Limitations of the Sherali-Adams Lift and Project System: Compromising Local and Global Arguments

Konstantinos Georgiou* Avner Magen*

Department of Computer Science, University of Toronto

{cgeorg,avner}@ cs.toronto.edu

Abstract

We consider several Optimization problems that are originated in the study of linear systems capturing Satisfaction problems. We obtain lower bounds for the rank of the associated polytopes, and integrality gaps after linear many liftings of the Sherali-Adams Lift and Project system. Finally, we show how some of these results can be extended to the stronger Lift and Project system of Lasserre.

Specifically, we show linear rank lower bounds for the Pigeonhole Principle, for Tseitin Tautology on random graphs, and for random k-LIN and k-CNF instances, $k \geq 3$. As an immediate corollary, we obtain tight integrality gaps for linear many rounds of Sherali-Adams for MAX-k-LIN and MAX-k-CNF, and a 7/6 integrality gap for VERTEXCOVER .

We refine the connection between the Sherali-Adams system and probabilistic reasoning, as was laid forth by Fernández de la Vega and Kenyon-Mathieu and by Charikar, Makarychev and Makarychev, and supply a methodology which we call "between local and global" that defines probability distributions on small sets using a simple distribution over large, yet not-too-large sets.

1 Introduction

A standard approach in approximating NP-hard problems is to formulate the problem as a 0-1 integer program and then relax the integrality condition to get a linear program which can be solved efficiently. The quality of such an approach is intimately related to the *integrality gap* of the relaxation, namely, the ratio between the optimum of the relaxation to that of the original problem. In some contexts, we are interested in linear relaxations of integer programs that may have *no solutions at all*. In these cases, the relevant information about the relaxation is under what conditions it is feasible, that it behaves differently than its integral counterpart.

In order to remedy the problem of discrepancy between integral and fractional solutions, a Linear Programming relaxation may be *tightened* by adding inequalities that are valid with respect to the original problem on one hand, and that exclude some fractional solutions on the other. Such tightenings can be done in an ad-hoc fashion, whereby relevant inequalities are identified "somehow" and are added to the relaxation. In

*Funded in part by NSERC

contrast, several methods (or procedures) were developed in order to obtain tightenings of relaxations in a systematic manner. These procedures, normally referred to as *Lift and Project* methods, use polynomial reasonings together with the fact that in the 0/1 domain, general polynomials can be reduced to multilinear polynomials (utilizing the identity $x^2 = x$), and then apply linearization which view of that information. Such procedures include the Lovász-Schrijver system [21], Sherali-Adams system [27], and Lasserre system [19] (see[20] for a comparison).

Lift and project systems are applied in rounds[1]. The bigger the number of rounds used, the more accurate the obtained tightened relaxation is. In fact, if as many rounds as the number of variables are used, the final relaxation is exact, all its feasible solutions are distributions of integral solutions, whence no integrality gap occur. More generally, the projection of a solution that is valid after t rounds, is a convex combination of integral solutions on these t variables. What distinguishes the different systems is the way they apply their nonlinear extension, and whether they pose some positive semidefinite conditions. Another common feature to all systems is an upper-bound on the time it takes to optimize over relaxations after a given number of rounds r, as a function of r. If the original relaxation admits a polynomial separation oracle, then optimizing over tightenings obtained after r of rounds can be done in time $n^{O(r)}$ [21]. Interestingly, in none of the systems does it seem possible to improve on the above bound (if one ignores the hidden constant in the exponent). This connection to time complexity has motivated researchers to study the (worst case) behaviour of the application of the systems to particular LP relaxations of NP-hard problems. Specifically, a major goal that emerges from the above discussion is to analyze the integrality gap after varying number of rounds of application of particular systems.

Do Lift and project systems in fact "deliver" and serve as an important ingredient in the domain of algorithm design? It seems like there are compelling evidences that the answer should essentially be positive. One type of algorithms is the one that uses SemiDefinite Programming, but adds some extra restrictions (described as "ad-hoc additions" at the beginning of this section). Examples for these are the Goemans-Williamson algorithm for Max-Cut [13], the Karloff-Zwick algorithm for Max-3-SAT [18], Karakostas algorithm for Vertex Cover [17], and the ARV algorithm for Sparsest-Cut [3]. While these algorithms are not presented as instances of Lift and Project systems, they can be viewed as a weaker version of tightened SDPs that are obtained by applying the Lovász-Schrijver system with positive semidefinite constraint for a constant number of rounds. Another type of algorithms are ones that make a "dynamic" use of the aforementioned accuracy/round tradeoffs. These are algorithms which lead to Polynomial-time Approximation-Schemes (PTAS) by using as many as $r(\epsilon)$ rounds to achieve approximation (or integrality-gap) $1 + \epsilon$ where $r(\epsilon)$ is a function that depends only on ϵ. Fernández de la Vega and Kenyon-Mathieu [11] have provided a PTAS for Max Cut in dense graphs using Sherali-Adams . In [22] it is shown how to get a Sherali-Adams based PTAS for VERTEXCOVER and max-INDEPENDENT-SET in minor-free graphs. Chlamtac and Singh [7] gave an approximation algorithm for INDEPENDENT-SET based on the Lasserre system, with the performance depending on the number of rounds applied. Interestingly, no PTAS are known for the weaker system of Lovász-Schrijver which highly motivates investigating the performance of Sherali-Adams and Lasserre over that of Lovász-Schrijver . Considerable effort was also invested in the lower bound department. See [2, 29, 28, 26], [5, 9, 1, 25, 12], and [11, 6], for the Lovász-Schrijver , Lovász-Schrijver with PSD constraints and Sherali-Adams systems respectively. Special credits should be also given to the very recent result by

[1]For the Sherali-Adams and Lasserre systems, where one carries out a number of subsequent liftings and only one projection at the end (and which are static systems in contrast to the Lovász-Schrijver system where one applies subsequent liftings and projections), it might sound more appropriate to refer to the number of tightenings as the level of the relaxation. Instead we choose to refer to this as the number of applied rounds, since we are thinking of the Sherali-Adams and Lasserre systems as systematic procedures for producing valid constraints for the integral hull, the same way cutting planes procedures do.

Schoenebeck [24], who independently from this work proved rank lower bounds and integrality gaps for the same random instances we do, for the stronger Lasserre system.

A related measure of efficiency of these methods is the *rank* of linear inequalities in a particular system: let ψ be a linear inequality that is satisfied for every integral solution of the problem. The rank of ψ with respect to a system is the number of rounds required before this inequality is implied. Equivalently, all points that satisfy the tightened relaxation after this many rounds (but no less) satisfy ψ. It is useful to notice that this essentially generalizes the first notion of integrality gaps after a certain number of rounds. Indeed, if the inequality that says that the objective function is bigger than α is not implied after r rounds, then there is a solution with objective value α after r many rounds. But the notion of rank is more versatile as it can be used even for the case where no integral solutions to the problem exist at all; in this case the inequality $1 \leq 0$ is valid to all integral solutions, and asking about the rank of this inequality is tantamount to asking how many rounds are needed before the tightened relaxation has no feasible solutions. Such settings are relevant in the context of proof complexity where the input is an (algebraization of an) unsatisfied propositional formula. We view the systems as a tool to derive valid inequalities, with the original inequalities being the axioms, and the goal is to "prove" the inequality $1 \leq 0$. The rank captures the depth of the shortest proof.

Results: In the current work we provide rank lower bounds to certain tautologies, that immediately imply tight integrality-gaps lower bounds for optimization problems. Specifically, we show rank lower bounds for the Pigeonhole Principle, the Tseitin Tautology, and for random $\mathrm{mod}\ 2$ constraints and CNF formulas. The rank lower bound for the Pigeonhole Principle is tight, while the rest of the rank lower bounds are linear. The rank lower bounds for mod 2 constraints and CNF formulas imply tight integrality gaps for the corresponding optimization problems MAX-k-LIN and MAX-k-CNF, with $k \geq 3$. The lower bound of the Tseitin Tautology is also extended to a Lasserre rank lower bound. Finally, the tight integrality gap for MAX-3-LIN implies a 7/6 integrality gap for VERTEXCOVER for linear many Sherali-Adams tightenings.

Techniques: Fernández de la Vega and Kenyon-Mathieu [11] gave a probabilistic interpretation to Sherali-Adams solutions. Roughly speaking, they consider an ensemble of probability spaces over 0/1 assignments of certain subsets of variables. These distributions should be supported on feasible integral solutions for the sets at hand (i.e., to constraints that are contained in these sets), and they should also be consistent with one another in the natural way: the probability of a partial 0/1 assignment to a set of variables is the same regardless of the superset over which the distribution is defined. Given such an ensemble, the vector with marginal probabilities of variables being 1 can be shown to be in the resulting Sherali-Adams lift and project after r rounds, where the subsets in the ensemble relate in some simple fashion to r. (This type of thinking was in fact implicit in [11] and was made explicit in [6], but both deal only with the case of MAXCUT . We formalize it later in Proposition (1).)

Notice that there is a very simple way to get a good ensemble: consider a distribution of (global) integral solutions, and define the different probability space as restrictions of that global distribution to the set they represent. Clearly such an ensemble satisfies the required conditions. Of course this construction is not very interesting as all it says is that points that are convex combination of integral solutions are in the Sherali-Adams system, which is obvious. In particular, no integrality gaps can be shown this way. When trying to show a lower bound, the goal then is to provide an ensemble of probability spaces that cannot be "explained" by a global probability space, and moreover, behaves quite differently with respect to the probabilities for singletons which correspond to the original variables of the problem.

The starting point of the main result of this this work (Section 5) is the paper of Buresh-Oppenheim et al. [5] that gives lower bounds to the same family of problems in the Lovász-Schrijver systems. A lower bound in the Lovász-Schrijver system can be viewed as a prover-adversary game and in the same time as an exposure process in a probabilistic domain. The fact that the Sherali-Adams system is stronger than the Lovász-Schrijver system is reflected by the fact that the exposure process in [5] is not enough for our need. In the Sherali-Adams world we need to supply an ensemble of probability spaces as was outlined above. Assuming, as is the case with all the problems we deal with, that locally the set of constraints are satisfiable, we may try to apply the following naive approach: for a set S of variables, define Ω_S as a uniform distribution over assignments to S that satisfy the constraints that contain S variables only. Local consistency (or local-agreement) of this approach is guaranteed if all distributions are uniform over all variables; but to be able to use such distributions we need to require that the sets never contain a single constraint. This type of reasoning is used in Sections (3) and (4) where we demonstrate simple lower bounds for Sherali-Adams .

As we later point out, there is no way to achieve the local agreement in the strategy defined above if we want to go beyond the size supported by a constraint. Roughly speaking, trying to define a distribution on a set just by looking at the constraints contained in it, but ignoring any other variables, is too local and bound to fail. Our main technical contribution is developing a methodology that goes beyond this barrier which we call "between local and global". For a set S we define a set $\bar{S}$ which we call the *advice-set* of S. Now, instead of defining a probability distribution that is uniform over valid solutions on S we take the uniform distributions over valid solutions of $\bar{S}$ (projected back to S). The precise way we obtain $\bar{S}$ is related to expansion properties of the constraints-variable graph that is associated with the system, similarly to the way that expansion is used in the Buresh-Openeheim et al. paper [5] and also in a result by Alekhnovich, Arora and Tourlakis [1]. The main technical challenge remains to show that such a method leads to distributions that have the local agreement property.

It is interesting to contrast our work to the recent and stronger result by Schoenebeck [24] that was obtained independently. Both results prove rank lower bounds for random k-LIN instances and subsequently give as corollary integrality gaps for a series of optimization problems. It is not surprising that both lower bounds rely on the same family of random instances that enjoy nice structural (expanding) properties. In both cases one has to define appropriate solutions of strong relaxations obtained after linear many tightenings. It seems that the most convenient and actually the only way we know for designing such solutions is to think of them as probability distributions of valid local solutions, a property which is inherent to both Sherali-Adams and Lasserre systems. Most importantly, both results stick with the uniform distribution over valid solutions of specially defined constraints. The way these constrains are defined lies in the heart of the problem, and individually to each of the two results. Our current work first tries to unify the way rank lower bounds for the Sherali-Adams system are proven. Proposition 1 provides such a framework which although was implicit in [11] and in particular in [6], it was not offered in its full and current generality. Given this tool, our work also tries to elucidate the (unavoidable) strategy followed almost by all previous negative results on lift and project systems; that is, given a set of variables over which a probability of local integral solutions is required, first define appropriate supersets and then consider the "simplest" possible distribution of valid solutions. These settled ideas give two instructive and simple applications. First in Section 3 we improve the best previous known rank lower bound for the Pigeonhole Principle, and the proof appears to be significantly simpler. Second, in the same concept Section 4 provides one more simple rank lower bound edifying on one hand the necessity of a "correction" phase, and on the other justifying that "simple" distributions of integral solutions seem to be the right choice. Given this instructive background, our work then, in Section 5, clarifies that the expanding properties of an instance is the natural obstacle that the Sherali-Adams system cannot overcome, by drawing the connections between expansion and advice-sets. In contrast, the same ideas are hidden in Shoenebeck's result who uses resolution as a black box in order to define what we call advice-sets. In particular, Schoenebeck shows that whenever a set of constraints eludes bounded width r resolution, it also fools $\Omega(r)$ Lasserre rounds. Why Lasserre fails to disprove the appropriately chosen set of constraints remains esoteric in the lower bound for resolution, however this elegant connection allows Schoenebeck to satisfy the extra positive semidefinitness conditions of the Lasserre system, that we fail to satisfy.

2 Preliminaries

2.1 The Sherali-Adams system

Denote by $\mathcal{P}_t^n$ all subsets of $[n]$ of size at most t.

Definition 1. Consider the polytope $P \subseteq \mathbb{R}^n$, defined by the linear constraints $\alpha_l^T \mathbf{x} - b \leq 0$, $l = 1, \ldots, m$, $0 \leq x_i \leq 1$, $i = 1, \ldots, n$. We define $SA^t(P) \subseteq \mathbb{R}^{\mathcal{P}_{t+1}^n - 1}$ to be the set of all $\mathbf{z} \in \mathbb{R}^{\mathcal{P}_{t+1}^n - 1}$, that satisfy the lifted linear system that is constructed as follows: for every constraint $\psi(\mathbf{x}) \leq 0$ of P, for every $U \in \mathcal{P}_t^n$, and for every $W \subseteq U$, consider the valid constraint

$$\psi(\mathbf{x}) \prod_{s \in W} x_s \prod_{s \in U \setminus W} (1 - x_s) \leq 0,$$

and for every $I \in \mathcal{P}_{|U|+1}^n$ replace $\prod_{s \in I} x_s$ by z_I. Add the resulting linear inequality to the set of constraints defining $SA^t(K)$.

The projection of $\mathbf{z}$ that satisfies the above linear system onto the singleton indices is what we call the t-th Sherali-Adams tightening of P. We denote this tightening as $S^t(K)$. Finally we define the Sherali-Adams rank of the polytope P as the minimum integer t such that $S^t(K)$ is the integral hull of K.

Next we give a condition for proving lower bounds in the Sherali-Adams system. A similar statement was first used in [11] and was stated explicitly in [6]. However, those version dealt specifically with MAXCUT and did not immediately give abstract treatment as the one we give here. Further, the proof below avoids the "inclusion-exclusion manipulations" that existed in the other variants.

Proposition 1. *Consider a polytope P defined on n variables. For any $I \subseteq [n]$, $|I| \leq r + 1$, and for any constraint of P with support J, consider some distributions $\mathcal{D}(I)$, $\mathcal{D}(I \cup J)$ of 0/1 assignments for I and $I \cup J$ respectively, and define the vector $\mathbf{z} \in \mathbb{R}^{\mathcal{P}_{r+1}^n}$, as $z_I = Pr_{\mathcal{D}(I)}$[all elements of I are 1]. Suppose that this collection of distributions satisfy the following properties.*

For every $U \subseteq [n]$, $|U| \leq r + 1$, and for every constraint of P with support J

1. ***(feasibility):*** *$\mathcal{D}(U \cup J)$ is a distribution of assignments that satisfy the constraint with support J*
2. ***(local agreement):*** *For every $i \in J$ and $H \subseteq U$, $\mathcal{D}(H \cup \{i\})$, $\mathcal{D}(U \cup J)$ agree on H, that is*

$$Pr_{\mathcal{D}(H \cup \{i\})}[\textit{all elements of } H \cup \{i\} \textit{ are 1}] = Pr_{\mathcal{D}(U \cup J)}[\textit{all elements of } H \cup \{i\} \textit{ are 1}]$$

Then $\mathbf{z} \in SA^r(P)$.

Proof. Consider some constraint $\alpha^T \mathbf{x} \leq b$ of P with support J. Fix also some $U \in \mathcal{P}_r^n$, and $W \subseteq U$. In the resulting lifted linear constraint, every term, that is every variable x_i, $i \in J$, and the constant term b will be multiplied by

$$\prod_{s \in W} x_s \prod_{s \in U \setminus W} (1 - x_s)$$

Now, for two disjoint sets $K, L \subseteq U \cup J$, define

$$y_{K,L} = Pr_{\mathcal{D}(U \cup J)}[\text{all elements of } K \text{ are assigned 1 \& all elements of } L \text{ are assigned 0}].$$

Notice that the mapping $P(K,L) = \prod_{s \in K} x_s \prod_{s \in L} (1 - x_s) \mapsto y_{K,L}$ is an homomorphism that sends $\prod_{i \in I} x_i$ to $y_{I,\emptyset}$. This can be verified inductively on the size of L by noticing that $P(K, L \cup i) = P(K,L) - P(K \cup i, L)$ and that $y_{K, L \cup i} = y_{K,L} - y_{K \cup i, L}$.

Now, notice that local agreement is precisely saying that $y_{K,\emptyset} = z_K$, and therefore the assignment z_I serves as a substitution of x_i with

$$y_{W \cup \{i\}, U \setminus W} \quad = \quad Pr_{\mathcal{D}(U \cup J)}[\text{all elements of } W \cup \{i\} \text{ are 1 \& all elements of } U \setminus W \text{ are 0}],$$

while constant b is replaced by $y_{W, U \setminus W} \cdot b$. All probabilities below are with respect to $\mathcal{D}(U \cup J)$. We have

$$\begin{aligned}
& \sum_{i \in J} a_i \cdot y_{W \cup \{i\}, U \setminus W} \\
= \;& \sum_{i \in J} a_i \sum \{Pr[\sigma] : \sigma \text{ is an assignment on } U \cup J \text{ that is } 0/1 \text{ on } W \cup \{i\}, U - W \text{ resp}\} \\
= \;& \sum \{Pr[\sigma] : \sigma \text{ is an assignment on } U \cup J \text{ that is } 0/1 \text{ on } W, U - W \text{ resp}\} \sum_{i : \sigma_i = 1} a_i \\
= \;& \sum \{Pr[\sigma] : \sigma \text{ is an assignment on } U \cup J \text{ that satisfies } \psi \text{ and that is } 0/1 \text{ on } W, U - W \text{ resp}\} \sum_{i : \sigma_i = 1} a_i \\
\leq \;& \sum \{Pr[\sigma] : \sigma \text{ is an assignment on } U \cup J \text{ that satisfies } \psi \text{ and that is } 0/1 \text{ on } W, U - W \text{ resp}\} \cdot b \\
= \;& y_{W, U \setminus W} \cdot b
\end{aligned}$$

where the second last equality follows from the feasibility condition and the inequality follows since by definition $\sum_{i : \sigma_i = 1} a_i \leq b$ for σ that satisfies ψ.

□

2.2 The Extension to the Lasserre System

The Lasserre system extends the Sherali-Adams system. Instead of supplying definition to Lasserre , we use the terminology of Definition (1). Consider the vector $\mathbf{z} \in SA^t(P)$. We say that $\mathbf{z}$ is *of PSD form* if we can associate every set I with a vector $\mathbf{v} \in \mathbb{R}^{\mathcal{P}_{t+1}^n}$ such that for every pair of sets I, J such that $|I \cup J| \leq t + 1$, it holds that

$$v_I \cdot v_J = z_{I \cup J}.$$

If in addition to the conditions of Definition (1) $\mathbf{z}$ is of PSD form, then $\mathbf{z} \in LA^t(P)$, that is $\mathbf{z}$ is a lifted solution for t many Lasserre tightenings of P. Analogously to Definition 1, we abbreviate the projection of $LA^t(P)$ onto the singleton indices by $L^t(P)$.

Notice now that the conditions of Proposition (1) along with the positive semidefiniteness of $\mathbf{z}$ is a sufficient condition for $\mathbf{z} \in LA^t(P)$.

2.3 Linear Relaxation of mod 2 Linear Systems

Consider a set M of $\mod 2$ equations on n variables. For $j \in M$ denote by $\Gamma(j)$ the set of variables that appear in equation j, so that constraint j is of the form $\sum_{i\in\Gamma(j)} x_i \equiv b \mod 2$, $b \in \{0,1\}$.

Any set M of $\mod 2$ equations can be naturally formulated as a CNF formula in the following way. For every x_i introduce the Boolean variable z_i (z_i is true iff x_i is assigned the value 1). Consider an equation $j \in M$, and all partitions Y, N of $\Gamma(j)$ with $|Y|$ being even. What we require is that $\neg\left((\wedge_{t\in Y} z_t) \wedge (\wedge_{t\in N} \bar{z}_t)\right)$. From there it is obvious how to formulate the problem as a linear program. The previous subformula for example will be converted to $\sum_{t\in Y} z_t - \sum_{t\in N} z_t \leq |Y| - 1$. The resulting linear program relaxation we denote by P_M.

3 The Pigeonhole Principle

The Pigeonhole Principle of order n, is the following infeasible linear system. Intuitively, we have $n+1$ pigeons and n holes. The semantics of x_{ij} is that pigeon i goes to pigeonhole j.

$$\begin{array}{rcll} x_{it} + x_{jt} & \leq & 1 & ,\forall i,j \in [n+1], i \neq j, \forall t \in [n] \\ \sum_{t=1}^{n} x_{it} & \geq & 1 & ,\forall i \in [n+1] \\ x_{it} & \in & \{0,1\} & ,\forall i \in [n+1], \forall t \in [n] \end{array}$$

By relaxing the variables to $x_{it} \in [0,1], \forall i \in [n+1], \forall t \in [n]$ we obtain a nonempty polytope that we denote by P_n.

P_n has provided separation between the performance of the Lovász-Schrijver system and the Gomory-Chvátal cutting planes procedure (see [5] for the definition). As it is implied by [15] and later proved in [5, 4], the Gomory-Chvátal rank of P_n is $O(\log n)$. In [5] it is shown that the rank of P_n with respect to the weakest version of Lovász-Schrijver system is exactly $n-1$, which directly implies that the Sherali-Adams rank is at most $n-1$. Further it is known [14, 21] that the rank of P_n with respect to the Lovász-Schrijver SDP system is 2. Altogether, in this section we provide a rank separation between Sherali-Adams , Gomory-Chvátal and Lovász-Schrijver SDP tightenings. Below we show that the rank of P_n is $n-1$, improving upon [23].

Denote by $\mathcal{E}_{it}$ the event that pigeon i goes to hole t, i.e. $x_{it} = 1$. For a set of pigeons $\{i_1, \ldots, i_s\}$, $s \leq n$, we define a distribution $\mathcal{U}(S)$ of assignments of S into the n holes: choose one of the $\binom{n}{s} s!$ different assignments of pigeons to different holes uniformly at random. This distribution of valid assignments satisfies a form of local agreement property as the next lemma suggests.

Lemma 2. *For every two sets $R = \{i_1, \ldots, i_r\}$, $S = \{j_1, \ldots, j_s\}$ of pigeons, such that $|R \cup S| \leq n$ and for every set of holes $T = \{t_1, \ldots, t_r\}$, distributions $\mathcal{U}(R), \mathcal{U}(R \cup S)$ agree on the event $\mathcal{E} = \mathcal{E}_{i_1 t_1} \cap \ldots \cap \mathcal{E}_{i_r t_r}$.*

Proof. Notice that the lemma states a simple fact, which is that a set of pigeons is matched to holes with probability that is not affected by the existence of other pigeons, as long as there are less than n pigeons in total of course. Here is a formal explanation.

Assume without loss of generality that R and S are disjoint, as otherwise we take $S \setminus R$ instead of S, arriving at the same conclusion. Define $\mathcal{Q}$ to be the event that none of pigeons in S goes to any of the holes in T. We have

$$Pr_{\mathcal{U}(R\cup S)}\left[\mathcal{E}\right] = Pr_{\mathcal{U}(R\cup S)}\left[\mathcal{Q}\right] Pr_{\mathcal{U}(R\cup S)}\left[\mathcal{E}|\mathcal{Q}\right] = \frac{\binom{n-r}{s}s!}{\binom{n}{s}s!}\cdot\frac{1}{\binom{n-s}{r}r!} = \frac{1}{\binom{n}{r}r!} = Pr_{\mathcal{U}(R)}\left[\mathcal{E}\right].$$

□

For every $I = \{i_1t_1, \ldots, i_rt_r\}$, $r \leq n-1$, denote $G_I = \{i_1, \ldots, i_r\}$ and define the distribution $\mathcal{D}(I) = \mathcal{U}(G_I)$. Let $N = n(n+1)$. Define also $\mathbf{z} \in \mathbb{R}^{\mathcal{P}^N_{n-1}-1}$ as

$$z_I = Pr_{\mathcal{D}(I)}[\mathcal{E}_{i_1t_1} \cap \ldots \cap \mathcal{E}_{i_rt_r}]. \tag{1}$$

Proposition 3. *For $\mathbf{z} \in \mathbb{R}^{\mathcal{P}^N_{n-1}}$ as defined in (1), we have $\mathbf{z} \in SA^{n-2}(P_n)$.*

Proof. We show that the distributions $D(I)$ satisfy the conditions of Proposition 1. A critical point is that the distribution that is defined for a set T only depends on G_T. Since there are at most 2 pigeons in any constraint of the polytope, the maximal number of pigeons in $U \cup J$ (using the terminology of Proposition 1) is not more than n. Further, the local agreement with respect to sets of pigeons immediately implies local agreement of the distributions of variables. Feasibility is guaranteed by the fact that any assignment of pigeons that does not assign more than one pigeon to the same hole, will satisfy the original constraints.

To conclude, Lemma 2 gives the local agreement property. Feasibility is implied by the definition of the experiment. All conditions of Proposition 1 are met, and the proposition follows. □

As an immediate application of Proposition 3 we get the desired rank lower bound.

Corollary 4. *For every n, the vector $(1/n, \ldots, 1/n)^T \in \mathbb{R}^n$ is in $S^{n-2}(P_n)$. Therefore, the Sherali-Adams rank of P_n is at least $n-1$.*

4 An Easy Rank Lower Bound for the Tseitin Tautology

Consider a graph $G = (V, E)$ on n vertices and the problem of assigning 0/1 values x_{ij} to all edges ij, such that

$$\sum_{j\in\Gamma(i)} x_{ij} = 1 \mod 2, \; \forall i \in V \tag{2}$$

where $\Gamma(i)$ are all adjacent vertices of i. For $i \in V$, we will refer to the constraint that involves all edges incident to i as the degree constraint of i.

The Tseitin tautology of a graph on odd number of vertices is the infeasible 0/1 system of constraints (2). The same system can be described as a 0/1 integer program (of exponential many constraints with respect to the degree of G) on $|E|$ many variables. By allowing the variables to assume values in $[0, 1]$ we obtain a nonempty polytope that we denote by TS_G; note that the all 1/2 vector is always a feasible solution.

As a warm up we show a simple rank lower bound of the Tseitin tautology that depends on the minimum degree of a graph.

Theorem 5. *Let $G = (V, E)$ be a graph on odd number of vertices, with minimum degree d. Then the Sherali-Adams rank of TS_G is at least $d - 1$.*

We need to show that the Sherali-Adams tightening of TS_G even after $d - 2$ liftings is nonempty. For this, we define $\mathbf{z} \in \mathbb{R}^{\mathcal{P}_{d-1}^{|E|}}$ such that $\mathbf{z} \in SA^{d-2}(TS_G)$. The indices of $\mathbf{z}$ will be again defined using a distribution of 0/1 assignments that enjoy the local agreement property.

We start with a graph $G = (V, E)$ of minimum degree d. Every set $J \subset E$ of size at most $2d - 2$ dominates at most one constraint, that is there is at most one edge constraint j with $\Gamma(j) \subseteq J$. For such a set $J \subset E$, consider the following distribution of 0/1 assignments $\mathcal{D}(J)$: choose uniformly at random one of the 0/1 assignments of J among those that do not contradict any dominated constraint. Note that if there is no dominated constraint, this is just the uniform distribution over all 0/1 assignments. An assignment that does not contradict a j dominated constraint we call j-consistent.

The next observation establishes the local agreement property. For $i \in V$, we denote by $E(i)$ the set of edges incident to i.

Claim 6. *For every $U \subseteq E$, $|U| \leq d - 2$, and for every $H \subseteq \Gamma(i) \cup U$, where $i \in V$ and $|H| \leq d - 1$ we have*

$$Pr_{\mathcal{D}(H)}[\textit{all edges in } H \textit{ are assigned } 1] = Pr_{\mathcal{D}(E(i)\cup U)}[\textit{all edges in } H \textit{ are assigned } 1] \tag{3}$$

The justification of claim 6 is very intuitive. Every 0/1 assignment of variables H can be extended to the same number of consistent assignments for $E(i) \cup U$. This property becomes formal in Lemmata 11 and 12.

Proof. (of Theorem 5) For $I \subseteq E$, $|I| \leq d - 1$, define $\mathbf{z} \in \mathbb{R}^{\mathcal{P}_{d-1}^{|E|}}$ as

$$z_I = \Pr_{\mathcal{D}(I)}[\text{all edges in } I \text{ are assigned } 1].$$

It is enough to argue that $\mathbf{z} \in SA^{d-2}(TS_G)$. This is implied by Proposition 1: local agreement follows from Claim 6 and feasibility from the experiment described above. □

The simplicity of the lower bound above, or more specifically, the fact that it hinges on uniform distribution of 0/1 assignments ($z_I = 2^{-|I|}$ for all I of size $< d - 1$) translates to a lower bound in the stronger Lasserre system.

Fix some orthonormal basis $\{\mathbf{u}_S\}_{S\subseteq[n]}$, and for $I \subseteq \mathcal{P}_{d-1}^n$ define

$$\mathbf{v}_I = \frac{1}{2^{|I|}} \sum_{S\subset I} \mathbf{u}_S.$$

For $I, J \subset \mathcal{P}_{d-1}^n$ we have,

$$\mathbf{v}_I \cdot \mathbf{v}_J = \frac{1}{2^{|I|+|J|}} \left(\sum_{S\subset I} \mathbf{u}_S\right) \cdot \left(\sum_{S\subset J} \mathbf{u}_S\right) = \frac{1}{2^{|I|+|J|}} \sum_{S\subseteq I\cap J} \mathbf{u}_S^2 = \frac{2^{|I\cap J|}}{2^{|I|+|J|}} = \frac{1}{2^{|I\cup J|}}.$$

As a consequence we get s a stronger version of Theorem 5:

Theorem 7. *Let $G = (V, E)$ be a graph on odd number of vertices, with minimum degree d. Then the Lasserre rank of TS_G is at least $d - 1$.*

For the special case when G is the complete graph on n (odd many) vertices, the rank of TS_G is at least $n-2$. It is also possible to achieve a rank lower bound of order $N^{1-\epsilon}$ for any $\epsilon > 0$; simply consider $n^{1/\epsilon}$-regular multigraphs on n vertices and apply Theorem 7.

Corollary 8. *For every $\epsilon > 0$, there exist graphs on n variables and $N = O(n^{1/\epsilon})$ edges, for which the Lasserre rank of the Tseitin Tautology is $\Omega(N^{1-\epsilon})$.*

Remark 1. Corollary 8 is interesting in the sense of pushing the rank lower bound nearly to being linear. However, the corresponding Tseitin tautology has a description of $n2^{n^{1/\epsilon}-1}$ many constraints and in this sense it does not seem very informative. In Section 5 we discuss polytopes defined on $O(n)$ constraints that still have linear Sherali-Adams rank.

5 Main Result: the "between Local and Global" Approach

In this section we are dealing with sets of $\bmod 2$ constraints on n variables. For a set M of $\bmod 2$ equations, define its constraint graph G_M, as the following bipartite graph from L to R. The left hand side L consists of a vertex for each constraint of P_M. The right hand side R consists of a vertex for every variable of P_M. There is an edge between a constraint-vertex j and a variable-vertex i, whenever variable i appears in constraint j. We abbreviate by G, instead of G_M, the corresponding constraint bipartite graph of P_M from the set of constraints L to the set of variables R. For $A \subseteq R$, we call a constraint $u \in L$, A-dominated if $\Gamma(u) \subseteq A$. We denote by G_{-A} the bipartite subgraph of G that we get after removing A and all A-dominated constraints.

Below we will be interested in polytopes defined by well expanding constraints.

Definition 2. Consider a bipartite graph $G = (V, E)$ with partition L, R.

Edge expansion: The expansion of $X \subset L$ is the value $|\Gamma(X)|/|X|$. The edge expansion[2] of G is the minimum expansion over subsets of size at most $|V|/2$. G is (r, ϵ) expanding if the minimum edge expansion over subsets of L of size at most r is at least ϵ.

Boundary expansion: The boundary expansion of $X \subset L$ is the value $|\partial X|/|X|$, where $\partial X = \{u \in R : |\Gamma(u) \cap X| = 1\}$. G is (r, ϵ) boundary expanding if the minimum boundary expansion over subsets of L of size at most r is at least ϵ.

The next fact is easy to verify.

Fact 9. *A bipartite graph $G = (V, E)$ of maximum degree d that is (r, ϵ) expanding is also $(r, 2\epsilon - d)$ boundary expanding.*

5.1 Establishing the Local Agreement Property

For every $A \subset R$, we denote by $\mathcal{U}(A)$ the uniform distribution over all 0/1 assignments of A that is consistent with A, namely that doesn't contradict any A-dominated constraint.

[2]Note that since G is a bipartite graph, the edge expansion of any subset of L is the same as the standard vertex expansion.

Theorem 10. *Let $A \subseteq B \subseteq R$ such that (i) G_{-A} is (ξ, ϵ) boundary expanding for some $\epsilon > 0$ and (ii) there are at most ξ many B-dominated constraints. Then $\mathcal{U}(A)$ and $\mathcal{U}(B)$ agree on A.*

Remark 2. Before we prove the theorem, we give an indication that in general we cannot expect the theorem to be true without conditions (i) and (ii), as was the case in the previous section. Indeed, consider two constraints $j_1, j_2 \in L$ that share a common variable $i \in R$. Set $B = \Gamma(j_1) \cup \Gamma(j_2)$, and $A = B \setminus \{i\}$. Assuming that the support of no other constraints is contained in B, we get that distribution $\mathcal{U}(A)$ sets any variable in A to 0/1 with probability 1/2 independently, but some of these assignments are not even extendible to B meaning that $\mathcal{U}(B)$ will assign them with probability zero.

The proof of Theorem 10 follows by Lemmas 11 and 12. Consult appendix for both proofs.

Lemma 11. *(Lemma 4.6 in [5]) Let $A \subseteq B \subseteq R$ as in the statement of Theorem 10. Then for every A-consistent assignment α there exist a B-consistent assignment β such that α, β agree on A.*

For an A-consistent assignment α we denote by $\mathcal{A}(\alpha)$ the set of B-consistent assignments that agree with α on A (which by Lemma 11 is nonempty). We have the following lemma.

Lemma 12. *(Lemma 4.5 in [5]) For any two A-consistent assignments α_1, α_2 we have $|\mathcal{A}(\alpha_1)| = |\mathcal{A}(\alpha_2)|$.*

Proof. (of Theorem 10) From Lemma 11, any A-consistent assignment is extendible to some B-consistent assignment. By Lemma 12 and for an A-consistent assignment α, the size of $\mathcal{A}(\alpha)$ does not depend on α. Hence, for any event $\mathcal{E}$ of A,

$$\frac{\sharp\ B\text{-consistent assignments that satisfy } \mathcal{E}}{\sharp\ B\text{-consistent assignments}} = \frac{\sharp\ A\text{-consistent assignments that satisfy } \mathcal{E}}{\sharp\ A\text{-consistent assignments}}$$

The first and second ratios equal $Pr_{\mathcal{U}(B)}[\mathcal{E}]$ and $Pr_{\mathcal{U}(A)}[\mathcal{E}]$ respectively. □

5.2 Towards Defining Appropriate Advice-Sets

We now need to define distributions over every set I of bounded size. Clearly, such sets will not always satisfy condition (i) of Theorem 10 (see Remark 2). Since on the one hand we can no longer use uniform distributions over assignments that are consistent with I, and on the other it seems hard to reason about a distribution that uses a totally different concept from uniformity, we take the following approach: For a set I we define a superset $\overline{I}$ such that $\overline{I}$ is "global enough" to contain sufficient information, or more specifically to satisfy condition (i) of Theorem 10, while it also is "local enough" to satisfy condition (ii) of the theorem. We then define $\mathcal{D}(I)$ as $\mathcal{U}(\overline{I})$. In other words, set I "inherits" the distribution from a simple uniform distribution over consistent assignments with respect to $\overline{I}$. The next algorithm is used to obtain this superset $\overline{I}$ which we call the *advice-set* of I, and is inspired by the correction procedure in [5].

Theorem 13. *Algorithm Advice, with internal parameters e_1, e_2, r, returns $\overline{Y} \subseteq R$ such that (a) $G|_{-\overline{Y}}$ is (ξ_Y, e_2) boundary expanding, (b) $\xi_Y > r - \frac{|Y|}{e_1 - e_2}$, and (c) $\overline{Y} \leq |Y| + e_2 r$,*

Proof. Suppose that the while loop terminates with $\xi = \xi_Y$. Then $\sum_{t=1}^{x} |M_t| = r - \xi_Y$. Since G is (r, e_1) boundary expanding, the set $M = \cup_{t=1}^{x} M_t$ has initially at least $e_1(r - \xi_Y)$ boundary neighbors. During the execution of the while loop, each set M_t has at most $e_2|M_t|$ boundary neighbors. Therefore, at the end of

Algorithm 1 Advice

The input is an (r, e_1) boundary expanding bipartite graph G, some $e_2 \in (0, e_1)$, and some $Y \subseteq R$, $|Y| < (e_1 - e_2)r$, with some order $Y = \{y_1, \ldots, y_i\}$.

Initially set $\overline{Y} \leftarrow \emptyset$, $t \leftarrow 0$ and $\xi \leftarrow r$
While $t \leq |Y|$ **do**
  $t \leftarrow t + 1$
  $\overline{Y} \leftarrow \overline{Y} \cup \{y_t\}$
  If $G_{-\overline{Y}}$ is not (ξ, e_2) boundary expanding **then**
    Find maximal $M_t \subset L$, $|M_t| \leq \xi$ such that $|\partial M_t| \leq e_2|M_t|$
    $\overline{Y} \leftarrow \overline{Y} \cup \partial M_t$
    $\xi \leftarrow \xi - |M_t|$
Return $\overline{Y}$

---

the procedure M has at most $e_2(r - \xi_Y)$ boundary neighbors. It follows that $|Y| + e_2(r - \xi_Y) > e_1(r - \xi_Y)$, which implies (b).

From the bound size of Y we know that $\xi_Y > 0$. In particular, ξ remains positive throughout the execution of the while loop. Next we identify a loop invariant: $G|_{-\overline{Y}}$ is (ξ, e_2) boundary expanding.

Indeed, note that the input graph G is (ξ, e_1) boundary expanding. At step t consider the set $\overline{Y} \cup \{y_t\}$, and suppose that $G_{-(\overline{Y} \cup \{y_t\})}$ is not (ξ, e_2) boundary expanding. We find maximal M_t, $|M_t| \leq \xi$, such that $|\partial M_t| \leq e_2|M_t|$. We claim that $G_{-(\overline{Y} \cup \{y_t\} \cup \partial M_t)}$ is $(\xi - |M_t|, e_2)$ boundary expanding (recall that since ξ remains positive, $|M_t| < \xi$). Now consider the contrary. Then, there must be $M' \subset L$ such that $|M'| \leq \xi - |M_t|$ and such that $|\partial M'| \leq e_2|M'|$. Consider then $M_t \cup M$ and note that $|M_t \cup M| \leq \xi$. More importantly $|\partial(M_t \cup M)| \leq e_2|M_t \cup M|$, and therefore we contradict the maximality of M_t; (a) follows.

Finally note that $\overline{Y}$ consists of Y union the boundary neighbors of all M_t. From the arguments above, the number of those neighbors does not exceed $e_2(r - \xi_Y)$ implying (c). □

5.3 Proving Rank Lower Bounds for Linear Systems with the Expansion Property

We are ready to state and prove the main theorem. For this we need to define appropriate distributions of 0/1 assignments. For a set $I \subseteq R$ we define

$$\mathcal{D}(I) = \mathcal{U}(\overline{I}),$$

where $\overline{I}$ is obtained by algorithm Advice of Section 5.2, and $\mathcal{U}$ is as defined in Section 5.1.

Theorem 14. *(Main) Consider a polytope P on n variables defined by constraints of support at most d, and such that the constraint bipartite graph G_P is (r, e) boundary expanding and $(r, 1)$ expanding. Then the Sherali-Adams rank of P is at least $cr - d - 1$, where $c = \min\{e/8, 1/8\}$.*

Proof. Consider $I \subseteq R$, $|I| \leq cr - d$, and a constraint of support $J \subseteq R$. Run algorithm Advice of Section 5.2 to obtain the advice-sets $\overline{I}$ and $\overline{I \cup J}$, with parameters $e_1 = e, e_2 = c$ (note that $|I \cup J| \leq cr < (e_1 - e_2)r$ as required by the algorithm). For every $I \subseteq R$, with $|I| \leq cr$, we define $\mathbf{z} \in \mathbb{R}^{\mathcal{P}^n_{cr}}$ as

$$z_I = \Pr_{\mathcal{D}(I)}[\text{all elements of } I \text{ are } 1].$$

The theorem follows after checking that conditions of Proposition 1 are met. For this we need to show that both local agreement and feasibility are satisfied. Feasibility is directly implied by the definition of distributions. To show that our distributions meet the local agreement property, we need to show that $\mathcal{D}(I)$ and $\mathcal{D}(I \cup J)$ agree on I.

First we show that $\mathcal{D}(I \cup J)$ and $\mathcal{U}(\overline{I} \cup \overline{I \cup J})$ agree on $\overline{I \cup J}$. By Theorem 13 we have that (a) $G_{-\overline{I \cup J}}$ is $(\xi_{I \cup J}, c)$ boundary expanding, (b) $\xi_{I \cup J} \geq r - \frac{|I \cup J|}{e-c} \geq \frac{6}{7}r$, and (c) $|\overline{I} \cup \overline{I \cup J}| < 2(|I \cup J| + cr) \leq 4cr < r/2$.

Set $A = \overline{I \cup J}$ and $B = \overline{I} \cup \overline{I \cup J}$. We claim that there exist at most $r/2$ many B-dominated constraints. Indeed, if there was $X \subseteq L$ with $\Gamma(X) \subseteq B$ and $|X| \geq r/2$, then there should be $X_0 \subseteq X$, with $|X_0| = r/2$ and $\Gamma(X_0) \subseteq B$. Since G is $(r, 1)$ expander, we should have $|X_0| \leq |B|$, or in other words $|X_0| < r/2$, a contradiction.

We are eligible to apply Theorem 10, because $A \subseteq B \subseteq R$, G_{-A} is $(6r/7, c)$ boundary expanding, and there are at most $r/2$ many B-dominated constraints. It follows that $\mathcal{D}(I \cup J)$ and $\mathcal{U}(\overline{I} \cup \overline{I \cup J})$ agree on $\overline{I \cup J}$. Similarly we show that $\mathcal{D}(I)$ and $\mathcal{U}(\overline{I} \cup \overline{I \cup J})$ agree on $\overline{I}$. It follows that $\mathcal{D}(I)$ and $\mathcal{D}(I \cup J)$ agree on $\overline{I} \cap \overline{I \cup J}$ and therefore on I as well. $\square$

All rank lower bounds follow immediately from Theorem 14 by considering appropriate instances. Following the lines of [5], we consider good expanders of bounded degree.

Definition 3. Consider the sample space of all $2\binom{n}{k}$ linear $\bmod 2$ equations on n variables, each involving exactly k variables. Define the probability distribution $\mathcal{M}_m^{k,n}$ on linear systems of equations mod 2 that chooses m such equations uniformly and independently.

Consider also the sample space of all $2^k\binom{n}{k}$ clauses on n variables, each involving exactly k literals. Define the probability distribution $\mathcal{N}_m^{k,n}$ of CNF that consists of m uniform and independent choices of such random clauses.

For every Δ, ϵ, k, there exist $\alpha > 0$ such that if F is chosen from $\mathcal{M}_{\Delta n}^{k,n}$, then the constraint bipartite graph G_F is almost surely $(\alpha n, k-1-\epsilon)$ expanding [8]. Consequently, what follows is an immediate corollary of Theorem 14 and follows the same line of arguments as in [5] (Corollary 4.10).

Corollary 15.

1. *Let H be a degree d expander with edge expansion $c \geq 1$. Then, the Sherali-Adams rank of TS_H is at least $\frac{n}{16} - d - 1$.*

2. *Let k, Δ be constants such that $k \geq 3$ and Δ is sufficiently large. Let F be a system of linear equalities drawn uniformly from $\mathcal{M}_{\Delta n}^{k,n}$. Then almost surely the linear relaxation associated with F has Sherali-Adams rank $\Omega(n)$.*

3. *Let k, Δ be constants such that $k \geq 3$ and Δ is sufficiently large. Let C be a system of CNF formulas drawn uniformly from $\mathcal{N}_{\Delta n}^{k,n}$. Then almost surely the linear relaxation associated with C has Sherali-Adams rank $\Omega(n)$.*

Proof.

1. Notice that when we view Tseitin Tautology over H as a set M of linear equations mod 2, G_M is just the incidence graph of H. This means that the size of the edge expansion of a subset of vertices S is exactly the boundary expansion of the set S in G_M (if e is cut by S, it means that the vertex associated with e is in the boundary of the set S in G_M, and vice versa). Therefore G_M is $(n/2, c)$ boundary expanding. Theorem 14 immediately implies the result now.

2. For big enough constant Δ, $F \sim \mathcal{M}^{k,n}_{\Delta n}$ is unsatisfiable with high probability. Since the graph associated with F is almost surely $(\alpha n, k-1-\epsilon)$ expanding, we get by Fact 9 that it is also $(\alpha n, k-2-2\epsilon)$ boundary expanding. For small enough ϵ and for $k \geq 3$, the boundary expansion remains positive. Again, this shows that Theorem 14 applies.

3. Start with a k-CNF formula $C \sim \mathcal{N}^{k,n}_{\Delta n}$ which for big enough Δ is unsatisfiable with high probability. Convert C to the k-CNF formula C' as follows: for every clause with odd (even) many positive literals, add all clauses on the same variables with odd (even) many positive literals. The resulting formula C' is unsatisfiable with high probability and it has Sherali-Adams rank not more than the Sherali-Adams rank of C. Now for C chosen from $\mathcal{N}^{k,n}_{\Delta n}$, we have that $G_{C'}$ is a bipartite constraint graph of some F chosen from $\mathcal{M}^{k,n}_{m}$. As before, $G_{C'}$ is $(\alpha n, k-2-2\epsilon)$ boundary expanding, and the claim follows.

□

6 Integrality Gaps for Linear Number of Rounds of Sherali-Adams

Theorem 14 is quite general, and allows us to obtain fairly easily a few integrality gap results.

6.1 Tight Integrality Gaps for MAX-k-LIN and MAX-k-SAT

For a system $\{f_1, \ldots, f_m\}$ of m many $\bmod 2$ equations over n variables and with support k, we denote by MAX-k-LIN the optimization problem of determining the maximum number of constraints that can be simultaneously satisfied. The corresponding decision problem is NP-hard and optimal inapproximability results are known [16].

Given a set $F = \{f_1, \ldots, f_m\}$ of m many $\bmod 2$ equations over n variables, let f_t be $\sum_{i \in I_t} x_i \equiv b_t \mod 2$, $I_t \subseteq [n]$. For every $t = 1, \ldots, m$ introduce a new variable y_t. Then the next optimization problem is an exact formulation of MAX-k-LIN.

$$\begin{array}{lll} \max & \sum_{t=1}^{m} y_t & \\ \text{s.t.} & y_t + \sum_{i \in I_t} x_i \equiv b_t + 1 \mod 2 & t = 1, \ldots m \\ & x_i, y_t \in \{0,1\} & i = 1, \ldots n,\ t = 1, \ldots m \end{array} \tag{4}$$

As explained in Section 2.3, the constraints of (4) can be naturally transformed to an equivalent 0/1 integer program. By further relaxing the variables x_i, y_t to assume values in $[0, 1]$, we obtain the natural linear program relaxation for F, that we denote by K_F. If instead we start with a system of k-CNF formulas, we define the corresponding optimization problem MAX-k-SAT. Following the same formulation, we obtain K_C, the linear program relaxation for MAX-k-SAT.

Given a set F of m many $\bmod 2$ equations, a random 0/1 assignment satisfies $m/2$ constraints. Similarly for a set C of m many k-CNF formulas, a random Boolean assignment satisfies $\left(1 - \frac{1}{2^k}\right) m$ many of them. Consequently, the integrality gap for K_F, K_C is at most 2 and $\frac{2^k}{2^k-1}$ respectively. As in [5], the next theorem establishes optimal and unconditional inapproximability lower bounds, (see Theorems 5.1 and 5.2 in [5] for complementary justification). Again, Theorem 16 follows as a corollary of Theorem 14.

Theorem 16.

1. *For all $k \geq 3$ and for every $\epsilon > 0$, there exist Δ, such that if $F \sim \mathcal{M}^{k,n}_{\Delta n}$, then almost surely the integrality gap of K_F after $\Omega(n)$ Sherali-Adams rounds is $2 - \epsilon$.*
2. *For all $k \geq 3$ and for every $\epsilon > 0$, there exist Δ, such that if $C \sim \mathcal{N}^{k,n}_{\Delta n}$, then almost surely the integrality gap of K_C after $\Omega(n)$ Sherali-Adams rounds is $\frac{2^k}{2^k-1} - \epsilon$.*

Proof.

1. As noted in [5] (using standard concentration arguments), for every $\epsilon > 0$ there exist $\Delta > 0$ such that for big enough n and for $F \sim \mathcal{M}^{k,n}_{\Delta n}$, no more than $1/2 + \epsilon$ equations are simultaneously satisfied in K_F. Set $y_t = 1$ and consider the distribution of valid solutions used in Theorem 14 and Corollary 15.It follows that the integrality gap of K_F is at least $2 - \epsilon'$ after $\Omega(n)$ Sherali-Adams tightenings.
2. As above, for every $\epsilon > 0$ there exist $\Delta > 0$ such that for big enough n and for $C \sim \mathcal{N}^{k,n}_{\Delta n}$, no more than $\frac{2^k-1}{2^k} + \epsilon$ clauses are simultaneously satisfied in C. Again set $y_t = 1$, and the same argument shows that the integrality gap is at least $\frac{2^k}{2^k-1} - \epsilon'$ after $\Omega(n)$ Sherali-Adams tightenings.

□

6.2 A 7/6 Integrality Gap for VERTEXCOVER

Given a graph $G = (V, E)$, a subset of V touching all edges is called a vertex cover. Determining the vertex cover of minimum size is one of the classic NP-hard problems. Our goal here is to study the performance of Sherali-Adams when applied to the standard LP for VERTEXCOVER (for the definition of the VERTEX-COVER polytope consult any of [2, 29, 26], [25, 12] and [6], who showed different types of integrality gaps in the Lovász-Schrijver , Lovász-Schrijver with PSDness and Sherali-Adams systems respectively).

Following the work of Feige and Ofek [10] and Schoenebeck et al. [25] we consider INDEPENDENT-SET instances obtained by random 3-LIN systems through the classical FGLSS reduction. Given a 3-LIN system σ we construct a graph H_σ as follows: for every equation C of σ we introduce four vertices corresponding to the partial assignments that satisfy C; two vertices are connected when their corresponding partial assignments are contradicting. Note that there is a natural correspondence between assignments variables to independent sets, where the number of equations satisfied equals the size of the independent set.

For $H_\sigma = (V, E)$ we define a good probability ensemble on small subsets of V that satisfy the conditions of Proposition 1. We obtain such an ensemble almost mechanically using (i) the reduction $\sigma \mapsto H_\sigma$, and (ii) the tools developed in Section 5. For a vertex i of H_σ, denote by $\lambda(i)$ the set of three variables of σ associated with i and, for a set of vertices I, let $\Lambda_I = \cup_{i \in I} \lambda(i)$. We now relate the above to the setting of

Section 5. Associate with σ the constraint bipartite graph G_σ on sets L and R. For $\Lambda_I \subseteq R$, consider the distribution $\mathcal{D}(\Lambda_I)$ of valid 0/1 assignments on R, as defined in Section 5.3. For the Sherali-Adams system applied to the standard LP relaxation of INDEPENDENT-SET we define

$$z_I = Pr_{\mathcal{D}(\Lambda_I)}[\text{ all variables in } \Lambda_I \text{ are assigned consistently with all } i \in I]. \tag{5}$$

As was mentioned, our starting point are random instances of 3-LIN systems. More precisely, we use the distribution $\mathcal{M}_m^{3,n}$ (see Definition 3), restricted to instances where every two equations share at most one variable; call this distribution $\mathcal{L}_m^{3,n}$. The same distribution was used by Schoenebeck et al. who showed (Lemma 5 in [25]) that a constant Δ can be chosen so that this restriction does not change significantly two important properties of $\mathcal{M}_{\Delta n}^{3,n}$: (i) expansion of the corresponding bipartite graph is still large enough and (ii) no more than $1/2+\epsilon$ constraints can be satisfied in σ, for some $\epsilon > 0$ that is arbitrary small when Δ is set to be sufficiently large.

Putting everything together, we start with a random 3-LIN system $\sigma \sim \mathcal{L}_{\Delta n}^{3,n}$, and we define $H_\sigma = (V, E)$ with $|V| = N = 4\Delta n$. By the above, H_σ has no vertex cover smaller than $N-(1/2+\epsilon)N/4$ almost surely. Also notice that $|\Lambda_I| \leq 3|I|$ for $I \subseteq V$, and therefore for $\beta > 0$, sets I of size $\leq \beta N$, we get that Λ_I is of size at most $3\beta N = 12\beta n$, and for β small enough this will allow us to use Theorem 14. It follows that the distribution ensemble used in defining the lifted variables (5) of the INDEPENDENT-SET polytope satisfies the local agreement property. Feasibility is implied by the reduction from σ to H_σ. To conclude we only need to notice that if $V = \{i_1, \ldots, i_N\}$, then the vector $(z_{\{i_1\}}, \ldots, z_{\{i_N\}})$ survives $\Omega(N)$ Sherali-Adams tightenings of the INDEPENDENT-SET polytope, or equivalently $(1-z_{\{i_1\}}, \ldots, 1-z_{\{i_N\}})$ survives $\Omega(N)$ Sherali-Adams tightenings of the VERTEXCOVER polytope.

Finally, consider $i \in V$. The expansion we guaranteed for σ implies that in the graph G_σ, we have that $G_{-\lambda(i)}$ has positive boundary expansion, and hence the advice-set $\overline{\lambda(i)}$ is $\lambda(i)$ itself. It follows that $z_{\{i\}} = 1/4$, since i is one of the 4 satisfying assignments of the (unique) $\lambda(i)$-dominated constraint of G_σ. We just sketched the proof of the next theorem.

Theorem 17. *For every $\epsilon > 0$, there exist a family of graphs on N vertices, for which the* VERTEXCOVER *polytope has integrality gap $7/6-\epsilon$ after $\Omega(N)$ Sherali-Adams tightenings.*

In a recent result, Charikar, Makarychev and Makarychev [6] showed that for every $\epsilon > 0$, there exist $\gamma > 0$, where γ goes to 0 with ϵ, and an infinite family of graphs on N vertices, such that the integrality gap of the VERTEXCOVER polytope is $2-\epsilon$ after $\Omega(N^\gamma)$ Sherali-Adams tightenings. Comparing this result to ours, the tradeoff is in the tightness of the integrality gap (achieved by Charikar et. al.) vs the asymptotically tight (linear) number of rounds (achieved by the current result).

References

[1] Mikhail Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovász-Schrijver hierarchy. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 294–303. ACM, 2005.

[2] 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.

[3] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231 (electronic), New York, 2004. ACM.

[4] Albert Atserias, Maria Luisa Bonet, and Jordi Levy. On Chvatal rank and cutting planes proofs. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR03-041, 2003.

[5] Joshua Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory of Computing*, 2(4):65–90, 2006.

[6] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. Manuscript, 2007.

[7] Eden Chlamtac and Gyanit Singh. Improved approximation guarantees through higher levels of SDP hierarchies. In *Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX) 2008*, 2008.

[8] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. Assoc. Comput. Mach.*, 35(4):759–768, 1988.

[9] Uriel Feige and Robert Krauthgamer. The probable value of the Lovász-Schrijver relaxations for maximum independent set. *SICOMP: SIAM Journal on Computing*, 32(2):345–370, 2003.

[10] Uriel Feige and Eran Ofek. Random 3cnf formulas elude the lovasz theta function. *CoRR*, abs/cs/0603084, 2006.

[11] Wenceslas Fernández de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–61, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[12] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Integrality gaps of $2 - o(1)$ for Vertex Cover SDPs in the Lovász-Schrijver hierarchy. In *Proceedings of the 47th IEEE Symposium on Foundations of Computer Science*, pages 702–712, 2007.

[13] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.

[14] Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Exponential lower bound for static semi-algebraic proofs. In *Automata, Languages and Programming, 29th International Colloquium, ICALP 2002, Malaga, Spain, July 8-13, 2002, Proceedings*, volume 2380 of *Lecture Notes in Computer Science*, pages 257–268. Springer, 2002.

[15] Mark Hartmann. *Cutting planes and the complexity of the integral hull*. PhD thesis, Cornell University, Department of Operations research and Industrial Engineering, 1988.

[16] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859 (electronic), 2001.

[17] George Karakostas. A better approximation ratio for the Vertex Cover problem. In *Proceedings of the Thirty-Second International Colloquium on Automata, Languages and Programming*, pages 1043–1050, 2005.

[18] Howard Karloff and Uri Zwick. A 7/8-approximation algorithm for MAX 3SAT? In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science, FOCS'97 (Miami Beach, Florida, October 20-22, 1997)*, pages 406–415, 1997.

[19] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Integer programming and combinatorial optimization (Utrecht, 2001)*, volume 2081 of *Lecture Notes in Comput. Sci.*, pages 293–303. Springer, Berlin, 2001.

[20] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003.

[21] László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, May 1991.

[22] Avner Magen and Mohammad Moharrami. Sherali-Adams based polynomial approximation schemes for NP-hard problems on planar and minor-free graphs. Manuscript, 2008.

[23] Mark Rhodes. Rank lower bounds for the Sherali-Adams operator. In *Computation and Logic in the Real World, Third Conference on Computability in Europe, CiE 2007, Siena, Italy, June 18-23, 2007, Proceedings*, volume 4497 of *Lecture Notes in Computer Science*, pages 648–659. Springer, 2007.

[24] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *IEEE Symposium on Foundations of Computer Science, FOCS'08*, 2008.

[25] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. A linear round lower bound for Lovász-Schrijver SDP relaxations of vertex cover. In *IEEE Conference on Computational Complexity*, pages 205–216. IEEE Computer Society, 2007.

[26] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovász-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.

[27] 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.

[28] Iannis Tourlakis. Towards optimal integrality gaps for hypergraph vertex cover in the Lovász-Schrijver hierarchy. In *8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005*, volume 3624, pages 233–244. Springer, 2005.

[29] Iannis Tourlakis. New lower bounds for vertex cover in the Lovász-Schrijver hierarchy. In *Proceedings of the 21st IEEE Conference on Computational Complexity*, pages 170–182. IEEE Computer Society, 2006.

Appendix

Proof. (of Lemma 11)

Consider a set X of ξ many B-dominated constraints. It suffices to show that X admits a 0/1 solution. Note that X has positive boundary expansion.

First we describe inductively an 1-1 correspondence from $X = \{c_1, \ldots, c_\xi\}$ to ∂X. At step $i-1$, we will have a matching $(c_1, v_1), \ldots, (c_{i-1}, v_{i-1})$, where $v_1, \ldots, v_{i-1} \in \partial X - \partial(X \setminus \{c_1, \ldots, c_{i-1}\})$. At step $i \leq \xi$, the set $X \setminus \{c_1, \ldots, c_{i-1}\}$ is not empty, and so there exist $v_i \in \partial(X \setminus \{c_1, \ldots, c_{i-1})$. Choose $c_i \in X \setminus \{c_1, \ldots, c_{i-1}\}$ whose neighbor is v_i (note that v_i does not have any other neighbor).

To that end, set arbitrarily all variables in $\partial X \setminus \{v_1, \ldots, v_\xi\}$; set also variables $v_1, \ldots, v_\xi$ to 0/1 so that X is satisfied. □

Proof. (of Lemma 12)

For two distinct 0/1 assignments α_1, α_2 of A, we describe an 1-1 mapping from $\mathcal{A}(\alpha_1)$ to $\mathcal{A}(\alpha_2)$. Let $\beta_1 \in \mathcal{A}(\alpha_1)$ and $\beta_2 \in \mathcal{A}(\alpha_2)$ and define $f : \mathcal{A}(\alpha_1) \rightarrow \mathcal{A}(\alpha_2)$ such that for $\beta \in \mathcal{A}(\alpha_1)$ and for every $i \in B$, $f \circ \beta(i) = (\beta(i) + \beta_1(i) + \beta_2(i)) \mod 2$. We claim that f meets the desired property.

First note that indeed for every $\beta \in \mathcal{A}(\alpha_1)$, we have $f \circ \beta \in \mathcal{A}(\alpha_2)$. To see this, consider a B-saturated constraint $j \in L$. Then

$$\sum_{i \in \Gamma(j)} f \circ \beta(i) = \sum_{i \in \Gamma(j)} (\beta(i) + \beta_1(i) + \beta_2(i)) \mod 2.$$

Each one of the individual sums for β, β_1, β_2 add up to some odd number, as these assignments do not contradict any B-saturated constraint. Hence, $f \circ \beta$ does not contradict j.

Next we claim that $f \circ \beta$ agrees with α_2 on I. Indeed, since $\beta, \beta_1 \in \mathcal{A}(\alpha_1)$, they agree on every $i \in I$. Therefore $f \circ \beta(i) = \beta_2(i)$ and since $\beta_2 \in \mathcal{A}(\alpha_2)$ the claim follows.

Finally, consider $\beta, \beta' \in \mathcal{A}(\alpha_1)$ that assign different value for at least a variable $i \in J$. Then

$$f \circ \beta(i) - f \circ \beta'(i) \equiv \big(\beta(i) - \beta'(i)\big) \mod 2 \not\equiv 0 \mod 2.$$

We have shown that $|\mathcal{A}(\alpha_1)| \leq |\mathcal{A}(\alpha_2)|$. Similarly we show the other direction. □