On Semi-Algebraic Proofs and Algorithms

Noah Fleming[†]
UCSD
Memorial University

Mika Göös
EPFL

Stefan Grosser
McGill University

Robert Robere[†]
McGill University

January 19, 2022

Abstract

We give a new characterization of the Sherali-Adams proof system, showing that there is a degree-d Sherali-Adams refutation of an unsatisfiable CNF formula C if and only if there is an $\varepsilon > 0$ and a degree-d conical junta J such that $\mathrm{viol}_C(x) - \varepsilon = J$, where $\mathrm{viol}_C(x)$ counts the number of falsified clauses of C on an input x. Using this result we show that the *linear separation complexity*, a complexity measure recently studied by Hrubeš (and independently by de Oliveira Oliveira and Pudlák under the name of *weak monotone linear programming gates*), monotone feasibly interpolates Sherali-Adams proofs.

We then investigate separation results for $\mathrm{viol}_C(x) - \varepsilon$. In particular, we give a family of unsatisfiable CNF formulas C which have polynomial-size and small-width resolution proofs, but for which any representation of $\mathrm{viol}_C(x) - 1$ by a conical junta requires degree $\Omega(n)$; this resolves an open question of Filmus, Mahajan, Sood, and Vinyals. Since Sherali-Adams can simulate resolution, this separates the non-negative degree of $\mathrm{viol}_C(x) - 1$ and $\mathrm{viol}_C(x) - \varepsilon$ for arbitrarily small $\varepsilon > 0$. Finally, by applying lifting theorems, we translate this lower bound into new separation results between extension complexity and monotone circuit complexity.

---

1 Introduction

The Sherali-Adams hierarchy is a well-studied method in optimization that provides an *automatic* way to convert a linear program with "weak" approximation guarantees into a linear program with "strong" approximation guarantees. Each "level" of the Sherali-Adams hierarchy is defined by systematically adding new variables and inequalities to the original linear program — "lifting" it to a higher dimensional space — and then "projecting" it back down to a polytope contained within the original linear program. The Sherali-Adams hierarchy at level r results in a linear program with roughly $\binom{n}{\leq r}$ constraints and variables, and it is known that at level n the hierarchy converges to the *integral hull* of the starting polytope. Owing to its strength and its generality, much work has been spent on understanding the strength of the Sherali-Adams hierarchy and its subsystems when applied to NP-Hard optimization problems [1, 6, 15, 19, 38, 44, 46, 47, 50, 57], as well as its strength when it is treated as a *refutation system* in propositional proof complexity [2, 5, 7, 18, 21, 34, 36].

A powerful generalization of the Sherali-Adams hierarchy is the notion of an *extended formulation*, which was originally formulated by Yannakakis [59] and extended by Braun et al. [9] to nested pairs of polytopes. Given a pair of polytopes $P \subseteq Q \subseteq \mathbb{R}^n$, an *extended formulation* of the pair (P, Q) is a polytope $K \subseteq \mathbb{R}^m$ with $m \geq n$ along with a projection π such that $P \subseteq \pi(K) \subseteq Q$. Indeed, the Sherali-Adams hierarchy gives many examples of extended formulations: if we are given some linear program $Q \subseteq [0, 1]^n$ which gives a "weak" approximation guarantee to any point in the integral hull

$$\text{int}(Q) := \text{conv}\left\{x \in \mathbb{Z}^n \mid x \in Q\right\},$$

then each level of the Sherali-Adams hierarchy produces an extended formulation (K, π) such that

$$\text{int}(Q) \subseteq \pi(K) \subseteq Q.$$

We regard this as giving a "tighter" relaxation to the integral hull, when compared to the starting polytope Q (indeed, at n-th level of the hierarchy it turns out that $\pi(K) = \text{int}(Q)$).

More generally, for some polytopes $P \subseteq \mathbb{R}^n$ with *exponentially* many facets, it turns out to be possible to find an extended formulation $K \subseteq \mathbb{R}^m$ such that $\pi(K) = P$ but $m = \text{poly}(n)$ and K has only polynomially many facets [59] (this is certainly a boon if $P = \text{int}(Q)$ for some combinatorial optimization problem encoded by Q!). Given a polytope pair $P \subseteq Q$ we therefore let $\text{xc}(P, Q)$ denote the size (= number of facets) of the smallest extended formulation for the pair P, Q. After a breakthrough result by Fiorini et. al. [25], strong lower bounds have been shown for the extension complexity of polytopes associated with many standard NP-Hard optimization problems [9–11, 14, 25, 31, 43, 55, 56, 59]. All of these lower bound results crucially rely on the close relationship between the number of facets in any extended formulation and the *non-negative rank* of a certain related matrix [9, 59].

Theorem 1.1 (Factorization Theorem). *Let $P \subseteq Q \subseteq \mathbb{R}^n$ be polyhedral sets, let $v_1, \ldots, v_n$ be the vertices of P and let $a_1 \cdot x \leq b_1, \ldots, a_m \cdot x \leq b_m$ be linear inequalities in $\mathbb{R}^n$ describing the facets of Q. The size of the smallest extended formulation of (P, Q) is $\text{rank}^+(S_{P,Q}) \pm 1$ where $S_{P,Q}$ is the $n \times m$ matrix defined by $S_{P,Q}(i, j) = b_j - a_j \cdot v_i$.*

Indeed, works of Chan et. al. [14] , Göös et. al. [33], and Kothari, Raghavendra, and Steurer [43] have shown that for certain NP-Hard optimization problems, lower bounds on the size of *arbitrary* extended formulations follow immediately from lower bounds on the Sherali-Adams hierarchy. These results join a long line of *lifting theorems* in communication, proof, and circuit complexity [16, 17, 22, 23, 26, 29, 30, 32, 35, 51, 53, 58] which systematically relate the complexity of computations in "complicated" computational models with complexity in "simple" computational models.

1.1 Sherali-Adams as Proofs and Extended Formulations as Circuits

In this work we further the study of the Sherali-Adams hierarchy and its relationship with extended formulations, but approach it from a different point of view. In particular, we are interested in the "dual view" of Sherali-Adams as a *propositional proof system* (first considered in [18]), and in extended formulations as a device for *computing boolean functions*. Before we formally state our results, let us first describe these two perspectives.

A *conical junta* $\mathcal{J}$ is a non-negative linear combination of conjunctions over a set of $\{0, 1\}$-valued variables. Given an unsatisfiable CNF formula $\mathcal{C} = C_1 \wedge C_2 \wedge \cdots \wedge C_m$, a *Sherali-Adams refutation*[*] of $\mathcal{C}$ is given by a list of $m + 1$ conical juntas $\mathcal{J}_1, \ldots, \mathcal{J}_{m+1}$ such that

$$\sum_{i=1}^{m} -\overline{C}_i \mathcal{J}_i + \mathcal{J}_{m+1} = -1$$

where all operations are done in multilinear polynomial arithmetic (so, $x^2 = x$) over $\mathbb{R}$ and $\overline{C}_i$ is the negation of the clause C_i. Indeed, if such a list of conical juntas exists then the original CNF formula must indeed be unsatisfiable, as if x was a satisfying assignment then $\overline{C}_i(x) = 0$ for all C_i and the above expression would reduce to $-1 = \mathcal{J}_{m+1}(x) \geq 0$, a contradiction. In this way, Sherali-Adams is naturally viewed as a *proof system* for refuting unsatisfiable formulas, and we can discuss complexity measures of its proofs such as the *degree* (i.e. the maximum degree of any product $\overline{C}_i \mathcal{J}_i$ as a multilinear polynomial) and its *size*[†] (the number of distinct monomials occurring in the proof after expanding all products and before cancellations).

It is also quite natural to study *extended formulations* as non-uniform computation devices for boolean functions. The next definition was formally introduced by Hrubeš [40].

Definition 1.2. A *separating polytope* for $f : \{0, 1\}^n \to \{0, 1, *\}$ is a polytope $P \in \mathbb{R}^n$ such that $\operatorname{conv} f^{-1}(1) \subseteq P$ and $P \cap f^{-1}(0) = \emptyset$. We say a polyhedron $P \subseteq \mathbb{R}^n$ is *monotone* if $x \in P \Rightarrow y \in P$ whenever $x \leq y$, and if $P \subseteq \mathbb{R}^n$ we let $P^* := \{y \in \mathbb{R}^n : \exists x \in P : x \leq y\} \supseteq P$ be the *monotone closure* of P. A *monotone separating polytope* for f is a polytope P such that P^* is a separating polytope for f.

Observe that if we have such a separating polytope K, then when given $x \in \{0, 1\}^n$ we can test if $f(x) = 1$ simply by testing if $x \in K$. One natural way to construct such separating polytope is as follows, and was introduced by Hrubeš [39] and studied independently by Göös, Jain and Watson [31]. Given a boolean function $f : \{0, 1\}^n \to \{0, 1\}$ define the polytope $Q_{f,1}$ by the $|f^{-1}(0)|$ linear inequalities

$$\forall y \in f^{-1}(0) : \sum_{i=1}^{n} x_i(1 - y_i) + (1 - x_i)y_i \geq 1$$

where we note that $\sum_i x_i(1 - y_i) + (1 - x_i)y_i =: h(x, y)$ is exactly the hamming distance between two $\{0, 1\}$-valued vectors x, y. It is easy to see that $x \in Q_{f,1}$ for every $x \in f^{-1}(1)$ since x must differ from every $y \in f^{-1}(0)$ on some coordinate; similarly, for any $y \in f^{-1}(0)$ we have that $y \notin Q_{f,1}$ as the hamming distance from y to itself is 0. It therefore follows that any extended formulation of the pair $(\operatorname{conv} f^{-1}(1), Q_{f,1})$ yields a separating polytope as we described above. One can further specialize

---

[*]This encoding of Sherali-Adams is slightly different from the "usual" definition of Sherali-Adams as a refutation system of linear inequalities or polynomial equations, but is easily seen to be equivalent (cf. Claim 3.32 in [27])

[†]This measure is sometimes called the *monomial size* in other works on the Sherali-Adams hierarchy to differentiate from the bit-length of the encoding of the proof. As this is the natural notion of size for our purposes we simply call it *size*.
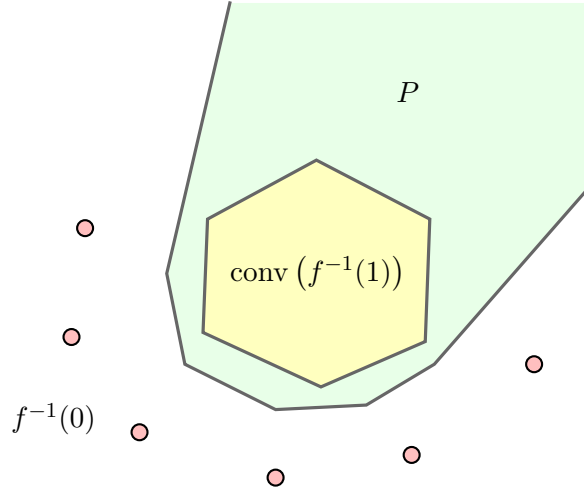
Figure 1: A monotone separating polytope.

this construction when the function $f : \{0, 1\}^n \to \{0, 1\}$ is *monotone* (recall f is monotone if $x \leq y$ $\Rightarrow f(x) \leq f(y)$, where the first inequality is interpreted coordinate-wise). In this case, we can simplify the description of $Q_{f,1}$ and still obtain a separating polytope. Specifically, if f is monotone then define the polytope $Q_{f,1}^+$ by the linear inequalities

$$\forall y \in f^{-1}(0) : \sum_{i=1}^n x_i(1 - y_i) \geq 1.$$

The fact that an extended formulation of $(\operatorname{conv} f^{-1}(1), Q_{f,1}^+)$ is a separating polytope (indeed, now we can even have a *monotone* separating polytopes) for monotone f follows by a similar argument as before, but uses the additional fact that for monotone functions, $f(x) = 1$ and $f(y) = 0$ if and only if there is a coordinate i such that $x_i = 1, y_i = 0$. We depict such a pair in Figure 1.

By using the Factorization Theorem (Theorem 1.1), we can relate the size of these extended formulations to the non-negative ranks of certain matrices. Namely, given $f : \{0, 1\}^n \to \{0, 1\}$ define the $f^{-1}(1) \times f^{-1}(0)$ matrix $S_f(x, y) := \sum_{i=1}^n x_i(1 - y_i) + (1 - x_i)y_i$, and its "monotone" counterpart $S_f^+(x, y) := \sum_{i=1}^n x_i(1 - y_i)$. Define the *separation complexity* quantities

$$\mathsf{sep}_1(f) := \operatorname{rank}^+(S_f - \mathbb{1}), \quad \mathsf{msep}_1(f) := \operatorname{rank}^+(S_f^+ - \mathbb{1})$$

where $\mathbb{1}$ represents the all-1s matrix of the appropriate complexity. By Theorem 1.1, these two quantities capture the size of the smallest extended formulations for the pairs $(\operatorname{conv} f^{-1}(1), Q_{f,1})$ and $(\operatorname{conv} f^{-1}(1), Q_{f,1}^+)$, respectively.

Now, it is natural to ask: are these interesting complexity measures of boolean computation? Hrubeš [39] and Göös, Jain and Watson [31] showed that the answer is *yes*: both of these quantities yield lower bounds on the *formula complexity* of computing f!

Theorem 1.3. *For any boolean function $f : \{0, 1\}^n \to \{0, 1\}$, we have $\mathsf{sep}_1(f) = O(n \cdot \mathsf{F}(f))$ where $\mathsf{F}(f)$ is the size of the smallest boolean formula computing f. Furthermore, if f is monotone, then $\mathsf{msep}_1(f) = O(n \cdot \mathsf{mF}(f))$ where $\mathsf{mF}(f)$ is the size of the smallest* monotone *boolean formula computing f.*
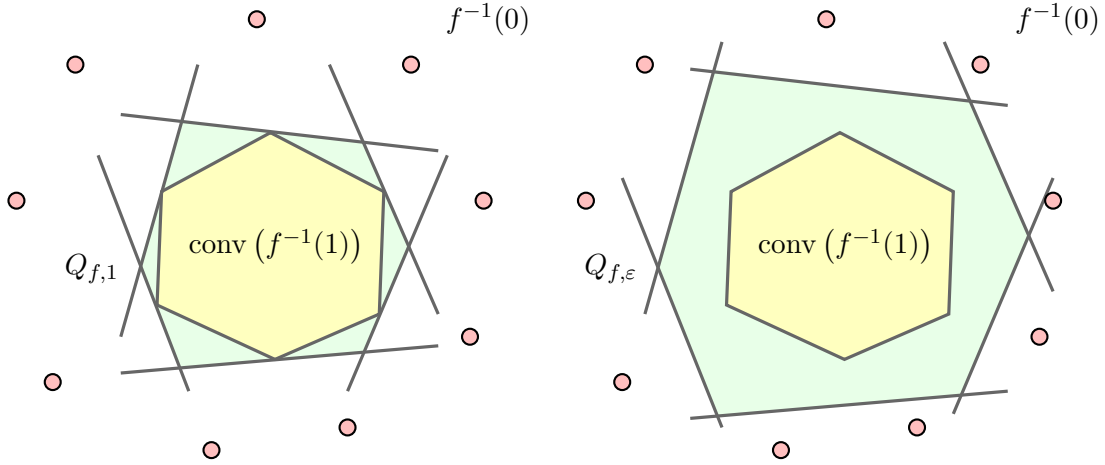
3

Figure 2: The left depicts the polytope pair $\operatorname{conv} f^{-1}(1)$ and $Q_{f,1}$; the right depicts $\operatorname{conv} f^{-1}(1)$ and $Q_{f,\varepsilon}$ for some $\varepsilon < 1$. By setting $\varepsilon < 1$ it is potentially easier to project a relatively low-facet polytope separating the 0s from the 1s of f.

Furthermore, observe the following. If our goal is just to separate the 1s of f from the 0s of f by means of a polytope we can weaken the inequalities

$$\sum_{i=1}^{n} x_i(1 - y_i) + (1 - x_i)y_i \geq 1, \quad \forall y \in f^{-1}(0).$$

Since $x \in f^{-1}(1)$ and $y \in f^{-1}(0)$ are boolean strings, by integrality it suffices to have the inequalities

$$\forall y \in f^{-1}(0) : \sum_{i=1}^{n} x_i(1 - y_i) + (1 - x_i)y_i \geq \varepsilon$$

for any $\varepsilon > 0$ (note we must have $\varepsilon > 0$, since if $\varepsilon = 0$ then $y \in f^{-1}(0)$ will occur inside of the polytope). With this in mind, define the quantities

$$\mathsf{sep}(f) := \min_{\varepsilon > 0} \operatorname{rank}^+(S_f - \varepsilon \mathbb{1}), \quad \mathsf{msep}(f) := \min_{\varepsilon > 0} \operatorname{rank}^+(S_f^+ - \varepsilon \mathbb{1}),$$

which generalize $\mathsf{sep}_1(f)$ and $\mathsf{msep}_1(f)$ in the natural way. These quantities were recently studied by Hrubeš [40], where it was shown that they lower bound the size of *boolean circuits* computing f.

Theorem 1.4. *For any boolean function $f : \{0,1\}^n \to \{0,1\}$ we have $\mathsf{sep}(f) = O(\mathsf{C}(f) + n)$, where $\mathsf{C}(f)$ is the size of the smallest boolean circuit computing f. Furthermore, if f is monotone, then $\mathsf{msep}(f) = O(\mathsf{mC}(f) + n)$, where $\mathsf{mC}(f)$ is the size of the smallest monotone boolean circuit computing f.*

We also note that the quantity $\mathsf{msep}(f)$ was independently studied by Pudlák and de Oliveira Oliveira, where it was captured by a model of computation they called *weak monotone linear programming gates* [20].

4

To summarize: the Sherali-Adams hierarchy can be interpreted both as a particular family of extended formulations in optimization, and also as a natural family of proof systems for refuting unsatisfiable formulas. If we consider extended formulations more generally, then it is natural to interpret them as a computational devices whose complexities are related to standard boolean circuit and formula models. In this way, the known "lifting theorems" from Sherali-Adams proofs to extended formulations fall naturally in line with other proof-to-circuit lifting theorems [22, 29, 53], as we will show next.

1.2 Our Results

Normal Form for Sherali-Adams. In this work, we systematically relate the complexity of Sherali-Adams proofs and its fragments with the linear separation quantities $\mathsf{sep}_1(f), \mathsf{msep}_1(f), \mathsf{sep}(f)$, and $\mathsf{msep}(f)$. Our first main result is the following novel "normal form" for Sherali-Adams proofs. Given a CNF formula $F = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ let $\mathsf{viol}_{\mathcal{C}}(x) = \sum_{i=1}^{m} \overline{C}_i(x)$ denote the number of falsified clauses on an input x. Furthermore, if $f : \{0,1\}^n \to \mathbb{R}_{\geq 0}$ is a non-negative real-valued boolean function then let $\deg^+(f)$ denote the minimum degree of a conical junta $\mathcal{J}$ such that $f = \mathcal{J}$.

Theorem 1.5. *For any unsatisfiable CNF formula $\mathcal{C}$, if there is a Sherali-Adams refutation of $\mathcal{C}$ with degree d and size s, then there is an $\varepsilon > 0$ and a degree d, size $s2^d$ conical junta $\mathcal{J}$ such that*

$$\mathsf{viol}_{\mathcal{C}} - \varepsilon = \mathcal{J}.$$

Consequently, if $\mathsf{SA}(\mathcal{C})$ is the minimum degree of any Sherali-Adams refutation of $\mathcal{C}$, then

$$\mathsf{SA}(\mathcal{C}) = \min_{\varepsilon > 0} \deg^+(\mathsf{viol}_{\mathcal{C}} - \varepsilon).$$

Several remarks on this theorem are in order. First, although we have stated this theorem for CNF formulas, a similar result immediately follows for arbitrary boolean CSPs. This is because if $P : \{0,1\}^k \to \{0,1\}$ is an arbitrary boolean CSP then we can represent P as a width-k unique DNF of its 1-inputs. This means that $\neg P = \sum_i D_i$ for some conjunctions D_i (specifically, conjunctions that recognize the 0s of P). By substituting these sums for the predicates we can immediately deduce a more general theorem for refuting arbitrary boolean CSPs.

Second, if there is a conical junta $\mathcal{J}$ such that $\mathsf{viol}_{\mathcal{C}} - \varepsilon = \mathcal{J}$ then note that we immediately obtain a Sherali-Adams refutation by rearranging the expression:

$$\sum_{i=1}^{m} -\frac{\overline{C}_i}{\varepsilon} + \frac{\mathcal{J}}{\varepsilon} = -1.$$

Thus this truly is a normal form that preserves the degree and the size (although, the size is only preserved up to a 2^d factor).

Third, one should note the similarities between Theorem 1.5 and the definitions of $\mathsf{sep}(f)$ and $\mathsf{msep}(f)$ — they are identical up to the substitution of "$\deg^+$" for "rank^+" and "$\mathsf{viol}_{\mathcal{C}}$" for "$S_f/S_f^+$". Looking forward, this similarity turns out to be quite important for the rest of our results.

Finally, we remark that the fragment of Sherali-Adams corresponding to $\deg^+(\mathsf{viol}_{\mathcal{C}} - 1)$ has also been studied in the literature as an object of interest. Göös, Jain and Watson studied it under the name of the "$\exists - 1$ Game", in which they proved degree lower bounds for Tseitin formulas [31]. Filmus, Mahajan, Sood, and Vinyals studied a further restriction in which the coefficients are required to be integers; they showed that this restriction was closely related to the complexity of MaxSAT resolution [24].

Feasible Interpolation for Sherali-Adams. Next, inspired by the normal-form theorem, we further develop the connection between Sherali-Adams and separation complexity by way of *monotone feasible interpolation*.

The feasible interpolation method relates the complexity of proofs to the complexity of computational models. Suppose we are given an unsatisfiable CNF formula of the form $A(x, z) \wedge B(y, z)$. Then, under an assignment $z \mapsto \alpha$, at least one of the formulas $A(x, \alpha)$, $B(y, \alpha)$ is unsatisfiable. Thus, we can associate with $A \wedge B$ a partial function $I : \{0, 1\}^z \to \{0, 1\}$, known as an *interpolant*, satisfying

$$I(\alpha) = \begin{cases} 0 & \text{if } A(x, \alpha) \text{ is satisfiable} \\ 1 & \text{if } B(y, \alpha) \text{ is satisfiable} \end{cases}$$

In its original form, introduced at this level of generality in the classic work of Krajíček [45], we say that a proof system P has *feasible interpolation* if we can extract a small computation of the interpolant I in some computational model from any small P-proof of $A \wedge B$. Furthermore, if $A \wedge B$ satisfies a certain monotonicity property, namely that all z-literals occur only negatively in A, then the interpolant function I is monotone and we say that P has *monotone* feasible interpolation if we can extract from a P-proof of $A \wedge B$ a computation in some monotone computational model. Instantiations of the method of (monotone) feasible interpolation have led to a number of important lower bounds in proof complexity. Razborov [52] showed that from proofs in certain fragments of bounded arithmetic one could extract Boolean circuits and used this to establish conditional unprovability of $P \neq \mathsf{NP}$ in these systems. The first lower bounds for Cutting Planes were established by Pudlák [48], who showed that proofs in this system gave rise to monotone real circuits; this built upon earlier work by Bonet, Pitassi and Raz [8] who showed that low-weight Cutting Planes gave rise to monotone circuits. Pudlák showed that span programs monotone feasibly interpolate Nullstellensatz [49], and de Oliveira Olivera and Pudlák showed that proofs in the Lovász-Schijver system convert to monotone linear programming circuits [20].

Recently, Hrubeš and Pudlák [41] and Fleming et al. [28] showed that the method of monotone feasible interpolation could be generalized to work for *arbitrary* unsatisfiable CNF formulas, not only for split formulas. They showed that from a small Cutting Planes proof of an unsatisfiable CNF formula $\mathcal{C}$ one could extract small monotone circuit computing an associated monotone function, termed the *unsatisfiability certificate* by Hrubeš and Pudlák [41]. (Fleming et. al. used a conceptually different, though ultimately equivalent, function [28]). If C is a clause and X is a subset of its variables we let C^X denote the subclause of C containing only literals over X.

Definition 1.6. Let $\mathcal{C} = C_1 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF formula and let (X, Y) be any partition of its variables. The *unsatisfiability certificate* associated with $\mathcal{C}$ and (X, Y) is the partial function $\mathrm{cert}_{\mathcal{C}}^{(X,Y)} : \{0, 1\}^m \to \{0, 1, *\}$ defined as

$$\mathrm{cert}_{\mathcal{C}}^{(X,Y)}(\alpha) = \begin{cases} 0 & \text{if } \{C_i^X : C_i \in \mathcal{C}, \alpha_i = 1\} \text{ is satisfiable,} \\ 1 & \text{if } \{C_i^Y : C_i \in \mathcal{C}, \alpha_i = 0\} \text{ is satisfiable,} \end{cases}$$

When it is clear from context, we may suppress the partition (X, Y) or the underlying CNF formula $\mathcal{C}$.

Our second main result is the following monotone feasible interpolation theorem for Sherali-Adams proofs[‡].

Theorem 1.7. *Let $\mathcal{C} = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ be any unsatisfiable CNF formula. If there is a Sherali-Adams proof of $\mathcal{C}$ with size s then for any partition (X, Y) of the variables of $\mathcal{C}$ we have $\mathsf{msep}(\mathsf{cert}_\mathcal{C}) = O(s^2)$.*

Several remarks on this theorem are in order. First, as was shown by Pudlák and Hrubeš [41] a standard interpolation theorem for split formulas follows from this interpolation (simply by "resolving away" the z variables and taking the natural partition of variables). Second, prior to this result there was no monotone feasible interpolation theorem known for Sherali-Adams. A recent work of Hakoniemi gave the first feasible interpolation theorem for Sherali-Adams [37], but, his feasible interpolation result only applied to the "standard" split formulas $A(x, z) \wedge B(y, z)$ and it only gave a small non-monotone Boolean circuit for the interpolant [37]. Finally, as we have remarked before, the model $\mathsf{msep}(f)$ was recently studied by de Oliveira Olivera and Pudlák under the name *weak monotone linear programming gates* [20]. They also introduced a model that they called *strong monotone linear programming gates*, and showed that "circuits" created out of strong monotone linear programming gates can monotone feasibly interpolate Lovász-Schrijver proofs. Pudlák and de Oliveira Olivera left as an open problem whether or not strong monotone linear programming gates can be efficiently simulated by weak monotone linear programming gates. We believe our feasible interpolation result makes this question more interesting, in light of the lack of separations between Lovász-Schrijver and Sherali-Adams proofs — the only separation known is due to Atserias and Ochremiak [3], who showed that degree-6 Lovász-Schrijver has polynomial-size refutations of Tseitin principles (and, furthermore, there are no good size lower bounds on Lovász-Schrijver proofs at all!).

Separation Results. Finally, we prove new separation results between the proof and circuit models described above. In particular, we are interested in the value of $\varepsilon > 0$ that is required in the definition of $\mathsf{msep}(f) := \min_{\varepsilon > 0} \mathrm{rank}^+(S_f^+ - \varepsilon \mathbb{1})$ and in our new characterization of Sherali-Adams degree $\min_{\varepsilon > 0} \deg^+(\mathsf{viol}_\mathcal{C} - \varepsilon)$. For instance: is it possible that we never need to take $\varepsilon < 1$? Or, in other words, is it possible that $\mathsf{msep}(f) = \mathsf{msep}_1(f)$, and that Sherali-Adams is already captured by the conical junta degree of $\mathsf{viol}_\mathcal{C} - 1$?

We give a negative answer to these questions. First, we show that if $\mathcal{C}$ is a *pebbling formula*, then $\deg^+(\mathsf{viol}_\mathcal{C} - 1)$ must be large. (In fact, our lower bound already holds for $\deg^+(\mathsf{viol}_{\mathrm{PEB}_G} - 0.99)$.)

Theorem 1.8. *There is a constant $\delta > 0$ such that for all sufficiently large m there is an in-degree-2 DAG G on m vertices such that*
$$\deg^+(\mathsf{viol}_{\mathrm{PEB}_G} - 1) = m^\delta,$$
where PEB_G is the pebbling formula associated with G.

Pebbling formulas are well-known to be refutable in Resolution width $O(1)$ and linear size on graphs of constant in-degree, and thus our result separates $\deg^+(\mathsf{viol}_\mathcal{C} - 1)$ from Resolution, resolving an open problem asked by Filmus, Mahajan, Sood and Vinyals [24]. Since Sherali-Adams can efficiently simulate Resolution [18], it follows from combining the previous theorem and Theorem 1.5 that $\deg^+(\mathsf{viol}_\mathcal{C} - 1)$ can be much larger than $\deg^+(\mathsf{viol}_\mathcal{C} - \varepsilon)$ for arbitrarily small $\varepsilon > 0$ (although, we note that the standard simulation of Resolution by Sherali-Adams results in ε that is exponentially small in the size of the proof!).

Next, by using known *lifting theorems* from conical junta degree to non-negative rank [33,43], we can lift the lower bound in the previous theorem to the following separation between $\mathsf{msep}_1(\mathrm{GEN}_n)$ and $\mathsf{mC}(\mathrm{GEN}_n)$, where GEN_n is the *Generation function*.

Theorem 1.9. *The GEN_n function is computable by polynomial-size monotone circuits, but, there is a $\delta > 0$ such that $\mathsf{msep}_1(\mathrm{GEN}_n) = 2^{\Omega(n^\delta)}$.*
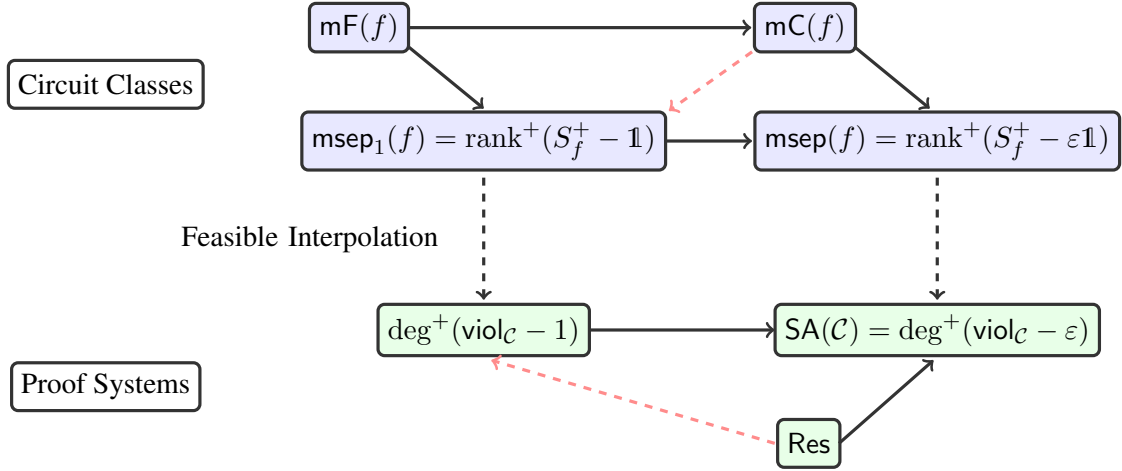
Figure 3: Semialgebraic circuit classes and proof systems considered in this paper. Solid arrows represent simulation results, while dashed red arrows represent new separations.

Since $\mathsf{msep}(f) \leq \mathsf{mC}(f)$, this provides the analogous separation between $\mathsf{msep}_1(f)$ and $\mathsf{msep}(f)$ (and, indeed, already between $\mathsf{msep}_1(f)$ and monotone circuit size). Taken together, these results imply that choosing smaller ε actually does increase the power of the corresponding proof systems and circuit models.

1.3 Paper Outline

The rest of this paper is outlined as follows. In Section 2 we carefully outline all important proof systems and non-uniform models of computation that we consider. In Section 3 we prove our new characterization of Sherali-Adams proofs, as well as our feasible interpolation results. In Section 4 we prove our new separations between fragments of Sherali-Adams and between separation complexities.

2   Preliminaries

2.1   Proof Systems

We recall some preliminaries from proof complexity. A *clause* C is a disjunction of boolean literals; the *width* of a clause, denoted $w(C)$, is the number of literals in C. If $\mathcal{C} = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ is a CNF formula then the *width* of $\mathcal{C}$ is the largest width of any clause in $\mathcal{C}$.

Resolution Proofs.   Fix an unsatisfiable CNF formula $\mathcal{C}$ over variables $x_1, \ldots, x_n$. A *Resolution refutation* of $\mathcal{C}$ is a sequence of clauses $D_1, \ldots, D_s$ ending in the empty clause $D_s = \emptyset$ such that for each $i \in [s]$, either D_i is in $\mathcal{C}$ or is derived from earlier clauses D_j, D_k with $j, k < i$ using the *resolution rule*

$$C \vee \ell, D \vee \bar{\ell} \vdash C \vee D$$

where the rule can only be applied if C and D do not contain literals of opposite sign. The *size* of the resolution proof is s, the number of clauses, and the *width* of the proof is the largest width of any clause

in the proof. Let $S_{\mathsf{Res}}(\mathcal{C})$ denote the minimum size of any resolution refutation of $\mathcal{C}$, and let $w_{\mathsf{Res}}(\mathcal{C})$ denote the smallest width of any resolution refutation of $\mathcal{C}$.

Sherali-Adams and Conical Junta Degree. In this paper we will regularly be doing arithmetic with real polynomials that represent boolean functions $f : \{0,1\}^n \to \mathbb{R}$. It is well known that each such function can be represented (uniquely) by *multilinear* polynomials (that is, the largest degree of any individual variable is 1). We will exclusively be operating using multilinear arithmetic with these polynomials — for example, we regard x_i^2 and x_i as representing the same boolean function. Formally, we work modulo the ideal $\langle x_i^2 - x_i \rangle_{i=1}^n$.

A d-*conjunction* J is a conjunction of d boolean literals. We will usually encode conjunctions as polynomials over the reals, and so if $S, T \subseteq [n]$ are subsets such that $S \cap T = \emptyset$ then we let

$$J_{S,T}(x) := \prod_{i \in S} x_i \prod_{j \in T} (1 - x_j)$$

be the conjunction that tests if all variables in S are 1 and all variables in T are 0. Similarly, if C is a clause, we will let $\overline{C}$ denote the conjunction such that $\overline{C}(x) = 1$ if and only if $C(x) = 0$ — that is, $\overline{C}$ tests if the clause C is falsified. To spell it out, if $C = \bigvee_{j \in T} x_j \vee \bigvee_{i \in S} \overline{x}_i$ then $\overline{C} = J_{S,T}$. Note that the empty conjunction $J_{\emptyset,\emptyset} = 1$.

A *conical junta* is any non-negative real combinations of conjunctions $\mathcal{J} = \sum_i \lambda_i J_i$, where $\lambda_i \geq 0$ is a real constant and J_i is a conjunction. The *degree* of a conical junta is the maximum degree of any conjunction in the conical junta representation, and the *size* of a conical junta is the number of monomials obtained after expanding all juntas but before cancellation. If $f : \{0,1\}^n \to \mathbb{R}_{\geq 0}$ is a non-negative real boolean function, then the *conical junta degree* of f, denoted $\deg^+(f)$, is the minimum integer d such that we can write

$$f = \sum_i \lambda_i J_i$$

where $\lambda_i \geq 0$ is a non-negative real constant and J_i is a $\leq d$-conjunction for each i. Note that, unlike polynomial degree, the representation of f as a conical junta is not necessarily unique.

A *Sherali-Adams refutation* of an unsatisfiable CNF formula $\mathcal{C} = C_1 \wedge \cdots \wedge C_m$ is given by a sequence of conical juntas $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_m, \mathcal{J}_{m+1}$ such that

$$\sum_{i=1}^{m} -\overline{C}_i \mathcal{J}_i + \mathcal{J}_{m+1} = -1.$$

The *degree* of the proof is the maximum degree of any of the polynomials $-\overline{C}_i p_i, \mathcal{J}$ before cancellation. The *size* of the proof is the number of distinct monomials that occurs in the proof (after expanding all polynomials and before cancellation). Let $\mathsf{SA}(\mathcal{C})$ denote the minimum degree of any Sherali-Adams refutation of $\mathcal{C}$. We note that CNF formulas are often encoded as linear inequalities rather than juntas; this is equivalent to the above definition up to increasing the degree by an additive factor of $w(C)$ (see e.g., [27, Claim 3.32]). If $\mathcal{C} = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ is a CNF formula then define $\mathsf{viol}_{\mathcal{C}} : \{0,1\}^n \to \mathbb{R}_{\geq 0}$ by

$$\mathsf{viol}_{\mathcal{C}}(x) := \sum_{i=1}^{m} \overline{C}_i(x).$$

That is, $\mathsf{viol}_{\mathcal{C}}(x)$ counts the number of falsified clauses of $\mathcal{C}$ on input x. It follows that $\mathcal{C}$ is unsatisfiable if and only if $\mathsf{viol}_{\mathcal{C}}(x) \geq 1$ for all $x \in \{0,1\}^n$ — indeed, if and only if $\mathsf{viol}_{\mathcal{C}}(x) \geq \varepsilon$ for any $\varepsilon > 0$. Thus, a conical junta representation of $\mathsf{viol}_{\mathcal{C}} - \varepsilon$ for any $\varepsilon > 0$ constitutes a proof that $\mathcal{C}$ is unsatisfiable.

In fact, conical junta representations of $\text{viol}_\mathcal{C} - \varepsilon$ correspond naturally to fragments of Sherali-Adams proofs. This is easy to see: suppose that $\text{viol}_\mathcal{C} - \varepsilon = \mathcal{J}$ where $\mathcal{J}$ is a conical junta. Explicitly, this means that

$$\sum_{j=1}^{m} -\overline{C}_j + \mathcal{J} = -\varepsilon$$

which, dividing by ε, yields

$$\sum_{j=1}^{m} -\frac{1}{\varepsilon}\overline{C}_j + \frac{\mathcal{J}}{\varepsilon} = -1.$$

Conversely, suppose that we have a Sherali-Adams refutation of $\mathcal{C}$ of the form

$$\sum_{j=1}^{m} -\alpha_j \overline{C}_j + \mathcal{J} = -1$$

where $0 \le \alpha_j \le 1/\varepsilon$ for all j. Since $\overline{C}_j$ is a conjunction and $1/\varepsilon \ge \alpha_j$, this means that

$$\sum_{j=1}^{m} -\frac{1}{\varepsilon}\overline{C}_j + \mathcal{J} + \sum_{j=1}^{m} \left(\frac{1}{\varepsilon} - \alpha_j\right)\overline{C}_j = -1$$

is also a Sherali-Adams refutation of $\mathcal{C}$. Multiplying through by ε yields the following:

Lemma 2.1. *Let* $\mathcal{C} = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ *be an unsatisfiable CNF formula. If there is a Sherali-Adams refutation of the form*

$$\sum_{j=1}^{m} -\alpha_j \overline{C}_j + \mathcal{J} = -1$$

where $\mathcal{J}$ *is a conical junta of degree* d *and size* s *and with* $0 \le \alpha_j \le 1/\varepsilon$ *for each* α_j, *then there is a degree-*$\max\{d, w(\mathcal{C})\}$ *conical junta* $\mathcal{J}'$ *such that* $\text{viol}_\mathcal{C} - \varepsilon = \mathcal{J}'$.

Conversely, if $\text{viol}_\mathcal{C} - \varepsilon = \mathcal{J}'$ *for some degree-*d, *size* s *conical junta* $\mathcal{J}'$, *then there is a Sherali-Adams proof of the above form with* $\alpha_j \le 1/\varepsilon$ *and where* $\mathcal{J} = \mathcal{J}'$.

As we will show in Section 3.1, it turns out that this fragment of Sherali-Adams is actually *complete*, in the sense that any Sherali-Adams proof can be put into this form without changing the degree and without (badly) changing the size.

2.2   Circuit Complexity and Extended Formulations.

Boolean Circuit Complexity.   A *partial* boolean function is a function $f : \{0,1\}^n \to \{0,1,*\}$ (the $*$ represents that we "don't care" what the function's output is). Given two partial boolean functions f, g we say that g *extends* f if for all x such that $f(x) \ne *$, $g(x) = f(x)$. A (total) boolean function $f : \{0,1\}^n \to \{0,1\}$ is *monotone* if $x \le y \Rightarrow f(x) \le f(y)$ (where the first inequality is taken coordinate-wise); a partial boolean function $f : \{0,1\}^n \to \{0,1,*\}$ is monotone if there is a total monotone boolean function g extending f.

We assume familiarity with standard boolean circuit models in complexity theory — e.g. boolean formulas, boolean circuits, etc. All circuits considered in this paper will be composed of AND ($\wedge$) and OR ($\vee$) gates, and will be *De-Morgan* in the sense that if any negations appear they appear at the inputs of the circuit. A boolean circuit is *monotone* if it does not contain any negation gates. We say

that a circuit C computes a partial function $f : \{0,1\}^n \to \{0,1,*\}$ if $f(x) = 1 \Rightarrow C(x) = 1$ and $f(x) = 0 \Rightarrow C(x) = 0$ for all $x \in \{0,1\}^n$ (in other words, the total function computed by C extends f). For any partial boolean f we let

- $\mathsf{C}(f) :=$ the size of the smallest boolean circuit computing f,

- $\mathsf{mC}(f) :=$ the size of the smallest monotone boolean circuit computing f,

- $\mathsf{F}(f) :=$ the size of the smallest boolean formula computing f,

- $\mathsf{mF}(f) :=$ the size of the smallest monotone boolean formula computing f.

Non-negative Rank and Extended Formulations. Let B be a non-negative real matrix. The *non-negative rank* of B, denoted $\mathrm{rank}^+(B)$, is the smallest integer r such that B can be written as the sum of r rank-1 non-negative matrices.

Non-negative rank is a very well-studied parameter in theoretical computer science, owing to its close relationship with polytopes in convex optimization. We, however, will be particularly interested in the connection between non-negative rank and circuit complexity.

Let $\mathbb{1}_{m,n}$ denote the $m \times n$ all-1s matrix (we may suppress the m, n when it is clear from context). Let $f : \{0,1\}^n \to \{0,1,*\}$ be a partial boolean function, let $U = f^{-1}(1)$ and let $V = f^{-1}(0)$. We let S_f be the $|U| \times |V|$ matrix defined by, for each $x \in U, y \in V$,

$$S_f(x,y) = \sum_{i=1}^n [\![x_i \neq y_i]\!]$$

where $[\![P]\!]$ is the $\{0,1\}$-indicator function for the predicate P. That is, the matrix S_f counts the number of witnesses that $x \neq y$ for each $x \in f^{-1}(1), y \in f^{-1}(0)$. If the function f is furthermore monotone, we define the matrix S_f^+ by

$$S_f^+ = \sum_{i=1}^n [\![x_i = 1 \wedge y_i = 0]\!].$$

Observe that both $S_f(x,y), S_f^+(x,y) \geq 1$ for all $x \in U, y \in V$. Because of this, we can say that for any $0 < \varepsilon \leq 1$ the matrices $S_f - \varepsilon\mathbb{1}$ and $S_f^+ - \varepsilon\mathbb{1}$ are both non-negative. With this in mind, we make the following definitions (the second two, of course, only make sense for monotone f):

$$\mathsf{sep}_1(f) := \mathrm{rank}^+(S_f - \mathbb{1})$$
$$\mathsf{sep}(f) := \min_{\varepsilon > 0} \mathrm{rank}^+(S_f - \varepsilon\mathbb{1})$$
$$\mathsf{msep}_1(f) := \mathrm{rank}^+(S_f^+ - \mathbb{1})$$
$$\mathsf{msep}(f) := \min_{\varepsilon > 0} \mathrm{rank}^+(S_f^+ - \varepsilon\mathbb{1})$$

We note the similarity between the definition of $\mathsf{sep}(f)$ and $\deg^+(\mathrm{viol}_C - \varepsilon)$ — the restriction of Sherali-Adams — which was introduced in the previous section. Furthermore, as we have discussed in detail in Section 1, each of these parameters can be interpreted as the minimum size of any extended formulation separating 0s of a function from 1s of a function, and they are closely related to circuit and formula size (cf. Theorem 1.3, Theorem 1.4).

3 Upper Bounds

3.1 A New Characterization of Sherali-Adams Degree

In this section we provide our new characterization of Sherali-Adams degree, which is recorded next.

Theorem 1.5. *For any unsatisfiable CNF formula $\mathcal{C}$, if there is a Sherali-Adams refutation of $\mathcal{C}$ with degree d and size s, then there is an $\varepsilon > 0$ and a degree d, size $s2^d$ conical junta $\mathcal{J}$ such that*

$$\mathsf{viol}_{\mathcal{C}} - \varepsilon = \mathcal{J}.$$

Consequently, if $\mathsf{SA}(\mathcal{C})$ is the minimum degree of any Sherali-Adams refutation of $\mathcal{C}$, then

$$\mathsf{SA}(\mathcal{C}) = \min_{\varepsilon > 0} \deg^+(\mathsf{viol}_{\mathcal{C}} - \varepsilon).$$

Proof. Let

$$\sum_{i=1}^{m} -\overline{C}_i \mathcal{J}_i + \sum_j \lambda_j J_j = -1$$

be a degree-d, size-s Sherali-Adams refutation of $\mathcal{C}$, where each $\mathcal{J}_i$ are conical juntas. By expanding each of the $\mathcal{J}_i$ into their monomials, we can re-write this proof as

$$\sum_{k=1}^{t} -\alpha_k \overline{C}_{i_k} m_k + \sum_j \lambda_j J_j = -1 \tag{1}$$

for some t where $\alpha_k \geq 0$ is a non-negative real constant and m_k is a monomial for each k.

Now, fix some $1 \leq u \leq t$. Assume w.l.o.g. that the variables occurring in m_u are $x_1, x_2, \ldots, x_\ell$ for some ℓ — explicitly, since m_u is a monomial we therefore have $m_u = \prod_{i=1}^{\ell} x_i = J_{[\ell],\emptyset}$. The main claim we use is the following.

Claim. Let $x_1, x_2, \ldots, x_n$ be any set of boolean variables. Then

$$\sum_{S \subseteq [n]} J_{S,[n]\setminus S} = 1.$$

Proof of Claim. By induction on n. If $n = 0$ then the only term in the sum is $J_{\emptyset,\emptyset}$, which is 1. If $n > 0$, then $\sum_{S \subseteq [n]} J_{S,[n]\setminus S} = x_n(\sum_{S \subseteq [n-1]} J_{S,[n-1]\setminus S}) + (1-x_n)(\sum_{S \subseteq [n-1]} J_{S,[n-1]\setminus S}) = x_1 + (1-x_1) = 1$ where we applied the induction hypothesis twice. □

Define the conical junta

$$\mathcal{J}_u := \sum_{S \subseteq [\ell], S \neq [\ell]} J_{S,[\ell]\setminus S} = \sum_{S \subseteq [\ell]} J_{S,[\ell]\setminus S} - J_{[\ell],\emptyset}.$$

By the Claim and the definition of $\mathcal{J}_u$ we have $m_u + \mathcal{J}_u = J_{[\ell],\emptyset} + \mathcal{J}_u = 1$. This means that

$$
\begin{aligned}
-\alpha_u m_u \overline{C}_{i_u} &= -\alpha_u \overline{C}_{i_u} J_{[\ell],\emptyset} + \alpha_u \overline{C}_{i_u} \mathcal{J}_u - \alpha_u \overline{C}_{i_u} \mathcal{J}_u \\
&= -\alpha_u \overline{C}_{i_u} + \alpha_u \overline{C}_{i_u} \mathcal{J}_u.
\end{aligned}
$$

Since $\alpha_u > 0$ and $\overline{C}_{i_u}$ is a conjunction it follows that $\alpha_u \overline{C}_{i_u} \mathcal{J}_u$ is a conical junta. By repeating this construction for every polynomial $-\alpha_k \overline{C}_{i_k} m_k$ it follows that we can rewrite the Sherali-Adams proof from Equation 1 as

$$\sum_{k=1}^{t} -\alpha_k \overline{C}_{i_k} + \left(\sum_j \lambda_j J_j + \sum_{k=1}^{t} \alpha_k \overline{C}_{i_k} \mathcal{J}_k \right) = -1.$$

Replacing $-\alpha_k \overline{C}_k m_k$ by $\alpha_k C_{i_k} \mathcal{J}_k$ preserves the degree but increases the monomial size of the proof by a 2^d factor. By applying Lemma 2.1 it follows that there is a $\varepsilon > 0$ such that $\mathsf{viol}_C - \varepsilon = \mathcal{J}$ for some size at most $s2^d$, degree-d conical junta $\mathcal{J}$. $\qquad\square$

3.2 Feasible Interpolation for Sherali-Adams

In this section, we give a monotone interpolation theorem for Sherali-Adams in terms of the *unsatisfiability certificate* (cf. Definition 1.6). For our purposes, the important property of the unsatisfiability certificate is the following connection it shares with viol_C.

Lemma 3.1. *Let C be an unsatisfiable CNF formula on m clauses, let (X, Y) be any partition of the variables of C, and let cert be the corresponding unsatisfiability certificate. There are surjections $\mu : \{0,1\}^X \to \mathsf{cert}^{-1}(1)$ and $\nu : \{0,1\}^Y \to \mathsf{cert}^{-1}(0)$ such that*

$$S_{\mathsf{cert}}^{+}(\mu(x), \nu(y)) = \mathsf{viol}_C(xy)$$

where xy denotes the concatenation of x, y.

Proof. We define μ and ν by their evaluations on assignments to the X and Y variables as follows. For any $x \in \{0,1\}^X$ define $\mu^x \in \{0,1\}^m$ by $\mu_i^x = 1 - C_i^X(x)$ — that is, μ_x^i is 1 iff $C_i^X(x)$ is falsified. Similarly, for any $y \in \{0,1\}^Y$ define $\nu^y \in \{0,1\}^m$ by $\mu_i^y = C_i^Y(y)$ — so, ν_y^i is 1 iff $C_i^Y(y)$ is satisfied. It follows by definition that $\mathsf{cert}(\mu^x) = 1$ and $\mathsf{cert}(\nu^y) = 0$, and furthermore it is clear that $x \mapsto \mu^x$ and $y \mapsto \nu^y$ are surjective maps since every 1 and 0 assignment to cert must have some witnessing assignments to the underlying variables. Finally, observe that if $\mu_i^x = 1$ and $\nu_i^y = 0$ then $C_i^X(x) = C_i^Y(y) = 0$, and thus the combined assignment xy violates the clause C_i. The converse also clearly holds, proving the lemma. $\qquad\square$

By combining this fact with Theorem 1.5, we can immediately prove the following monotone feasible interpolation theorem for Sherali-Adams proofs. This next theorem is slightly weaker (in that we lose a 2^d factor in the size) than the interpolation theorem stated in Section 1, but, has the benefit of admitting a completely transparent proof. We show how to remove the 2^d factor in Appendix A.

Theorem 3.2. *For any unsatisfiable CNF C and any partition of its variables (X, Y), if there is a conical junta $\mathcal{J}$ of size s and an $\varepsilon > 0$ such that $\mathsf{viol}_C - \varepsilon = J$ then*

$$\mathrm{rank}^{+}(\mathsf{cert}_C - \varepsilon) \leq s.$$

Consequently, if there is a degree-d, size-s Sherali-Adams refutation of C then $\mathsf{msep}(\mathsf{cert}_C) \leq s2^d$.

Proof. The "Consequently" statement of the theorem is just an application of Theorem 1.5, so we focus on proving the non-negative rank upper bound for cert_C.

Write $\mathsf{viol}_C - \varepsilon = \sum_i \lambda_i J_i$ where each J_i is a conjunction. Under the partition of variables (X, Y), we regard this as a $\{0,1\}^X \times \{0,1\}^Y$ matrix, where the (x, y)th entry of the matrix is exactly $\mathsf{viol}_C(xy) - \varepsilon$. By Lemma 3.1 it follows immediately that $\mathrm{rank}^{+}(\mathsf{cert}_C - \varepsilon) = \mathrm{rank}^{+}(\mathsf{viol}_C - \varepsilon)$,

since the existence of the two surjections implies that the matrix $\text{viol}_{\mathcal{C}} - \varepsilon$ is exactly the matrix $\text{cert}_{\mathcal{C}} - \varepsilon$ with some extra copies of rows and columns padded in. Thus if we show $\text{rank}^+(\text{viol}_{\mathcal{C}} - \varepsilon) \le s$ the proof of the theorem will be completed.

To show this, we use the expression of $\text{viol}_{\mathcal{C}} - \varepsilon$ as a conical junta. Observe that if J_i is a conjunction over X and Y variables, then we can interpret any term z_i or $1 - z_i$ in the product J_i as a rank-1 $2^{|X|} \times 2^{|Y|}$ non-negative matrix (indeed, it is rank-1 since each term depends only on the x assignment or the y assignment). If we let $A \circ B$ denote the Hadamard (i.e. entrywise) product of two matrices A and B, it follows that the conjunction J_i is simply the Hadamard product of a number of rank-1 non-negative matrices. Since $\text{rank}^+(A \circ B) \le \text{rank}^+(A)\,\text{rank}^+(B)$ — a fact easily verified by taking non-negative rank decompositions — it follows that $\text{rank}^+(J_i) = 1$. Therefore, if we can write $\text{viol}_{\mathcal{C}} - \varepsilon = \mathcal{J}$ where $\mathcal{J}$ is a non-negative combination of s conjunctions, it immediately follows by the subadditivity of rank^+ that $\text{rank}^+(\text{viol}_{\mathcal{C}} - \varepsilon) \le s$, proving the theorem. $\qquad\square$

4 Lower Bounds

In this section we prove our new separation between fragments of Sherali-Adams and Resolution and then show how to lift this separation to a new separation between extension complexity (particularly, msep_1) and monotone circuit complexity. First, we show that the well-studied *Pebbling formulas* PEB_G require large conical junta degree to represent $\text{viol}_{\text{PEB}_G} - 1$ (cf. Theorem 1.8). However, our eventual goal is to apply a *lifting theorem* to get a strong lower bound on $\text{msep}_1(f) := \text{rank}^+(S_f^+ - \mathbb{1})$ from the lower bound on $\deg^+(\text{viol}_{\text{PEB}_G} - 1)$ for some specially tailored monotone boolean function f. Unfortunately, the known lifting theorems from conical junta degree to non-negative rank [33, 43] only apply to *approximate* conical junta degree, which complicates the lower bound slightly.

Definition 4.1 (Approximate conical junta degree). For a non-negative function $f \colon \{0,1\}^n \to \mathbb{R}_{\ge 0}$ we define its ϵ-*approximate conical junta degree* $\widetilde{\deg}_\epsilon^+(f)$, as the minimum degree of a conical junta $\mathcal{J}$ such that $\mathcal{J}(x) \in f(x) \pm \epsilon$ for all inputs x.

For convenience, instead of proving lower bounds on the conical junta degree of PEB_G directly, we will instead work with the slightly more flexible *Iteration problem* ITER_n (this is also sometimes known as the *Housesitting Principle*) [4, 12]. An easy reduction from ITER_n to PEB_G will then yield the degree lower bound for Pebbling formulas.

4.1 Conical Junta Lower Bound for Iteration

We study the *Iteration problem* ITER_n on an n-by-n grid. An input to this problem consists of n^2 input variables x_v, one for each grid node $v \in [n] \times [n]$. The variables take values from an alphabet of size $O(n)$. Namely, for each $v \in [n-1] \times [n]$, we have $x_v \in [n] \cup \{\bot\}$. If $x_v \in [n]$, we say v is *active*, and it points to the x_v-th node on the next row. If $x_v = \bot$, we say v is *disabled*. For each node on the final row $v \in \{n\} \times [n]$ there is a boolean variable $x_v \in \{\top, \bot\}$ where $x_v = \top$ means x_v is *active* and $x_v = \bot$ means v is *disabled*. The goal is to output one of the following types of solution.

  (i) the *distinguished source node* $(1,1)$ is disabled ($x_{1,1} = \bot$), or
  (ii) an active node that points to a disabled node ($x_{i,j} = k \in [n]$ and $x_{i+1,k} = \bot$), or
 (iii) an active node on the final row ($x_{n,j} = \top$).

For larger-than-boolean alphabets, we generalise the definition of a conjunction J naturally to be a partial assignment of pointers to nodes; for example, $J = [x_{1,1} = 3; x_{2,3} = \bot; x_{4,1} = 6]$ is a
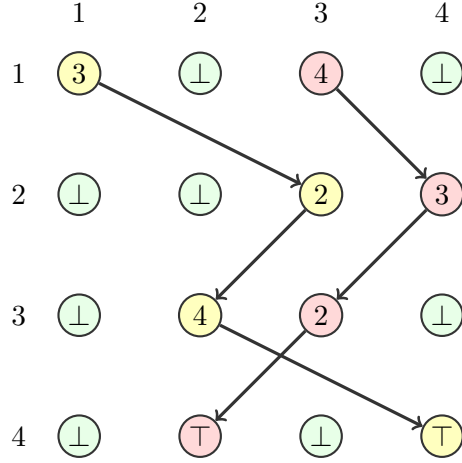
Figure 4: An example outcome of the random variable X_2 on ITER_4.

conjunction of degree 3. We prove strong lower bounds against the conical junta degree of $\text{viol}_{\text{ITER}_n} - 1$; indeed, we are even able to prove lower bounds against the conical junta degree of $\text{viol}_{\text{ITER}_n} - 0.99$, which will turn out to be crucial for later results.

Theorem 4.2. $\widetilde{\deg}^+_{1/20}(\text{viol}_{\text{ITER}_n} - 1) \geq \Omega(n)$.

Proof. We define three random inputs X_1, X_2, X_3 that admit 1, 2, or 3 solutions, all of type (iii). The inputs are constructed from given data (v_i^j), $j \in [2]$, $i \in [n]$, that specifies for each row $i \in [n]$, two (or three in the case of X_3) nodes $v_i^1, v_i^2 \in \{i\} \times [n]$. The resulting input, call it $X(v_i^j)$, is obtained by activating all the nodes v_i^j, pointing v_i^j to v_{i+1}^j, and disabling all other nodes.

$X_1 = $ *"two random paths merging at the last row."* We choose v_i^j for $j \in [2]$, $i \in [n]$ uniformly at random subject to $v_1^1 = (1,1)$, $v_i^1 \neq v_i^2$ for $i \in [n-1]$, $v_n^1 = v_n^2$. Then $X_1 = X(v_i^j)$.

$X_2 = $ *"two random disjoint paths."* We choose v_i^j for $j \in [2]$, $i \in [n]$ uniformly at random subject to $v_1^1 = (1,1)$, $v_i^1 \neq v_i^2$ for $i \in [n]$. Then $X_2 = X(v_i^j)$.

$X_3 = $ *"three random disjoint paths."* We choose v_i^j for $j \in [3]$, $i \in [n]$ uniformly at random subject to $v_1^1 = (1,1)$, $v_i^j \neq v_i^{j'}$ for $j \neq j'$, $i \in [n]$. Then $X_3 = X(v_i^j)$.

Suppose for the sake of contradiction that $\mathcal{J} = \sum_i \lambda_i J_i$ is a degree-$o(n)$ conical junta that for every input x outputs $\mathcal{J}(x) \in \text{viol}_{\text{ITER}_n}(x) - 1 \pm \epsilon$ where $\epsilon := 1/20$. We may assume that each conjunction J in $\mathcal{J}$ is such that if it reads an active node v on row $n-1$ (meaning that the conjunction contains $[x_{n-1,j} = k]$ for some k and v is node $(n-1,j)$) then it also reads the boolean variable associated with node (n, x_v) on the last row. (Indeed, we may always replace J containing $[x_{n-1,j} = k]$ with the equivalent conic combination $J \cdot [x_{n,k} = \top] + J \cdot [x_{n,k} = \bot]$.) We call a conjunction J *paired* if it reads at least two active nodes on the last row, that is, J witnesses at least two solutions of type (iii).

Claim 4.3. If J is not paired, then $\mathbb{E}[J(X_2)] \leq 2 \cdot \mathbb{E}[J(X_1)]$.

Claim 4.4. If J is paired, then $\mathbb{E}[J(X_3)] \geq (3 - o(1)) \cdot \mathbb{E}[J(X_2)]$.

15

We prove the claims shortly after we first complete the proof of the theorem assuming them. We write $\mathcal{J} = \sum_i \lambda_i J_i = \mathcal{J}_{\text{not}} + \mathcal{J}_{\text{pair}}$ where $\mathcal{J}_{\text{pair}}$ consists of conjunctions that are paired and $\mathcal{J}_{\text{not}}$ consists of those that are not. We calculate

$$
\begin{aligned}
1 - \epsilon \ &\leq \ \mathbb{E}[\mathcal{J}(X_2)] && (\mathcal{J} \text{ approximates viol} - 1) \\
&= \ \mathbb{E}[\mathcal{J}_{\text{not}}(X_2)] + \mathbb{E}[\mathcal{J}_{\text{pair}}(X_2)] && \\
&\leq \ 2\mathbb{E}[\mathcal{J}_{\text{not}}(X_1)] + \mathbb{E}[\mathcal{J}_{\text{pair}}(X_2)] && (\text{Claim 4.3 and linearity of } \mathbb{E}) \\
&\leq \ 2\epsilon + \mathbb{E}[\mathcal{J}_{\text{pair}}(X_2)] && (\mathcal{J} \text{ approximates viol} - 1) \\
&\leq \ 2\epsilon + \mathbb{E}[\mathcal{J}_{\text{pair}}(X_3)]/(3 - o(1)) && (\text{Claim 4.4 and linearity of } \mathbb{E}) \\
&\leq \ 2\epsilon + (2 + \epsilon)/(3 - o(1)) && (\mathcal{J} \text{ approximates viol} - 1) \\
&< \ 1 - \epsilon. && (\epsilon < 1/10)
\end{aligned}
$$

This is the desired contradiction that concludes the proof (modulo the Claims). $\qquad\square$

Proof of Claim 4.3. The claim says that $\Pr[J(X_1) = 1] \geq \Pr[J(X_2) = 1]/2$ for a non-paired J. Consider any fixed setting of the v_i^j such that $X(v_i^j)$ is in the support of X_2 and such that J accepts $X(v_i^j)$. Since J is not paired, it does not read the nodes $\{v_{n-1}^b, v_n^b\}$ for some $b \in [2]$, and hence if we pick $b \in [2]$ uniformly at random, $\Pr_b[J((v_i^j)_{i \in [n-1]}, v_n^b, v_n^b) = 1] \geq 1/2$. Moreover, if $(v_i^j)_{i \in [n]}$ is distributed as in the process for X_2, then $((v_i^j)_{i \in [n-1]}, v_n^b, v_n^b)$ (for random $b \in [2]$) is distributed as in the process for X_1. Hence J's acceptance probability can decrease by at most a factor of $1/2$. $\qquad\square$

Proof of Claim 4.4. The claim says that $\Pr[J(X_3) = 1] \geq (3 - o(1))\Pr[J(X_2) = 1]$ for a paired J. Since J reads at most $o(n)$ nodes, there is some middle row $i^* \in [n/3, 2n/3]$ from which J reads no nodes. Split any input x into three parts so that $x = (x^{\text{T}}, x^{\text{M}}, x^{\text{B}})$ where x^{T} is the pointer assignment for nodes in topmost $i^* - 1$ rows, x^{M} is for nodes in row i^*, and x^{B} is for nodes in the remaining bottommost rows. Write also $J(x) = J^{\text{T}}(x^{\text{T}}) \cdot J^{\text{B}}(x^{\text{B}})$. Note that for both $i = 2, 3$, the variables X_i^{T} and X_i^{B} are independent and thus

$$
\Pr[J(X_i) = 1] \ = \ \Pr[J^{\text{T}}(X_i^{\text{T}}) = 1 \wedge J^{\text{B}}(X_i^{\text{B}}) = 1] \ = \ \Pr[J^{\text{T}}(X_i^{\text{T}}) = 1] \cdot \Pr[J^{\text{B}}(X_i^{\text{B}}) = 1]. \quad (2)
$$

We will prove the following estimates, which, when plugged into (2), would complete the proof:

$$
\begin{aligned}
\Pr[J^{\text{T}}(X_3^{\text{T}}) = 1] \ &\geq \ (1 - o(1))\Pr[J^{\text{T}}(X_2^{\text{T}}) = 1], && (3) \\
\Pr[J^{\text{B}}(X_3^{\text{B}}) = 1] \ &\geq \ (3 - o(1))\Pr[J^{\text{B}}(X_2^{\text{B}}) = 1]. && (4)
\end{aligned}
$$

Let us prove (3). Consider generating a sample from X_3^{T} as follows: (i) sample $x^{\text{T}} \sim X_2^{\text{T}}$, and (ii) add to x^{T} a random third path which is disjoint from the existing two. We argue that if the sample x^{T} in step (i) is accepted by J^{T}, then the conjunction continues to accept after step (ii) with high probability. Indeed, conditioned on the first step being accepting for J^{T}, we note that the probability that any one node is picked to lie on the third path is at most $1/(n-2)$ and hence (by a union bound) the probability that some node read by J^{T} lies on the third path is at most $\deg(J)/(n - 2) = o(n)/(n - 2) = o(1)$. Hence J^{T} continues to accept with probability $1 - o(1)$, which proves (3).

Let us finally prove (4). Note how X_2^{B} (resp. X_3^{B}) is distributed: it consists of two (three) uniformly random disjoint paths down the grid. Consider the following bipartite graph $(L \cup R, E)$ where

- Left vertices L are outcomes of X_2^{B} (pairs of disjoint paths).
- Right vertices R are outcomes of X_3^{B} (triples of disjoint paths).

- There is an edge $(l, r) \in E$ iff the two paths of l are a subset of the three paths of r.

Note that this graph is biregular with right degree $d_R := 3$ and the left degree $d_L := 3|R|/|L|$. Denote by $L' \subseteq L$ (resp. $R' \subseteq R$) the set of $l \in L$ (resp. $r \in R$) accepted by J^{B}. Let $E(L', R') := L' \times R' \cap E$ denote the set of edges between L' and R'. Note that if $r \in R'$, then at most one of r's neighbours is in L'; this is because J^{B} is paired and hence it requires a prescribed pair of the three paths in r to be present. Consequently $|E(L', R')| \leq |R'|$. On the other hand, if $l \in L'$, then $1 - o(1)$ fraction of its neighbours are in R'; this is because of the same argument as in the preceding paragraph (condition on $X_2^{\text{B}} = l \in L'$ in step (i) and choose a random neighbour of l in step (ii)). Hence $|E(L', R')| \geq (1 - o(1))d_L|L'|$. We now put these observations together to prove (4):

$$
\begin{aligned}
\Pr[J^{\text{B}}(X_3^{\text{B}}) = 1] &= |R'|/|R| \\
&\geq |E(L', R')|/|R| \\
&\geq (1 - o(1))d_L|L'|/|R| \\
&= (3 - o(1))|L'|/|L| \\
&= (3 - o(1))\Pr[J^{\text{B}}(X_2^{\text{B}}) = 1]. \qquad \square
\end{aligned}
$$

4.2 Lower bound for Pebbling

The *Pebbling problem* PEB_G is defined relative to a DAG $G = (V, E, v^*)$ where v^* is a *distinguished source node* (in-degree 0) and where every node has out-degree at most 2 (but in-degree may be unbounded). The input to PEB_G is $y \in \{0, 1\}^E$, that is, an boolean assignment to the edges of G. Such an assignment y naturally defines a subgraph G_y of G consisting of all the edges e such that $y_e = 1$, which we call the *active* edges. The goal is to output one of the following types of solution.

(i) the distinguished source node is a sink (out-degree 0) in G_y.
(ii) a node $v \in V$ that is a *proper sink* in G_y, meaning v has in-degree ≥ 1 and out-degree 0.

Note that the presence of any solution can be certified by reading at most 3 bits, and hence PEB_G corresponds to the canonical search problem of an unsatisfiable 3-CNF formula. Moreover, if G has m edges then this 3-CNF formula has at most $m + 1$ clauses, as every possible violation in PEB_G is either the violation of the first type, or can be identified with an incoming edge to a node. It follows that for all $z \in \{0, 1\}^m$

$$
\text{viol}_{\text{PEB}_G}(z) - 1 \leq m, \tag{5}
$$

which we have recorded for later use.

We prove the following theorem, which is a strengthening of Theorem 1.8.

Theorem 4.5. *There is a DAG G with m edges such that $\widetilde{\deg}_{1/20}^+(\text{viol}_{\text{PEB}_G} - 1) \geq m^{\Omega(1)}$.*

Proof. We describe a simple reduction from ITER_n to PEB_G where G has $O(n^3)$ edges. The goal of the reduction is to show that any degree-d conical junta approximating $\text{viol}_{\text{PEB}_G} - 1$ can be translated into a degree-d conical junta approximating $\text{viol}_{\text{ITER}_n} - 1$. Hence the theorem follows from Theorem 4.2.

The DAG $G = (V, E, v^*)$ is defined as follows. First, V includes all nodes in the grid $[n] \times [n]$ underlying ITER_n. We naturally set $v^* := (1, 1)$. Observe that naively connecting each grid node to all its possible n successors on the subsequent row would result in a DAG of unbounded out-degree. To circumvent this, we instead include in G for every node $v = (i, j) \in [n - 1] \times [n]$, a $\log n$-depth binary tree T_v with n leaves $\ell_1^v, \ldots, \ell_n^v$. We identify the root of T_v with v and we identify each leaf ℓ_k^v with the

17

grid node $(i + 1, k) \in [n] \times [n]$. Finally, we include in V an extra bottom row of nodes $\{n + 1\} \times [n]$. For every grid node (n, j) on the n-th row, we include the single edge $e_v := ((n, j), (n + 1, j))$. This completes the description of G.

Given an input x to ITER_n we define an input $y = y(x)$ to PEB_G as follows. Consider a node $v \in [n - 1] \times [n]$. If $x_v = k \in [n]$ we activate all edges on the unique root-to-ℓ_k^v path in T_v and disable the other edges in T_v. If $x_v = \perp$ we disable all edges in T_v. Finally, for a node $v = (n, j)$ on the last row, we activate its associated edge e_v iff $x_v = \top$.

Note that $y = y(x)$ faithfully "models" the input x in the sense that the solutions of x appear associated with the same nodes as solutions in y. In particular, $\text{viol}_{\text{ITER}_n}(x) = \text{viol}_{\text{PEB}_G}(y)$. Moreover, each variable y_e of PEB_G is a function of a single variable of ITER_n. Thus every degree-d conjunction $J(y)$ can be translated into a degree-d conjunction $J'(x)$ such that $J'(x) = J(y(x))$. Applying this translation to the conjunctions in a conical junta, we conclude that if there is a conical junta approximating $\text{viol}_{\text{PEB}_G} - 1$, there is one for approximating $\text{viol}_{\text{ITER}_n} - 1$ of the same degree. $\qquad \square$

4.3 Separating Extended Formulations from Monotone Circuits

As we have seen in Section 2, if $f : \{0, 1\}^n \to \{0, 1, *\}$ is a partial monotone boolean function then

$$\text{msep}(f) := \min_{\varepsilon > 0} \text{rank}^+(S_f^+ - \varepsilon \mathbb{1}) \leq \text{mC}(f),$$

and

$$\text{msep}_1(f) := \text{rank}^+(S_f^+ - \mathbb{1}) \leq \text{mF}(f),$$

where $\text{mC}(f)$ ($\text{mF}(f)$) is the size of the smallest monotone circuit (formula, resp.) computing f. It is natural to wonder if it is necessary to chose $\varepsilon < 1$ in order to simulate monotone circuits — in other words, whether or not $\text{msep}_1(f)$ is already upper bounded by $\text{mC}(f)$. In this section, we show that the separation proven in Section 4.2 implies that such a bound is not possible. In particular, we prove the following theorem:

Theorem 1.9. *The* GEN_n *function is computable by polynomial-size monotone circuits, but, there is a* $\delta > 0$ *such that* $\text{msep}_1(\text{GEN}_n) = 2^{\Omega(n^\delta)}$.

Let us begin by defining the GEN_n function, which was originally introduced by Raz and McKenzie [51] and has now been the subject of a number of works in circuit complexity [13, 22, 23, 34, 51, 54].

Definition 4.6. Let n be a positive integer. A set of triples $T \subseteq [n]^3$ is said to *generate* $k \in [n]$ if $k = 1$ or if there is a triple $(i, j, k) \in T$ such that T generates both i and j.

The GEN_n function is a monotone boolean function defined as follows. As input, the function receives a list of triples $T \subseteq [n]^3$ encoded as a binary string of length $\{0, 1\}^{n^3}$. It then outputs 1 if and only if n can be generated from T.

It is well known that GEN_n has polynomial-size monotone circuits [51], so, we focus on proving the lower bound in Theorem 1.9. To do this we apply the following lifting theorem from conical junta degree to non-negative rank due to Kothari, Meka, and Raghavendra [43] (we note that applying the lifting theorem from Göös et. al. would also yield an essentially identical result [33]). In order to state the theorem we introduce the important notion of a *pattern matrix*:

Definition 4.7. Let $f : \{0, 1\}^n \to \mathbb{R}$ be a boolean function, and let $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be any function. Define the *pattern matrix* $f \circ g^n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$ for any $x \in \mathcal{X}^n, y \in \mathcal{Y}^n$ by

$$(f \circ g^n)(x, y) := f(g(x_1, y_1), g(x_2, y_2), \ldots, g(x_n, y_n)).$$

We note that even though we have defined it as a function, we think of $f \circ g^n$ as a $\mathcal{X}^n \times \mathcal{Y}^n$ matrix in the natural way.

The next theorem is a slight modification of Theorem 1.7 in [43], and follows immediately from the proof of [43].

Theorem 4.8. *Let $f : \{0,1\}^n \to \mathbb{R}_{\geq 0}$ be any boolean function with $\mathbb{E}[f] = 1$, and suppose that $\deg^+(f + 1/100n) \geq 9 \deg(f)$. For any $b \geq 100 \log n$ there is a gadget function $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ such that*

$$\mathrm{rank}^+(f \circ g^n) = 2^{\Omega(b \deg^+(f + 1/100n))}.$$

We combine this theorem with our lower bound for PEB_G from the previous section (Theorem 4.5) and an embedding argument to prove the Theorem 1.9.

Theorem 4.9. *There is a constant $\delta > 0$ such that for all sufficiently large n, $\mathsf{msep}_1(\mathrm{GEN}_n) \geq 2^{\Omega(n^\delta)}$.*

Proof. Let G be the DAG with m edges guaranteeing the approximate conical junta degree lower bound from Theorem 4.5, and let g be the gadget function guaranteed by Theorem 4.8. We begin by showing that

$$\mathrm{rank}^+((\mathsf{viol}_{\mathrm{PEB}_G} - 1) \circ g^m) = 2^{\Omega(m^\delta)}$$

for some universal constant $\delta > 0$. Once we have this, we show that the pattern matrix $(\mathsf{viol}_{\mathrm{PEB}_G} - 1) \circ g^m$ can be embedded as a submatrix of the slack matrix $S_{U,V}^+$ associated with GEN_n (cf. Section 2); since non-negative rank is monotone decreasing with respect to taking submatrices this immediately implies $\mathsf{msep}_1(\mathrm{GEN}_n) \geq 2^{\Omega(n^\delta)}$, proving the theorem.

Let $f = \mathsf{viol}_{\mathrm{PEB}_G} - 1$, with the goal of applying the previous lifting theorem, and note $f : \{0,1\}^m \to \mathbb{R}_{\geq 0}$ and furthermore that $f(x) \leq m$ by Equation 5. It follows that $\mathbb{E}f \leq m$, and thus there is some $\varepsilon \leq 1/100$ such that

$$\begin{aligned}
\deg^+(f/\mathbb{E}f + 1/100m) &= \deg^+(\mathsf{viol}_{\mathrm{PEB}_G} - 1 + \mathbb{E}f/100m) \\
&= \deg^+(\mathsf{viol}_{\mathrm{PEB}_G} - (1 - \varepsilon)) \\
&\geq \widetilde{\deg}_\varepsilon^+(\mathsf{viol}_{\mathrm{PEB}_G} - 1) \geq m^\delta
\end{aligned}$$

where we have used the fact that $\deg^+$ is invariant under multiplying by positive constants, and also Theorem 4.5. Finally, since $\mathsf{viol}_{\mathrm{PEB}_G}$ can be represented as a sum of violations of a 3-CNF formula it follows that $\deg(f) = 3$. We can therefore apply Theorem 4.8 and conclude that

$$\mathrm{rank}^+((\mathsf{viol}_{\mathrm{PEB}_G} - 1) \circ g^m) = 2^{\Omega(m^\delta)}.$$

So, in the remainder of the proof we show, for *any* DAG G with m edges, $(\mathsf{viol}_{\mathrm{PEB}_G} - 1) \circ g^m$ can be embedded as a submatrix of the slack matrix $S_{\mathrm{GEN}_n}^+$ associated with the GEN_n function for some $n = \mathsf{poly}(m)$. This is the same embedding argument that has been used in many works relating "lifted" PEB_G to GEN_n [22, 34, 51, 54]; for this reason, we only sketch the argument for the sake of completeness. Recall from Section 2 the definition of S_f^+: if $f : \{0,1,*\}^n \to \{0,1\}$ is a partial monotone boolean function and $U = f^{-1}(1), V = f^{-1}(0)$, then S_f^+ is the $|U| \times |V|$ matrix defined by

$$S_f^+(x,y) = \sum_{i=1}^n [\![x_i = 1 \wedge y_i = 0]\!]$$

19

for all $x \in U, y \in V$. To construct our embedding we create two mappings $\mu : \{0,1\}^{mb} \to \mathrm{GEN}_n^{-1}(1)$ and $\rho : \{0,1\}^{mb} \to \mathrm{GEN}_n^{-1}(0)$ such that $S_{\mathrm{GEN}_n}^+(\mu(x), \rho(y)) = (\mathsf{viol}_{\mathrm{PEB}_G} \circ g^m)(x,y)$.

Let G be any DAG with m edges, and assume without loss of generality that G has a single distinguished source node s^* with a single outgoing edge, and sink nodes $u_1, \ldots, u_t$. We define an *auxiliary graph* G^{aux} as follows. The nodes of the auxiliary graph G^{aux} are the edges of G. Then, if $(u,v), (v,w)$ are edges of G that share a node v, we add a directed edge $((v,w), (u,v))$ to G^{aux} (note the reverse in the edge direction). Note the sink node of G^{aux} corresponds to the unique edge leaving the source node of s^*, and the source nodes of G^{aux} correspond to the edges that enter sinks of G.

Letting $N = 2^b$, for each node in G^{aux} we create N points in the resulting GEN instance, and we index each such point as e_x for each $x \in \{0,1\}^b$. We also create a designated start node $\mathbf{1}$ and a designated target node $\mathbf{n}$.

Now we describe the functions μ, ρ in the embedding.

- *Definition of* $\mu : \{0,1\}^{mb} \to \mathrm{GEN}_n^{-1}(1)$. Let $x \in \{0,1\}^{mb}$ and write $x = x_1 x_2 \cdots x_m$ where $x_i \in \{0,1\}^b$. Order the edges $e^1, e^2, \ldots, e^m$, and we regard e^1 as the unique edge leaving the distinguished source node. We regard each $x_i \in \{0,1\}^b$ as selecting the point $e_{x_i}^i$ from the resulting GEN instance; we then plant a copy of G^{aux} inside the GEN instance on these points. That is, whenever a node $u = e^i$ in G^{aux} has two incoming edges from $v = e^j, w = e^k$, we plant the triple $(e_{x_i}^i, e_{x_j}^j, e_{x_k}^k)$; when u has only a single incoming edge from v we plant the triple $(e_{x_j}^j, e_{x_j}^j, e_{x_i}^i)$. Finally, for each of source nodes $u = e^i$ of G^{aux} we add the triple $(\mathbf{1}, \mathbf{1}, e_{x_i}^i)$. It is easy to see that the result is in $\mathrm{GEN}^{-1}(1)$, since we have explicitly planted the graph G^{aux} connecting the source node $\mathbf{1}$ to the target node $\mathbf{n}$.

- *Definition of* $\rho : \{0,1\}^{mb} \to \mathrm{GEN}_n^{-1}(0)$. Given $y \in \{0,1\}^{mb}$ we similarly write $y = y_1 y_2 \cdots y_m$ and add triples to the instance as follows. For any constraint of PEB_G enforcing that a node v in G is not a proper sink (for example, by preventing the case where $e^i = (u,v)$ is in G and $e^j = (v, w_1), e^k = (v, w_2)$ are not in G), and for any $x^i, x^j, x^k \in \{0,1\}^b$, we add the triple $(e_{x_j}^j, e_{x_k}^k, e_{x_i}^i)$ to the instance iff the constraint is satisfied by the assignment $e^i = g(x_i, y_i), e^j = g(x_j, y_j), e^k = g(x_k, y_k)$. We apply a similar construction for each of the constraints corresponding to the source node of s^* having an outgoing edge and set all other triples to 0. This is a 0-instance of GEN_n since if it was a 1-instance it must contain an embedding of G^{aux} (as in the construction of μ above), but this would imply that the corresponding assignment x embedding G^{aux} would satisfy every constraint of PEB_G, which is a contradiction.

Finally, to see that $S_{\mathrm{GEN}_n}^+(\mu(x), \mu(y)) = \mathsf{viol}_{\mathrm{PEB}_G}(g^m(x,y))$ observe that, by construction, each triple in $\mu(x)$ which does not occur in $\mu(y)$ corresponds exactly to a unique violated constraint of PEB_G. $\quad\square$

A Monotone Feasible Interpolation for Sherali-Adams

In this appendix we prove a refinement of Theorem 3.2, showing that Sherali-Adams is interpolated by *non-negative span programs*, a model introduced by de Oliveira Oliveira and Publák [20] as an extension of the classic span program of Karchmer and Wigderson [42].

Non-negative Span Programs. A *non-negative span program* (over the reals) is defined by a matrix $M \in \mathbb{R}^{k \times t}$, a labelling $\rho \in \{1, z_i, \neg z_i\}^k$ of the rows, and a "target" vector $b \in \mathbb{R}^t$. The program is *monotone* if no row is labelled by $\neg z_i$. The non-negative span program *accepts* an input $z \in \{0,1\}^n$ if

b is a conic combination of the rows of M labelled by 1 under z. That is, letting $M^{\rho(z)}$ be the matrix obtained from M by replacing each column $c \in [k]$ for which $\rho_j(z) = 0$ by the all-0 row, the non-negative span program accepts iff there is non-negative vector $y \in \mathbb{R}^k$ such that $y^\top M^{\rho(z)} = b$. A span program computes a function $f : \{0,1\}^n \to \{0,1\}$ if

$$f(z) = 1 \iff \exists y \in \mathbb{R}^k_{\geq 0}, \quad y^\top M^{\rho(z)} = b.$$

The *size* of a non-negative span program is the size of the matrix M.

It is known that non-negative span programs are simulated by msep. Therefore, the following interpolation theorem improves upon Theorem 3.2 in two ways: it provides a tighter connection of Sherali-Adams proofs to a computational model, and it removes the pesky blow-up of 2^d.

Theorem A.1. *Let $\mathcal{C}$ be any unsatisfiable CNF formula and (X, Y) be any partition of its variables. If there is a Sherali-Adams refutation of $\mathcal{C}$ of size s then there is a $O(s^2)$-size non-negative span program computing* $\mathsf{cert}_{\mathcal{C}}^{(X,Y)}$.

We begin with some definitions that will be used throughout this section. A set of polynomials P is *unsatisfiable* (modulo the ideal $\langle x_i^2 = x_i \rangle_{i=1}^n$) if there is no $x \in \{0,1\}^n$ such that $p(x) \geq 0$ for all $p \in P$. For this section it will be convenient to define Sherali-Adams refutations for more general unsatisfiable sets of polynomials, rather than only CNF formulas. We state this generalization next, in a manner that will be convenient for this section.

A *Sherali-Adams* refutation of an unsatisfiable set of polynomials P on variables $x_1, \ldots, x_n$ is given by a set of polynomials R such that

- R contains only polynomials of the form pJ, or J, for $p \in P$ and J a non-negataive junta.

- The constant -1 is in the cone generated by taking non-negative linear combinations of the polynomials in R.

The *size* of a Sherali-Adams derivation is the number of distinct monomials occurring in the (monomial expansion of) the polynomials in R.

We will work with a slight generalization of interpolant functions for polynomials. For an unsatisfiable set of polynomials $A(x, z) \cup B(y, z)$ let

$$I_{A \cup B}(\alpha) = \begin{cases} 0 & \text{if } A(x, \alpha) \text{ is satisfiable,} \\ 1 & \text{if } B(y, \alpha) \text{ is satisfiable.} \end{cases}$$

be the associated *interpolant* function.

The proof of Theorem A.1 will proceed as follows. First, we will reduce the task of constructing a computation of an unsatisfiability certificate $\mathsf{cert}_{\mathcal{C}}$ from a refutation of a CNF formula $\mathcal{C}$ to the task of constructing an interpolant $I_{A_{\mathcal{C}} \wedge B_{\mathcal{C}}}$ from a refutation of an associated split formula $A_{\mathcal{C}} \wedge B_{\mathcal{C}}$. Following this, we show how to extract a small non-negative span program computing the monotone interpolant. We will do this in two steps. First, we will show that a *feasible disjunction* property holds for Sherali-Adams: from any refutation of an unsatisfiable system of polynomials $A(x, z) \wedge B(y, z)$ and a $\{0,1\}$-assignment α to the z-variables, we can extract a refutation of $A(x, \alpha)$ or $B(y, \alpha)$ of roughly the same size. Second, we will use this feasible disjunction property to construct a monotone non-negative span program which has the same size as the Sherali-Adams refutation and searches for a refutation of $A(x, \alpha)$.

Reduction to Interpolant Formulas. Given any unsatisfiable CNF formula $\mathcal{C}$ we will describe a transformation of $\mathcal{C}$ into a monotone split formula due to Hrubeš and Pudlák [41]. Then, we will argue that any computation of the interpolant $I_{\mathcal{C}}$ of the split formula associated with $\mathcal{C}$ is also a computation of $\text{cert}_{\mathcal{C}}$. Recall that we say that a split (CNF) formula $A(x,z) \wedge B(y,z)$ is *monotone* if every z-variable in A occurs negatively.

From a CNF formula $\mathcal{C} = C_1 \wedge \ldots \wedge C_m$ and a partition (X, Y) of its variables, we will define the monotone CNF formula $A_{\mathcal{C}}(X, Z) \wedge B_{\mathcal{C}}(Y, Z)$ as follows. Introduce m new variables $z_1, \ldots, z_m$ and define

$$A_{\mathcal{C}}(x,z) := (C_1^X \vee \neg z_1) \wedge \ldots \wedge (C_m^X \vee \neg z_m)$$
$$B_{\mathcal{C}}(y,z) := (C_1^Y \vee z_1) \wedge \ldots \wedge (C_m^Y \vee z_m)$$

Let $I_{\mathcal{C}}^{(X,Y)}$ be the interpolant associated with the split formula for $A_{\mathcal{C}} \wedge B_{\mathcal{C}}$:

$$I_{\mathcal{C}}^{(X,Y)}(\alpha) = \begin{cases} 0 & \text{if } \{C_i^X : \alpha_i = 1\} \text{ is satisfiable,} \\ 1 & \text{if } \{C_i^Y : \alpha_i = 0\} \text{ is satisfiable.} \end{cases}$$

Observe that $I_{\mathcal{C}}^{(X,Y)}$ is exactly $\text{cert}_{\mathcal{C}}^{(X,Y)}$. Finally, we record the following observation, which relates the complexity of $\mathcal{C}$ and $A_{\mathcal{C}}(x,z) \wedge B_{\mathcal{C}}(y,z)$.

Observation A.2. *There is a Sherali-Adams derivation of $\mathcal{C}$ from $A_{\mathcal{C}}(x,z) \wedge B_{\mathcal{C}}(y,z)$ of size $|\mathcal{C}|$.*

Thus, a lower bound on the size of refutations of the split formula would imply a lower bound on the size of refutations of $\mathcal{C}$.

Feasible Disjunction. We will show that from a Sherali-Adams refutation R of an unsatisfiable set of polynomials $A(x,z) \cup B(y,z)$ and assignment $\alpha \mapsto z$ we can extract a refutation of either $A(x,\alpha)$ or $B(y,\alpha)$ of the same size. In fact, we will first argue that simply restricting the polynomials in $R(x,y,z)$ to either the x-variables or the y-variables will constitute a refutation of either $A(x,\alpha)$ or $B(y,\alpha)$. Let $R_A(x,z)$ be the set of polynomials obtained by projecting R onto the (x,y)-variables. That is, R_A is formed by first removing any polynomial involving an axiom from B (i.e., polynomials of the form $p(y,z)J(x,y,z)$ for $p \in B$) and discarding the y-juntas from the remaining polynomials (i.e., $p(x,z)J(x,z)J'(y)$ is replaced by $p(x,z)J(x,z)$ for $p \in A$. Let $R_B(y,z)$ be defined analogously.

Lemma A.3. *Let α be any $\{0,1\}$-assignment to the variables z. If there is a Sherali-Adams refutation R of an unsatisfiable set of polynomials $A(x,z) \cup B(y,z)$ of size s then there is a refutation of $A(x,\alpha)$ or $B(y,\alpha)$ using only the polynomials in $R_A(x,\alpha)$ or $R_B(y,\alpha)$ respectively.*

The proof of this lemma is similar to a main lemma from [37]. It will be convenient to recall the following notion of a pseudo-expectation for a set of polynomials $R \subseteq \mathbb{R}[x]$.

Definition A.4. Let R be a set of polynomials. A *pseudo-expectation for R* is a linear function $\tilde{\mathbb{E}} : \mathbb{R}[x] \to \mathbb{R}$ satisfying

1. $\tilde{\mathbb{E}}[1] = 1$,

2. $\tilde{\mathbb{E}}[p] \geq 0$ for every $p \in R$.

It is easy to see that -1 is a conic combination of the polynomials in R if and only if there is no pseudo-expectation for R.

Claim. R is a Sherali-Adams refutation if and only if there is no pseudo-expectation for R.

Proof of Claim. Suppose for contradiction that R is a Sherali-Adams refutation and that there exists a pseudo-expectation $\tilde{\mathbb{E}}$ for R. By the first assumption, there exists $\lambda_1, \ldots, \lambda_{|R|} \in \mathbb{R}^{\geq 0}$ such that

$$\sum_{p_i \in R} \lambda_i p_i = -1.$$

Applying $\tilde{\mathbb{E}}$ to both sides of this equation yields a contradiction.

For the other direction, define the following sets

$$T_1 := \{p \in \mathbb{R}[x] : p \text{ is a conic combination of polynomials } R\},$$
$$T_2 := \{-1\}.$$

Since these sets are disjoint and convex, the hyperplane separation theorem yields a linear function $L : \mathbb{R}[x] \to \mathbb{R}$ such that $L[p] \geq 0$ for every $p \in T_1$ and $L[-1] < 0$. By linearity of L, $L[1] > 0$. Define $\tilde{\mathbb{E}}$ as follows: for any polynomial $p \in \mathbb{R}[x]$ let

$$\tilde{\mathbb{E}}[p] := L[p]/L[1]$$

and observe that the desired properties hold. $\qquad\square$

We are now ready to prove Lemma A.3.

Proof of Lemma A.3. Let $R(x, y, z)$ be a Sherali-Adams refutation of $A(x, z) \cup B(y, z)$ and hence $R(x, y, \alpha)$ is a refutation of $A(x, \alpha) \cup B(y, \alpha)$. Suppose that neither $R_A(x, \alpha)$ nor $R_y(y, \alpha)$ is a refutation; that is, -1 cannot be formed as a conic combination of the polynomials R_A or R_B. By the previous claim there exists pseudo-expectations $\tilde{\mathbb{E}}_x$ and $\tilde{\mathbb{E}}_y$ for $R_A(x, \alpha)$ and $R_B(y, \alpha)$ respectively. Define a linear function $\tilde{\mathbb{E}} : \mathbb{R}[x, y] \to \mathbb{R}$ as follows: for every monomial $m(x)m'(y)$ let

$$\tilde{\mathbb{E}}[m(x)m'(y)] := \tilde{\mathbb{E}}_x[m(x)] \cdot \tilde{\mathbb{E}}_y[m'(y)],$$

and extend $\tilde{\mathbb{E}}$ linearly. We claim that $\tilde{\mathbb{E}}$ is a pseudo-expectation for $R(x, y, \alpha)$. Indeed, the following holds by construction

$$\tilde{\mathbb{E}}[p(x)J(x)J'(y)] = \tilde{\mathbb{E}}_x[p(x)J(x)] \cdot \tilde{\mathbb{E}}_y[J(y)] \geq 0,$$

for every $p(x)J(x)J'(y) \in R(x, y, \alpha)$, and similarly for $q(y)J(x)J'(y) \in R(x, y, \alpha)$. Thus, $\tilde{\mathbb{E}}$ is a pseudo-expectation for $R(x, y, \alpha)$, which by the previous claim is a contradiction. $\qquad\square$

Interpolation by Non-negative Span Programs. We now turn to showing that Sherali-Adams refutations of split formulas are interpolated by (monotone) non-negative span programs.

Recall that we say that a CNF formula $A(x, z) \wedge B(y, z)$ on variables is *monotone* if only negative z-literals appear in A. More generally, we say an unsatisfiable set of polynomials $A(x, z) \cup B(y, z)$ is *monotone* if every $p \in A(x, z)$ can be written as polynomial

$$p(x, z) := \prod_{i \in I} z_i p(x)$$

for some (potentially empty) set of indices I and $p \in \mathbb{R}[x]$.

Theorem A.5. *Let $A(x, z) \cup B(y, z)$ be an unsatisfiable set of polynomials. If there is a Sherali-Adams refutation of $A(x, z) \cup B(y, z)$ of size s then there is a non-negative span program M of size $\mathsf{poly}(s)$ such that M computes the interpolant $I_{A \cup B}$. Furthermore, if $A(x, z) \cup B(y, z)$ is monotone then so is M.*

We will first prove our interpolation theorem for sets of polynomials $A(x, z) \cup B(y, z)$ satisfying the following normal form from [49]. Afterwards, we will argue that this normal form assumption is without loss of generality.

Definition A.6. A set of polynomials $A(x, z) \cup B(y, z)$ is in *normal form* if every polynomial either does not contain a variable from z or is of the form $z_i p(x) + p'(x)$ for a variable z_i.

The high-level idea is that from a refutation R of $A \cup B$ we will construct a non-negative span program which, on input α, searches for a refutation in $R_A(x, \alpha)$. The columns of the non-negative span program will be labelled by all monomials restricted to the x-variables in R and each row of the non-negative span program will be a polynomial from $R_A(x, \alpha)$. Therefore, there will exist a non-negative linear combination of the rows equalling $(-1, 0, \dots, 0)$, where the first column is corresponds to the monomial 1, if and only if $R_A(x, \alpha)$ is a Sherali-Adams refutation.

Lemma A.7. *If there is a Sherali-Adams refutation of an (monotone) unsatisfiable set of polynomials $A(x, z) \cup B(y, z)$ in normal form of size s then there is a (monotone) non-negative span program M of size $O(s^2)$ computing the interpolant $I_{A \cup B}$.*

Proof. Let R be a refutation of $A(x, z) \cup B(y, z)$. We will construct a non-negative span program (M, ρ, b) which, given an assignment α, searches for a refutation in $R_A(x, \alpha)$. Denote by S_x the set of monomials in the variables x that occur in the polynomials R, including the constant 1. That is, $m(x) \in S_x$ iff there exists a monomial $m'(y, z)$ and polynomial $p \in R_A$ such that mm' occurs in the expansion of p as a sum of monomials. The columns of the matrix M will index all monomials in S_x.

The rows of M will be all polynomials in $R_A(x, y)$, more formally the coefficients vectors of those polynomials. We will accomplish this as follows:

- For each polynomial $(z_i p(x) + p'(x)) J(x) J'(z)$ in R_A, if $z_i = 1 \nRightarrow J'(z) = 0$ create a row containing the coefficient vector of the polynomial $(p(x) + p'(x)) J(x)$ and label the row with z_i. As well, if $z_i = 0 \nRightarrow J'(z) = 0$ create a row containing the coefficient vector of $p'(x) J(x)$ and label it $\neg z_i$.

- For each polynomial $p(x) J(x) J'(y) \in R_A$ which does not depend on a z-variable, create a row containing the coefficient vector of $p(x) J(x)$ and label it 1.

Let ρ be the labelling that results, and define $b := (-1, 0, \dots, 0)$, where we have assumed that the first column of M corresponds to the monomial 1. Observe that the non-negative span program (M, ρ, b) accepts if and only if -1 is a non-negative linear combination of the rows labelled 1 under an assignment $\alpha \in \{0, 1\}^n$. That is, the non-negative span program tests whether there is a Sherali-Adams refutation of $A(x, \alpha)$ using the polynomials R_A. Furthermore, if A is monotone, then we no longer have any rows labelled by $\neg z_i$ and the non-negative span program is monotone.

It remains to argue that (M, ρ, b) computes $I_{A \cup B}$ correctly. That is, if it accepts $\alpha \in \{0, 1\}^n$ then $A(x, \alpha)$ is unsatisfiable and if it rejects α then $B(y, \alpha)$ is unsatisfiable. This follows from Lemma A.3, noting that the rows of $M^{\rho(\alpha)}$ are exactly coefficient vectors of the polynomials in $R(x, \alpha)$. The non-negative span program has at most $|S_x|$ columns and $|R_A|$ rows, and is therefore of size of $O(|R|^2)$. $\quad\square$

To prove Theorem A.5 it remains to show that the normal form assumption is without loss of generality; for this, we will use the transformations of Pudlák and Sgall [49]. In particular, any set of polynomials $A(x, z)$ can be transformed into normal form via the following process:

1. Introduce $|z|$ new variables w.

2. Replace each polynomial $p(x, z)$ for $p \in A$ by the polynomial $p(x, w)$.

3. For each z_i, introduce the constraint $w_i - z_i = 0$.

It is easy to see that the resulting formula is satisfiable iff the original formula is satisfiable. Next, we show that Sherali-Adams can efficiently deduce the original formula from this normal form.

Claim. The constraints of $A(x, z)$ can be deduced in Sherali-Adams in size $\text{poly}(|A|)$ from the constraints which result from the aforementioned transformation.

Proof of Claim. Consider any monomial $m(x, w)$ that is present in the monomial expansion of the constraints resulting from the transformation; we will deduce $m(x, z)$ as follows. Denote by $m(x, w \setminus w_i)$ the monomial obtained by discarding variable w_i. Sequentially, for each variable w_i in $m(x, w, z)$, replace $m(x, w)$ by $m(x, w) + m(x, w \setminus w_i)(x_i - w_i)$ if m occurs positively in $A(x, z)$ and by $m(x, w) + m(x, w \setminus w_i)(w_i - x_i)$ if m occurs negatively. $\square$

Unfortunately, for *monotone* sets of polynomials $A(x, z)$ this transformation destroys the monotonicity. We will circumvent this by using an alternative transformation for monotone formulas:

1. For each monotone polynomial $\prod_{i=1}^{k} z_i p(x)$ in A, introduce new variables w_i for $i \in [k]$.

2. Replace $\prod_{i=1}^{k} z_i p(x)$ by the following set of polynomial inequalities

$$z_1 \Big(p(x) + \sum_{i \in 2}^{k} w_i \Big) = 0, \quad z_2 w_2 = 0, \quad \dots, z_k w_k = 0.$$

It is not difficult to see that this transformation preserves satisfiability. One direction follows by soundness and the observation that the constraints of the monotone normal form can be derived from the constraints of the original formula in Sherali-Adams. For the other direction, consider any satisfying assignment (x, z) to $\prod_{i=1}^{k} z_i p(x)$ and note that either some $z_i = 0$ or $p(x) = 0$. In the first case, set $w_i = -p(x)$ and $w_j = 0$ for $j \neq i$. In the second case, set $w_j = 0$ for all j.

Claim. The constraints of $A(x, z)$ can be deduced in Sherali-Adams from the constraints which result from the transformation in size $\text{poly}(|A|)$.

Proof of Claim. We will prove that each polynomial $\prod_{i=1}^{k} z_i p(x) \; A(x, z)$ can be deduced in polynomial size from the constraints that result from its monotone transformation as follows: For each $2 \leq i \leq k$ deduce $z_1 \dots z_k w_i = 0$ and add it to $z_1 \dots z_k (p(x) + w_1 + \dots + w_k) = 0$. $\square$

We can now prove Theorem A.5.

Proof of Theorem A.5. Let $A(x, z) \wedge B(y, z)$ be a (monotone) unsatisfiable set of polynomials. By the previous claims we can assume without loss of generality that these formulas are in normal form. Applying Lemma A.7 completes the proof. $\square$

25

Finally, we are ready to prove the main theorem of this section.

Proof of Theorem A.1. Suppose that there is a Sherali-Adams refutation of $\mathcal{C}$ of size s. Then, by Observation A.2, there is a refutation of the associated split formula $A_{\mathcal{C}} \wedge B_{\mathcal{C}}$ of size $O(s)$. Applying Theorem A.5 to $A_{\mathcal{C}} \cup B_{\mathcal{C}}$, produces a monotone non-negative span program of size $\text{poly}(s)$ computing $I_{\mathcal{C}}^{(X,Y)}$, which is equivalent to $\text{cert}_F^{(X,Y)}$. $\square$

References

[1] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory Comput.*, 2(2):19–51, 2006.

[2] Albert Atserias and Elitza N. Maneva. Sherali-adams relaxations and indistinguishability in counting logics. In Shafi Goldwasser, editor, *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, pages 367–379. ACM, 2012.

[3] Albert Atserias and Joanna Ochremiak. Proof complexity meets algebra. *ACM Trans. Comput. Log.*, 20(1):1:1–1:46, 2019.

[4] Paul Beame, Stephen A. Cook, Jeff Edmonds, Russell Impagliazzo, and Toniann Pitassi. The relative complexity of NP search problems. In Frank Thomson Leighton and Allan Borodin, editors, *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing, 29 May-1 June 1995, Las Vegas, Nevada, USA*, pages 303–314. ACM, 1995.

[5] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for lov[a-acute]sz–schrijver systems and beyond follow from multiparty communication complexity. *SIAM J. Comput.*, 37(3):845–869, 2007.

[6] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory Comput.*, 8(1):269–289, 2012.

[7] Christoph Berkholz. The relation between polynomial calculus, sherali-adams, and sum-of-squares proofs. In Rolf Niedermeier and Brigitte Vallée, editors, *35th Symposium on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018, Caen, France*, volume 96 of *LIPIcs*, pages 11:1–11:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[8] Maria Luisa Bonet, Toniann Pitassi, and Ran Raz. Lower bounds for cutting planes proofs with small coefficients. *J. Symb. Log.*, 62(3):708–728, 1997.

[9] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation limits of linear programs (beyond hierarchies). *Math. Oper. Res.*, 40(3):756–772, 2015.

[10] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. *Algorithmica*, 76(3):597–629, 2016.

[11] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 161–170. ACM, 2013.

[12] Samuel R. Buss. Lower bounds on nullstellensatz proofs via designs. In Paul Beam and Samuel R. Buss, editors, *Proof Complexity and Feasible Arithmetics, Proceedings of a DIMACS Workshop, New Brunswick, New Jersey, USA, April 21-24, 1996*, volume 39 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 59–71. DIMACS/AMS, 1996.

[13] Siu Man Chan and Aaron Potechin. Tight bounds for monotone switching networks via fourier analysis. *Theory Comput.*, 10:389–419, 2014.

[14] 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, 2016.

[15] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. *SIAM J. Comput.*, 39(6):2487–2512, 2010.

[16] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting using low-discrepancy gadgets. *SIAM J. Comput.*, 50(1):171–210, 2021.

[17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017.

[18] Stefan S. Dantchev, Barnaby Martin, and Mark Nicholas Charles Rhodes. Tight rank lower bounds for the sherali-adams proof system. *Theor. Comput. Sci.*, 410(21-23):2054–2063, 2009.

[19] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In Nikhil Bansal, Kirk Pruhs, and Clifford Stein, editors, *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007*, pages 53–61. SIAM, 2007.

[20] Mateus de Oliveira Oliveira and Pavel Pudlák. Representations of monotone boolean functions by linear programs. *ACM Trans. Comput. Theory*, 11(4):22:1–22:31, 2019.

[21] Susanna F. de Rezende, Mika Göös, Jakob Nordström, Toniann Pitassi, Robert Robere, and Dmitry Sokolov. Automating algebraic proof systems is np-hard. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 209–222. ACM, 2021.

[22] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, Robert Robere, and Marc Vinyals. Lifting with simple gadgets and applications to circuit and proof complexity. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 24–30. IEEE, 2020.

[23] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication (and what it means for proof and circuit complexity). 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 295–304. IEEE Computer Society, 2016.

[24] Yuval Filmus, Meena Mahajan, Gaurav Sood, and Marc Vinyals. Maxsat resolution and subcube sums. In Luca Pulina and Martina Seidl, editors, *Theory and Applications of Satisfiability Testing - SAT 2020 - 23rd International Conference, Alghero, Italy, July 3-10, 2020, Proceedings*, volume 12178 of *Lecture Notes in Computer Science*, pages 295–311. Springer, 2020.

[25] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *J. ACM*, 62(2):17:1–17:23, 2015.

[26] Noah Fleming, Mika Göös, Russell Impagliazzo, Toniann Pitassi, Robert Robere, Li-Yang Tan, and Avi Wigderson. On the power and limitations of branch and cut. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021*, volume 200 of *LIPIcs*, pages 6:1–6:30. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[27] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Electron. Colloquium Comput. Complex.*, 26:106, 2019.

[28] Noah Fleming, Denis Pankratov, Toniann Pitassi, and Robert Robere. Random $\Theta(\log n)$-CNFs are hard for cutting planes. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 109–120, 2017.

[29] Ankit Garg, Mika Göös, Pritish Kamath, and Dmitry Sokolov. Monotone circuit lower bounds from resolution. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 902–911. ACM, 2018.

[30] Mika Goos. *Communication Lower Bounds via Query Complexity*. PhD thesis, University of Toronto (Canada), 2016.

[31] Mika Göös, Rahul Jain, and Thomas Watson. Extension complexity of independent set polytopes. *SIAM J. Comput.*, 47(1):241–269, 2018.

[32] Mika Göös, Sajin Koroth, Ian Mertz, and Toniann Pitassi. Automating cutting planes is np-hard. In Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy, editors, *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 68–77. ACM, 2020.

[33] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM J. Comput.*, 45(5):1835–1869, 2016.

[34] Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. *SIAM J. Comput.*, 47(5):1778–1806, 2018.

[35] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *SIAM J. Comput.*, 49(4), 2020.

[36] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[37] Tuomas Hakoniemi. Feasible interpolation for polynomial calculus and sums-of-squares. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 63:1–63:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[38] Samuel B. Hopkins, Tselil Schramm, and Luca Trevisan. Subexponential lps approximate max-cut. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 943–953. IEEE, 2020.

[39] Pavel Hrubes. On the nonnegative rank of distance matrices. *Inf. Process. Lett.*, 112(11):457–461, 2012.

[40] Pavel Hrubes. On ϵ-sensitive monotone computations. *Comput. Complex.*, 29(2):6, 2020.

[41] Pavel Hrubes and Pavel Pudlák. Random formulas, monotone circuits, and interpolation. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 121–131. IEEE Computer Society, 2017.

[42] Mauricio Karchmer and Avi Wigderson. On span programs. In *Proceedings of the Eigth Annual Structure in Complexity Theory Conference, San Diego, CA, USA, May 18-21, 1993*, pages 102–111. IEEE Computer Society, 1993.

[43] Pravesh Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps. *CoRR*, abs/1610.02704, 2016.

[44] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 132–145. ACM, 2017.

[45] Jan Krajícek. Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic. *J. Symb. Log.*, 62(2):457–486, 1997.

[46] Claire Mathieu and Alistair Sinclair. Sherali-adams relaxations of the matching polytope. 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 293–302. ACM, 2009.

[47] Ryan O'Donnell and Tselil Schramm. Sherali - adams strikes back. In Amir Shpilka, editor, *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA*, volume 137 of *LIPIcs*, pages 8:1–8:30. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[48] Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *J. Symb. Log.*, 62(3):981–998, 1997.

[49] Pavel Pudlák and Jirí Sgall. Algebraic models of computation and interpolation for algebraic proof systems. In Paul Beam and Samuel R. Buss, editors, *Proof Complexity and Feasible Arithmetics, Proceedings of a DIMACS Workshop, New Brunswick, New Jersey, USA, April 21-24, 1996*, volume 39 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 279–295. DIMACS/AMS, 1996.

[50] Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of UNIQUE GAMES. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 575–585. IEEE Computer Society, 2009.

[51] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Comb.*, 19(3):403–435, 1999.

[52] Alexander A Razborov. Unprovability of lower bounds on circuit size in certain fragments of bounded arithmetic. *Izvestiya: mathematics*, 59(1):205, 1995.

[53] Robert Robere. *Unified lower bounds for monotone computation*. PhD thesis, University of Toronto (Canada), 2018.

[54] Robert Robere, Toniann Pitassi, Benjamin Rossman, and Stephen A. Cook. Exponential lower bounds for monotone span programs. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016*, pages 406–415. IEEE Computer Society, 2016.

[55] Thomas Rothvoß. Some 0/1 polytopes need exponential size extended formulations. *Math. Program.*, 142(1-2):255–268, 2013.

[56] Thomas Rothvoss. The matching polytope has exponential extension complexity. *J. ACM*, 64(6):41:1–41:19, 2017.

[57] 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.

[58] Xiaodi Wu, Penghui Yao, and Henry S. Yuen. Raz-mckenzie simulation with the inner product gadget. *Electron. Colloquium Comput. Complex.*, page 10, 2017.

[59] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991.