suppose that G is a regular, simple, n-vertex graph and assume that $|\lambda| < \rho$ for all eigenvalues of G's random walk matrix $D^{-1}A$ (excluding the trivial 1 eigenvalue). Then for any $1 \le c \le \Omega(\log(1/\rho))$, Sherali–Adams with $n^{O(c/\log(1/\rho))}$ rounds certifies that* $\textsc{independent-set}(G) \le 2^{-c}$.

Because the maximum independent set value is related to the minimum vertex cover value via a proof that is captured by the Sherali–Adams proof system, and because the maximum clique problem is equivalent to maximum independent set on the complement graph, this also implies that Sherali–Adams certifies nontrivial bounds on the value of the maximum clique and minimum independent set. We leave the precise consequences for clique and vertex cover as an exercise for the interested reader.

1.2 Prior work

It is a folklore result that in random graphs with average degree n^δ, 3-round Sherali–Adams (SA) certifies a MAX-CUT value of at most $\max(1 - \Omega(\delta), \frac{2}{3})$ (observed for the special case of $\delta > \frac{1}{2}$ in [6, 39]); this is simply because of concentration phenomena, since most edges participate in roughly the same number of odd cycles of length $O(\frac{1}{\delta})$, after which one can apply the triangle inequality. However, this observation does not allow one to take the refutation strength independent of the average degree.

There is some prior work examining the performance of Sherali–Adams hierarchies on random CSPs. De la Vega and Mathieu [23] show that in *dense* graphs, with average degree $\Omega(n)$, $O(1)$ rounds of Sherali–Adams certify tight bounds on MAX-CUT. This result heavily leverages the density of the graph, and is consistent with the fact that dense MAX-CUT is known to admit a PTAS [27]. However, their work does not give $O(1)$-round approximations better than $2 - o(1)$ for graphs with $o(n^2)$ edges.

Another relevant line of work is a series of LP hierarchy lower bounds (both for Sherali–Adams and for the weaker Lovász–Schrijver hierarchy) for problems such as MAX-CUT, vertex cover, and sparsest cut, including [1, 3, 23, 44], and culminating in the already mentioned result of Charikar, Makarychev and Makarychev; in [13], they give subexponential lower bounds on the number of rounds of Sherali–Adams required to strongly refute MAX-CUT in random regular graphs. Initially, one might expect that this result

[8] Though for 2-XOR and MAX-CUT we have done this explicitly, for higher-arity random CSPs we have left this for the interested reader.

[9] Theorem 3.11 actually gives nontrivial bounds for irregular graphs as well, but what we are able to control is the size of the clique as weighted by the vertex degree, rather than weighting all vertices equally.

could be strengthened to prove that sparse random graphs require almost-exponential-sized LPs to refute MAX-CUT; our result demonstrates instead that [13] is almost tight.

We also mention the technique of *global correlation rounding* in the sum-of-squares hierarchy, which was used to give subexponential time algorithms for unique games [8] and polynomial-time approximations to max-bisection [42]. One philosophical similarity between these algorithms and ours is that they both relate local properties (correlation among edges) to global properties (correlation of uniformly random pairs). But [8, 42] use the fact that the relaxation is an SDP (whereas our result is interesting *because* it is in the LP-only setting), and the "conditioning" steps that drive their algorithm are a fundamentally different approach.

There is much prior work concerned with certifying bounds on random CSPs, and we survey only some of them here, referring the interested reader to the discussion in [2]. The sequence of articles [30, 43, 34] establishes sum-of-squares lower bounds for refuting any random constraint satisfaction problem, and these results are tight via the SOS algorithms of [2, 40]. The upshot is that for k-SAT and k-XOR,[10] $\omega(1)$ rounds of SOS are necessary to strongly refute an instance with $m = o(n^{k/2})$ clauses, and $O(1)$ rounds of SOS suffice when $m = \tilde{\Omega}(n^{k/2})$. Because SOS is a tighter relaxation than Sherali–Adams, the lower bounds [30, 43, 34] apply; our work can be seen to demonstrate that SA does not lag far behind SOS, strongly refuting with $O(1)$ rounds as soon as $m = \Omega(n^{\lceil k/2 \rceil + \delta})$ for any $\delta > 0$.

In a way, our result is part of a trend in anti-separation results for SDPs and simpler methods for pseudorandom and structured instances. For example, we have for planted clique that the SOS hierarchy performs no better than the Lovász–Schrijver+ hierarchy [20, 7], and also no better than a more primitive class of estimation methods based on local statistics (see, e. g., [41] for a discussion). Similar results hold for problems relating to estimating the norms of random tensors [32]. Further, in [32] an equivalence is shown between SOS and spectral algorithms for a large class of average-case problems. Our result shows that for random CSPs, the guarantees of linear programs are surprisingly not far from the guarantees of SOS.

Finally, we mention related work in extended formulations. The sequence of articles [12, 33] shows that SA lower bounds for CSPs imply lower bounds for any LP relaxation which is an extended formulation; the stronger (and later) statement is due to Kothari et al. [33], who show that subexponential-round integrality gaps for CSPs in the Sherali–Adams hierarchy imply subexponential-size lower bounds for any LP extended formulation. The techniques of these papers are then applied in conjunction with the techniques of [30, 43, 13] to give subexponential lower bounds against CSPs for any LP. Our results give an upper limit to the mileage one can get from these lower bounds in the case of MAX-CUT, as we show that the specific construction of [13] cannot be strengthened much further.

1.3 Techniques

Our primary insight is that while Sherali–Adams is unable to reason about spectral properties *globally*, it does enforce that every set of R variables behave *locally* according to the marginals of a valid distribution, which induces *local* spectral constraints on every subset of up to R variables.

At first, it is unclear how one harnesses such local spectral constraints. But now, suppose that we are in a graph with a small spectral radius. This implies that random walks mix rapidly, in say t steps, to a

[10] This is more generally true for any CSP that supports a k-wise independent distribution over satisfying assignments.

close-to-uniform distribution. Because a typical pair of vertices at distance t is distributed roughly as a uniformly random pair of vertices, any subset of R vertices which contains a path of length t already allows us to relate global and local graph properties.

To see why this helps, we take for a moment the "pseudoexpectation" view, in which we think of the R-round Sherali–Adams as providing a proxy for the degree-R moments of a distribution over MAX-CUT solutions $x \in \{\pm 1\}^n$, with MAX-CUT value

$$\text{MAX-CUT}(G) = \tfrac{1}{2} - \tfrac{1}{2} \mathop{\mathbf{E}}_{(u,v)\in E(G)} \widetilde{\mathbf{E}}[\mathsf{X}_u \mathsf{X}_v] \tag{1.3}$$

where $\widetilde{\mathbf{E}}[\mathsf{X}_u\mathsf{X}_v]$ is the "pseudo-correlation" of variables $\mathsf{X}_u, \mathsf{X}_v$. Because there is no globally consistent assignment, the pseudo-correlation $\widetilde{\mathbf{E}}[\mathsf{X}_u\mathsf{X}_v]$ for vertices u,v sampled uniformly at random will be close to 0.[11] But in any fixed subgraph of size $\Omega(t)$, enforcing $\widetilde{\mathbf{E}}[\mathsf{X}_u\mathsf{X}_v] \approx 0$ for pairs u,v at distance t has consequences, and limits the magnitude of correlation between pairs of adjacent vertices as well. In particular, for any X_S which is the restriction of X to a set S of up to R vertices, the pseudo-second moment matrix $\widetilde{\mathbf{E}}[\mathsf{X}_S\mathsf{X}_S^\top]$ must be PSD, forcing some entries to 0 gives a constraint on the magnitude of edge correlations.

For example, suppose for a moment that we are in a graph G with $t = 2$, and that S is a star graph in G, given by one "root" vertex r with $k \le R-1$ children $U = \{u_1, \dots, u_k\}$, and define $\mathbb{X} = \widetilde{\mathbf{E}}[\mathsf{X}_S\mathsf{X}_S^\top] \succeq 0$. Notice that pairs of distinct children u_i, u_j are at distance $t = 2$ in S. If we suppose for a moment that $\widetilde{\mathbf{E}}[\mathsf{X}_{u_i}\mathsf{X}_{u_j}] = 0$ for every $u_i \ne u_j$, the only nonzero entries of $\mathbb{X}$ are the diagonals (which are all $\widetilde{\mathbf{E}}[\mathsf{X}_i^2] = 1$), and the entries corresponding to edges from the root to its children, (r, u_i), which are $\widetilde{\mathbf{E}}[\mathsf{X}_r\mathsf{X}_{u_i}]$. Now defining the vector $c \in \mathbb{R}^S$ with a 1 at the root r, $c_r = 1$, and α (to be chosen later) on each child $u \in U$, $c_u = \alpha$, we have from the PSDness of $\mathbb{X}$ that

$$0 \le c^\top \mathbb{X} c = \|c\|_2^2 + \sum_{u\in U} 2c_r c_u \cdot \widetilde{\mathbf{E}}[\mathsf{X}_r X_u] = (1+\alpha^2 k) + 2\alpha k \mathop{\mathbf{E}}_{(u,v)\in E(S)} \widetilde{\mathbf{E}}[\mathsf{X}_u\mathsf{X}_v].$$

Choosing $\alpha = k^{-1/2}$, we have that

$$\mathop{\mathbf{E}}_{(u,v)\in E[S]} \widetilde{\mathbf{E}}[\mathsf{X}_u\mathsf{X}_v] \ge -k^{-1/2}.$$

To deduce this lower bound on the edge correlation, we required that $\widetilde{\mathbf{E}}[\mathsf{X}_{u_i}\mathsf{X}_{u_j}] = 0$ for all $u_i \ne u_j \in S$; this will not necessarily hold for each star S (or even for any star S). But if we take a weighted average over *all* k-stars according to a well-chosen distribution D, this will (approximately) hold on average: we will have

$$\mathop{\mathbf{E}}_{S\sim D} \widetilde{\mathbf{E}}\left[\mathsf{X}_{u_i^{(S)}}\mathsf{X}_{u_j^{(S)}}\right] \approx 0$$

for all $i \ne j$. Our strategy is to take a carefully-chosen average over specific subgraphs S of G with $|S| = \Omega(t)$, according to a distribution D. By our choice of distribution and subgraph, the fact that the subgraphs *locally* have PSD pseudocorrelation matrices has consequences for the *global* average pseudocorrelation across edges, which in turn gives a bound on the objective value eq. (1.3). This allows us to show that Sherali–Adams certifies much better bounds than we previously thought possible, by aggregating local spectral information across many small subgraphs.

[11]This is implicit in our proof, but intuitively it should be true because, e. g., u, v should be connected by equally many even- and odd-length paths.

Organization

We begin with technical preliminaries in Section 2. In Section 3 we prove our main result. Section 4 establishes a mild lower bound for arbitrary graphs. Finally, Section 5 applies Theorem 1.1 to the refutation of arbitrary Boolean CSPs.

2 Setup and preliminaries

We begin by recalling preliminaries and introducing definitions that we will rely upon later.

2.1 Random walks on undirected graphs

Here, we recall some properties of random walks in undirected graphs that will be of use to us.

Definition 2.1. Let $G = (V, E)$ be an undirected finite graph, with parallel edges and self-loops allowed,[12] and with no isolated vertices. The *standard random walk on* G is the Markov chain on V in which at each step one follows a uniformly random edge out of the current vertex. For $u \in V$, we use the notation $\boldsymbol{v} \sim u$ to denote that $\boldsymbol{v}$ is the result of taking one random step from u.

Definition 2.2. We write K for the transition operator of the standard random walk on G. That is, K is obtained from the adjacency matrix of G by normalizing the uth row by a factor of $1/\deg(u)$.

Definition 2.3. We write π for the probability distribution on V defined by $\pi(v) = \frac{\deg(v)}{2|E|}$. As is well known, this is an invariant distribution for the standard random walk on G, and this Markov chain is reversible with respect to π. For $\boldsymbol{u} \sim \pi$ and $\boldsymbol{v} \sim \boldsymbol{u}$, the distribution of $(\boldsymbol{u}, \boldsymbol{v})$ is that of a uniformly random (directed) edge from E. We will also use the notation $\pi_* = \min_{v \in V} \{\pi(v)\}$.

Definition 2.4. For $f, g : V \to \mathbb{R}$ we use the notation $\langle f, g \rangle_\pi$ for $\mathbf{E}_{\boldsymbol{u} \sim \pi}[f(\boldsymbol{u}) g(\boldsymbol{u})]$. This is an inner product on the vector space $\mathbb{R}^V$; in case G is regular and hence π is the uniform distribution, it is the usual inner product scaled by a factor of $1/|V|$. It holds that

$$\langle f, Kg \rangle_\pi = \langle Kf, g \rangle_\pi = \mathop{\mathbf{E}}_{(\boldsymbol{u}, \boldsymbol{v}) \sim E} [f(\boldsymbol{u}) g(\boldsymbol{v})] \,. \tag{2.1}$$

Definition 2.5. A *stationary d-step walk* is defined to be a sequence $(\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ formed by choosing an initial vertex $\boldsymbol{u}_0 \sim \pi$, and then taking a standard random walk, with $\boldsymbol{u}_t \sim \boldsymbol{u}_{t-1}$. Generalizing Equation (2.1), it holds in this case that

$$\mathbf{E}[f(\boldsymbol{u}_0) g(\boldsymbol{u}_d)] = \langle f, K^d g \rangle_\pi \,.$$

[12] Self-loops count as "half an edge," and contribute 1 to a vertex's degree.

2.2 Tree-indexed random walks

To prove our main theorem we define a class of homomorphisms we call *tree-indexed random walks*.

Definition 2.6. Suppose we have a finite undirected tree with vertex set T. A *stationary T-indexed random walk in G* is a random homomorphism $\boldsymbol{\phi}: T \to V$ defined as follows: First, root the tree at an arbitrary vertex $i_0 \in T$. Next, define $\boldsymbol{\phi}(i_0) \sim \pi$. Then, independently for each "child" j of i_0 in the tree, define $\boldsymbol{\phi}(j) \sim \boldsymbol{\phi}(i_0)$; that is, define $\boldsymbol{\phi}(j) \in V$ to be the result of taking a random walk step from $\boldsymbol{\phi}(i_0)$. Recursively repeat this process for all children of i_0's children, etc., until each vertex $k \in T$ has been assigned a vertex $\boldsymbol{\phi}(k) \in V$.

We note that the homomorphism $\boldsymbol{\phi}$ defining the T-indexed random walk need not be injective. Consequently, if T is a tree with maximum degree D, we can still have a T-indexed random walk in a d-regular graph with $d < D$.

The following fact is simple to prove; see, e. g., [35].

Fact 2.7. *The definition of $\boldsymbol{\phi}$ does not depend on the initially selected root $i_0 \in T$. Further, for any two vertices $i, j \in T$ at tree-distance d, if $i = i_0, i_1, \ldots, i_d = j$ is the unique path in the tree between them, then the sequence $(\boldsymbol{\phi}(i_0), \boldsymbol{\phi}(i_1), \ldots, \boldsymbol{\phi}(i_d))$ is distributed as a stationary d-step walk in G.*

2.3 2XOR and signed random walks

The 2-XOR constraint satisfaction problem is defined by instances of linear equations in $\mathbb{F}_2^n$. For us it will be convenient to associate with these instances a graph with signed edges, and on such graphs we perform a slightly modified random walk.

Definition 2.8. We assume that for each pair of vertices (u, v) which are connected by an edge in G, there is an associated *sign* $\xi_{uv} = \xi_{vu} \in \{\pm 1\}$.[13] We arrange these signs into a symmetric matrix $\Xi = (\xi_{uv})_{uv}$. If G has no (u, v) edge then the entry Ξ_{uv} will not matter; we can take it to be 0.

Definition 2.9. We write $\overline{K} = \Xi \circ K$ for the *signed transition operator*.[14] The operator $\overline{K}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_\pi$, and hence has real eigenvalues. It also holds that

$$\langle f, \overline{K} g \rangle_\pi = \langle \overline{K} f, g \rangle_\pi = \mathop{\mathbf{E}}_{(\boldsymbol{u},\boldsymbol{v}) \sim E} [\xi_{\boldsymbol{uv}} f(\boldsymbol{u}) g(\boldsymbol{v})] \,. \tag{2.2}$$

Definition 2.10. We may think of G and Ξ as defining a 2-XOR constraint satisfaction problem (CSP), in which the task is to find a labeling $f: V \to \{\pm 1\}$ so as to maximize the fraction of edges $(u, v) \in E$ for which the constraint $f(u) f(v) = \xi_{uv}$ is satisfied. The fraction of satisfied constraints is

$$\mathop{\mathbf{E}}_{(\boldsymbol{u},\boldsymbol{v}) \sim E} \left[\tfrac{1}{2} + \tfrac{1}{2} \xi_{\boldsymbol{uv}} f(\boldsymbol{u}) f(\boldsymbol{v}) \right] = \tfrac{1}{2} + \tfrac{1}{2} \langle f, \overline{K} f \rangle_\pi \,. \tag{2.3}$$

We will typically ignore the $\frac{1}{2}$'s and think of the 2-XOR CSP as maximizing the quadratic form $\langle f, \overline{K} f \rangle_\pi$. When all signs in the matrix Ξ are -1, we refer to this as the MAX-CUT CSP.

[13] If G has multiple (u, v) edges, we think of them as all having the same sign.

[14] Here $\circ$ denotes the Hadamard product.

Definition 2.11. A *signed* stationary d-step walk is a sequence of pairs $(\boldsymbol{u}_t, \boldsymbol{\sigma}_t) \in V \times \{\pm 1\}$ for $0 \le t \le d$, chosen as follows: first, we choose a stationary d-step walk $(\boldsymbol{u}_0, \boldsymbol{u}_1, \dots, \boldsymbol{u}_d)$ in G; second, we choose $\boldsymbol{\sigma}_0 \in \{\pm 1\}$ uniformly at random; finally, we define $\boldsymbol{\sigma}_t = \boldsymbol{\sigma}_{t-1} \xi_{\boldsymbol{u}_{t-1}\boldsymbol{u}_t}$. Generalizing Equation (2.2), it holds in this case that

$$\mathbf{E}[\boldsymbol{\sigma}_0 f(\boldsymbol{u}_0) \boldsymbol{\sigma}_d g(\boldsymbol{u}_d)] = \langle f, \overline{K}^d g \rangle_\pi \,.$$

Definition 2.12. We extend the notion from Theorem 2.6 to that of a *signed* stationary T-indexed random walk in G. Together with the random homomorphism $\boldsymbol{\phi} : T \to V$, we also choose a random signing $\boldsymbol{\sigma} : T \to \{\pm 1\}$ as follows: for the root i_0, the sign $\boldsymbol{\sigma}(i_0) \in \{\pm 1\}$ is chosen uniformly at random; then, all other signs are deterministically chosen—for each child j of i_0 we set $\boldsymbol{\sigma}(j) = \xi_{i_0 j} \boldsymbol{\sigma}(i_0)$, and in general $\boldsymbol{\sigma}(k) = \xi_{k'k} \boldsymbol{\sigma}(k)$ where k' is the parent of k. Again, it is not hard to show that the definition of $(\boldsymbol{\phi}, \boldsymbol{\sigma})$ does not depend on the choice of root i_0, and that for any path $i_0, i_1, \dots, i_d$ of vertices in the tree, the distribution of $(\boldsymbol{\phi}(i_0), \boldsymbol{\sigma}(i_0)), (\boldsymbol{\phi}(i_1), \boldsymbol{\sigma}(i_1)), \dots (\boldsymbol{\phi}(i_d), \boldsymbol{\sigma}(i_d))$ is that of a signed stationary d-step walk in G.

2.4 Proof systems

Our central object of study is the Sherali–Adams proof system, but our results also apply to a weaker proof system which we introduce below.

Definition 2.13. We define the *R-local, degree-D (static) sum of squares (SOS) proof system* over indeterminates $\mathsf{X}_1, \dots, \mathsf{X}_n$ as follows. The "lines" of the proof are real polynomial inequalities in $\mathsf{X} = (\mathsf{X}_1, \dots, \mathsf{X}_n)$. The "default axioms" are any real inequalities of the form $p(\mathsf{X}_{u_1}, \dots, \mathsf{X}_{u_R})^2 \ge 0$, where p is a polynomial in at most R variables and of degree at most $D/2$. The "deduction rules" allow one to derive any nonnegative linear combination of previous lines. This is a sound proof system for inequalities about n real numbers $\mathsf{X}_1, \dots, \mathsf{X}_n$.

In addition to the default axioms, one may also sometimes include problem-specific "equalities" of the form $q(\mathsf{X}) = 0$. In this case, one is allowed additional axioms of the form $q(\mathsf{X})s(\mathsf{X}) \gtreqless 0$ as long as the polynomial $q(\mathsf{X})s(\mathsf{X})$ depends on at most R indeterminates and has degree at most D.

Fact 2.14. *The case of $R = \infty$ (equivalently, $R = n$) corresponds to the well-known* degree-D SOS proof system.

Definition 2.15. Suppose one includes the *Boolean equalities*, meaning $\mathsf{X}_u^2 - 1 = 0$ for all $1 \le i \le n$.[15] In this case $D = \infty$ is equivalent to $D = R$, and the corresponding proof system is the well-known *R-round Sherali–Adams proof system*. It is well known that every inequality $p(\mathsf{X}_{u_1}, \dots, \mathsf{X}_{u_R}) \ge 0$ that is true for ± 1-valued $\mathsf{X}_{u_1}, \dots, \mathsf{X}_{u_R}$ is derivable in this system.

Fact 2.16. *There is a* $\mathrm{poly}(n^R, L)$*-time algorithm based on Linear Programming for determining whether a given polynomial inequality $p(\mathsf{X}) \ge 0$ of degree at most R (and rational coefficients of total bit-complexity L) is derivable in the R-round Sherali–Adams proof system.*

[15] Or alternatively, $\mathsf{X}_u^2 - \mathsf{X}_u = 0$ for all i.

For example, in our primary case of MAX-CUT on the graph $G = (V,E)$ with variables $\{\mathsf{X}_u\}_{u\in V(G)}$, our axioms will be the Boolean inequalities $\{\mathsf{X}_u^2 - 1 = 0\}_{u\in V(G)}$, and we wish to certify the best upper bound possible on the degree-2 polynomial that defines the objective value, $\frac{1}{2} - \frac{1}{2}\mathbf{E}_{(u,v)\sim E(G)}\mathsf{X}_u\mathsf{X}_v$. By now there are many references which treat the use of the Sherali–Adams proof system in the optimization of constraint satisfaction problems; for example see [12, 13].

Fact 2.17. *We will often be concerned with the R-local, degree-2 SOS proof system, where all lines are quadratic inequalities. In this case, we could equivalently state that the default axioms are all those inequalities of the form*

$$\mathsf{x}^\top P \mathsf{x} \geq 0\,, \tag{2.4}$$

where $\mathsf{x} = (\mathsf{X}_{u_1},\dots,\mathsf{X}_{u_R})$ *is a length-R subvector of* X*, and P is an* $R\times R$ *positive semidefinite (PSD) matrix.*

Remark 2.18. In fact, we will often be concerned with the R-round, degree-2 Sherali–Adams proof system. Despite the restriction to $D = 2$, we only know the $\mathrm{poly}(n^R, L)$-time algorithm for deciding derivability of a given quadratic polynomial $p(\mathsf{X}) \geq 0$ (of bit-complexity L).

2.5 Linear Algebra

We will use the following basic linear algebraic notions. We will denote the identity matrix by $\mathbb{1}$. For an $n\times m$ real matrix M, we use $\|M\|$ or $\|M\|_{op}$ to denote M's operator norm, which is equal to the maximum singular value. When M is square and $m = n$, we use $\rho(M)$ to denote the *spectral radius* of M, $\rho(M) = \max\{|\lambda| \mid \exists v\in\mathbb{R}^n \text{ s.t. } Mv = \lambda v\}$. We will also use the following facts:

Fact 2.19. *If* $M\in\mathbb{R}^{n\times n}$*, then for all* $i\in[n]$*,* $\sum_{j\in[n]} M_{ji}^2 \leq \|M\|^2$*, and* $\sum_{j\in[n]} M_{ij}^2 \leq \|M\|^2$.

Proof. By definition of the operator norm,

$$\|M\| = \max_{\|v\|=1}\|Mv\| \geq \|Me_i\| = \sqrt{\sum_{j\in[n]} M_{ji}^2}\,.$$

where e_i is the ith standard basis vector. Repeating the same argument using the fact that $\|M\| = \|M^\top\|$ proves the second inequality. □

Fact 2.20. *Let* $e\in\mathbb{R}^n$ *be the all-1's vector. If* $K\in\mathbb{R}^{n\times n}$ *is a transition matrix with stationary distribution* $\pi\in\mathbb{R}^n$*, then* $K' = K - e\pi^\top$ *is the projection of K orthogonal to the trivial eigenspace with left eigenvector* π *and right eigenvector e.*

Proof. This is a consequence of the spectral theorem: viewing K as a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^n$, notice that K is self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_\pi$. Thus we can write $Kv = \sum_{i=1}^n \lambda_i\phi_i\langle\phi_i, v\rangle_\pi$, for orthonormal eigenfunctions $\phi_1,\dots,\phi_n$ with corresponding real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Since K is a transition operator, $\lambda_i \leq 1$ for all i. Because e is a right-eigenvector of K with eigenvalue 1, we can take $\phi_1 = e$, $\lambda_1 = 1$, and then $(K - e\pi^\top)v = \sum_{i=2}^n \lambda_i\phi_i\langle\phi_i, v\rangle_\pi$. From this we see that $K - e\pi^\top$ has the same spectral decomposition as K, except λ_1 is replaced by 0. □

3 2XOR certifications from spider walks

In this section, we prove our main theorem: given a 2-XOR or MAX-CUT instance on a graph G with small spectral radius, we will show that the R-local degree-2 SOS proof system gives nontrivial refutations with R not too large. We also prove a similar result for maximum independent set.

Our strategy is as follows: we select a specific tree T of size $\propto R$, and we consider the distribution over copies of T in our graph given by the T-indexed stationary random walk. We will use this distribution to define the coefficients for a degree-2, R-local proof that bounds the objective value of the CSP. We will do this by exploiting the uniformity of the graph guaranteed by the small spectral radius, and the fact that degree-2 R-local SOS proofs can certify positivity of quadratic forms $c^\top \mathsf{X}|_S \mathsf{X}|_S^\top c$, where $X|_S$ is the restriction of X to a set S of variables with $|S| \le R$ and $c \in \mathbb{R}^{|S|}$.

Intuitively, in the "pseudoexpectation" view, the idea of our proof is as follows. When there is no globally consistent assignment, a uniformly random pair of vertices $u, v \in V$ will have pseudocorrelation close to zero. On the other hand, if t-step random walks mix to a roughly uniform distribution over vertices in the graph, then pairs of vertices at distance t will also have pseudocorrelation close to zero. But also, in our proof system the degree-2 pseudomoments of up to R variables obey a positive-semidefiniteness constraint. By choosing the tree T with diameter at least t, while also choosing T to propagate the effect of the low-pseudocorrelation at the diameter to give low-pseudocorrelation on signed edges, we show that the proof system can certify that the objective value is small. Specifically, we will choose T to be a *spider graph*:

Definition 3.1. For integers $k, \ell \in \mathbb{N}^+$, we define a (k,ℓ)*-spider graph* to be the tree formed by gluing together k paths of length ℓ at a common endpoint called the *root*. This spider has $k\ell + 1$ vertices and diameter 2ℓ.

While we were not able to formally prove that the spider is the optimal choice of tree, intuitively, we want to choose a tree that maximizes the ratio of the number of pairs at maximum distance (since such pairs relate the local properties to the global structure) to the number of vertices in the tree (because we need to take our number of rounds R to be at least the size of the tree). Among trees, the spider is the graph that maximizes this ratio.

Let us henceforth fix a (k,ℓ)-spider graph, where the parameters k and ℓ will be chosen later. We write S for the vertex set of this tree (and sometimes identify S with the tree itself).

Definition 3.2. For $0 \le d \le 2\ell$, we define the matrix $A^{(d)} \in \mathbb{R}^{S \times S}$ to be the "distance-d" adjacency matrix of the spider; i. e., $A^{(d)}_{ij}$ is 1 if $\mathrm{dist}_S(i,j) = d$ and is 0 otherwise. (We remark that $A^{(0)}$ is the identity matrix.)

The following key technical theorem establishes the existence of a matrix Ψ which will allow us to define the coefficients in our R-local degree-2 SOS proof. It will be proven in Section 3.3:

Theorem 3.3. *For any parameter $\alpha \in \mathbb{R}$, there is a PSD matrix $\Psi = \Psi_\alpha \in \mathbb{R}^{S \times S}$ with the following* *properties:*

$$\begin{aligned}
\langle \Psi, A^{(0)} \rangle &= 1 + \frac{1}{2k}\alpha^{2\ell} + \frac{1}{k-1}\frac{\alpha^{2\ell}-\alpha^2}{\alpha^2-1}\,, \\
\langle \Psi, A^{(1)} \rangle &= \alpha, \\
\langle \Psi, A^{(d)} \rangle &= 0 \quad \textit{for } 1 < d < 2\ell\,, \\
\langle \Psi, A^{(2\ell)} \rangle &= \frac{1-1/k}{2}\alpha^{2\ell}\,.
\end{aligned}$$

Here we are using the notation $\langle B, C \rangle$ for the "matrix (Frobenius) inner product" $Tr(B^\top C)$.

The crucial property of Ψ is that, for a well-chosen value of α, it has value $\Theta(1)$ on $A^{(0)}$, value 0 on $A^{(d)}$ for $1 < d < 2\ell$, and the values on $A^{(1)}$ and $A^{(2\ell)}$ are related by a 2ℓth power:

Corollary 3.4. *Assuming that $k \ge 3^\ell$ and taking $\alpha = k^{1/2\ell}$, the PSD matrix Ψ satisfies*

$$3/2 \le \langle \Psi, A^{(0)} \rangle \le 2, \quad \langle \Psi, A^{(1)} \rangle = k^{1/2\ell}, \quad \langle \Psi, A^{(d)} \rangle = 0 \,\textit{for}\; 1 < d < 2\ell, \quad \langle \Psi, A^{(2\ell)} \rangle = \tfrac{1}{2}(k-1)\,.$$

We will also use the following small technical lemma, which will let us relate the correlation across vertex pairs when weighted by some matrix M to M's spectral norm; for example, when M is a power of the transition matrix of our instance.

Lemma 3.5. *Let $M \in \mathbb{R}^{V \times V}$ and recall $\pi_* = \min_{v \in V}\{\pi(v)\} > 0$. Then the 2-local, degree-2 SOS proof system can derive*

$$\mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi} \sum_{v \in V} M_{\boldsymbol{u}v} \mathsf{X}_{\boldsymbol{u}} \mathsf{X}_v \le \pi_*^{-1/2} \|M\| \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi} \mathsf{X}_{\boldsymbol{u}}^2\,.$$

Proof. The proof system can derive the following inequality for any $\gamma > 0$, since the difference of the two sides is a perfect square:

$$M_{uv}\mathsf{X}_u\mathsf{X}_v \le \frac{M_{uv}^2}{2\gamma\pi(v)}\mathsf{X}_u^2 + \frac{\gamma\pi(v)}{2}\mathsf{X}_v^2\,.$$

Thus it can derive

$$\mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi} \sum_{v \in V} M_{\boldsymbol{u}v}\mathsf{X}_{\boldsymbol{u}}\mathsf{X}_v \le \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi} \mathsf{X}_{\boldsymbol{u}}^2 \sum_{v\in V} \frac{M_{\boldsymbol{u}v}^2}{2\gamma\pi(v)} + \frac{\gamma}{2}\mathop{\mathbf{E}}_{\boldsymbol{v}\sim\pi}\mathsf{X}_{\boldsymbol{v}}^2\,. \tag{3.1}$$

We'll take $\gamma = \pi_*^{-1/2}\|M\|$. Since we can certainly derive $a\mathsf{X}_u^2 \le b\mathsf{X}_u^2$ whenever $a \le b$, we see that it suffices to establish

$$\sum_{v\in V} \frac{M_{\boldsymbol{u}v}^2}{2\gamma\pi(v)} \le \frac{\gamma}{2}$$

for every outcome of $\boldsymbol{u}$. But this is implied by $\sum_v M_{\boldsymbol{u}v}^2 \le (\pi(v)/\pi_*)\|M\|^2$ for all $v \in V$, which is indeed true. $\square$

We can now prove the following main theorem:

Theorem 3.6. *Suppose we have a graph $G = (V,E)$ with signed transition operator $\overline{K}$ and $\pi_* = \min_{v\in V} \frac{\deg(v)}{2|E|}$. Given parameters $k \geq 3^\ell$, let $R = k\ell + 1$ and define*

$$\beta = \frac{k\pi_*^{-1/2}}{2k^{1/2\ell}}\rho(\overline{K})^{2\ell} + \frac{2}{k^{1/2\ell}}\,.$$

Then R-local, degree-2 SOS can deduce the bound "$\rho(\overline{K}) \leq \beta$"; more precisely, it can deduce the two inequalities

$$-\beta\langle \mathsf{X},\mathsf{X}\rangle_\pi \leq \langle \mathsf{X},\overline{K}\mathsf{X}\rangle_\pi \leq \beta\langle \mathsf{X},\mathsf{X}\rangle_\pi\,.$$

Before proving this theorem, let us simplify the parameters. For any $\varepsilon \in (0,\varepsilon_0]$ for ε_0 a universal constant, we can choose ℓ to be the smallest integer so that $(\frac{1}{\varepsilon}\rho(\overline{K}))^{2\ell}\pi_*^{-1/2} \leq \varepsilon$, and $k = \lceil(\frac{1}{\varepsilon})^{2\ell}\rceil$. This gives the corollary:

Corollary 3.7. *Suppose we have a graph $G = (V,E)$ with signed transition operator $\overline{K}$ and $\pi_* = \min_{v\in V} \frac{\deg(v)}{2|E|}$. Given $\varepsilon \in (\rho(\overline{K}),\varepsilon_0]$ for ε_0 a universal constant, take*

$$\ell = \left\lceil \frac{1}{4}\frac{\log(\varepsilon^2\pi_*)}{\log(\rho(\overline{K})/\varepsilon)}\right\rceil \qquad \textit{and} \qquad k = \left\lceil \left(\frac{1}{\varepsilon}\right)^{2\ell}\right\rceil.$$

Then for $R = k\ell+1$, it holds that R-local degree-2 SOS can deduce the bound $\rho(\overline{K}) \leq \frac{5}{2}\varepsilon$. In particular, if we think of G,Ξ as a 2-XOR CSP, it holds that R-round Sherali–Adams can deduce the bound OBJ $\leq \frac{1}{2} + \frac{5}{4}\varepsilon$.

Proof. We take the parameters as outlined above. The constraints $\mathsf{X}_u^2 = 1$ imply that R-round Sherali–Adams can deduce that $\langle \mathsf{X},\mathsf{X}\rangle_\pi = 1$ whenever $R \geq 2$. Further, as noted in eq. (2.3), $\mathrm{OBJ}(\mathsf{X}) = \frac{1}{2} + \frac{1}{2}\langle \mathsf{X},\overline{K}\mathsf{X}\rangle_\pi$. Combining these observations with Theorem 3.6 gives the result. □

Theorem 3.7 implies the 2-XOR version of Theorem 1.1 since in simple graphs, $\log\frac{1}{\pi^*} = \Theta(\log n)$.

Proof of Theorem 3.6. For our (k,ℓ)-spider graph on S, let $(\boldsymbol{\phi},\boldsymbol{\sigma})$ be a signed stationary S-indexed random walk in G. Define $\overline{\mathsf{x}}$ to be the S-indexed vector with $\overline{\mathsf{x}}_i = \boldsymbol{\sigma}(i)\mathsf{X}_{\boldsymbol{\phi}(i)}$. Then letting Ψ be the PSD matrix from Theorem 3.4, the R-local, degree-2 SOS proof system can derive

$$\langle \Psi, \overline{\mathsf{x}}\overline{\mathsf{x}}^\top\rangle = \overline{\mathsf{x}}^\top\Psi\overline{\mathsf{x}} \geq 0\,.$$

(This is in the form of Inequality (2.4) if we take $P = \mathrm{diag}(\boldsymbol{\sigma})\Psi\,\mathrm{diag}(\boldsymbol{\sigma})$.) Furthermore, the proof system can deduce this inequality in expectation; namely,

$$\langle \Psi,\mathsf{Y}\rangle \geq 0, \text{ where } \mathsf{Y} = \mathbf{E}[\overline{\mathsf{x}}\overline{\mathsf{x}}^\top]\,. \tag{3.2}$$

Now by the discussion in Theorems 2.11 and 2.12,

$$\mathsf{Y}_{ij} = \mathbf{E}[\boldsymbol{\sigma}(i)\mathsf{X}_{\boldsymbol{\phi}(i)}\boldsymbol{\sigma}(j)\mathsf{X}_{\boldsymbol{\phi}(j)}] = \langle \mathsf{X},\overline{K}^{\mathrm{dist}_S(i,j)}\mathsf{X}\rangle_\pi\,. \tag{3.3}$$

Thus recalling the notation $A^{(d)}$ from Theorem 3.2,

$$\mathsf{Y} = \sum_{d=0}^{2\ell} \langle \mathsf{X}, \overline{K}^d \mathsf{X} \rangle_\pi A^{(d)}, \tag{3.4}$$

and hence from Inequality (3.2) we get that R-local, degree-2 SOS can deduce

$$0 \le \sum_{d=0}^{2\ell} \langle \Psi, A^{(d)} \rangle \langle \mathsf{X}, \overline{K}^d \mathsf{X} \rangle_\pi = c_0 \langle \mathsf{X}, \mathsf{X} \rangle_\pi + k^{1/2\ell} \langle \mathsf{X}, \overline{K} \mathsf{X} \rangle_\pi + \tfrac{1}{2}(k-1) \langle \mathsf{X}, \overline{K}^{2\ell} \mathsf{X} \rangle_\pi, \tag{3.5}$$

for some constant $3/2 \le c_0 \le 2$ (here we used Theorem 3.4). Regarding the last term, we have:

$$\langle \mathsf{X}, \overline{K}^{2\ell} \mathsf{X} \rangle_\pi = \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi} \sum_{v \in V} (\overline{K}^{2\ell})_{\boldsymbol{u} v} \mathsf{X}_{\boldsymbol{u}} \mathsf{X}_v. \tag{3.6}$$

If we cared only about the Sherali–Adams proof system with Boolean equalities, we would simply now deduce

$$\mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi} \sum_{v \in V} (\overline{K}^{2\ell})_{\boldsymbol{u} v} \mathsf{X}_{\boldsymbol{u}} \mathsf{X}_v \le \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi} \sum_{v \in V} \left| (\overline{K}^{2\ell})_{\boldsymbol{u} v} \right| \le \sqrt{|V|} \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi} \| \overline{K}^{2\ell}_{\boldsymbol{u},\cdot} \| \le \sqrt{|V|} \rho(\overline{K}^{2\ell}) = \sqrt{|V|} \rho(\overline{K})^{2\ell},$$

and later combine this with $c_0 \langle \mathsf{X}, \mathsf{X} \rangle_\pi = c_0$. But proceeding more generally, we instead use Theorem 3.5 to show that our proof system can derive

$$\mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi} \sum_{v \in V} (\overline{K}^{2\ell})_{\boldsymbol{u} v} \mathsf{X}_{\boldsymbol{u}} \mathsf{X}_v \le \pi_*^{-1/2} \rho(\overline{K})^{2\ell} \langle \mathsf{X}, \mathsf{X} \rangle_\pi.$$

Putting this into Equation (3.6) and Inequality (3.5) we get

$$\langle \mathsf{X}, \overline{K} \mathsf{X} \rangle_\pi \ge -\frac{c_0 + \frac{1}{2}(k-1) \pi_*^{-1/2} \rho(\overline{K})^{2\ell}}{k^{1/2\ell}} \langle \mathsf{X}, \mathsf{X} \rangle_\pi \ge -\beta \langle \mathsf{X}, \mathsf{X} \rangle_\pi.$$

Repeating the derivation with $-\overline{K}$ in place of $\overline{K}$ completes the proof. □

3.1 Max-Cut

The following theorem is quite similar to Theorem 3.6. In it, we allow K to have the large eigenvalue 1, and only certify that it has no large-magnitude negative eigenvalue. The subsequent corollary is deduced identically to Theorem 3.7.

Theorem 3.8. *Given transition operator K for the standard random walk on G, let $K' = K - e\pi^\top$, where e is the all-1's vector and π is the stationary measure. For parameters $k \ge 3^\ell$, let $R = k\ell + 1$ and define*

$$\beta = \frac{k \pi_*^{-1/2}}{2k^{1/2\ell}} \rho(K')^{2\ell} + \frac{2}{k^{1/2\ell}}.$$

Then $2R$-local, degree-2 SOS can deduce the bound "$\lambda_{\min}(K) \ge -\beta$"; more precisely, it can deduce the inequality

$$\langle \mathsf{X}, K' \mathsf{X} \rangle_\pi \ge -\beta \langle \mathsf{X}, \mathsf{X} \rangle_\pi.$$

Remark 3.9. Notice that $\rho(K')$ is equal to the absolute value of the maximum-magnitude eigenvalue of K when the trivial 1 eigenvalue is excluded (see Theorem 2.20).

Corollary 3.10. *Suppose we have a graph $G=(V,E)$ with transition operator K and centered transition operator $K' = K - e\pi^\top$ with e the all-1's vector, and $\pi_* = \min_{v\in V} \frac{\deg(v)}{2|E|}$. Given $\varepsilon > \rho(K')$, take*

$$\ell = \left\lceil \frac{1}{4} \frac{\log(\varepsilon^2 \pi_*)}{\log(\rho(K')/\varepsilon)} \right\rceil \qquad \text{and} \qquad k = \left\lceil \left(\frac{1}{\varepsilon}\right)^{2\ell} \right\rceil .$$

Then for $R = k\ell + 1$, it holds that $2R$-local degree-2 SOS can deduce the bound $\rho(K') \le \frac{5}{2}\varepsilon$. In particular, if we think of G as a MAX-CUT *CSP, it holds that R-round Sherali–Adams can deduce the bound* $\mathrm{OBJ} \le \frac{1}{2} + \frac{5}{4}\varepsilon$.

Again, Theorem 3.10 implies Theorem 1.1 since in simple graphs, $\log \frac{1}{\pi^*} = \Theta(\log n)$.

Proof of Theorem 3.8. The proof is a modification of the proof of Theorem 3.6. Letting S be the (k,ℓ)-spider vertices, instead of taking a signed stationary S-indexed random walk in G, we take two independent *unsigned* stationary S-indexed random walks, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. For $j \in \{1,2\}$, define x_j to be the S-indexed vector with ith coordinate equal to $\mathsf{X}_{\boldsymbol{\phi}_j(i)}$, and write $\dot{\mathsf{x}}$ for the concatenated vector $(\mathsf{x}_1, \mathsf{x}_2)$. Also, for $0 < \theta < 1$ a parameter[16] slightly less than 1, let Ψ be the PSD matrix from Theorem 3.4, and define the PSD block-matrix

$$\dot{\Psi} = \frac{1}{2} \left(\begin{array}{c|c} \frac{1}{\theta}\Psi & -\Psi \\ \hline -\Psi & \theta\Psi \end{array} \right) .$$

Then as before, the $2R$-local, degree-2 SOS proof system can derive

$$0 \le \langle \dot{\Psi}, \mathbf{E}\, \dot{\mathsf{x}}\dot{\mathsf{x}}^\top \rangle = \iota \langle \Psi, \mathsf{Y} \rangle - \langle \Psi, \mathsf{Z} \rangle, \quad \text{where} \quad \iota = \tfrac{1/\theta + \theta}{2}, \quad \mathsf{Y} = \mathbf{E}[\mathsf{x}\mathsf{x}^\top], \quad \mathsf{Z} = \mathbf{E}[\mathsf{x}_1 \mathsf{x}_2^\top], \tag{3.7}$$

and x (which will play the role of $\overline{\mathsf{x}}$) denotes the common distribution of x_1 and x_2. Similar to Equations (3.3) and (3.4), we now have

$$\mathsf{Y} = \sum_{d=0}^{2\ell} \langle \mathsf{X}, K^d \mathsf{X} \rangle_\pi A^{(d)} ,$$

and by independence of x_1 and x_2 we have

$$\mathsf{Z} = \langle 1, \mathsf{X} \rangle_\pi^2 \cdot J = \langle 1, \mathsf{X} \rangle_\pi^2 \cdot \sum_{d=0}^{2\ell} A^{(d)} .$$

Thus applying Theorem 3.4 to Inequality (3.7), our proof system can derive

$$0 \le \iota \cdot \left(c_0 \langle \mathsf{X}, \mathsf{X} \rangle_\pi + k^{1/2\ell} \langle \mathsf{X}, K\mathsf{X} \rangle_\pi + \tfrac{1}{2}(k-1) \langle \mathsf{X}, K^{2\ell} \mathsf{X} \rangle_\pi \right) - \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[\mathsf{X}_{\boldsymbol{u}}]^2 \cdot \left(c_0 + k^{1/2\ell} + \tfrac{1}{2}(k-1) \right) . \tag{3.8}$$

[16] This parameter is introduced to fix a small annoyance; the reader might like to imagine $\theta = 1$ at first.

By selecting θ appropriately, we can arrange for the factor $c_0 + k^{1/2\ell} + \frac{1}{2}(k-1)$ on the right to equal $\iota \cdot \frac{1}{2}(k-1)$. Inserting this choice into Inequality (3.8) and then dividing through by ι, we conclude that the proof system can derive

$$0 \le c_0 \langle \mathsf{X}, \mathsf{X} \rangle_\pi + k^{1/2\ell} \langle \mathsf{X}, K\mathsf{X} \rangle_\pi + \tfrac{1}{2}(k-1)\Big(\langle \mathsf{X}, K^{2\ell}\mathsf{X} \rangle_\pi - \langle 1, X \rangle_\pi^2 \Big),$$

cf. Inequality (3.5). Recalling now that K has the constantly-1 function as a right eigenvector with eigenvalue 1, since K is self-adjoint under $\langle \cdot, \cdot \rangle_\pi$ (see Theorem 2.20) we have the identity

$$\langle \mathsf{X}, K^{2\ell}\mathsf{X} \rangle_\pi - \langle 1, X \rangle_\pi^2 = \langle \mathsf{X}, K^{2\ell}\mathsf{X} \rangle_\pi - \langle \mathsf{X}, e\pi^\top \mathsf{X} \rangle_\pi = \langle \mathsf{X}, (K - e\pi^\top)^{2\ell} \mathsf{X} \rangle_\pi .$$

Now the remainder of the proof is just as in Theorem 3.6, with $K - e\pi^\top$ in place of $\overline{K}$, except we do not have the step of repeating the derivation with $-\overline{K}$ in place of $\overline{K}$. □

3.2 Independent Set

Given a graph $G = (V, E)$, consider the Sherali–Adams LP relaxation for maximum independent set using formal variables Z_u for each vertex $u \in V$ (which represent the 0/1 indicators for membership in the independent set):[17]

$$\text{INDEPENDENT-SET}(G) \;\le\; \max \frac{1}{|V|} \sum_{u \in V} \mathsf{Z}_u \qquad \text{s.t.} \qquad \begin{array}{ll} \mathsf{Z}_u^2 = \mathsf{Z}_u & \text{(for all } u \in V) \\ \mathsf{Z}_u \mathsf{Z}_v = 0 & \text{(for all } (u,v) \in E) \end{array}$$

We show that in regular and near-regular graphs with bounded spectral radius, The Sherali–Adams LP can certify nontrivial bounds on the value of the maximum independent set.

Theorem 3.11. *Given transition operator K for the standard random walk on G, let $K' = K - e\pi^\top$, where e is the all-1's vector. For parameters $k \ge 3^\ell$, let $R = k\ell + 1$ and define*

$$\beta = \frac{k\pi_*^{-1/2}}{2k^{1/2\ell}} \rho(K')^{2\ell} + \frac{2}{k^{1/2\ell}} .$$

Then $2R$-local, degree-2 SOS can deduce the bound "$\mathbf{E}_{\boldsymbol{u} \sim \pi}[\mathsf{Z}_{\boldsymbol{u}}] \le \beta$."

Again, in the following corollary we select parameters for convenience:

Corollary 3.12. *Suppose we have a* regular *graph $G = (V, E)$ with transition operator K and centered transition operator $K' = K - e\pi^\top$ for e the all-1's vector. Given $\varepsilon > \rho(K')$, take*

$$\ell = \left\lceil \frac{1}{4} \frac{\log(\varepsilon^2 n)}{\log(\rho(K')/\varepsilon)} \right\rceil \qquad \text{and} \qquad k = \left\lceil \left(\frac{1}{\varepsilon}\right)^{2\ell} \right\rceil .$$

Then for $R = k\ell + 1$, it holds that $2R$-local degree-2 SOS can deduce the bound INDEPENDENT-SET$(G) \le \frac{5}{2}\varepsilon$.

[17] We have changed to work with Z variables to emphasize that they are $0/1$ valued rather than ± 1-valued.

Proof of Theorem 3.11. The proof is a slight modification of the proof of Theorem 3.8. Letting S be the (k,ℓ)-spider vertices, we again take two independent *unsigned* stationary S-indexed random walks, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. For $j \in \{1,2\}$, define z_j to be the S-indexed vector with ith coordinate equal to $\mathsf{Z}_{\boldsymbol{\phi}_j(i)}$, and write $\dot{\mathsf{z}}$ for the concatenated vector $(\mathsf{z}_1,\mathsf{z}_2)$. Also, define the PSD block matrix

$$\dot{\Psi} = \frac{1}{2}\left(\begin{array}{c|c} \Psi & -\Psi \\ \hline -\Psi & \Psi \end{array}\right).$$

Then as before, the $2R$-local, degree-2 SOS proof system can derive

$$0 \le \langle \dot{\Psi}, \mathbf{E}\,\dot{\mathsf{z}}\dot{\mathsf{z}}^\top \rangle = \langle \Psi, \mathsf{A}\rangle - \langle \Psi, \mathsf{B}\rangle, \quad \text{where} \quad \mathsf{A} = \mathbf{E}[\mathsf{z}\mathsf{z}^\top], \quad \mathsf{B} = \mathbf{E}[\mathsf{z}_1\mathsf{z}_2^\top], \tag{3.9}$$

where z denotes the common distribution of z_1 and z_2. Similar to Equations (3.3) and (3.4), we have

$$\mathsf{A} = \sum_{d=0}^{2\ell} \langle \mathsf{Z}, K^d \mathsf{Z}\rangle_\pi A^{(d)},$$

with the additional twist that our proof system certifies $\langle \mathsf{Z}, K\mathsf{Z}\rangle_\pi = 0$, since we have the constraints that $\mathsf{Z}_u\mathsf{Z}_v = 0$ on edges. By the independence of z_1 and z_2,

$$\mathsf{B} = \langle 1, \mathsf{Z}\rangle_\pi^2 \cdot J = \langle 1, \mathsf{Z}\rangle_\pi^2 \cdot \sum_{d=0}^{2\ell} A^{(d)}.$$

Thus applying Theorem 3.4 to Inequality (3.9), our proof system can derive

$$0 \le \left(c_0 \langle \mathsf{Z}, \mathsf{Z}\rangle_\pi + \tfrac{1}{2}(k-1)\langle \mathsf{Z}, K^{2\ell}\mathsf{Z}\rangle_\pi\right) - \langle 1, \mathsf{Z}\rangle_\pi^2 \left(c_0 + k^{1/2\ell} + \tfrac{1}{2}(k-1)\right). \tag{3.10}$$

Recalling now that K has the constantly-1 function e as a right eigenvector with eigenvalue 1, since K is self-adjoint under $\langle \cdot, \cdot \rangle_\pi$ (see Theorem 2.20) we have the identity

$$\langle \mathsf{Z}, K^{2\ell}\mathsf{Z}\rangle_\pi - \langle 1, \mathsf{Z}\rangle_\pi^2 = \langle \mathsf{Z}, K^{2\ell}\mathsf{Z}\rangle_\pi - \langle \mathsf{Z}, e\pi^\top \mathsf{Z}\rangle_\pi = \langle \mathsf{Z}, (K - e\pi^\top)^{2\ell}\mathsf{Z}\rangle_\pi.$$

Now the remainder of the proof is similar to that of Theorem 3.8. We observe that $\langle \mathsf{Z}, \mathsf{Z}\rangle_\pi = \mathbf{E}_{\boldsymbol{u}\sim\pi}\,\mathsf{Z}_{\boldsymbol{u}}$ from the constraint $\mathsf{Z}_u = \mathsf{Z}_u^2$, and $\langle 1, \mathsf{Z}\rangle_\pi^2 = \mathbf{E}_{\boldsymbol{u}\sim\pi}[\mathsf{Z}_{\boldsymbol{u}}]^2$. We apply these observations along with Theorem 3.5 and to Inequality (3.10) to get

$$0 \le c_0 \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[\mathsf{Z}_{\boldsymbol{u}}] + \tfrac{1}{2}(k-1)\pi_*^{-1/2}\rho(K-J)^{2\ell} \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[\mathsf{Z}_{\boldsymbol{u}}] - (c_0 + k^{1/2\ell}) \cdot \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[\mathsf{Z}_{\boldsymbol{u}}]^2$$

and then algebraic manipulations along with the bound $0 \le c_0 \le 2$ give us the result. □

3.3 A technical construction of coefficients on the spider

Here, we prove Theorem 3.3. The construction of Ψ is technical, and we cannot offer strong intuition for it. Roughly speaking, we build Ψ as a sum of $\ell+1$ quadratic forms of vectors, which ensures that $\Psi \succeq 0$. The first vector adds mass to Ψ's correlation with $A^{(0)}, A^{(1)}$ and $A^{(2)}$; each successive tth vector for $t \in \{0,\dots,\ell-1\}$ adds mass to the $A^{(2t)}$ and $A^{(2t+4)}$ correlations while subtracting mass from the $A^{(2t+2)}$ correlation. By tuning everything carefully, and finally making a small rank-1 adjustment, we ensure that the only nonzero correlations are on $A^{(d)}$ for $d = 0, 1, 2\ell$.

Proof of Theorem 3.3. We are considering the (k,ℓ)-spider graph on vertex set S. We write V_t for the set of all vertices at distance t from the root (so $|V_0| = 1$ and $|V_t| = k$ for $1 \le t \le \ell$). We will be considering vectors in $\mathbb{R}^S$, with coordinates indexed by the vertex set S. For $0 \le t \le \ell$ define the vector

$$\mu_t = \operatorname*{avg}_{i \in V_t}\{\alpha^t e_i\},$$

where $e_i = (0,\dots,0,1,0,\dots,0)$ is the vector with the 1 in the ith position. Further define vectors

$$\begin{aligned}\chi &= \mu_0 + \mu_1\,,\\ \psi_t &= \mu_t - \mu_{t+2} \text{ for } 0 \le t < \ell\,,\end{aligned}$$

with the understanding that $\mu_{\ell+1} = 0$. Next, define the PSD matrix

$$\widetilde{\Psi} = \chi\chi^\top + \sum_{t=0}^{\ell-1} \psi_t \psi_t^\top\,.$$

This will almost be our desired final matrix Ψ. Let us now compute

$$\langle \widetilde{\Psi}, A^{(d)}\rangle = \chi^\top A^{(d)}\chi + \sum_{t=0}^{\ell-1} \psi_t^\top A^{(d)} \psi_t\,.$$

To do this, we observe that

$$\mu_s^\top A^{(d)} \mu_t = \alpha^{s+t} \Pr_{\boldsymbol{i}\sim V_s,\ \boldsymbol{j}\sim V_t}[\mathrm{dist}_S(\boldsymbol{i},\boldsymbol{j}) = d]\,,$$

and

$$\mu_0^\top A^{(d)}\mu_t = \mu_t^\top A^{(d)}\mu_0 = \begin{cases}\alpha^t & \text{if } d = t\,,\\ 0 & \text{else;}\end{cases}$$

$$\text{and for } s,t > 0, \quad \mu_s^\top A^{(d)}\mu_t = \begin{cases}(1/k)\alpha^{s+t} & \text{if } d = |s-t|\,,\\ (1-1/k)\alpha^{s+t} & \text{if } d = s+t\,,\\ 0 & \text{else}\,.\end{cases}$$

From this we can compute the following (with a bit of effort):

$$\begin{aligned}\langle \widetilde{\Psi}, A^{(0)}\rangle &= 2 + (2/k)\alpha^2 + (2/k)\alpha^4 + \cdots + (2/k)\alpha^{2\ell-2} + (1/k)\alpha^{2\ell}\\ \langle \widetilde{\Psi}, A^{(1)}\rangle &= 2\alpha\\ \langle \widetilde{\Psi}, A^{(2)}\rangle &= -(2/k)\alpha^2 - (2/k)\alpha^4 - (2/k)\alpha^6 - \cdots - (2/k)\alpha^{2\ell-2}\\ \langle \widetilde{\Psi}, A^{(2t+1)}\rangle &= 0, \quad 1 \le t < \ell\\ \langle \widetilde{\Psi}, A^{(2t)}\rangle &= 0, \quad 1 < t < \ell\\ \langle \widetilde{\Psi}, A^{(2\ell)}\rangle &= (1-1/k)\alpha^{2\ell}\end{aligned}$$

Now, for a parameter $\eta > 0$ to be chosen shortly, we finally define the PSD matrix

$$\Psi = \frac{1}{2}\widetilde{\Psi} + \eta\mu_1\mu_1^\top .$$

We have

$$\langle \eta\mu_1\mu_1^\top, A^{(d)} \rangle = \begin{cases} \eta(1/k)\alpha^2 & \text{if } d = 0, \\ \eta(1-1/k)\alpha^2 & \text{if } d = 2. \end{cases}$$

Therefore by carefully choosing

$$\eta = \frac{1}{k-1}\left(\frac{\alpha^{2\ell-2}-1}{\alpha^2-1}\right),$$

we get all of the desired inner products in the theorem statement. □

4 Lower bounds

In this section, we show that degree-R Sherali–Adams cannot refute a random 2-XOR or MAX-CUT instance better than $\frac{1}{2}+\Omega(\frac{1}{R})$. This is a straightforward application of the framework of Charikar, Makarychev and Makarychev [13]. In that article, the authors show that if every subset of r points in a metric space can be locally embedded into the unit sphere, then Goemans–Williamson rounding can be used to give a $\Theta(r)$-round Sherali–Adams feasible point. The upshot is the following theorem appearing in [13] (where it is stated in slightly more generality, for the $0/1$ version of the cut polytope):

Theorem 4.1 (Theorem 3.1 in [13])**.** *Let (X,ρ) be a metric space, and assume that every $r = 2R+3$ points of (X,ρ) isometrically embed in the Euclidean sphere of radius 1. Then the following point is feasible for R-rounds of the Sherali–Adams relaxation for the cut polytope:*[18]

$$\widetilde{\mathbf{E}}[\mathsf{X}_i\mathsf{X}_j] = 1 - \frac{2}{\pi}\arccos\left(1-\frac{1}{2}\rho(i,j)^2\right).$$

Proposition 4.2. *In any 2-XOR or MAX-CUT instance, R-rounds of Sherali–Adams cannot certify that*

$$\mathrm{OBJ}(x) < \frac{1}{2} + \frac{1}{2\pi R} - \frac{1}{3R^2}.$$

Proof. Suppose that we are given a 2-XOR (equivalently, MAX-CUT) instance on the graph G, so that on each edge $(i,j) \in E(G)$ we have the constraint $\mathsf{X}_i\mathsf{X}_j b_{ij} = 1$ for some $b_{ij} \in \{\pm 1\}$. Define the metric space on (X,ρ) as follows: let $X = \{1,\dots,n\}$ have a point for each vertex of G, and set $\rho(i,j) = \sqrt{2\left(1-b_{ij}\frac{1}{r}\right)}$ for $r = 2R+3$.

We claim that any r points of X embed isometrically into the Euclidean sphere of radius 1. To see this, fix a set $S \subset X$, and define the $|S|\times|S|$ matrix B_S so that

$$(B_S)_{ij} = \begin{cases} \frac{b_{ij}}{r} & \text{if } (i,j) \in E(G), \\ 0 & \text{otherwise}. \end{cases}$$

[18] We are here using the standard branch of arccos, returning values in $[0,\pi)$.

So long as $|S| \le r$, the matrix $M_S = \mathbb{1} + B_S$ is diagonally dominant, and therefore positive semidefinite, so in particular there must exist $V \in \mathbb{R}^{|S|\times|S|}$ with $M_S = VV^\top$. For each $i \in S$, let v_i correspond to the row of V indexed by i. Since for each $i \in S$, $1 = (M_S)_{i,i} = \|v_i\|^2$, we assign to $i \in S$ the unit vector v_i, and further for every pair $i, j \in S$, $\|v_i - v_j\|^2 = 2 - 2b_{ij}\frac{1}{r} = \rho(i,j)^2$.

Applying Theorem 4.1, we have that the solution

$$\widetilde{\mathbf{E}}[\mathsf{X}_i\mathsf{X}_j] = 1 - \frac{2}{\pi}\arccos\left(1 - \frac{1}{2}\cdot 2\left(1 - b_{ij}\frac{1}{r}\right)\right) = 1 - \frac{2}{\pi}\arccos\left(b_{ij}\frac{1}{r}\right)$$

is feasible. For convenience, let $f(z) = 1 - \frac{2}{\pi}\arccos(z)$. We use the following properties of f:

Claim 4.3. *The function $f(z) = 1 - \frac{2}{\pi}\arccos(z)$ exhibits the rotational symmetry $f(z) = -f(-z)$, and further $f(z) \ge \frac{2}{\pi}z$ for $z \in [0,1]$.*

We give the proof of the claim (using straightforward calculus) below. Now, because $f(z) = -f(-z)$, we have that

$$b_{ij}\cdot\widetilde{\mathbf{E}}[\mathsf{X}_i\mathsf{X}_j] = b_{ij}\cdot f\left(b_{ij}\frac{1}{r}\right) = f\left(\frac{1}{r}\right) \ge \frac{2}{\pi}\cdot\frac{1}{r},$$

where to obtain the final inequality we have used that for $z \in [0,1]$, $f(z) \ge \frac{2}{\pi}z \ge 0$. We conclude that R-round Sherali–Adams is unable to certify that OBJ $< \frac{1}{2} + \frac{1}{\pi r} = \frac{1}{2} + \frac{1}{\pi}\frac{1}{2R+3}$, as desired. □

Proof of Theorem 4.3. The rotational symmetry follows from simple manipulations:

$$f(z) - (-f(-z)) = 2 - \frac{2}{\pi}(\arccos(z) + \arccos(-z)) = 2 - \frac{2}{\pi}\arccos(-1) = 0.$$

For the second claim, we use that the derivative of $f(z) - \frac{2}{\pi}z$ is positive in the interval $[0, \frac{1}{2}]$:

$$\frac{\partial}{\partial z}\left(f(z) - \frac{2}{\pi}z\right) = \frac{2}{\pi}\frac{1}{\sqrt{1-z^2}} - \frac{2}{\pi} > 0 \text{ for } |z| < 1,$$

and that at $z = 0$, $f(z) - \frac{2}{\pi}z = 0$. □

5 Refutation for any Boolean CSP

In this section, we argue that Sherali–Adams can also refute non-trivial Boolean CSPs. For integers $n, t \ge 1$, we will use the notation $[n]^t$ to denote the set of sequences of length t in elements of $[n]$. First, for any predicate $P : \{\pm 1\}^k \to \{0,1\}$ we define a parameterized distribution over the CSP with constraints from P:

Definition 5.1. Let $P : \{\pm 1\}^k \to \{0,1\}$ be a predicate. Then we define a *random instance of P on n vertices with m expected clauses* to be an instance sampled as follows: define $p = \frac{m}{n^k}$, and for each sequence $S \in [n]^k$, independently with probability p we sample a uniformly random string $\zeta_S \in \{\pm 1\}^k$ and add the constraint that $P(x^S \odot \zeta_S) = 1$, where $\circ$ denotes the entry-wise (or Hadamard) product.

This is one of several popular models, and in our case it is the most convenient to work with. By employing some manipulations, results from this model transfer readily to the others (see for example Appendix D of [2] for details).

Our result is as follows:

Theorem 5.2. *Suppose that $P : \{\pm 1\}^k \to \{0,1\}$ and that $\delta, \varepsilon > 0$ are fixed constants. Let $\mathbf{E}[P]$ be the probability that a random $x \in \{\pm 1\}^k$ satisfies P. Then with high probability, for a random instance $\mathcal{I}$ of P on n variables with $m \geq n^{\lceil k/2 \rceil + \delta}$ expected clauses, the R-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}}(x) \leq \mathbf{E}[P] + \varepsilon$ with $R = O_{\varepsilon,\delta,k}(1)$ rounds. More specifically,*

$$R = k\ell \left(\frac{3 \cdot 2^{k/2-1}}{\varepsilon} \right)^{2\ell} + k \qquad \textit{for} \qquad \ell = \lceil \lceil \tfrac{k}{2} \rceil \tfrac{1}{2\delta} \rceil .$$

We can also prove a more fine-grained result, to obtain strong refutation at lower clause densities when the predicate has certain properties.

Definition 5.3. We say that a predicate $P : \{\pm 1\}^k \to \{0,1\}$ is *η-far from t-wise supporting* if every t-wise uniform distribution has probability mass at least η on the set of unsatisfying assignments $P^{-1}(0)$.

Theorem 5.4. *Suppose that $P : \{\pm 1\}^k \to \{0,1\}$ is η-far from t-wise supporting, and that $\delta, \varepsilon > 0$. Then with high probability, for a random instance $\mathcal{I}$ of P on n variables and $m \geq n^{\lceil t/2 \rceil + \delta}$ expected clauses, the R-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}}(x) \leq 1 - \eta + \varepsilon$ with $R = O_{\varepsilon,\delta,t}(1)$ rounds. More specifically,*

$$R = t\ell \left(\frac{3 \cdot 2^{t/2-1}}{\varepsilon} \right)^{2\ell} + t \qquad \textit{for} \qquad \ell = \lceil \lceil \tfrac{t}{2} \rceil \tfrac{1}{2\delta} \rceil .$$

Following the strategy introduced in [2], we will do this by first refuting weighted random instances of k-XOR for $k \geq 1$. After this, any predicate $P : \{\pm 1\}^k \to \{0,1\}$ can be decomposed according to its Fourier decomposition, which will yield a weighted sum of t-XOR instances for $t \leq k$, and our proof system will refute each individually.

5.1 Higher-arity XOR

Ultimately, we will reduce each k-CSP to a sum over weighted t-XOR instances with $t \leq k$:

Definition 5.5. Let $\mathcal{W}$ be a distribution over signed integers. We say that $\mathcal{I}$ is a *random k-XOR instance weighted according to $\mathcal{W}$* if it is sampled as follows: for each sequence $S \in [n]^k$, we take a b_S to be equal to a uniformly random sample from $\mathcal{W}$, and finally set the objective function to be $\sum_S b_S \cdot x^S$.

Following the standard strategy introduced by [29, 25] and subsequently honed in many articles, we will reduce refuting these t-XOR instances to refuting 2-XOR instances.

5.1.1 Even k-XOR

In this case, we perform a standard transformation to view the k-XOR instance as a 2-XOR instance on super-vertices given by subsets of vertices of size $k/2$.

Definition 5.6. Suppose $k > 1$ is an integer and $\mathcal{I}$ is a $2k$-XOR instance on n variables $x_1, \ldots, x_n$, with objective $\sum_{U \in [n]^{2k}} b_U \cdot x^U$. Then we let its *flattening*, $\mathcal{I}_{\text{flat}}$, be the 2-XOR instance on n^k variables given by associating a new variable y_S for each sequence $S \in [n]^k$, and for each $U \in [n]^{2k}$, choosing the partition of U into the sequences S, T with S containing the first k elements and T containing the last k, taking the objective function $\sum_{S,T} b_U \cdot y_S y_T$.

Lemma 5.7. *Suppose that $\mathcal{I}$ is a $2k$-XOR instance, and let $\mathcal{I}_{\text{flat}}$ be the 2-XOR instance given by its flattening. Then if the R-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}_{\text{flat}}}(x) \leq c$, then the $k \cdot R$-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}}(x) \leq c$.*

Proof. Every degree-R Sherali–Adams proof for $\mathcal{I}_{\text{flat}}$ can be transformed into a Sherali–Adams proof of degree at most kR for $\mathcal{I}$ by applying the transformation $y_S = \prod_{i \in S} x_i = x^S$. Further, this transformation exactly relates the objective functions of $\mathcal{I}_{\text{flat}}$ and $\mathcal{I}$. This proves the claim. $\square$

If the $2k$-XOR instances that we start with are random weighted instances, then their flattenings are also random weighted 2-XOR instances.

Claim 5.8. *Suppose that $\mathcal{I}$ is a random $2k$-XOR instance on n vertices weighted according to $\mathcal{W}$. Then the flattening $\mathcal{I}_{\text{flat}}$ is a random 2-XOR instance on n^k vertices weighted according to $\mathcal{W}$.*

Proof. This fact is immediate, since the sequences $U \in [n]^{2k}$ are in bijection with ordered pairs of sequences $S, T \in [n]^k$. $\square$

We will require the following proposition, which applies our main theorem in the context of random k-XOR instances with random weights from well-behaved distributions.

Proposition 5.9. *Suppose that $\mathcal{W}$ is a distribution over integers which is symmetric about the origin, and let k be a fixed positive integer, and n be an integer sufficiently large. Let $\mathbf{E}$ denote the expectation under the measure $\mathcal{W}$, and let $\sigma^2 \geq \mathbf{E}\, w^2$ be a bound on the variance. Furthermore, suppose that*

- *The expected absolute value is at least $\mathbf{E}\,|w| = \omega\left(\sigma\sqrt{\frac{\log n}{n^k}}\right)$,*
- *With high probability over n^{2k} i.i.d. samples $w_1, \ldots, w_{n^{2k}} \sim \mathcal{W}$, $\max_{i \in [n^{2k}]} |w_i| \leq M = o(\sigma^2 n^k)$.*

Now, define

$$\rho = O\left(\frac{\sigma \log n}{\mathbf{E}\,|w| \sqrt{n^k}} \cdot \max(1, \tfrac{M}{\sqrt{n^k}})\right).$$

Then if $\mathcal{I}$ is a random $2k$-XOR instance on n variables weighted according to $\mathcal{W}$, with high probability $\mathcal{I}$ has $\mathbf{E}\,|w| \cdot n^{2k} \pm O(\sigma n^k \sqrt{\log n})$ constraints. Further, choosing $\ell \in \mathbb{N}_+$ large enough so that $n^{k/4\ell}\rho \leq \frac{1}{2}\varepsilon^{2\ell}$ and setting $R = 2k \cdot \ell \cdot \left(\frac{1}{\varepsilon}\right)^{2\ell}$, R rounds of Sherali–Adams can deduce the bound $\mathrm{OBJ}_{\mathcal{I}}(x) \leq \frac{1}{2} + \frac{3}{2}\varepsilon$.

To prove the above, we require the following standard matrix Bernstein inequality:

Theorem 5.10 (Theorem 6.6.1 in [48])**.** *Let $A_1,\dots,A_m \in \mathbb{R}^{N\times N}$ be independent random matrices, with $\mathbf{E}A_i = 0$ for all $i\in[m]$ and $\|A_i\| \le M$ for all $i\in[m]$. Let $A = \sum_{i\in[m]} A_i$ denote their sum, and suppose that $\|\mathbf{E}AA^\top\|, \|\mathbf{E}A^\top A\| \le \sigma^2$. Then*

$$\Pr\left(\|A\| \ge t\right) \le N\cdot\exp\left(\frac{1}{2}\frac{-t^2}{\sigma^2+\frac{1}{3}Mt}\right).$$

Proof of Theorem 5.9. Given a weighted $2k$-XOR instance on n variables with weights from $\mathcal{W}$, we consider its flattening $\mathcal{I}_{\text{flat}}$ with objective function $\mathrm{OBJ}(x) = \frac{1}{m}\sum_{i,j\in[N]}\frac{1}{2}(1+b_{ij}x_ix_j)$ for m the absolute sum of weights, we construct its signed adjacency matrix as follows: first take the matrix W defined so that $W_{i,j} = b_{ij}$, and obtain a new matrix $B = \frac{1}{2}(W+W^\top)$. For any x, applying Theorem 5.7 we have that $\frac{1}{2}+\frac{1}{2m}x^\top Bx = \mathrm{OBJ}_{\mathcal{I}}(x)$.

Since $\mathcal{W}$ is a distribution over integers, $2B$ has signed integer entries. We think of $2B$ as defining a multigraph G on n^k vertices with signed edges, so that there are $2\cdot|B_{ij}|$ multiedges between $i,j\in[n^k]$, each with sign $\mathrm{sgn}(B_{ij})$. Let $2\cdot D$ be the degree matrix of G, let $A = |2B|$ be the adjacency matrix of G, let $\Xi = \mathrm{sgn}(B)$ be the matrix of signs of B, and let $K = (2D^{-1})A = D^{-1}B\circ\Xi$ be the transition matrix for the random walk on G.

To apply Theorem 3.7, we must upper bound the spectral radius of $\Xi\circ K = D^{-1}B$, as well as bound the minimum degree of G and the total number of edges. We will use the bound

$$\|D^{-1}B\|_{op} \le \|D^{-1}\|_{op}\cdot\|B\|_{op} \le \frac{1}{\pi^*}\|B\|_{op}.$$

First, we bound $\|B\|_{op}$. Take B' to be the truncated version of B, so that $B'_{i,j} = \mathrm{sgn}(B_{i,j})\cdot\max(|B_{i,j}|,M)$. Thinking of the matrix B' as the sum of $\binom{n^k}{2}+n^k$ symmetric matrices, one for each pair $i,j\in[n^k]$, satisfies the requirements of Theorem 5.10. We have that $\mathbf{E}[B'B'^\top] \preceq n^k\sigma^2\cdot\mathbb{1}$, so applying Theorem 5.10 with $t = O\max(\sqrt{\sigma^2 n^k\log n}, M\log n)$ we get that with high probability,

$$\|B\|_{op} \le O\left(\max\left(\sqrt{\sigma^2 n^k\log n}, M\log n\right)\right),$$

where we have also used that with high probability $B = B'$ by the properties of $\mathcal{W}$.

Now, we bound the sum of degrees $2m$ and the minimum degree $d_{\min}$. We have that the total sum of the degrees is given by $2m = \sum_{i,j\in[n^k]}|b_{ij}|$ with $b_{ij}\sim\mathcal{W}$. By a Bernstein inequality,

$$\Pr\left(\left|2m - n^{2k}\mathbf{E}|w|\right| \ge s\right) \le 2\exp\left(-\frac{1}{2}\frac{s^2}{n^{2k}\cdot\mathbf{E}w^2+\frac{1}{3}Ms}\right),$$

so since by assumption $\sigma^2 n^{2k} \gg M$, setting $s = O(\sigma n^k\sqrt{\log n})$ we have that with high probability

$$2m = n^{2k}\mathbf{E}|w| \pm O(\sigma n^k\sqrt{\log n}). \tag{5.1}$$

By our assumptions on $\mathcal{W}$ we have that for every $i \in [n^k]$, $\mathbf{E}\deg_G(i) = n^k\,\mathbf{E}|w|$. Applying a Bernstein inequality gives that

$$\Pr[\deg_G(i) \le n^k\,\mathbf{E}|w| - t] \le \exp\left(-\frac{1}{2}\cdot\frac{t^2}{n^k\sigma^2 + \frac{1}{3}Mt}\right),$$

so using that $M \le n^k\sigma^2$ and taking $t = O\left(\sqrt{n^k\sigma^2\log n}\right)$ we get that $d_{\min} = n^k \cdot \mathbf{E}|w| \pm O(\sqrt{n^k\sigma^2\log n})$ with high probability. This gives that with high probability,

$$\pi^* = \frac{d_{\min}}{2m} \ge \frac{n^k\cdot\mathbf{E}|w| - O(\sigma\sqrt{n^k\log n})}{n^{2k}\cdot\mathbf{E}|w| + O(\sigma n^k\sqrt{\log n})} \ge \frac{1}{n^k}\cdot(1-o(1))$$
$$\|\Xi\circ K\|_{op} \le \frac{\|B\|_{op}}{d_{\min}} \le \frac{O(\sigma\sqrt{n^k}\log n)\cdot\max(1,\frac{M}{\sqrt{n^k}})}{n^k\cdot\mathbf{E}|w| - O(\sigma\sqrt{n^k\log n})} \le O\left(\frac{\sigma\log n}{\mathbf{E}|w|\sqrt{n^k}}\right)\cdot\max(1,\tfrac{M}{\sqrt{n^k}}),$$

where we have used that $\sigma\sqrt{\log n} \ll \sqrt{n^k}\,\mathbf{E}|w|$.

Now, the result follows by applying Theorem 3.7 and Theorem 5.7. □

5.1.2 Odd k-XOR

For odd integers k, k-XOR instances do not have the same natural, symmetric flattenings. Instead, we define what we call a *lift*:

Definition 5.11. Suppose $k \ge 1$ is an integer and $\mathcal{I}$ is a $(2k+1)$-XOR instance on n variables $x_1,\dots,x_n$, with objective $\sum_{U\in[n]^{2k+1}} b_U\cdot x^U$. Then we let its *lift*, $\mathcal{I}_{lift}$, be the bipartite 2-XOR instance on parts each containing n^{k+1} variables created as follows:

- Create new variables $w_1,\dots,w_n$
- For each $U\in[n]^{2k+1}$, choose a random index $i_U\in[n]$ and modify the objective to $\sum_{U\in[n]^{2k+1}} b_U\cdot x^U\cdot w_{i_U}$
- For each sequence S associate a new variable y_S, and for each sequence $T\in[n]^k$ and index $i\in[n]$ associate a new variable $z_{T,i}$. We understand $y_S = \prod_{i\in S} x_i$, and $z_{T,i} = \left(\prod_{j\in T} x_j\right)\cdot w_i$.
- For each $U\in[n]^{2k+1}$, we take the sequence $V=(U,i_U)$ and assign it a new coefficient $b'_V = b_U$. For the remaining b'_V, we set $b'_V = 0$.

Finally, $\mathcal{I}_{lift}$ is the instance with the objective function $\sum_{S\in[n]^{k+1},T\in[n]^k,i\in[n]} b'_{S\cup T\cup i}\cdot y_S z_{T,i}$.

We obtain a statement analogous to Theorem 5.7 for odd k-XOR:

Lemma 5.12. *Suppose that $\mathcal{I}$ is a weighted $(2k+1)$-XOR instance, and let $\mathcal{I}_{\text{flat}}$ be the bipartite 2-XOR instance given by its flattening. Then if the R-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}_{\text{flat}}}(x)\le c$, then the $(k+1)\cdot R$-round Sherali–Adams proof system can certify that $\mathrm{OBJ}_{\mathcal{I}}(x)\le c$.*

Proof. The only modification to the proof of Theorem 5.7 is that for $z_{T,i}$ we substitute $z_{T,i} = x^T$ (where we have implicitly substituted $w_i = 1$ for all $i \in [n]$). □

However, the lifting procedure does not preserve the weighting distribution $\mathcal{W}$, because of the step in which a random index i_U is chosen to lift U. For this reason, we prove an analog of Theorem 5.9:

Proposition 5.13. *Suppose that $\mathcal{W}$ is a distribution over integers which is symmetric about the origin, and let k be a fixed positive integer, and let n be an integer sufficiently large. Let $\mathbf{E}$ denote the expectation under the measure $\mathcal{W}$, and let $\sigma^2 \geq Ew^2$ be a bound on the variance. Furthermore, suppose that*

- *The expected absolute value is at least $\mathbf{E}|w| \gg \sigma\sqrt{\frac{\log n}{n^k}}$,*
- *With high probability over n^{2k+1} i.i.d. samples $w_1, \ldots, w_{n^{2k+1}} \sim \mathcal{W}$, $\max_{i\in[n^{2k+1}]}|w_i| \leq M \ll \sigma^2 n^k$.*

Now, define

$$\rho = O\left(\frac{\sigma \log n}{\mathbf{E}|w|\sqrt{n^k}} \cdot \max(1, \tfrac{M}{\sqrt{n^k}})\right).$$

Then if $\mathcal{I}$ is a random $(2k+1)$-XOR instance on n variables weighted according to $\mathcal{W}$, with high probability it has $\mathbf{E}|w| \cdot n^{2k+1} \pm O(\sigma n^k \sqrt{\log n})$ constraints. Furthermore, choosing $\ell \in \mathbb{N}_+$ large enough so that $n^{(k+1)/4\ell}\rho \leq \frac{1}{2}\varepsilon^{2\ell}$ and setting $R = (2k+2)\ell \cdot \left(\frac{1}{\varepsilon}\right)^{2\ell}$, R rounds of Sherali–Adams can deduce the bound $\mathrm{OBJ}_{\mathcal{I}}(x) \leq \frac{1}{2} + \frac{3}{2}\varepsilon$.

Proof. The thread of the proof is the same as that of Theorem 5.9. We will refute $\mathcal{I}_{lift}$, since by Theorem 5.12 this is sufficient. We begin by associating with $\mathcal{I}_{lift}$ a multigraph G (which we may do because $\mathcal{W}$ is a distribution over integers). The multigraph G is a bipartite graph, with one bipartition corresponding to variables y_S for $S \in [n]^{k+1}$, and one bipartition corresponding to variables $z_{T,i}$ for $T \in [n]^k$ and $i \in [n]$. We let the block matrix B be (the $\frac{1}{2}$-scaled) signed adjacency matrix of G, let Ξ be the matrix of signs, D be the diagonal degree matrix, and $K \circ \Xi = D^{-1}B$ be the signed transition matrix of the random walk on G. In order to apply Theorem 3.7 we must bound $\|K \circ \Xi\|$ and $\pi^* = d_{\min}(G)/2m$.

First, we bound the vertex degrees. For a vertex of the form (T,i), the expected value of the incident edge $(S, T\cup i)$ is $b_{S,T} \cdot \frac{1}{n}$, where $b_{S,T}$. The degree of $T \cup i$ is simply the sum

$$\deg_G(T\cup i) = \sum_{S\in[n]^{k+1}} |b_{S,T\cup i}|,$$

a sum of independent random variables with expectation $\frac{1}{n}\mathbf{E}|w|$ and variances $\frac{1}{n}\sigma^2$. Applying a Bernstein inequality, we have that

$$\Pr\left(\left|\deg_G(T\cup i) - \frac{1}{n}\cdot n^{k+1}\,\mathbf{E}[|w|]\right| \geq t\right) \leq 2\exp\left(\frac{1}{2}\frac{-t^2}{n^k\sigma^2 + \frac{1}{3}Mt}\right),$$

so taking $t = O(\sqrt{\sigma^2 n^k \log n})$ (and using that $M \ll n^k\sigma^2$), we have that the degree of $T \cup i$ vertices is $\deg_G(T\cup i) = n^k\,\mathbf{E}|w| \pm O(\sqrt{\sigma^2 n^k \log n})$ with high probability.

A similar argument applies to the S vertices; the total degree of such a vertex is

$$\deg_G(S) = \sum_{T\in[n]^k} \left| \sum_{i\in[n]} b_{S,T\cup i} \right| ,$$

since only one of the $b_{S,T\cup i}$ will be nonzero. The inner sums are independent random variables with mean $\mathbf{E}\,|w|$ and variance σ^2, therefore

$$\Pr\left(\left| \deg_G(S) - n^k\,\mathbf{E}\,|w| \right| \geq t \right) \leq 2\exp\left(\frac{1}{2} \frac{-t^2}{n^k\sigma^2 + \frac{1}{3}Mt} \right) ,$$

so taking $t = O(\sqrt{\sigma^2 n^k \log n})$ (and using that $M \ll n^k\sigma^2$), we have that the degree of S vertices is also $\deg_G(S) = n^k\,\mathbf{E}\,|w| \pm O(\sqrt{\sigma^2 n^k \log n})$ with high probability.

We finally bound $\|K\circ\Xi\| \leq \|D^{-1}\|\cdot\|B\|$. As before, B is a sum of independent symmetric matrices, one for each coefficient b_U from $\mathcal{I}$. That is, we can define matrices B_U for each $U\in[n]^{2k+1}$ with $U = S,T$ for $S\in[n]^{k+1}, T\in[n]^k$ where B_U has a number $b_U \sim \mathcal{W}$ in one off-diagonal block entry $(S, T\cup i)$ and the other off-diagonal block entry $(T\cup i, S)$ for a randomly chosen $i\in[n]$. Thus, $\mathbf{E}\,B_U B_U^\top$ is a diagonal matrix with $\frac{1}{n}\sigma^2$ on each diagonal of the form $(T\cup i, T\cup i)$ and σ^2 on each block diagonal of the form (S,S). We then have that $\mathbf{E}\,BB^\top \preceq n^k\sigma^2\mathbb{1}$, since for each S there is a sum over n^k matrices B_U and for each $T\cup i$ there is a sum over n^{k+1} matrices B_U. Applying Theorem 5.10 by using the same truncation trick again, we have that

$$\|B\| \leq O\left(\max\left(\sqrt{\sigma^2 n^k \log n}, M\log n \right) \right) ,$$

and from this we have that with high probability,

$$\|K\circ\Xi\| \leq \frac{1}{\deg_{\min}(G)} \leq O\left(\frac{\sigma\log n}{\mathbf{E}\,|w|\sqrt{n^k}} \right) \cdot \max(1, \tfrac{M}{\sqrt{n^k}}) , \tag{5.2}$$

$$\pi^* = \frac{\deg_{\min}(G)}{2m} = \frac{1}{n^{k+1}} \cdot (1 \pm o(1)) , \tag{5.3}$$

after which we can apply Theorem 3.7. □

5.2 From Boolean CSPs to k-XOR

Following [2], we prove Theorem 5.2 via reduction to XOR.

Proof of Theorem 5.2. Given a random instance of the CSP defined by the predicate P, and $p = n^{-\lfloor k/2\rfloor+\delta}$, a Bernstein inequality gives us that the number of constraints m is with high probability given by $m = n^{\lceil k/2\rceil+\delta} \pm 10\sqrt{n^{\lceil k/2\rceil+\delta}\log n}$. Set $\ell = \lceil \frac{1}{2\delta} \rceil$.

Since P is a Boolean predicate, we can write P in its Fourier expansion:

$$P(z) = \sum_{\alpha\subseteq[k]} \widehat{P}(\alpha) \prod_{i\in\alpha} z_i .$$

Using this expansion, we re-write the objective function. Recall that $[n]^k$ is the set of all sequences of k elements of $[n]$. For each $S \in [n]^k$, let b_S be the $0/1$ indicator that there is a constraint on S, and let $\zeta_S \in \{\pm 1\}^S$ be the literal signing on the variables for the constraint on S. Then, if the total number of constraints is m,

$$\begin{aligned}
\mathrm{OBJ}_{\mathcal{I}}(x) &= \frac{1}{m} \sum_{S \in [n]^k} b_S \cdot P(x^S \circ \zeta_S) \\
&= \frac{1}{m} \sum_{S=\{i_1,\dots,i_k\} \in [n]^k} \sum_{\alpha \subseteq [k]} b_S \cdot \widehat{P}(\alpha) \prod_{a \in \alpha} x_{i_a} (\zeta_S)_{i_a} \\
&= \frac{1}{m} \sum_{\alpha \subseteq [k]} \widehat{P}(\alpha) \cdot \sum_{T \in [n]^{|\alpha|}} \left(\sum_{S \in [n]^k, S|_\alpha = T} b_S \cdot \prod_{a \in \alpha} (\zeta_S)_{i_a} \right) \cdot x^T .
\end{aligned} \tag{5.4}$$

Now, define for each $\alpha \subseteq [k]$ with $|\alpha| = t > 0$ the t-XOR instance

$$\mathcal{I}^\alpha(x) = \frac{1}{m} \sum_{T \in [n]^t} \left(\sum_{S \in [n]^k, S|_\alpha = T} b_S \cdot \prod_{a \in \alpha} (\zeta_S)_{i_a} \right) \cdot x^T = \frac{1}{m} \sum_{T \in [n]^t} w_T \cdot x^T ,$$

where we have taken $w_T = \sum_{S \in [n]^k, S|_\alpha = T} b_S \cdot \prod_{a \in \alpha} (\zeta_S)_{i_a}$. So that from Equation (5.4),

$$\mathrm{OBJ}_{\mathcal{I}}(x) = \sum_{\alpha \subseteq [k]} \widehat{P}(\alpha) \cdot \mathcal{I}^\alpha(x) . \tag{5.5}$$

Let $\mathcal{W}_{n^{k-t}}$ be the distribution defined so that $w \sim \mathcal{W}_{n^{k-t}}$ is a sum of n^{k-t} independent variables taking value $\{\pm 1\}$ with probability p and value 0 otherwise. Since for each $S \supseteq T$, the quantity $\prod_{a \in \alpha} (\zeta_S)_{i_a}$ is an independent uniform sign in $\{\pm 1\}$ and b_S is an independent Bernoulli-p variable, we have that the coefficients w_T in $\mathcal{I}^\alpha$ are i.i.d. from $\mathcal{W}_{n^{k-t}}$. The following lemma establishes some properties of $\mathcal{W}_N$ (we will prove the lemma in Appendix A):

Lemma 5.14. *Let $\mathcal{W}_N(p)$ be the distribution defined so that $X \sim \mathcal{W}_N$ is given by $X = \sum_{t=1}^N Y_t \cdot Z_t$, where the $\{Y_t\}_t, \{Z_t\}_t$ are i.i.d with $Y_t \sim \mathrm{Ber}(p)$ and $Z_t \sim \{\pm 1\}$. Then for $X \sim \mathcal{W}_N(p)$, $\mathbf{E}\, X = 0$ and $\mathbf{E}\, X^2 = pN$. Further, so long as $pN \geq 1$, $\mathbf{E}\,|X| \geq \frac{2}{e^{3/2}} \sqrt{pN}$, and $\Pr(|X| > 2t\sqrt{pN}) \leq 2 \exp\left(-t^2\right)$. Otherwise, if $pN \leq 1$, $\mathbf{E}\,|X| \geq \frac{1}{2e} \log \frac{1}{1-pN}$, and $\Pr(|X| \geq 1+t) \leq \exp\left(-\frac{1}{2} t\right)$.*

From Theorem 5.14, we have that $\mathbf{E}\, w_T^2 = pn^{k-t}$, and by Cauchy–Schwarz, $\mathbf{E}\,|w_T| \leq \sqrt{\mathbf{E}\, w_T^2}$. Let m_α be the total absolute weight of constraints in $\mathcal{I}^\alpha$,

$$m_\alpha = \sum_T |w_T| .$$

Notice that in all cases, $m_\alpha \leq m$.

Now, we show that Sherali–Adams can certify upper bounds on $|\mathcal{I}^\alpha(x)|$ for every α. First, consider α with $|\alpha| = t = 1$. In this case, Sherali–Adams with $R = 1$ can certify that

$$\mathcal{I}^\alpha(x) = \frac{1}{m} \sum_{i \in [n]} w_i \cdot x_i \leq \frac{1}{m} \sum_{i \in [n]} |w_i| = \frac{m_\alpha}{m} ,$$

From an application of Bernstein's inequality (the same as in the proof of Theorem 5.9),

$$m_\alpha \le n \cdot \sqrt{\mathbf{E}\, w_T^2} + \sqrt{pn^k \log n}$$

with high probability whenever $pn^k \gg \sqrt{pn^{k-1}}$, and applying our bound on m we conclude that with high probability SA will certify that

$$\mathcal{J}^\alpha(x) \le \frac{n \cdot \sqrt{pn^{k-1}}(1+o(1))}{pn^k(1 \pm o(1))} \le \frac{2}{\sqrt{pn^{k-1}}}\,.$$

The lower bound on $\mathcal{J}^\alpha(x)$ follows from identical reasoning with its negation, so we can conclude that with high probability SA can certify that

$$|\mathcal{J}^\alpha(x)| \le \frac{2}{\sqrt{pn^{k-1}}} = o(1)\,.$$

Now, we tackle α with $|\alpha| = t$ for $2 \le t \le k$. We will verify that the conditions of Theorems 5.9 and 5.13 hold. First, consider the α with $|\alpha| = t$ for $pn^{k-t} \ge 1$. From Theorem 5.14, in this case we have that

$$\mathbf{E}\,|w_T| \ge \frac{2}{\mathrm{e}^{3/2}}\sqrt{pn^{k-t}}\,,$$

and with high probability, $|w_T| \le O(\sqrt{t\,pn^{k-t}\log n})$ for all $T \in [n]^t$. Letting $M = O(\sqrt{t\,pn^{k-t}\log n})$, and $\sigma^2 = pn^{k-t}$, we meet the conditions for Theorem 5.9:

$$\begin{aligned} M \le O(\sqrt{pn^{k-t}}) &\ll pn^{k-t} \cdot n^{\lfloor t/2 \rfloor} = \sigma^2 n^{\lfloor t/2 \rfloor}\,, \\ \mathbf{E}\,|w_T| \ge \frac{2}{\mathrm{e}^{3/2}}\sqrt{pn^{k-t}} &\gg \sqrt{\frac{pn^{k-t}\log n}{n^{\lfloor t/2 \rfloor}}} = \sqrt{\frac{\sigma^2 \log n}{n^{\lfloor t/2 \rfloor}}}\,, \end{aligned}$$

so long as $pn^{k-t} \ge 1$, which we have assumed. So applying Theorems 5.9 and 5.13 to both $\frac{m}{m_\alpha}\mathcal{J}^\alpha$ and $-\frac{m}{m_\alpha}\mathcal{J}^\alpha$, we have

$$\rho = O\left(\frac{\sigma \log n}{\mathbf{E}\,|w_T|\sqrt{n^{\lfloor t/2 \rfloor}}}\right) \cdot \max(1, \tfrac{M}{\sqrt{n^{\lfloor t/2 \rfloor}}}) \le O\left(\frac{\log n}{\sqrt{n^{\lfloor t/2 \rfloor}}}\right) \cdot \max\left(1, \frac{\sqrt{pn^{k-t}}}{n^{\lfloor t/2 \rfloor}}\right)\,,$$

and so long as $\frac{m_\alpha}{m} \cdot n^{\lceil t/2 \rceil/4\ell}\rho \le \frac{1}{2}\varepsilon^{2\ell}$, with high probability over $\mathcal{J}^\alpha$, $t(\ell r+1)$ rounds of Sherali–Adams certify that $|\mathcal{J}^\alpha(x)| \le \frac{3}{2}\varepsilon$. We confirm that

$$\frac{m_\alpha}{m} \cdot n^{\lceil t/2 \rceil/4\ell} \cdot \rho = \frac{\sqrt{pn^{k-t}} \cdot n^t}{pn^k} \cdot n^{\lceil t/2 \rceil/4\ell} \cdot \rho \ll o(1)\,,$$

whenever $t \ge 1$ and $\ell \ge 1$.

Finally, we handle α with $|\alpha| = t$ for $2 \le t$ and $pn^{k-t} < 1$. From Theorem 5.14 we have

$$\mathbf{E}\,|w_T| \ge \frac{1}{\mathrm{e}} \log \frac{1}{1 - pn^{k-t}}\,,$$

and with high probability, $|w_T| \le 4\log n$ for all $T \in [n]^t$. Taking $M = 4\log n$, we meet the conditions of Theorems 5.9 and 5.13:

$$M \le 4\log n \ll n^{\lceil k/2\rceil - \lceil t/2\rceil + \delta} = pn^{k-t} n^{\lfloor t/2\rfloor} = \sigma^2 n^{\lfloor t/2\rfloor},$$
$$\mathbf{E}|w_T| \ge \frac{1}{\mathrm{e}} \log \frac{1}{1-pn^{k-t}} \ge \frac{1}{\mathrm{e}} pn^{k-t} \gg \sqrt{\frac{pn^{k-t}\log n}{n^{\lfloor t/2\rfloor}}} = \sqrt{\frac{\sigma^2 \log n}{n^{\lfloor t/2\rfloor}}},$$

where the last inequality is true whenever $pn^{k-t+\lfloor t/2\rfloor} = n^{\lceil k/2\rceil - \lceil t/2\rceil + \delta} \gg \log n$, which we have by assumption. So applying Theorems 5.9 and 5.13 to both $\frac{m}{m_\alpha}\mathcal{J}^\alpha$ and $-\frac{m}{m_\alpha}\mathcal{J}^\alpha$, we have that for

$$\rho = O\left(\frac{\sqrt{pn^{k-t}}\log n}{pn^{k-t}\sqrt{n^{\lfloor t/2\rfloor}}}\right) = O(\log n) \cdot \sqrt{\frac{1}{pn^{k-\lceil t/2\rceil}}} = O\left(\frac{\log n}{\sqrt{n^{\lceil k/2\rceil - \lceil t/2\rceil + \delta}}}\right),$$

so long as $\frac{m_\alpha}{m} n^{\lceil t/2\rceil/4\ell}\rho \le \frac{1}{2}\varepsilon^{2\ell}$, $R = t(\ell r + 1)$ rounds of Sherali–Adams certify that $|\mathcal{J}^\alpha(x)| \le \frac{3}{2}\varepsilon$ with high probability. Verifying,

$$\frac{m_\alpha}{m} \cdot n^{\lceil t/2\rceil/4\ell} \cdot O\left(\frac{\log n}{\sqrt{n^{\lceil k/2\rceil - \lceil t/2\rceil + \delta}}}\right) = \frac{1}{\sqrt{pn^{k-t}}} \cdot n^{\lceil t/2\rceil/4\ell} \cdot O\left(\frac{\log n}{\sqrt{n^{\lceil k/2\rceil - \lceil t/2\rceil + \delta}}}\right),$$

which is maximized at $t = k$. By our choice of ℓ, the condition holds.

We therefore have (using Parseval's identity and $\|x\|_1 \le \sqrt{k}\|x\|_2$ for $x \in \mathbb{R}^k$ to simplify Equation (5.5)) that the same number of rounds certifies that

$$\mathrm{OBJ}_{\mathcal{J}}(x) \le \sum_{\alpha \subset [k]} \widehat{P}(\alpha) \cdot \mathcal{J}^\alpha(x) \le \widehat{P}(\emptyset) + \sqrt{2^k}\frac{3}{2}\varepsilon,$$

as desired. □

Using arguments analogous to the above along with the reasoning outlined in Theorem 4.9, proof 2 and Claim 6.7 from [2], we can also prove Theorem 5.4.

A Characteristics of distributions of XOR-subformula coefficients

We now prove Theorem 5.14. We will use the following estimate for the mean absolute deviation of a binomial random variable.

Lemma A.1 (e. g., [9]). *If X is distributed according to the binomial distribution $X \sim \mathrm{Bin}(n,p)$, then*

$$\mathbf{E}|X - \mathbf{E}X| = \sqrt{\frac{2}{\pi} np(1-p)} + O\left(\frac{1}{\sqrt{n}}\right).$$

Proof of theorem 5.14. We calculate the absolute value directly. Given that there are exactly k nonzero Y_t, the absolute value of X is distributed according to $|\mathrm{Bin}(k, \frac{1}{2}) - \frac{1}{2}k|$. Using the method of conditional expectations,

$$\mathbf{E}|X| = \sum_{k=0}^{N} \Pr[k \text{ nonzero } Y_t\text{'s}] \cdot \mathbf{E}|\mathrm{Bin}(k, \tfrac{t}{2}) - \tfrac{1}{2}k| \ge \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} \cdot \sqrt{\frac{1}{2\pi}k},$$

where we have applied the estimate from Theorem A.1. Letting $D(a\|b) = a\ln\frac{a}{b} + (1-a)\ln\frac{1-a}{1-b}$ be the relative entropy, we then have from Stirling's inequality that

$$\mathbf{E}\,|X| \geq \sum_{k=1}^{N}\sqrt{\frac{2\pi}{\mathrm{e}^2}\frac{N}{k(N-k)}}\cdot\exp\left(-N\cdot D\left(\tfrac{k}{N}\|p\right)\right)\cdot\sqrt{\frac{1}{2\pi}k} \geq \frac{1}{\mathrm{e}}\sum_{k=1}^{N}\exp\left(-N\cdot D\left(\tfrac{k}{N}\|p\right)\right). \tag{A.1}$$

Now, if $pN < 1$, we take

$$(\text{A.1}) \geq \frac{1}{\mathrm{e}}\sum_{k=1}^{N}\exp\left(k\log\left(\frac{pN}{k}\right)\right) = \frac{1}{\mathrm{e}}\sum_{k=1}^{N}\left(\frac{pN}{k}\right)^k \geq -\frac{1}{\mathrm{e}}\log(1-pN) - O(p^N), \tag{A.2}$$

as desired.

If $pN \geq 1$, applying the change of variables $\ell = k - \lfloor pN \rfloor$ and $\delta = \frac{\ell}{N}$,

$$(\text{A.1}) \geq \frac{1}{\mathrm{e}}\sum_{\substack{\ell=1-\lfloor pN\rfloor \\ \delta=\frac{\ell}{N}}}^{\lfloor(1-p)N\rfloor}\exp\left(-N\cdot D\left(p+\delta\|p\right)\right). \tag{A.3}$$

Now using the Taylor expansion for $\log(1-x)$ to simplify and restricting the sum over the range $\ell = [-\lfloor\sqrt{pN}\rfloor, \lfloor\sqrt{pN}\rfloor]$, we get the bound

$$(\text{A.3}) \geq \frac{1}{\mathrm{e}}\sum_{\substack{\ell=-\lfloor\sqrt{pN}\rfloor \\ \delta=\frac{\ell}{N}}}^{\lfloor\sqrt{pN}\rfloor}\exp\left(-N\frac{\delta^2}{2}\right) \geq \frac{1}{\mathrm{e}}\cdot 2\sqrt{pN}\cdot\exp(-\tfrac{p}{2}) \geq \frac{2}{\mathrm{e}^{3/2}}\sqrt{pN}, \tag{A.4}$$

as desired.

The first and second moment we can also obtain by calculation; the Z_t ensure that the summands have mean 0, and the Y_t give that the variance of the summands is p, which gives the result.

The tail bound $\Pr(|X| \geq (1+2t)\sqrt{pN}) \leq 2\exp(-t^2)$ comes from an application of Bernstein's inequality if $pN \geq 1$; when $pN < 1$, we again apply Bernstein's inequality, in which case we have

$$\Pr\left(|X| - \mathbf{E}\,|X| \geq s\right) \leq \exp\left(-\frac{1}{2}\frac{s^2}{pN + \frac{1}{3}s}\right),$$

and choosing $s = 1+t$ gives the result. □

Acknowledgments

We thank Luca Trevisan for helpful comments, Sam Hopkins for feedback on an earlier written version, and Boaz Barak for suggesting the title. We also thank the Schloss Dagstuhl Leibniz Center for Informatics (and more specifically the organizers of the CSP Complexity and Approximability workshop), as well as the Casa Mathemática Oaxaca (and more specifically the organizers of the Analytic Techniques in TCS workshop); parts of this paper came together during discussions at these venues. T.S. also thanks the Oberwolfach Research Institute for Mathematics (and the organizers of the Proof Complexity and Beyond workshop), the Simons Institute (and the organizers of the Optimization semester program), and the Banff International Research Station (and the organizers of the Approximation Algorithms and Hardness workshop) where she tried to prove the opposite of the results in this paper, as well as Sam Hopkins, with whom some of those efforts were made.

References

[1] MIKHAIL ALEKHNOVICH, SANJEEV ARORA, AND IANNIS TOURLAKIS: Towards strong nonapproximability results in the Lovász–Schrijver hierarchy. *Comput. Complexity*, 20(4):615–648, 2011. [doi:10.1007/s00037-011-0027-z] 6

[2] SARAH R. ALLEN, RYAN O'DONNELL, AND DAVID WITMER: How to refute a random CSP. In *Proc. 56th FOCS*, pp. 689–708. IEEE Comp. Soc., 2015. [doi:10.1109/FOCS.2015.48] 5, 6, 7, 23, 28, 31

[3] SANJEEV ARORA, BÉLA BOLLOBÁS, LÁSZLÓ LOVÁSZ, AND IANNIS TOURLAKIS: Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006. [doi:10.4086/toc.2006.v002a002] 5, 6

[4] SANJEEV ARORA, LÁSZLÓ LOVÁSZ, ILAN NEWMAN, YUVAL RABANI, YURI RABINOVICH, AND SANTOSH VEMPALA: Local versus global properties of metric spaces. *SIAM J. Comput.*, 41(1):250–271, 2012. [doi:10.1137/090780304] 5

[5] FRANCISCO BARAHONA AND ALI RIDHA MAHJOUB: On the cut polytope. *Math. Programming*, 36(2):157–173, 1986. [doi:10.1007/BF02592023] 2

[6] BOAZ BARAK, MORITZ HARDT, THOMAS HOLENSTEIN, AND DAVID STEURER: Subsampling mathematical relaxations and average-case complexity. In *Proc. 22nd Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'11)*, pp. 512–531. SIAM, 2011. [doi:10.1137/1.9781611973082.41] 6

[7] BOAZ BARAK, SAMUEL B. HOPKINS, JONATHAN A. KELNER, PRAVESH K. KOTHARI, ANKUR MOITRA, AND AARON POTECHIN: A nearly tight Sum-of-Squares lower bound for the Planted Clique problem. *SIAM J. Comput.*, 48(2):687–735, 2019. Preliminary version in FOCS'16. [doi:10.1137/17M1138236] 7

[8] BOAZ BARAK, PRASAD RAGHAVENDRA, AND DAVID STEURER: Rounding semidefinite programming hierarchies via global correlation. In *Proc. 52nd FOCS*, pp. 472–481. IEEE Comp. Soc., 2011. [doi:10.1109/FOCS.2011.95] 7

[9] COLIN R. BLYTH: Expected absolute error of the usual estimator of the binomial parameter. *The American Statistician*, 34(3):155–157, 1980. [doi:10.1080/00031305.1980.10483022] 31

[10] RAVI B. BOPPANA: Eigenvalues and graph bisection: An average-case analysis. In *Proc. 28th FOCS*, pp. 280–285. IEEE Comp. Soc., 1987. [doi:10.1109/SFCS.1987.22] 3

[11] ANDREI Z. BRODER, ALAN M. FRIEZE, STEPHEN SUEN, AND ELI UPFAL: Optimal construction of edge-disjoint paths in random graphs. *SIAM J. Comput.*, 28(2):541–573, 1998. [doi:10.1137/S0097539795290805] 4

[12] SIU ON CHAN, JAMES R. LEE, PRASAD RAGHAVENDRA, AND DAVID STEURER: Approximate constraint satisfaction requires large LP relaxations. *J. ACM*, 63(4):34:1–22, 2016. [doi:10.1145/2811255] 2, 7, 12

[13] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Integrality gaps for Sherali–Adams relaxations. In *Proc. 41st STOC*, pp. 283–292. ACM Press, 2009. [doi:10.1145/1536414.1536455] 2, 3, 5, 6, 7, 12, 21

[14] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Local global tradeoffs in metric embeddings. *SIAM J. Comput.*, 39(6):2487–2512, 2010. [doi:10.1137/070712080] 5, 6

[15] MOSES CHARIKAR AND ANTHONY WIRTH: Maximizing quadratic programs: Extending Grothendieck's Inequality. In *Proc. 45th FOCS*, pp. 54–60. IEEE Comp. Soc., 2004. [doi:10.1109/FOCS.2004.39] 3, 4

[16] NICHOLAS COOK, LARRY GOLDSTEIN, AND TOBIAS JOHNSON: Size biased couplings and the spectral gap for random regular graphs. *Ann. Probab.*, 46(1):72–125, 2018. [doi:10.1214/17-AOP1180] 4

[17] CHARLES DELORME AND SVATOPLUK POLJAK: Laplacian eigenvalues and the maximum cut problem. *Math. Programming*, 62(1–3):557–574, 1993. ACM DL. 3

[18] MICHEL MARIE DEZA AND MONIQUE LAURENT: *Geometry of Cuts and Metrics*. Volume 15 of *Algorithms and Combinatorics*. Springer, 1997. [doi:10.1007/978-3-642-04295-9] 2

[19] WILLIAM E. DONATH AND ALAN J. HOFFMAN: Lower bounds for the partitioning of graphs. *IBM J. Res. Dev.*, 17(5):420–425, 1973. Included in "Selected Papers of Alan J. Hoffman: With Commentary" (World Scientific, 2003, pp. 437–442). [doi:10.1147/rd.175.0420] 3

[20] URIEL FEIGE AND ROBERT KRAUTHGAMER: The probable value of the Lovász–Schrijver relaxations for maximum independent set. *SIAM J. Comput.*, 32(2):345–370, 2003. [doi:10.1137/S009753970240118X] 7

[21] URIEL FEIGE AND MICHAEL LANGBERG: The RPR^2 rounding technique for semidefinite programs. *J. Algorithms*, 60(1):1–23, 2006. Preliminary version in ICALP'01. [doi:10.1016/j.jalgor.2004.11.003] 3, 4

[22] URIEL FEIGE AND ERAN OFEK: Spectral techniques applied to sparse random graphs. *Random Struct. Algor.*, 27(2):251–275, 2005. [doi:10.1002/rsa.20089] 3

[23] WENCESLAS FERNANDEZ DE LA VEGA AND CLAIRE MATHIEU: Linear programming relaxations of maxcut. In *Proc. 18th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'07)*, pp. 53–61. SIAM, 2007. ACM DL. 2, 3, 6

[24] MIROSLAV FIEDLER: Algebraic connectivity of graphs. *Czechoslovak Math. J.*, 23(2):298–305, 1973. [doi:10.21136/CMJ.1973.101168] 3

[25] JOEL FRIEDMAN, ANDREAS GOERDT, AND MICHAEL KRIVELEVICH: Recognizing more unsatisfiable random k-SAT instances efficiently. *SIAM J. Comput.*, 35(2):408–430, 2005. [doi:10.1137/S009753970444096X] 23

[26] JOEL FRIEDMAN, JEFF KAHN, AND ENDRE SZEMERÉDI: On the second eigenvalue of random regular graphs. In *Proc. 21st STOC*, pp. 587–598. ACM Press, 1989. [doi:10.1145/73007.73063] 4

[27] ALAN FRIEZE AND RAVI KANNAN: The regularity lemma and approximation schemes for dense problems. In *Proc. 37th FOCS*, pp. 12–20. IEEE Comp. Soc., 1996. [doi:10.1109/SFCS.1996.548459] 6

[28] MICHEL X. GOEMANS AND DAVID P. WILLIAMSON: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. [doi:10.1145/227683.227684] 3

[29] ANDREAS GOERDT AND MICHAEL KRIVELEVICH: Efficient recognition of random unsatisfiable k-SAT instances by spectral methods. In *Proc. 18th Symp. Theoret. Aspects of Comp. Sci. (STACS'01)*, pp. 294–304. Springer, 2001. [doi:10.1007/3-540-44693-1_26] 23

[30] DIMA GRIGORIEV: Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoret. Comput. Sci.*, 259(1–2):613–622, 2001. [doi:10.1016/S0304-3975(00)00157-2] 5, 7

[31] JOHAN HÅSTAD: An NP-complete problem – some aspects of its solution and some possible applications. Technical Report 16, Inst. Mittag-Leffler, 1984. 3

[32] 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 *Proc. 58th FOCS*, pp. 720–731. IEEE Comp. Soc., 2017. [doi:10.1109/FOCS.2017.72] 7

[33] PRAVESH K. KOTHARI, RAGHU MEKA, AND PRASAD RAGHAVENDRA: Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In *Proc. 49th STOC*, pp. 590–603. ACM Press, 2017. [doi:10.1145/3055399.3055438] 2, 5, 7

[34] PRAVESH K. KOTHARI, RYUHEI MORI, RYAN O'DONNELL, AND DAVID WITMER: Sum of squares lower bounds for refuting any CSP. In *Proc. 49th STOC*, pp. 132–145. ACM Press, 2017. [doi:10.1145/3055399.3055485, arXiv:1701.04521] 5, 7

[35] DAVID A. LEVIN AND YUVAL PERES: Counting walks and graph homomorphisms via Markov chains and importance sampling. *Amer. Math. Monthly*, 124(7):637–641, 2017. [doi:10.4169/amer.math.monthly.124.7.637] 10

[36] BOJAN MOHAR AND SVATOPLUK POLJAK: Eigenvalues in combinatorial optimization. In *Combinatorial and Graph-Theoretical Problems in Linear Algebra*, pp. 107–151. Springer, 1993. [doi:10.1007/978-1-4613-8354-3_5] 3

[37] RYAN O'DONNELL AND TSELIL SCHRAMM: Sherali–Adams strikes back. In *Proc. 34th Comput. Complexity Conf. (CCC'19)*, pp. 8:1–30. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.8] 1

[38] SVATOPLUK POLJAK: Polyhedral and eigenvalue approximations of the max-cut problem. In *Sets, Graphs and Numbers (Budapest, 1991)*, volume 60 of *Colloq. Math. Soc. János Bolyai*, pp. 569–581. North-Holland, 1992. LINK. 2

[39] SVATOPLUK POLJAK AND ZSOLT TUZA: The expected relative error of the polyhedral approximation of the max-cut problem. *Oper. Res. Lett.*, 16(4):191–198, 1994. [doi:10.1016/0167-6377(94)90068-X] 2, 6

[40] PRASAD RAGHAVENDRA, SATISH RAO, AND TSELIL SCHRAMM: Strongly refuting random CSPs below the spectral threshold. In *Proc. 49th STOC*, pp. 121–131. ACM Press, 2017. [doi:10.1145/3055399.3055417] 7

[41] PRASAD RAGHAVENDRA, TSELIL SCHRAMM, AND DAVID STEURER: High-dimensional estimation via sum-of-squares proofs. In *Proc. Internat. Congress of Mathematicians (ICM'18)*, volume 4, pp. 3389–3424. World Scientific, 2019. [doi:10.1142/9789813272880_0186, arXiv:1807.11419] 7

[42] PRASAD RAGHAVENDRA AND NING TAN: Approximating CSPs with global cardinality constraints using SDP hierarchies. In *Proc. 23rd Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'12)*, pp. 373–387. SIAM, 2012. [doi:10.1137/1.9781611973099.33, arXiv:1110.1064] 7

[43] GRANT SCHOENEBECK: Linear level Lasserre lower bounds for certain k-CSPs. In *Proc. 49th FOCS*, pp. 593–602. IEEE Comp. Soc., 2008. [doi:10.1109/FOCS.2008.74] 5, 7

[44] GRANT SCHOENEBECK, LUCA TREVISAN, AND MADHUR TULSIANI: Tight integrality gaps for Lovász–Schrijver LP relaxations of vertex cover and max cut. In *Proc. 39th STOC*, pp. 302–310. ACM Press, 2007. [doi:10.1145/1250790.1250836] 2, 3, 6

[45] HANIF D. SHERALI AND WARREN P. ADAMS: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discr. Math.*, 3(3):411–430, 1990. [doi:10.1137/0403036] 2

[46] KONSTANTIN TIKHOMIROV AND PIERRE YOUSSEF: The spectral gap of dense random regular graphs. *Ann. Probab.*, 47(1):362–419, 2019. [doi:10.1214/18-AOP1263, arXiv:1610.01765] 3, 4

[47] LUCA TREVISAN: Max cut and the smallest eigenvalue. *SIAM J. Comput.*, 41(6):1769–1786, 2012. Preliminary version in STOC'09. [doi:10.1137/090773714] 3

[48] JOEL A. TROPP: An Introduction to Matrix Concentration Inequalities. *Foundations and Trends® in Machine Learning*, 8(1–2):1–230, 2015. [doi:10.1561/2200000048, arXiv:1501.01571] 25

[49] URI ZWICK: Outward rotations: A tool for rounding solutions of semidefinite programming relaxations, with applications to max cut and other problems. In *Proc. 31st STOC*, pp. 679–687. ACM Press, 1999. [doi:10.1145/301250.301431] 3, 4

AUTHORS

Ryan O'Donnell
Professor
Department of Computer Science
Carnegie Mellon University
odonnell@cs.cmu.edu.

Tselil Schramm
Assistant professor
Department of Statistics
Stanford University
schramm@stanford.edu

ABOUT THE AUTHORS

RYAN O'DONNELL received a B. Sc. from the University of Toronto in 1999 and a Ph. D. from the MIT Mathematics Department in 2003. His Ph. D. advisor was Madhu Sudan. Following this he was a postdoc at IAS for a year in Avi Wigderson's group, and a postdoc at Microsoft Research for two years in Jennifer Chayes's group. Since 2006 he has been a professor in the Computer Science Department at Carnegie Mellon University. Ryan's research interests include Analysis of Boolean Functions, Constraint Satisfaction Problems, Quantum Complexity, and Probability. He enjoys his spare time.

TSELIL SCHRAMM received a B. Sc. from Harvey Mudd College in 2012 and a Ph. D. from the UC Berkeley Computer Science Department in 2017, co-advised by Prasad Raghavendra and Satish Rao. Following this she spent two years as a postdoc at Harvard and MIT, hosted by Boaz Barak, Jonathan Kelner, Ankur Moitra, and Pablo Parrilo. In January of 2021 she began an appointment as an assistant professor in the Stanford University Statistics Department. Her research interests include average-case complexity, algorithms, and semidefinite programming hierarchies.