

Sherali–Adams Strikes Back

Ryan O’Donnell

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
<https://www.cs.cmu.edu/~odonnell/>
odonnell@cs.cmu.edu

Tselil Schramm

Harvard University, Cambridge, MA, USA
Massachusetts Institute of Technology, Cambridge, MA, USA
<https://tselilschramm.org/>
tselil@seas.harvard.edu

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 $\exp(c \frac{\log n}{\log \Delta})$ -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 $n \cdot \text{polylog}(n)$ 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.

2012 ACM Subject Classification Theory of computation → Linear programming; Theory of computation → Convex optimization; Theory of computation → Network optimization

Keywords and phrases Linear programming, Sherali–Adams, max-cut, graph eigenvalues, Sum-of-Squares

Digital Object Identifier 10.4230/LIPIcs.CCC.2019.8

Related Version A version of the paper is available at <https://arxiv.org/abs/1812.09967>.

Funding *Ryan O’Donnell*: 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).

Tselil Schramm: This work was supported by NSF grants CCF 1565264 and CNS 1618026.

Acknowledgements We thank Luca Trevisan for helpful comments, 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 Matemática 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.



© Ryan O’Donnell and Tselil Schramm;
licensed under Creative Commons License CC-BY
34th Computational Complexity Conference (CCC 2019).

Editor: Amir Shpilka; Article No. 8; pp. 8:1–8:30



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



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:

$$\begin{aligned} \text{MAX-CUT}(G) \leq \max \quad & \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{|E|} \sum_{uv \in E} Y_{uv} \\ \text{s.t.} \quad & -1 \leq Y_{uv} \leq 1 && \text{(for all } u, v \in V) \\ & -Y_{uv} - Y_{vw} - Y_{wu} \leq 1 && \text{(for all } u, v, w \in V) \\ & -Y_{uv} + Y_{vw} + Y_{wu} \leq 1 && \text{(for all } u, v, w \in V) \end{aligned}$$

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 $Y_{uv} \equiv -1/3$ is always a valid solution. Worse, there are graphs G with $\text{MAX-CUT}(G)$ 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 for overcoming 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 [20, 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 [13, 35] has shown that for any CSP (such as MAX-CUT), the R -round Sherali–Adams LP relaxation achieves essentially 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 [18, 44] concluding in a result of Charikar–Makarychev–Makarychev [15], it was demonstrated that for any constant $\epsilon > 0$, there are graphs (random Δ -regular ones, $\Delta = O(1)$) for which the $n^{\Omega(1)}$ -round Sherali–Adams LP has a MAX-CUT integrality gap of $2 - \epsilon$. 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 $1/2$: *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

$$\text{MAX-CUT}(G) \leq \frac{|V|}{4|E|} \lambda_{\max}(D - A) \quad (1)$$

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

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 Fiedler and Donath–Hoffman [25, 21] among others (Inequality (1) is specifically from Mohar–Poljak [37]). It was recognized early on that eigenvalue methods work particularly well for solving planted-random instances (e.g., of 2-XOR [32] and MIN-BISECTION [11]) and for certifying MAX-CUT values near $1/2$ for truly random instances. Indeed, as soon as one knows (as we now do [46, 24]) 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 Inequality (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 Inequality (1) above to

$$\text{MAX-CUT}(G) \leq \min_{\substack{U \text{ diagonal} \\ \text{tr}(U)=0}} \frac{|V|}{4|E|} \lambda_{\max}(D - A + U) \quad (3)$$

(as suggested by Delorme and Poljak [19], following [21, 11]), one obtains the polynomial-time computable *semidefinite programming* (SDP) bound. Goemans and Williamson [29] showed this bound has integrality ratio less than $1.14 \approx 1/.88$ for *worst-case* G , and it was subsequently shown [50, 23, 14] 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 [18, 44]¹ 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 One would also add the subsequently written [15] here.

1.1 Our results

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

► **Theorem 1** (Simplified version of Theorem 30 and Corollary 31). *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 $1 \leq c \leq \Omega(\log(1/\rho))$, Sherali–Adams with $n^{O(c/\log(1/\rho))}$ rounds certifies that $\text{MAX-CUT}(G) \leq 1/2 + 2^{-c}$.*

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 $\text{MAX-CUT}(G) \leq 50.1\%$ 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.

► **Remark 2.** We wish to emphasize that it is not the resulting $n^{O(1)}$ running time that is surprising; one can (and should) already achieve this with the eigenvalue bound. What is surprising is the inherent power of the Sherali–Adams relaxation itself.

► **Remark 3.** One might ask whether Theorem 1 even requires the assumption of small eigenvalues. That is, *perhaps $n^{o(1)}$ -round Sherali–Adams can certify $\text{MAX-CUT} \leq 1/2 + o(1)$ whenever this is true.* We speculate that this may in fact be the case. As mentioned earlier, the basic SDP relaxation Inequality (3) has this property [50, 23, 14], meaning that whenever graph G has $\text{MAX-CUT}(G) \leq 1/2 + o(1)$, there is a traceless diagonal matrix U such that the eigenvalue bound applied to $A - U$ certifies the maxcut bound. It seems possible that our proof might be adapted to work with this $A - U$ rather than A , in which case Sherali–Adams would also have the property.

We add that the plain eigenvalue bound does *not* have this property: there exist graphs with large (nontrivial) eigenvalues even though the maximum cut is close to $1/2$.³

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 [27, 12, 17], the following was recently shown:

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

Combining the above with Theorem 1, we have the following consequence for MAX-CUT on random regular graphs:

► **Corollary 4.** *Let n , $3 \leq \Delta \leq n/2$, and $1 \leq c \leq \Omega(\log \Delta)$ 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 $\text{MAX-CUT}(G) \leq \frac{1}{2} + 2^{-c}$.*

² Actually, there is a slight mismatch between our result and Inequality (2): in Theorem 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 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.

For example, if $\Delta \geq C \cdot 10^6$ (for C the constant in the bound on $\lambda(G)$), then $n^{1/3}$ -rounds of Sherali–Adams can certify $\text{MAX-CUT}(G) \leq .51$. This result serves as a partial converse to [15]:

► **Theorem** ([15, Theorem 5.3]). *For every fixed integer $\Delta \geq 3$, with high probability over the choice of an n -vertex Δ -regular random graph G ,⁴ the $n^{\Theta(1/f(\Delta))}$ -round Sherali–Adams relaxation for MAX-CUT has value at least $\text{MAX-CUT}(G) \geq 0.99$, where $f(\Delta)$ is a function that grows with Δ .*

While [15] show that Δ -regular random graphs require Sherali–Adams (and by [35], 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 [15], by tracing back through citations (e.g. [3, 4, 16]) 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 bounded by .99 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 5** (Simplified version of Theorem 38). *Suppose that $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ is a k -ary Boolean predicate, and that $\delta, \epsilon > 0$. 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 $m \geq n^{\lceil k/2 \rceil + \delta}$ expected clauses, with high probability Sherali–Adams can certify that $\text{OBJ}_{\mathcal{I}}(x) \leq \mathbf{E}[P] + \epsilon$ using $R = O_{\epsilon, \delta, k}(1)$ rounds.*

This almost matches the comparable performance of Sum-of-Squares (SOS) and spectral algorithms [2], which are known to require $m \geq n^{k/2}$ clauses to certify comparable bounds in polynomial time [31, 43, 34].⁵ Prior to our work it was known that Sherali–Adams admits *weak* refutations (i.e. a certificate that $\text{OBJ} \leq 1 - o(1)$) when $m \geq n^{k/2}$, but it was conceivable (and even conjectured) that $O(1)$ -rounds could not certify $\text{OBJ} \leq 1 - \delta$ 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 .⁶

⁴ In [15], 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 $o(1)$.

⁵ 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 .

⁶ 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.

1.2 Prior work

It is a folklore result that in random graphs with average degree n^δ , 3-round Sherali–Adams 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 $O(\frac{1}{\delta}) \geq 3$, 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 are some prior works examining the performance of Sherali–Adams hierarchies on random (and otherwise “locally dense”) CSPs. Building on works showing a PTAS for fully dense MAX-CUT [28], the work of de la Vega and Mathieu [18] shows that in graphs with average degree $\Omega(n)$, Sherali–Adams with $O(1)$ rounds certifies tight bounds on MAX-CUT. Subsequent works extended this to give a density/rounds tradeoff [49, 6]; the best of these shows that Sherali–Adams accurately estimates the MAX-CUT in graphs of average degree Δ using $O(n/\Delta)$ rounds. One may compare this to our theorem, which uses $n^{O(1/\log \Delta)}$ rounds for random graphs of average degree Δ .

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, 18, 44], and culminating in the already mentioned result of Charikar, Makarychev and Makarychev; in [15], 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 could be strengthened to prove that sparse random graphs require almost-exponential-sized LPs to refute MAX-CUT; our result demonstrates instead that [15] 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 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 are many prior works 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 works [31, 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,⁷ $\omega(1)$ 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 [31, 43, 34] apply; our work can be seen to demonstrate that Sherali–Adams 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 [22, 7], and also no better than a more primitive class of estimation methods based on local statistics (see

⁷ This is more generally true for any CSP that supports a k -wise independent distribution over satisfying assignments.

e.g. [41] for a discussion). Similar results hold for problems relating to estimating the norms of random tensors [33]. Further, in [33] 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 works in extended formulations. The sequence of works [13, 35] show that Sherali–Adams lower bounds for CSPs imply lower bounds for any LP relaxation; the stronger (and later) statement is due to [35], who show that subexponential-round integrality gaps for CSPs in the Sherali–Adams hierarchy imply subexponential-size lower bounds for any LP. These works are then applied in conjunction with [31, 43, 15] 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 [15] 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 whose adjacency matrix has a small spectral radius (excluding the trivial eigenvalue). This implies that random walks mix rapidly, in say t steps, to a 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

$$\text{MAX-CUT}(G) = \frac{1}{2} - \frac{1}{2} \mathbf{E}_{(u,v) \in E(G)} \tilde{\mathbf{E}}[x_u x_v], \tag{4}$$

where $\tilde{\mathbf{E}}[x_u x_v]$ is the “pseudo-correlation” of variables x_u, x_v . Because there is no globally consistent assignment, the pseudo-correlation $\tilde{\mathbf{E}}[x_u x_v]$ for vertices u, v sampled uniformly at random will be close to 0.⁸ But in any fixed subgraph of size $\Omega(t)$, enforcing $\tilde{\mathbf{E}}[x_u x_v] \approx 0$ for pairs u, v at distance t has consequences, and limits the magnitude of correlation between pairs of adjacent vertices as well. In particular, because the pseudo-second moment matrix $\tilde{\mathbf{E}}[x_S x_S^\top]$ for x_S the restriction of x to a set S of up to R vertices 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 $k \leq R - 1$ children $U = \{u_1, \dots, u_k\}$, and call $X = \tilde{\mathbf{E}}[x_S x_S^\top] \succeq 0$. Notice that pairs of distinct children u_i, u_j are at distance $t = 2$ in S . If we then require $\tilde{\mathbf{E}}[x_{u_i} x_{u_j}] = 0$ for every $u_i \neq u_j$, the only nonzero entries of X are the diagonals (which are all $\tilde{\mathbf{E}}[x_i^2] = 1$), and the entries corresponding to edges from the root to its children, (r, u_i) , which are $\tilde{\mathbf{E}}[x_r x_{u_i}]$. Now defining the vector $c \in \mathbb{R}^S$ with a 1 at the root r , $c_r = 1$ and α on each child $u \in U$, $c_u = \alpha$, we have from the PSDness of X that

$$0 \leq c^\top X c = \|c\|_2^2 + \sum_{u \in U} 2c_r c_u \cdot \tilde{\mathbf{E}}[x_r x_u] = (1 + \alpha^2 k) + 2\alpha k \mathbf{E}_{(u,v) \in E(S)} \tilde{\mathbf{E}}[x_u x_v].$$

⁸ 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.

Choosing $\alpha = k^{-1/2}$, this implies that within S , the average edge correlation is lower bounded by $\mathbf{E}_{(u,v) \in E[S]} \tilde{\mathbf{E}}[x_u x_v] \geq -k^{-1/2}$. Of course, for a given star S we cannot know that $\tilde{\mathbf{E}}[x_{u_i} x_{u_j}] = 0$, but if we take a well-chosen weighted average over *all* stars, this will (approximately) hold on average.

Our strategy is to take a carefully-chosen average over specific subgraphs S of G with $|S| = \Omega(t)$. 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 Equation (4). 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.

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 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 6.** Let $G = (V, E)$ be an undirected finite graph, with parallel edges and self-loops allowed⁹, 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 $\mathbf{v} \sim u$ to denote that \mathbf{v} is the result of taking one random step from u .

► **Definition 7.** 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 u th row by a factor of $1/\deg(u)$.

► **Definition 8.** 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 $\mathbf{u} \sim \pi$ and $\mathbf{v} \sim \mathbf{u}$, the distribution of (\mathbf{u}, \mathbf{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 9.** For $f, g : V \rightarrow \mathbb{R}$ we use the notation $\langle f, g \rangle_\pi$ for $\mathbf{E}_{\mathbf{u} \sim \pi} [f(\mathbf{u})g(\mathbf{u})]$. 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

$$\langle f, Kg \rangle_\pi = \langle Kf, g \rangle_\pi = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [f(\mathbf{u})g(\mathbf{v})]. \quad (5)$$

⁹ Self-loops count as “half an edge”, and contribute 1 to a vertex’s degree.

► **Definition 10.** A stationary d -step walk is defined to be a sequence $(\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_d)$ formed by choosing an initial vertex $\mathbf{u}_0 \sim \pi$, and then taking a standard random walk, with $\mathbf{u}_t \sim \mathbf{u}_{t-1}$. Generalizing Equation (5), it holds in this case that

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

2.2 Tree-indexed random walks

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

► **Definition 11.** Suppose we have a finite undirected tree with vertex set T . A stationary T -indexed random walk in G is a random homomorphism $\phi : T \rightarrow V$ defined as follows: First, root the tree at an arbitrary vertex $i_0 \in T$. Next, define $\phi(i_0) \sim \pi$. Then, independently for each “child” j of i_0 in the tree, define $\phi(j) \sim \phi(i_0)$; that is, define $\phi(j) \in V$ to be the result of taking a random walk step from $\phi(i_0)$. Recursively repeat this process for all children of i_0 's children, etc., until each vertex $k \in T$ has been assigned a vertex $\phi(k) \in V$.

We note that the homomorphism ϕ 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., [36].

► **Fact 12.** The definition of ϕ 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 $i = i_0, i_1, \dots, i_d = j$ is the unique path in the tree between them, then the sequence $(\phi(i_0), \phi(i_1), \dots, \phi(i_d))$ 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 13.** We assume that for each vertex pair (u, v) where G has edge, there is an associated sign $\xi_{uv} = \xi_{vu} \in \{\pm 1\}$.¹⁰ 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 14.** We write $\bar{K} = \Xi \circ K$ for the signed transition operator. The operator \bar{K} is self-adjoint with respect to $\langle \cdot, \cdot \rangle_\pi$, and hence has real eigenvalues. It also holds that

$$\langle f, \bar{K}g \rangle_\pi = \langle \bar{K}f, g \rangle_\pi = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [\xi_{\mathbf{u}\mathbf{v}} f(\mathbf{u})g(\mathbf{v})]. \quad (6)$$

► **Definition 15.** 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 $f : V \rightarrow \{\pm 1\}$ 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

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[\frac{1}{2} + \frac{1}{2} \xi_{\mathbf{u}\mathbf{v}} f(\mathbf{u})f(\mathbf{v}) \right] = \frac{1}{2} + \frac{1}{2} \langle f, \bar{K}f \rangle_\pi. \quad (7)$$

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

¹⁰If G has multiple (u, v) edges, we think of them as all having the same sign.

► **Definition 16.** We say that a signed stationary d -step walk is a sequence of pairs $(\mathbf{u}_t, \sigma_t) \in \{\pm 1\} \times V$ for $0 \leq t \leq d$, chosen as follows: first, we choose a stationary d -step walk $(\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_d)$ in G ; second, we choose $\sigma_0 \in \{\pm 1\}$ uniformly at random; finally, we define $\sigma_t = \sigma_{t-1} \xi_{\sigma_{t-1} \sigma_t}$. Generalizing Equation (6), it holds in this case that

$$\mathbf{E}[\sigma_0 f(\mathbf{u}_0) \sigma_d g(\mathbf{u}_d)] = \langle f, \overline{K}^d g \rangle_\pi.$$

► **Definition 17.** We extend the notion from Definition 11 to that of a signed stationary T -indexed random walk in G . Together with the random homomorphism $\phi : T \rightarrow V$, we also choose a random signing $\sigma : T \rightarrow \{\pm 1\}$ as follows: for the root i_0 , the sign $\sigma(i_0) \in \{\pm 1\}$ is chosen uniformly at random; then, all other signs are deterministically chosen – for each j of i_0 we set $\sigma(j) = \xi_{i_0 j} \sigma(i_0)$, and in general $\sigma(k) = \xi_{k' k} \sigma(k')$ where k' is the parent of k . Again, it is not hard to show that the definition of (ϕ, σ) 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 $(\phi(i_0), \sigma(i_0)), (\phi(i_1), \sigma(i_1)), \dots, (\phi(i_d), \sigma(i_d)))$ 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, although our results also apply to a weaker proof system (see Remark 24). We first define Sherali–Adams in this “proof system” format (as opposed to the original optimization format); see, e.g., [9] for some commentary on this perspective.

► **Definition 18.** Let X_1, \dots, X_n be indeterminates that are supposed to stand for real numbers ± 1 . The R -round Sherali–Adams proof system [45] may be defined as follows: The “lines” of the proof are real polynomial inequalities in X_1, \dots, X_n (where the polynomials may as well be multilinear). The allowed “axioms” are any real inequalities of the form $q(X_{u_1}, \dots, X_{u_R}) \geq 0$, where the inequality is true for every ± 1 assignment to the indeterminates X_{u_i} . The “deduction rules” allow one to derive any nonnegative linear combination of previous lines. This is a sound proof system for inequalities about ± 1 numbers X_1, \dots, X_n .

► **Fact 19.** There is a $\text{poly}(n^R, L)$ -time algorithm based on Linear Programming for determining whether a given polynomial inequality $p(\mathbf{X}) \geq 0$ of degree at most R (and rational coefficients of total bit-complexity L) is derivable in the R -round Sherali–Adams proof system.

As mentioned earlier, it will be helpful for us to take a “Sum-of-Squares” perspective on Sherali–Adams. The well-known fact here is that a multilinear polynomial $q(X_{u_1}, \dots, X_{u_R})$ is nonnegative for all ± 1 assignments if and only if it can be represented as the (multilinearization of) a squared polynomial p^2 on R indeterminates. (This p will be the unique “Fourier expansion” for the function $\sqrt{q} : \{\pm 1\}^R \rightarrow \mathbb{R}$; again, see [9] for some discussion.) Let us now define a proof system that can encapsulate both Sherali–Adams and SOS:

► **Definition 20.** We define the R -local, degree- D (static) Sum-of-Squares (SOS) proof system over indeterminates X_1, \dots, X_n as follows. The “lines” of the proof are real polynomial inequalities in $\mathbf{X} = (X_1, \dots, X_n)$. The default “axioms” are any real inequalities of the form $p(X_{u_1}, \dots, X_{u_R})^2 \geq 0$, 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 X_1, \dots, X_n .

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

► **Fact 21.** *The case of $R = \infty$ (equivalently, $R = n$) corresponds to the well-known degree- D SOS proof system.*

► **Definition 22.** *Suppose one includes the Boolean equalities, meaning $X_u^2 - 1 = 0$ for all $1 \leq i \leq n$.¹¹ In this case $D = \infty$ is equivalent to $D = R$, and the corresponding proof system is equivalent to R -round Sherali–Adams.*

► **Fact 23.** *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*

$$x^\top P x \geq 0, \tag{8}$$

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

► **Remark 24.** In fact, we will often be concerned with the R -round, degree-2 Sherali–Adams proof system, which is strictly weaker than the general R -round Sherali–Adams proof system. Despite this restriction to $D = 2$, we only know the $\text{poly}(n^R, L)$ -time algorithm for deciding derivability of a given quadratic polynomial $p(X) \geq 0$ (of bit-complexity L).

3 Proof of Main Theorem (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.

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 $c^\top X|_S X|_S^\top c$, where $X|_S$ is the restriction of X to a set S of variables with $|S| \leq R$ 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 25.** *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 .*

¹¹ Or alternatively, $X_u^2 - X_u = 0$ for all i .

8:12 Sherali–Adams Strikes Back

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 26.** For $0 \leq d \leq 2\ell$, we define the matrix $A^{(d)} \in \mathbb{R}^{S \times S}$ to be the “distance- d ” adjacency matrix of the spider; i.e., $A_{ij}^{(d)}$ is 1 if $\text{dist}_S(i, j) = d$ 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 3.2:

► **Theorem 27.** 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:

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

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

► **Corollary 28.** Assuming that $k \geq 3^\ell$ and taking $\alpha = k^{1/(2\ell)}$, the PSD matrix Ψ satisfies the following four statements:

$$\begin{aligned} 3/2 \leq \langle \Psi, A^{(0)} \rangle \leq 2, & \quad \langle \Psi, A^{(1)} \rangle = k^{1/(2\ell)}, \\ \langle \Psi, A^{(d)} \rangle = 0 \text{ for } 1 < d < 2\ell, & \quad \langle \Psi, A^{(2\ell)} \rangle = \frac{1}{2}(k-1). \end{aligned}$$

We will also use the following small technical lemma:

► **Lemma 29.** 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

$$\mathbf{E}_{\mathbf{u} \sim \pi} \sum_{v \in V} M_{uv} X_u X_v \leq \pi_*^{-1/2} \|M\|_2 \mathbf{E}_{\mathbf{u} \sim \pi} X_u^2.$$

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

$$M_{uv} X_u X_v \leq \frac{M_{uv}^2}{2\gamma\pi(v)} X_u^2 + \frac{\gamma\pi(v)}{2} X_v^2.$$

Thus it can derive

$$\mathbf{E}_{\mathbf{u} \sim \pi} \sum_{v \in V} M_{uv} X_u X_v \leq \mathbf{E}_{\mathbf{u} \sim \pi} X_u^2 \sum_{v \in V} \frac{M_{uv}^2}{2\gamma\pi(v)} + \frac{\gamma}{2} \mathbf{E}_{\mathbf{v} \sim \pi} X_v^2. \quad (9)$$

We’ll take $\gamma = \pi_*^{-1/2} \|M\|_2$. Since we can certainly derive $aX_u^2 \leq bX_v^2$ whenever $a \leq b$, we see that it suffices to establish

$$\sum_{v \in V} \frac{M_{uv}^2}{2\gamma\pi(v)} \leq \frac{\gamma}{2}$$

for every outcome of \mathbf{u} . But this is implied by $\sum_v M_{\mathbf{u}v}^2 \leq (\pi(v)/\pi_*)\|M\|_2^2$ for all $v \in V$, which is indeed true. \blacktriangleleft

We can now prove the following main theorem:

► **Theorem 30.** *Given parameters $k \geq 3^\ell$, let $R = k\ell + 1$ and define*

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

where $\rho(\overline{K})$ denotes the spectral radius of the signed transition matrix \overline{K} . Then R -local, degree-2 SOS can deduce the bound “ $\rho(\overline{K}) \leq \beta$ ”; more precisely, it can deduce the two inequalities

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

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

► **Corollary 31.** *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 $\epsilon > \min(\pi_*^{-1/2}, \rho(\overline{K}))$, take $\ell = \lceil \frac{1}{4} \frac{\log(\epsilon^2 \pi_*)}{\log(\rho(\overline{K})/\epsilon)} \rceil$, and take $k = \lceil (\frac{1}{\epsilon})^{2\ell} \rceil$. Then for $R = k\ell + 1$, it holds that R -local degree-2 SOS can deduce the bound $\rho(\overline{K}) \leq \frac{5}{2}\epsilon$. In particular, if we think of G, Ξ as a 2-XOR CSP, it holds that R -round Sherali–Adams can deduce the bound $\text{OBJ} \leq \frac{1}{2} + \frac{5}{4}\epsilon$.*

Proof. Taking the parameters as above, and using that the constraints $X_u^2 = 1$ imply that R -round Sherali–Adams can deduce that $\langle \mathbf{X}, \mathbf{X} \rangle_\pi = 1$ whenever $R \geq 2$, and that as noted in Equation (7), $\text{OBJ}(\mathbf{X}) = \frac{1}{2} + \frac{1}{2}\langle \mathbf{X}, \overline{K}\mathbf{X} \rangle_\pi$, so Theorem 30 gives the result. \blacktriangleleft

Corollary 31 implies the 2-XOR version of Theorem 1 since in simple graphs, $\log \frac{1}{\pi_*} = \Theta(\log n)$.

Proof of Theorem 30. For our (k, ℓ) -spider graph on S , let (ϕ, σ) be a signed stationary S -indexed random walk in G . Define $\overline{\mathbf{x}}$ to be the S -indexed vector with $\overline{\mathbf{x}}_i = \sigma(i)X_{\phi(i)}$. Then letting Ψ be the PSD matrix from Corollary 28, the R -local, degree-2 SOS proof system can derive

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

(This is in the form of Inequality (8) if we take $P = \text{diag}(\sigma)\Psi \text{diag}(\sigma)$.) Furthermore, the proof system can deduce this inequality in expectation; namely,

$$\langle \Psi, \mathbf{Y} \rangle \geq 0, \text{ where } \mathbf{Y} = \mathbf{E}[\overline{\mathbf{x}}\overline{\mathbf{x}}^\top]. \quad (10)$$

Now by the discussion in Definitions 16 and 17,

$$\mathbf{Y}_{ij} = \mathbf{E}[\sigma(i)X_{\phi(i)}\sigma(j)X_{\phi(j)}] = \langle \mathbf{X}, \overline{K}^{\text{dist}_S(i,j)}\mathbf{X} \rangle_\pi. \quad (11)$$

Thus recalling the notation $A^{(d)}$ from Definition 26,

$$\mathbf{Y} = \sum_{d=0}^{2\ell} \langle \mathbf{X}, \overline{K}^d \mathbf{X} \rangle_\pi A^{(d)}, \quad (12)$$

8:14 Sherali–Adams Strikes Back

and hence from Inequality (10) we get that R -local, degree-2 SOS can deduce

$$0 \leq \sum_{d=0}^{2\ell} \langle \Psi, A^{(d)} \rangle \langle \mathbf{X}, \overline{K}^d \mathbf{X} \rangle_\pi = c_0 \langle \mathbf{X}, \mathbf{X} \rangle_\pi + k^{1/(2\ell)} \langle \mathbf{X}, \overline{K} \mathbf{X} \rangle_\pi + \frac{1}{2}(k-1) \langle \mathbf{X}, \overline{K}^{2\ell} \mathbf{X} \rangle_\pi, \quad (13)$$

for some constant $3/2 \leq c_0 \leq 2$ (here we used Corollary 28). Regarding the last term, we have:

$$\langle \mathbf{X}, \overline{K}^{2\ell} \mathbf{X} \rangle_\pi = \mathbf{E}_{\mathbf{u} \sim \pi} \sum_{v \in V} (\overline{K}^{2\ell})_{uv} \mathbf{X}_u \mathbf{X}_v. \quad (14)$$

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

$$\begin{aligned} \mathbf{E}_{\mathbf{u} \sim \pi} \sum_{v \in V} (\overline{K}^{2\ell})_{uv} \mathbf{X}_u \mathbf{X}_v &\leq \mathbf{E}_{\mathbf{u} \sim \pi} \sum_{v \in V} \left| (\overline{K}^{2\ell})_{uv} \right| \\ &\leq \sqrt{|V|} \mathbf{E}_{\mathbf{u} \sim \pi} \|\overline{K}_{\mathbf{u}, \cdot}^{2\ell}\|_2 \leq \sqrt{|V|} \rho(\overline{K}^{2\ell}) = \sqrt{|V|} \rho(\overline{K})^{2\ell}, \end{aligned}$$

and later combine this with $c_0 \langle \mathbf{X}, \mathbf{X} \rangle_\pi = c_0$. But proceeding more generally, we instead use Lemma 29 to show that our proof system can derive

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

Putting this into Equation (14) and Inequality (13) we get

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

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

3.1 Max-Cut

The following theorem is quite similar to Theorem 30. 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 Corollary 31.

► **Theorem 32.** *Given transition operator K for the standard random walk on G , let $K' = K - J$, where J is the all-1's matrix. For parameters $k \geq 3^\ell$, let $R = k\ell + 1$ and define*

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

(Note that $\rho(K')$ is equal to maximum-magnitude eigenvalue of K when the trivial 1 eigenvalue is excluded.) Then $2R$ -local, degree-2 SOS can deduce the bound “ $\lambda_{\min}(K) \geq -\beta$ ”; more precisely, it can deduce the inequality

$$\langle \mathbf{X}, \overline{K} \mathbf{X} \rangle_\pi \geq -\beta \langle \mathbf{X}, \mathbf{X} \rangle_\pi.$$

► **Corollary 33.** *Suppose we have a graph $G = (V, E)$ with transition operator K and centered transition operator $K' = K - J$, and $\pi_* = \min_{v \in V} \frac{\deg(v)}{2|E|}$. Given $\epsilon > \min(\pi_*^{-1/2}, \rho(K'))$, take $\ell = \left\lceil \frac{1}{4} \frac{\log(\epsilon^2 \pi_*)}{\log(\rho(\overline{K})/\epsilon)} \right\rceil$, and take $k = \lceil (\frac{1}{\epsilon})^{2\ell} \rceil$. Then for $R = k\ell + 1$, it holds that $2R$ -local degree-2 SOS can deduce the bound $\rho(K') \leq \frac{5}{2}\epsilon$. In particular, if we think of G as a MAX-CUT CSP, it holds that R -round Sherali–Adams can deduce the bound $\text{OBJ} \leq \frac{1}{2} + \frac{5}{4}\epsilon$.*

Again, Corollary 33 implies Theorem 1 since in simple graphs, $\log \frac{1}{\pi^*} = \Theta(\log n)$.

Proof of Theorem 32. The proof is a modification of the proof of Theorem 30. 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, ϕ_1 and ϕ_2 . For $j \in \{1, 2\}$, define \mathbf{x}_j to be the S -indexed vector with i th coordinate equal to $X_{\phi_j(i)}$, and write $\dot{\mathbf{x}}$ for the concatenated vector $(\mathbf{x}_1, \mathbf{x}_2)$. Also, for $0 < \theta < 1$ a parameter¹² slightly less than 1, let Ψ be the PSD matrix from Corollary 28, and define the PSD block-matrix

$$\dot{\Psi} = \frac{1}{2} \begin{pmatrix} \frac{1}{\theta} \Psi & -\Psi \\ -\Psi & \theta \Psi \end{pmatrix}.$$

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

$$0 \leq \langle \dot{\Psi}, \mathbf{E} \dot{\mathbf{x}} \dot{\mathbf{x}}^\top \rangle = \iota \langle \Psi, \mathbf{Y} \rangle - \langle \Psi, \mathbf{Z} \rangle, \quad \text{where } \iota = \frac{1/\theta + \theta}{2}, \quad \mathbf{Y} = \mathbf{E}[\mathbf{x} \mathbf{x}^\top], \quad \mathbf{Z} = \mathbf{E}[\mathbf{x}_1 \mathbf{x}_2^\top], \quad (15)$$

and \mathbf{x} (which will play the role of $\bar{\mathbf{x}}$) denotes the common distribution of \mathbf{x}_1 and \mathbf{x}_2 . Similar to Equations (11) and (12), we now have

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

and by independence of \mathbf{x}_1 and \mathbf{x}_2 we have

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

Thus applying Corollary 28 to Inequality (15), our proof system can derive

$$0 \leq \iota \cdot \left(c_0 \langle \mathbf{X}, \mathbf{X} \rangle_\pi + k^{1/(2\ell)} \langle \mathbf{X}, K \mathbf{X} \rangle_\pi + \frac{1}{2}(k-1) \langle \mathbf{X}, K^{2\ell} \mathbf{X} \rangle_\pi \right) - \mathbf{E}_{\mathbf{u} \sim \pi} [\mathbf{X}_{\mathbf{u}}]^2 \cdot \left(c_0 + k^{1/(2\ell)} + \frac{1}{2}(k-1) \right). \quad (16)$$

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 (16) and then dividing through by ι , we conclude that the proof system can derive

$$0 \leq c_0 \langle \mathbf{X}, \mathbf{X} \rangle_\pi + k^{1/(2\ell)} \langle \mathbf{X}, K \mathbf{X} \rangle_\pi + \frac{1}{2}(k-1) (\langle \mathbf{X}, K^{2\ell} \mathbf{X} \rangle_\pi - \langle 1, X \rangle_\pi^2),$$

cf. Inequality (13). Recalling now that K has the constantly-1 function as an eigenvector, with eigenvalue 1, we have the identity

$$\langle \mathbf{X}, K^{2\ell} \mathbf{X} \rangle_\pi - \langle 1, X \rangle_\pi^2 = \langle \mathbf{X}, (K - J)^{2\ell} \mathbf{X} \rangle_\pi.$$

Now the remainder of the proof is just as in Theorem 27, with $K - J$ in place of \bar{K} , except we do not have the step of repeating the derivation with $-\bar{K}$ in place of \bar{K} . ◀

¹²This parameter is introduced to fix a small annoyance; the reader might like to imagine $\theta = 1$ at first.

3.2 A technical construction of coefficients on the spider

Proof of Theorem 27. 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 $1 \leq t \leq \ell$). We will be considering vectors in \mathbb{R}^S , with coordinates indexed by the vertex set S . For $0 \leq t \leq \ell$ define the vector

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

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

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

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

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

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

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

To do this, we observe that

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

and

$$\begin{aligned} \mu_0^\top A^{(d)} \mu_t &= \mu_t^\top A^{(d)} \mu_0 = \begin{cases} \alpha^t & \text{if } d = t, \\ 0 & \text{else;} \end{cases} \\ \text{and for } s, t > 0, \quad \mu_s^\top A^{(d)} \mu_t &= \begin{cases} (1/k)\alpha^{s+t} & \text{if } d = |s - t|, \\ (1 - 1/k)\alpha^{s+t} & \text{if } d = s + t, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

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

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

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

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

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 [15]. In that work, the authors show that if every subset of r points in a metric 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 [15] (where it is stated in slightly more generality, for the 0/1 version of the cut polytope):

► **Theorem 34** (Theorem 3.1 in [15]). *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:*

$$\tilde{\mathbf{E}}[x_i x_j] = 1 - \frac{2}{\pi} \arccos \left(1 - \frac{1}{2} \rho(i, j)^2 \right).$$

► **Proposition 35.** *In any 2-XOR or MAX-CUT instance, R -rounds of Sherali–Adams cannot certify that*

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

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 $x_i x_j b_{ij} = 1$ for some $b_{ij} \in \{\pm 1\}$. Define the metric space on (X, ρ) as follows: let $X = \{x_1, \dots, x_n\}$ have a point for each vertex of G , and set $\rho(x_i, x_j) = \sqrt{2 \left(1 - b_{ij} \frac{1}{R} \right)}$.

We claim that any $r = 2R + 3$ 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

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

So long as $|S| \leq r$, the matrix $M_S = \mathbf{1} + B_S$ is diagonally dominant, and therefore positive semidefinite, so from the Cholesky decomposition of M_S we assign to each $x_i \in S$ a vector v_i so that $\|v_i\|_2 = 1$, and so that for every pair $x_i, x_j \in S$, $\|v_i - v_j\|^2 = 2 - 2b_{ij} \frac{1}{r} = \rho(i, j)^2$.

Applying Theorem 34, we have that the solution

$$\tilde{\mathbf{E}}[x_i x_j] = 1 - \frac{2}{\pi} \arccos \left(1 - \frac{1}{2} \cdot 2 \left(1 - b_{ij} \frac{1}{r} \right) \right) = 1 - \frac{2}{\pi} \arccos \left(b_{ij} \frac{1}{r} \right)$$

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

8:18 Sherali–Adams Strikes Back

▷ **Claim 36.** The function $f(z) = 1 - \frac{2}{\pi} \arccos(z)$ exhibits the rotational symmetry $f(z) = -f(-z)$, and further $f(z) \geq \frac{2}{\pi}z$ 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

$$b_{ij} \cdot \tilde{\mathbf{E}}[x_i x_j] = b_{ij} \cdot f\left(b_{ij} \frac{1}{r}\right) = f\left(\frac{1}{r}\right),$$

and using that for $z \in [0, 1]$, $f(z) \geq \frac{2}{\pi}z \geq 0$,

$$\geq \frac{2}{\pi} \cdot \frac{1}{r}.$$

We conclude that $R = \frac{1}{2}(r - 3)$ rounds of Sherali–Adams are unable to certify that $\text{OBJ} < \frac{1}{2} + \frac{2}{\pi} \frac{1}{2R+3}$, as desired. ◀

Proof of Claim 36. 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 $[0, \frac{1}{2}]$:

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

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

5 Refutation for any Boolean CSP

In this section, we argue that R -round Sherali–Adams can also refute any non-trivial Boolean CSP. First, for any predicate $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ we define a parameterized distribution over the CSP with constraints from P :

► **Definition 37.** Let $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ 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 ordered multiset $S \subset [n]$ with $|S| = k$, 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 \odot 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 38.** Suppose that $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ and that $\delta, \epsilon > 0$ 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 $\text{OBJ}_{\mathcal{I}}(x) \leq \mathbf{E}[P] + \epsilon$ when $R = O_{\epsilon, \delta, k}(1)$ rounds. More specifically, $R = k\ell \left(\frac{3 \cdot 2^{k/2-1}}{\epsilon} \right)^{2\ell} + k$ for $\ell = \lceil \lceil \frac{k}{2} \rceil \frac{1}{2\delta} \rceil$.

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

► **Definition 39.** We say that a predicate $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ 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 40.** Suppose that $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ is η -far from t -wise supporting, and that $\delta, \epsilon > 0$. 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 $\text{OBJ}_{\mathcal{I}}(x) \leq 1 - \eta + \epsilon$ with $R = O_{\epsilon, \delta, t}(1)$ rounds. More specifically, $R = t\ell \left(\frac{3 \cdot 2^{t/2-1}}{\epsilon} \right)^{2\ell} + t$ for $\ell = \lceil \lceil \frac{t}{2} \rceil \frac{1}{2\delta} \rceil$.

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 $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ 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 41.** 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 ordered multiset $S \subset [n]$ with $|S| = k$, 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 [30, 26] and subsequently honed in many works, 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 42.** Suppose $k > 1$ is an integer and \mathcal{I} is a $2k$ -XOR instance on n variables x_1, \dots, x_n , with objective $\sum_{U \in [n]^{2k}} b_U \cdot x^U$ where the sum is over ordered multisets $U \subset [n]$, $|U| = 2k$. 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 ordered multiset $S \subset [n]$, $|S| = k$, and for each $U \subset [n]$ with $|U| = 2k$, choosing the partition of U into the ordered multisets 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 43.** 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 $\text{OBJ}_{\mathcal{I}_{\text{flat}}}(x) \leq c$, then the $k \cdot R$ -round Sherali–Adams proof system can certify that $\text{OBJ}_{\mathcal{I}}(x) \leq c$.

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. ◀

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

▷ **Claim 44.** Suppose that \mathcal{I} is a random $2k$ -XOR instance on n vertices weighted according to \mathcal{W} . Then the flattening \mathcal{I}_{flat} is a random 2 -XOR instance on n^k vertices weighted according to \mathcal{W} .

Proof. This fact is immediate, since the ordered multisets $U \subset [n]$, $|U| = 2k$ are in bijection with ordered pairs of multisets $S, T \subset [n]$, $|S| = |T| = k$. ◁

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 45.** *Suppose that \mathcal{W} is a distribution over integers which is symmetric about the origin, and let $n, k \geq 1$ be positive integers. 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 $\mathbf{E}|w| \gg \sigma \sqrt{\frac{\log n}{n^k}}$,*
- *With high probability over n^{2k} i.i.d. samples $w_1, \dots, w_{n^{2k}} \sim \mathcal{W}$, $\max_{i \in [n^{2k}]} |w_i| \leq M \ll \sigma^2 n^k$.*

Now, define

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

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 $n^{k/4\ell} \rho \leq \frac{1}{2} \epsilon^{2\ell}$ and setting $R = 2k \cdot \ell \cdot \left(\frac{1}{\epsilon}\right)^{2\ell}$, R rounds of Sherali–Adams can deduce the bound $\text{OBJ}_{\mathcal{I}}(x) \leq \frac{1}{2} + \frac{3}{2}\epsilon$.

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

► **Theorem 46** (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 $\|A_i\| \leq M$ for all $i \in [m]$. Let $A = \sum_{i \in [m]} A_i$ denote their sum, and suppose that $\|\mathbf{E}AA^\top\|, \|\mathbf{E}A^\top A\| \leq \sigma^2$. Then*

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

Proof of Proposition 45. Given a weighted $2k$ -XOR instance on n variables with weights from \mathcal{W} , we consider its flattening \mathcal{I}_{flat} with objective function $\text{OBJ}(x) = \frac{1}{m} \sum_{i,j \in [N]} \frac{1}{2}(1 + b_{ij}x_i x_j)$ 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 Lemma 43 we have that $\frac{1}{2} + \frac{1}{2m}x^\top Bx = \text{OBJ}_{\mathcal{I}}(x)$.

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 $\text{sgn}(B_{ij})$. Let $2 \cdot D$ be the degree matrix of G , let $A = |2B|$ be the adjacency matrix of G , let $\Xi = \text{sgn}(B)$ be the matrix of signs of B , and let $K = (2D^{-1})A = D^{-1}B \otimes \Xi$ be the transition matrix for the random walk on G .

To apply Corollary 31, 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

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

First, we bound $\|B\|_{op}$. Take B' to be the truncated version of B , so that $B'_{i,j} = \text{sgn}(B_{i,j}) \cdot \max(|B_{i,j}|, M)$. 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 46. We have that $\mathbf{E}[B'B'^\top] \preceq n^k \sigma^2 \cdot \mathbf{1}$, so applying Theorem 46 with $t = O(\max(\sqrt{\sigma^2 n^k \log n}, M \log n))$ we get that with high probability,

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

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,

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

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

$$2m = n^{2k} \mathbf{E}|w| \pm O(\sigma n^k \sqrt{\log n}). \quad (17)$$

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

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

so using that $M \leq n^k \sigma^2$ and taking $t = O(\sqrt{n^k \sigma^2 \log n})$ 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,

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

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

Now, the result follows by applying Corollary 31 and Lemma 43. \blacktriangleleft

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 47.** *Suppose $k \geq 1$ 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, $\mathcal{I}_{\text{lift}}$, 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 add modify the objective to $\sum_{U \in [n]^{2k+1}} b_U \cdot x^U \cdot w_{i_U}$

- For each ordered multiset S associate a new variable y_S , and for each ordered multiset $T \in [n]^k$ 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 ordered multiset $V = (U, i_U)$ 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 Lemma 43 for odd k -XOR:

► **Lemma 48.** *Suppose that \mathcal{I} is a weighted $(2k+1)$ -XOR instance, and let \mathcal{I}_{flat} be the bipartite 2-XOR instance given by its flattening. Then if the R -round Sherali–Adams proof system can certify that $\text{OBJ}_{\mathcal{I}_{flat}}(x) \leq c$, then the $(k+1) \cdot R$ -round Sherali–Adams proof system can certify that $\text{OBJ}_{\mathcal{I}}(x) \leq c$.*

Proof. The only modification to the proof of Lemma 48 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 Proposition 45:

► **Proposition 49.** *Suppose that \mathcal{W} is a distribution over integers which is symmetric about the origin, and let $n, k \geq 1$ be integers. 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 $\mathbf{E}|w| \gg \sigma \sqrt{\frac{\log n}{n^k}}$,*
- *With high probability over n^{2k+1} i.i.d. samples $w_1, \dots, w_{n^{2k+1}} \sim \mathcal{W}$, $\max_{i \in [n^{2k+1}]} |w_i| \leq M \ll \sigma^2 n^k$.*

Now, define

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

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 $n^{(k+1)/4\ell} \rho \leq \frac{1}{2} \epsilon^{2\ell}$ and setting $R = (2k+2)\ell \cdot \left(\frac{1}{\epsilon}\right)^{2\ell}$, R rounds of Sherali–Adams can deduce the bound $\text{OBJ}_{\mathcal{I}}(x) \leq \frac{1}{2} + \frac{3}{2}\epsilon$.

Proof. The thread of the proof is the same as that of Proposition 45. We will refute \mathcal{I}_{lift} , since by Lemma 48 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 Corollary 31 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

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

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

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

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 $\mathbf{E} B B^\top \preceq n^k \sigma^2 \mathbf{1}$, 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 46 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,

$$\|K \circ \Xi\| \leq \frac{1}{\deg_{\min}(G)} \leq O \left(\frac{\sigma \log n}{\mathbf{E} |w| \sqrt{n^k}} \right) \cdot \max \left(1, \frac{M}{\sqrt{n^k}} \right), \quad (18)$$

$$\pi^* = \frac{\deg_{\min}(G)}{2m} = \frac{1}{n^{k+1}} \cdot (1 \pm o(1)) \quad (19)$$

After which we can apply Corollary 31. ◀

5.2 From Boolean CSPs to k-XOR

Following [2], we prove Theorem 38 via reduction to XOR.

Proof of Theorem 38. 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 $m = n^{\lfloor k/2 \rfloor + \delta} \pm 10 \sqrt{n^{\lfloor k/2 \rfloor + \delta} \log n}$. Set $\ell = \lceil \frac{1}{2\delta} \rceil$.

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

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

Using this expansion, we re-write the objective function. Recall that $[n]^k$ is the set of all ordered multisets 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 . Then, if the total number of constraints is m ,

$$\begin{aligned} \text{OBJ}_{\mathcal{I}}(x) &= \frac{1}{m} \sum_{S \in [n]^k} b_S \cdot P(x^S \odot \zeta_S) \\ &= \frac{1}{m} \sum_{S=\{i_1, \dots, i_k\} \in [n]^k} \sum_{\alpha \subseteq [k]} b_S \cdot \widehat{P}(\alpha) \prod_{a \in \alpha} x_{i_a} (\zeta_S)_{i_a} \\ &= \frac{1}{m} \sum_{\alpha \subseteq [k]} \widehat{P}(\alpha) \cdot \sum_{T \in [n]^{|\alpha|}} \left(\sum_{S \in [n]^k, S|_{\alpha}=T} b_S \cdot \prod_{a \in \alpha} (\zeta_S)_{i_a} \right) \cdot x^T. \end{aligned} \quad (20)$$

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 (20),

$$\text{OBJ}_{\mathcal{I}}(x) = \sum_{\alpha \subseteq [k]} \widehat{P}(\alpha) \cdot \mathcal{I}^{\alpha}(x). \quad (21)$$

Let $\mathcal{W}_{n^{k-t}}$ be the distribution defined so that $w \sim \mathcal{W}_t$ 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}^{α} 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 50.** *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 $Y_t \sim \text{Ber}(p)$ 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 $\Pr(|X| > 2t\sqrt{pN}) \leq 2 \exp(-t^2)$. Otherwise, if $pN \leq 1$, $\mathbf{E} |X| \geq \frac{1}{2e} \log \frac{1}{1-pN}$, and $\Pr(|X| \geq 1+t) \leq \exp(-\frac{1}{2}t)$.*

From Lemma 50, 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}^{α} ,

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

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

Now, we show that SA 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 Proposition 45), $m_{\alpha} \leq n \cdot \sqrt{\mathbf{E} w_T^2} + \sqrt{pn^k \log n}$ 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

$$\mathcal{I}^{\alpha}(x) \leq \frac{n \cdot \sqrt{pn^{k-1}}(1 + o(1))}{pn^k(1 \pm o(1))} \leq \frac{2}{\sqrt{pn^{k-1}}}$$

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

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

Now, we tackle α with $|\alpha| = t$ for $2 \leq t \leq k$. We will verify that the conditions of Propositions 45 and 49 hold. First, consider the α with $|\alpha| = t$ for $pn^{k-t} \geq 1$. From Lemma 50, in this case we have that $\mathbf{E}|w_T| \geq \frac{2}{e^{3/2}} \sqrt{pn^{k-t}}$, and with high probability, $|w_T| \leq O(\sqrt{tpn^{k-t} \log n})$ for all $T \in [n]^t$. Letting $M = O(\sqrt{tpn^{k-t} \log n})$, and $\sigma^2 = pn^{k-t}$, we meet the conditions for Proposition 45:

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

so long as $pn^{k-t} \geq 1$, which we have assumed. So applying Propositions 45 and 49 to both $\frac{m}{m_\alpha} \mathcal{I}^\alpha$ and $-\frac{m}{m_\alpha} \mathcal{I}^\alpha$, we have

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

and so long as $\frac{m_\alpha}{m} \cdot n^{\lfloor t/2 \rfloor / 4\ell} \rho \leq \frac{1}{2} \epsilon^{2\ell}$, with high probability over \mathcal{I}^α , $t(\ell + 1)$ rounds of Sherali–Adams certify that $|\mathcal{I}^\alpha(x)| \leq \frac{3}{2} \epsilon$. We confirm that

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

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

Finally, we handle α with $|\alpha| = t$ for $2 \leq t$ and $pn^{k-t} < 1$. From Lemma 50 we have $\mathbf{E}|w_T| \geq \frac{1}{e} \log \frac{1}{1-pn^{k-t}}$, and with high probability, $|w_T| \leq 4 \log n$ for all $T \in [n]^t$. Taking $M = 4 \log n$, we have that we meet the conditions of Propositions 45 and 49

$$\begin{aligned} M &\leq 4 \log n \ll n^{\lceil k/2 \rceil - \lceil t/2 \rceil + \delta} = pn^{k-t} n^{\lfloor t/2 \rfloor} = \sigma^2 n^{\lfloor t/2 \rfloor}, \\ \mathbf{E}|w_T| &\geq \frac{1}{e} \log \frac{1}{1-pn^{k-t}} \geq \frac{1}{e} pn^{k-t} \gg \sqrt{\frac{pn^{k-t} \log n}{n^{\lfloor t/2 \rfloor}}} = \sqrt{\frac{\sigma^2 \log n}{n^{\lfloor t/2 \rfloor}}}, \end{aligned}$$

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 Propositions 45 and 49 to both $\frac{m}{m_\alpha} \mathcal{I}^\alpha$ and $-\frac{m}{m_\alpha} \mathcal{I}^\alpha$, we have that for

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

so long as $\frac{m_\alpha}{m} n^{\lfloor t/2 \rfloor / 4\ell} \rho \leq \frac{1}{2} \epsilon^{2\ell}$, $R = t(\ell + 1)$ rounds of Sherali–Adams certify that $|\mathcal{I}^\alpha(x)| \leq \frac{3}{2} \epsilon$ with high probability. Verifying,

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

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

We therefore have (using Parseval’s identity and $\|x\|_1 \leq \sqrt{k}\|x\|_2$ for $x \in \mathbb{R}^k$ to simplify Equation (21)) that the same number of rounds certifies that

$$\text{OBJ}_{\mathcal{I}}(x) \leq \sum_{\alpha \subset [k]} \widehat{P}(\alpha) \cdot \mathcal{I}^\alpha(x) \leq \widehat{P}(\emptyset) + \sqrt{2^k} \frac{3}{2} \epsilon,$$

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 40.

References

- 1 Mikhail Alekhovich, Sanjeev Arora, and Iannis Tourlakis. Towards Strong Nonapproximability Results in the Lovász-Schrijver Hierarchy. *Computational Complexity*, 20(4):615–648, 2011. doi:10.1007/s00037-011-0027-z.
- 2 Sarah R. Allen, Ryan O’Donnell, and David Witmer. How to Refute a Random CSP. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 689–708. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.48.
- 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.
- 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 Francisco Barahona and Ali Ridha Mahjoub. On the cut polytope. *Math. Programming*, 36(2):157–173, 1986. doi:10.1007/BF02592023.
- 6 Boaz Barak, Moritz Hardt, Thomas Holenstein, and David Steurer. Subsampling Mathematical Relaxations and Average-case Complexity. In Dana Randall, editor, *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 512–531. SIAM, 2011. doi:10.1137/1.9781611973082.41.
- 7 Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 428–437. IEEE Computer Society, 2016. doi:10.1109/FOCS.2016.53.
- 8 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding Semidefinite Programming Hierarchies via Global Correlation. In Rafail Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 472–481. IEEE Computer Society, 2011. doi:10.1109/FOCS.2011.95.
- 9 Christoph Berkholz. The relation between polynomial calculus, Sherali-Adams, and sum-of-squares proofs. In *35th Symposium on Theoretical Aspects of Computer Science (STACS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- 10 Colin R Blyth. Expected Absolute Error of the Usual Estimator of the Binomial Parameter. *American Statistician*, pages 155–157, 1980.
- 11 Ravi Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science*, pages 280–285. IEEE, 1987.
- 12 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, 1999. doi:10.1137/S0097539795290805.

- 13 Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate Constraint Satisfaction Requires Large LP Relaxations. *J. ACM*, 63(4):34:1–34:22, October 2016. doi:10.1145/2811255.
- 14 M Charikar and A Wirth. Maximizing quadratic programs: extending Grothendieck's inequality. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 54–60. IEEE, 2004.
- 15 Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 283–292. ACM, 2009. doi:10.1145/1536414.1536455.
- 16 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.
- 17 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.
- 18 Wenceslas Fernandez de la Vega and Claire Mathieu. Linear Programming Relaxations of Maxcut. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07*, pages 53–61, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics. URL: <http://dl.acm.org/citation.cfm?id=1283383>. 1283390.
- 19 Charles Delorme and Svatopluk Poljak. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62(1-3):557–574, 1993.
- 20 Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer, Heidelberg, 1997. doi:10.1007/978-3-642-04295-9.
- 21 William E Donath and Alan J Hoffman. Lower bounds for the partitioning of graphs. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 437–442. World Scientific, 2003.
- 22 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.
- 23 Uriel Feige and Michael Langberg. The RPR 2 rounding technique for semidefinite programs. In *International Colloquium on Automata, Languages, and Programming*, pages 213–224. Springer, 2001.
- 24 Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures Algorithms*, 27(2):251–275, 2005. doi:10.1002/rsa.20089.
- 25 Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.
- 26 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.
- 27 Joel Friedman, Jeff Kahn, and Endre Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 587–598, 1989.
- 28 Alan Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 12–20. IEEE, 1996.
- 29 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.
- 30 Andreas Goerdt and Michael Krivelevich. Efficient Recognition of Random Unsatisfiable k-SAT Instances by Spectral Methods. In Afonso Ferreira and Horst Reichel, editors, *STACS 2001, 18th Annual Symposium on Theoretical Aspects of Computer Science, Dresden, Germany, February 15-17, 2001, Proceedings*, volume 2010 of *Lecture Notes in Computer Science*, pages 294–304. Springer, 2001. doi:10.1007/3-540-44693-1_26.

- 31 Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001. doi:10.1016/S0304-3975(00)00157-2.
- 32 Johan Håstad. An NP-complete problem – some aspects of its solution and some possible applications. Master’s thesis, Uppsala University, 1984.
- 33 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 Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 720–731. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.72.
- 34 Pravesh Kothari, Ryuhei Mori, Ryan O’Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, pages 132–145, 2017.
- 35 Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 590–603, 2017. doi:10.1145/3055399.3055438.
- 36 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.
- 37 Bojan Mohar and Svatopluk Poljak. Eigenvalues in combinatorial optimization. In *Combinatorial and graph-theoretical problems in linear algebra*, pages 107–151. Springer, 1993.
- 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*, pages 569–581. North-Holland, Amsterdam, 1992.
- 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.
- 40 Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 121–131, 2017. doi:10.1145/3055399.3055417.
- 41 Prasad Raghavendra, Tselil Schramm, and David Steurer. High-dimensional estimation via sum-of-squares proofs. *CoRR*, abs/1807.11419, 2018. arXiv:1807.11419.
- 42 Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 373–387. SIAM, 2012. doi:10.1137/1.9781611973099.
- 43 Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.74.
- 44 Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In David S. Johnson and Uriel Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 302–310. ACM, 2007. doi:10.1145/1250790.1250836.
- 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. Discrete Math.*, 3(3):411–430, 1990. doi:10.1137/0403036.
- 46 Konstantin Tikhomirov and Pierre Youssef. The spectral gap of dense random regular graphs. *arXiv preprint*, 2016. arXiv:1610.01765.

- 47 Luca Trevisan. Max Cut and the Smallest Eigenvalue. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 263–272, New York, NY, USA, 2009. ACM. doi:10.1145/1536414.1536452.
- 48 Joel Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.
- 49 Yuichi Yoshida and Yuan Zhou. Approximation schemes via Sherali-Adams hierarchy for dense constraint satisfaction problems and assignment problems. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 423–438. ACM, 2014.
- 50 Uri Zwick. Outward Rotations: A Tool for Rounding Solutions of Semidefinite Programming Relaxations, with Applications to MAX CUT and Other Problems. In *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, STOC '99, pages 679–687, New York, NY, USA, 1999. ACM. doi:10.1145/301250.301431.

A Characteristics of distributions of XOR-subformula coefficients

We now prove Lemma 50. We will use the following estimate for the mean absolute deviation of a binomial random variable.

► **Lemma 51** (e.g. [10]). *If X is distributed according to the binomial distribution $X \sim \text{Bin}(n, p)$, then $\mathbf{E}|X - \mathbf{E}X| = \sqrt{\frac{2}{\pi}np(1-p)} + O(\frac{1}{\sqrt{n}})$.*

Proof of Lemma 50. We calculate the absolute value directly. Given that there are exactly k nonzero Y_t , the absolute value of X is distributed according to $|\text{Bin}(k, \frac{1}{2}) - \frac{1}{2}k|$. Using the method of conditional expectations,

$$\mathbf{E}|X| = \sum_{k=0}^N \Pr[k \text{ nonzero } Y_t \text{'s}] \cdot \mathbf{E}|\text{Bin}(k, \frac{1}{2}) - \frac{1}{2}k| \geq \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \cdot \sqrt{\frac{1}{2\pi}} k,$$

where we have applied the estimate from Lemma 51. 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

$$\mathbf{E}|X| \geq \sum_{k=1}^N \sqrt{\frac{2\pi}{e^2} \frac{N}{k(N-k)}} \cdot \exp(-N \cdot D(\frac{k}{N}||p)) \cdot \sqrt{\frac{1}{2\pi}} k \geq \frac{1}{e} \sum_{k=1}^N \exp(-N \cdot D(\frac{k}{N}||p)), \quad (22)$$

Now, if $pN < 1$, we take

$$\text{Equation (22)} \geq \frac{1}{e} \sum_{k=1}^N \exp\left(k \log\left(\frac{pN}{k}\right)\right) = \frac{1}{e} \sum_{k=1}^N \left(\frac{pN}{k}\right)^k \geq -\frac{1}{e} \log(1-pN) - O(p^N) \quad (23)$$

as desired.

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

$$\text{Equation (22)} \geq \frac{1}{e} \sum_{\substack{\ell=1-\lfloor pN \rfloor \\ \delta=\frac{\ell}{N}}}^{\lfloor (1-p)N \rfloor} \exp(-N \cdot D(p+\delta||p)) \quad (24)$$

8:30 Sherali–Adams Strikes Back

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

$$\text{Equation (24)} \geq \frac{1}{e} \sum_{\substack{\ell=-\lfloor\sqrt{pN}\rfloor \\ \delta=\frac{\ell}{N}}}^{\lfloor\sqrt{pN}\rfloor} \exp\left(-N\frac{\delta^2}{2}\right) \geq \frac{1}{e} \cdot 2\sqrt{pN} \cdot \exp\left(-\frac{p}{2}\right) \geq \frac{2}{e^{3/2}}\sqrt{pN}, \quad (25)$$

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

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

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