Electronic Colloquium on Computational Complexity, Report No. 107 (2007)

Exponential lower bounds and Integrality Gaps for Tree-like Lovász-Schrijver Procedures

Toniann Pitassi and Nathan Segerlind

October 25, 2007

Abstract

The matrix cuts of Lovász and Schrijver are methods for tightening linear relaxations of zero-one programs by the addition of new linear inequalities. We address the question of how many new inequalities are necessary to approximate certain combinatorial problems with strong guarantees, and to solve certain instances of Boolean satisfiability.

We show that relaxations of linear programs, obtained by tightening via any subexponential-size semidefinite Lovász-Schrijver derivation tree, cannot approximate max-k-SAT to a factor better than $1+\frac{1}{2^k-1}$, max-k-XOR to a factor better than $2-\varepsilon$, nor vertex cover to a factor better than $7/6$.

We prove exponential size lower bounds for tree-like Lovász-Schrijver proofs of unsatisfiability for several prominent unsatisfiable CNFs, including random 3-CNF formulas, random systems of linear equations, and the Tseitin graph formulas. Furthermore, we prove that tree-like LS_+ cannot polynomially simulate tree-like cutting planes, and that tree-like LS_+ cannot polynomially simulate unrestricted resolution.

All of our size lower bounds for derivation trees are based upon connections between the size and height of the derivation tree (its *rank*). The primary method is a tree-size/rank trade-off for Lovász-Schrijver refutations: Small tree size implies small rank. Surprisingly, this does not hold for derivations of arbitrary linear inequalities. We show that for LS_0 and LS, there are examples with polynomial-size tree-like derivations, but requiring linear rank.

1 Introduction

The method of semidefinite relaxations has emerged as a powerful tool for approximating NP-complete problems. Central among these techniques are the lift-and-project methods of Lovász and Schrijver [23] for tightening a linear relaxation of a zero-one programming problem. For several optimization problems, a small number of applications of the semidefinite Lovász-Schrijver operator transforms a simple linear programming relaxation into a tighter linear program that better approximates the zero-one program and yields a state-of-the-art approximation algorithm. For example, one round of the semidefinite tightening, starting from the natural linear programming formulation of the independent set problem gives the Lovász Theta functions [22], one round starting from the natural linear programming formulation of the max cut problem gives the famous Goemans-Williamson relaxation for approximating the maximum cut in a graph [15], and three rounds gives the breakthrough Arora Rao Vazirani relaxation for approximating the sparsest cut problem [6] (for a discussion of these algorithms in the context of Lovász-Schrijver tightenings of linear relaxations, see [26]). When used for solving the Boolean satisfiability problem, one round of semidefinite

ISSN 1433-8092

tightening followed by a linear programming test for feasibility efficiently solves satisfiability for CNFs such as the propositional pigeonhole principle, which are known to require exponential runtimes when processed by resolution based solvers [17, 20]. Given the power of Lovász-Schrijver tightening, it is natural to ask what it *cannot* do.

The Lovász-Schrijver operators proceed by iteratively adding new inequalities to the linear relaxation of a zero-one program, and each new inequality satisfies all zero-one solutions to the original program. In this article, we prove lower bounds for the number of inequalities that must be added in order to approximate combinatorial optimization problems and to solve certain instances of the Boolean satisfiability problem. These are unconditional negative results for an important model of computation that includes the best known approximation algorithms for several fundamental problems and an approach to solving satisfiability instances that can be exponentially more efficient than resolution-based solvers.

Most prior results studying the limitations of Lovász-Schrijver tightened linear relaxations have focused on "rank", that is, the number of rounds of tightening that must be applied in order to obtain some approximation guarantee. If the intermediate inequalities are arranged as the nodes of a tree, with the parents of an inequality being the previous inequalities from which it is derived, then the rank of an inequality is the minimum height of a derivation tree for that inequality. We study the *size* of the derivation trees needed to provide good approximations to combinatorial optimization problems and to solve instances of the Boolean satisfiability problem (hence the term "tree-size"). By Caratheodory's theorem we can bound the branching factor of a derivation tree as $O(n^2)$, where n is the number of variables, and thus lower bounds for tree-size imply lower bounds for rank via $rank = \Omega(\log(treesize)/\log n)$. In this way, lower bounds for tree-size are stronger than lower bounds for rank.

1.1 Tightening linear relaxations, an approach to approximation and solving Boolean satisfiability

The linear relaxation of a zero-one program is simply the shift from optimizing an objective function over the zero-one points of a polytope to optimizing over all points of a polytope. A *tightening* of a linear relaxation is the addition of new linear inequalities that are satisfied by all zero-one points of the polytope. Lovász and Schrijver introduced several methods for tightening linear relaxations, among them the *non-commutative (LS_0)*, *linear (LS)*, and *semidefinite (LS_+)* operators [23]. (Definition 2.9 defines these precisely.)

Sometimes by optimizing over all points of a polytope (or ones of its tightenings) we can obtain a decent approximation to the zero-one optimization problem. An *integrality gap* for a polytope is a measure of the quality of such an approximation: For simplicity, we consider only objective functions that take strictly positive values on non-trivial instances. For a minimization problem, the integrality gap of a polytope is the ratio of the minimum of the objective function over the zero-one points of the polytope to the minimum of the objective function over the entire polytope. For maximization problems, it is the ratio of the maximum of the objective function over the entire polytope to the maximum of the objective function over the zero-one points of the polytope. In both cases, the integrality gap is at least one, and the closer the integrality gap is to one, the better the approximation guarantee.

The Lovász-Schrijver operators can be viewed as a way to improve the integrality gap of a zero-one programming problem. When using these methods, the hope is that by adding derived inequalities, fractional solutions that are poor approximations to the zero-one optimum will be eliminated, and the integrality gap of the polytope will become closer to one.

Relaxation and tightening methods can also be used to certify that propositional formulas are unsatisfiable. In this framework, a formula in conjunctive normal form is translated into a system of linear inequalities in a standard way (eg. $x \vee \neg y \vee z$ translates into $x+1-y+z \geq 1$). Derived inequalities are added via one of the Lovász-Schrijver methods. If linear programming reveals that the tightened polytope is empty, that proves that the input CNF is unsatisfiable.

1.2 Summary of results

The first result of the paper is a general tree-size/rank tradeoff for LS_0, LS and LS_+ refutations[1]. In particular, Theorem 3.10 demonstrates that for any LS_0, LS or LS_+ refutation of a system of inequalities I, $\text{rank}(I) \leq 3\sqrt{n \ln S_T(I)}$, where $S_T(I)$ denotes the minimum tree-size of a refutation of I. This implies that $S_T(I) \geq 2^{\Omega((\text{rank}(I))^2/9n)}$. We show that the trade-off of Theorem 3.10 is asymptotically tight (up to a logarithmic factor) for the non-commutative (LS_0) and linear (LS) Lovász-Schrijver operators (Theorem 3.12). For the semidefinite operator (LS_+), we do not know whether or not Theorem 3.10 is asymptotically tight.

Theorem 3.10 allows us to quickly deduce tree-size lower bounds from known rank lower bounds for LS_+ refutations of several well-known “sparse and expanding” systems: Random 3-CNFs, random systems of linear equations, and the Tseitin principles on a constant-degree expander. These results are presented in Section 4.

The trade-off of Theorem 3.10 does not hold for derivations of arbitrary linear inequalities. For LS_0 and LS, such an extension of Theorem 3.10 fails outright: Theorem 3.14 demonstrates sets of inequalities I and a target inequality $a^T X \geq b$ so that $a^T X \geq b$ has polynomial tree-size LS_0 derivations from I but all derivations of $a^T X \geq b$ from I require linear LS rank. At the heart of this is an interesting observation: The deduction theorem in LS_0 and LS can require a linear increase in rank. Whether or not there is a rank tree-size trade-off for arbitrary derivations in LS_+ is still open, as is the question of whether or not the deduction theorem for LS_+ requires an increase in rank.

Despite our lack of a general tree-size/rank trade-off for derivations of arbitrary linear inequalities, we prove integrality gaps for LS_+ tightenings of small tree-size by using ad-hoc modifications of the technique. For several combinatorial optimization problems, we show that there are instances for which every polytope that is obtained by applying an LS_+ tightening of sub-exponential tree-size has a large integrality gap: For max-k-SAT, the integrality gap is $1+\frac{1}{2^k-1}$, for max-k-LIN have integrality gap $2-\varepsilon$, and for vertex cover, the integrality gap is $7/6$. These results are presented in Section 5.

In Section 6, we address how well LS_+ stacks up as a propositional proof system. In particular, we show that tree-like LS_+ refutations require an exponential increase in size to simulate tree-like Gomory-Chvatal cutting planes refutations, Theorem 6.10, and that tree-like LS_+ refutations require an exponential increase in size to simulate DAG-like resolution refutations, Theorem 6.27. In the language of propositional proof complexity [12], we show that LS_+ does not p-simulate tree-like cutting planes nor does it p-simulate DAG-like resolution.

[1] A *refutation* is a derivation that shows a zero-one program has no feasible solutions.

1.3 Comparisons with previous work

The technique of applying a partial assignment to reduce the rank of a tree-like Lovász-Schrijver derivation is inspired by a line of work due to Grigoriev and his coauthors [16, 18, 17, 19] and a paper by Kojevnikov and Itsykson [21] that prove lower bounds on the tree-sizes of LS_+ refutations by proving lower bounds on the tree-sizes of *static positivstellensatz refutations*. (Static positivstellensaz refutations can efficiently simulate tree-like LS_+ derivations, so LS_+ tree-size bounds follow immediately from these size bounds.) A technique frequently used in those analyses is to show that given a small static positivstellensatz refutation, one can construct a small assignment to the variables that will cause all monomials of large multilinear degree to vanish, yet static positivstellensatz refutations of the restricted system of inequalities still require large multilinear degree. Grigoriev et al used this technique to show that static positivstellensatz refutations of a system of inequalities known as the *fractional knapsack* require exponential size [17]. Kojevnikov and Itsykson used a variant of it to show an exponential size lower bound for static positivstellensatz refutations of the Tseitin principle [21].

In this paper, we apply partial assignments that eliminate all paths in an LS_+ derivation that lift on many different variables, thereby creating low-rank derivations that contradict known rank bounds[2]. This technique is somewhat easier to apply than one based upon the static positivstellensatz, simply because there are many more rank lower bounds known for LS_+ than there are multilinear degree bounds known for static positivstellensatz refutations[3].

Our results focus on the Lovász-Schrijver systems, and eliminate reasoning about the (apparently) more complicated and powerful static positivstellensatz system. For example, our size lower bound for tree-like LS_+ refutations of the Tseitin principle is self-contained in that it follows only from a simple rank lower bound for LS_+ refutations of the Tseitin principle and a general tree-size/rank trade-off for LS_+ refutations. Our tree-size lower bounds for refuting random 3-CNFs and random systems of linear equations are new, as are our separations of tree-like cutting planes and unrestricted resolution from tree-like LS_+.

To the best of our knowledge, all integrality gaps shown earlier for Lovász-Schrijver tightenings of linear relaxations applied only to tightenings of low rank, so our results for tree-size-based integrality gaps are new. However, this work on integrality gaps falls squarely within the philosophy delineated by Arora, Bollobas and Lovász [5]. Hardness of approximation results based upon PCP technology are wanting in three ways. First, such results are conditional upon complexity theoretic conjectures such as $P \neq NP$ or $NP \neq ZPP$ or some such thing. Second, because of the heavy use of reductions that increase input size by polynomial factors, PCP results do not rule out the possibility of slightly-subexponential time approximation algorithms that run in time $2^{n^{\varepsilon}}$ (with $\varepsilon < 1$). Third, for many problems, there is a nagging gap between known PCP based hardness of approximation results and the best known approximation algorithms. By considering a concrete approach, Lovász-Schrijver tightenings, we establish *unconditional* limits to approximation possible with current algorithmic techniques. Furthermore, the bounds we obtain are of the form $2^{\Omega(n)}$ where n is the input size, so we rule out the possibility of weakly sub-exponential algorithms (of a particular form).

The proof technique that we employ explicitly uses pre-existing rank bounds. In particular, our tree-size-based integrality gaps for max-k-SAT and max-k-LIN directly extend the rank-based integrality gaps shown in [9], and our our tree-size-based integrality gap for vertex cover extends the rank-based integrality gap shown in [24]. Our refutation tree-size bounds for Tsetin principles and random linear equations extend the

[2]The distinction between paths that lift on many different variables and paths that lift many times upon a small set of variables is addressed in Subsection 3.2.

[3]One advantage of working with static positivstellensatz derivations is closure under certain local reductions, see Subsection 6.1.

rank bounds of [9] and our refutation tree-size bounds for random 3-CNFs extend the rank bounds of [2]. The separation of tree-like GC cutting planes from tree-like LS_+ builds upon a rank bound for the counting mod two principles that is implicit in the work of Grigoriev [16] and Kojevnikov and Itsykson [21], and the separation of DAG-like resolution from tree-like LS_+ begins with an extension of the LS_0 rank bound for the GT_n principles proved in [9]. The asymptotic optimality of Theorem 3.10 for LS_0 and LS, and the "deduction requires an increase in rank" result for LS_0 and LS, uses the $\Omega(n)$ rank bound proved for refutations of the propositional pigeonhole principle by Grigoriev Hirsch and Pasechnik [17].

1.4 Outline

The rest of the paper is organized as follows. In Section 2 we present some elementary background material, and define the Lovász-Schrijver proof systems (also known as matrix-cut proof systems), and prove some basic properties of these systems. In Section 3 we prove the tree-size/rank tradeoff for LS_0, LS and LS_+ refutations, and prove that such a tradeoff is false for LS_0 and LS derivations of arbitrary linear inequalities. In Section 4, we combine the tree-size/rank tradeoff with existing rank bounds to obtain new tree-size bounds for refutations of sparse, exanding formulas. In Section 5, we prove the integrality gaps for subexponential tree-size LS_+ tightenings of max-k-SAT, max-k-LIN, and vertex cover. In Section 6, we show that tree-like LS_+ cannot polynomially simulate tree-like Gomory-Chvatal Cutting Planes proofs, nor can it polynomially simulate unrestricted resolution. We end our journey in Section 7 with discussion and open problems.

2 Background

A *literal* is a propositional variable or its negation. A *clause* is a disjunction of literals. A *CNF* is a conjunction of clauses, specified as a set of clauses. A *k-CNF* is a CNF whose clauses are each of width at most k. When processed by zero-one programming methods, clauses are converted into inequalities in the usual way, eg. $X_1 \vee \neg X_2 \vee X_3$ is converted to $X_1 + (1 - X_2) + X_3 \geq 1$. Notice that the $0/1$ solutions to the inequality are exactly the satisfying assignments to the clause. Variables are written with upper case letters, ie. $X_1, \ldots X_n$, whereas points in $\mathbb{R}$ are written with lower case letters, eg. $x_1, \ldots x_n \in \mathbb{R}$. Vectors of variables are written simply as X and elements of $\mathbb{R}^n$ are written as x.

A *restriction* ρ is a map from a set of variables to $\{0, 1, *\}$. For a polynomial $f(X)$, *the restriction of* $f(X)$ *by* ρ, $f(X) \restriction_\rho$ is is defined by substituting 1 for each X_i with $\rho(X_i) = 1$, and substituting 0 for each X_i with $\rho(X_i) = 0$. The restriction of a polynomial inequality, $(f(X) \geq g(X)) \restriction_\rho$ is defined to be $f(X) \restriction_\rho \geq g(X) \restriction_\rho$.

We make heavy use of the affine Farkas lemma as a kind of "completeness theorem" for linear programming.

Lemma 2.1. *(Affine Farkas Lemma) Let* $I = \{a_i^T X \geq b_i \mid i = 1, \ldots, m\}$ *be a system of inequalities so that for all* x *satisfying each inequality in* I, $c^T x \leq b$. *Then there exists* $\alpha_1, \ldots, \alpha_m$, *each* $\alpha_i \geq 0$, *such that* $d - c^T X = \sum_{i=1}^m \alpha_i (b_i - a_i^T X)$.

2.1 Expansion basics

Many of the tree-size lower bounds obtained in Section 4 and Section 5 depend upon expansion in the constraints of the problems.

Definition 2.2. *Let $e(V_1,V_2)$ be the number of edges (v_1,v_2) with $v_i \in V_i$. The edge-expansion of a graph $G=(V,E)$ is*

$$\min_{\substack{S\subseteq V \\ 0<|S|\leq|V|/2}} \frac{e(S,V\setminus S)}{|S|}.$$

Definition 2.3. *A bipartite graph from V to U is an (r,c)-expander if, for all subsets $X\subset V$ where $|X|\leq r$, we have $\Gamma(X)\geq c|X|$. The expansion of a set $X\subseteq V$, $e(X)$, is the value $|\Gamma(X)|/|X|$.*

Definition 2.4. *Let G be a bipartite graph from V to U. The* boundary *of a set $X\subset V$, ∂X, is defined as $\partial X=\{u\in U \ : \ |\Gamma(u)\cap X|=1\}$. G is an (r,c)-boundary expander if for all subsets $X\subset V$ where $|X|\leq r$, we have $|\partial X|\geq c|X|$. The boundary expansion of a set $X\subset V$ is the value $|\partial X|/|X|$.*

The following fact relates bipartite expansion with boundary-expansion.

Fact 2.5. *If G is a bipartite graph from V to U where V has maximal degree d and if G is an (r,c)-expander, then G is a $(r,2c-d)$-boundary expander.*

2.2 Matrix-cut proof systems

Our results prove a connection between tree-size, a concept that is inherently syntactic, and rank, a concept that is more often studied from a dual perspective that characterizes the points that survive all possible cuts (via "protection matrices"). To bridge these perspectives, we must use two equivalent formulations of the Lovász-Schrijver systems, and the requisite notation to handle both.

When manipulating the Lovász-Schrijver systems syntactically, we reason about points in $\mathbb{R}^n$, but when we take the dual perspective of protection matrices, we reason about points in $\mathbb{R}^{n+1}$. Don't blame us! These perspectives and methods of notation are standard. For fixed $n\in\mathbb{N}$, *elements of $\mathbb{R}^n$ are indexed by $\{1,\dots n\}$ and elements of $\mathbb{R}^{n+1}$ are indexed by $\{0,\dots n\}$.*

Definition 2.6. *A* cone *is a subset of $\mathbb{R}^{n+1}$ that is closed under addition and multiplication by positive scalars. A* polyhedral cone *is the set of solutions to a family of homogenized linear inequalities, $\{x\in\mathbb{R}^{n+1} \mid Ax\geq 0\}$. A* face *of polyhedral cone $\{x\in\mathbb{R}^{n+1}\mid Ax\geq 0\}$ is a set of the form $\{x\in\mathbb{R}^{n+1}\mid A'x\geq 0,\ A''x=0\}$ where A' and A'' partition the rows of A.*

Let $I=\{a_i^TX\geq b_i \mid i=1,\dots,m\}$ be a system of linear inequalities in the variables $X_1,\dots X_n$. Define the polytope of *I as $P_I=\{x\in\mathbb{R}^n\mid \forall i\in[m],\ a_i^Tx\geq b_i\}$. Define* the homogenized cone of *I as $K_I=\{x\in\mathbb{R}^{n+1}\mid \forall i\in[m],\ a_i^Tx-b_ix_0\geq 0\}$.*

Definition 2.7. *For $S\subseteq\mathbb{R}^{n+1}$, let $S\restriction_{x_0=1}=\{x\in\mathbb{R}^n\mid (1,x_1,\dots x_n)\in S\}$. For each $i\in[n]$, $\varepsilon\in\{0,1\}$, let $\{X_i=\varepsilon\}$ denote either the set $\{x\in\mathbb{R}^n\mid x_i=\varepsilon\}$, or the set $\{x\in\mathbb{R}^{n+1}\mid x_i=\varepsilon\}$, as context dictates.*

Definition 2.8. *Let $x\in[0,1]^n$. $Supp(x)$ are those indices/coordinates i such that x_i is equal to 0 or 1. $E(x)$ are the other indices–those indices j such that x_j is not integral. Of course $[n]=Supp(x)\cup E(x)$.*

There are several cutting planes proof systems defined by Lovász and Schrijver, collectively referred to as matrix cuts [23]. In these proof systems, we begin with a polytope P defined by the linear relaxation of the zero-one programming problem.

Definition 2.9. *Given a polytope $P\subseteq[0,1]^n$ defined by $a_i^TX\geq b_i$ for $i=1,2,\dots,m$:*

(1) *An inequality $d - c^T X \geq 0$ is called an N-cut for P if*

$$d - c^T X = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{i,j}(b_i - a_i^T X)X_j + \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_{i,j}(b_i - a_i^T X)(1 - X_j) + \sum_{j=1}^{n} \lambda_j (X_j^2 - X_j) \quad (1)$$

where $\alpha_{i,j}, \beta_{i,j} \geq 0$ and $\lambda_j \in \mathbb{R}$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$.

(2) *An N-cut is called an N_0-cut if Equation 1 holds when we view $X_i X_j$ as distinct from $X_j X_i$ for all $1 \leq i < j \leq n$. (For this reason, N_0-cuts are called* non-commutative cuts.*)*

(3) *An inequality $d - c^T X$ is called an N_+-cut if*

$$d - c^T X = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{i,j}(b_i - a_i^T X)X_j + \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_{ij}(b_i - a_i^T X)(1 - X_j) + \sum_{j=1}^{n} \lambda_j (X_j^2 - X_j) + \sum_k (g_k + h_k^T X)^2$$

where $\alpha_{i,j}, \beta_{i,j} \geq 0$, $\lambda_j \in \mathbb{R}$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$ and each $(g_k + h_k^T X)$ is an affine function.

For each of the above cuts, we say that the inequality $a_i^T \geq b_i$ is a hypothesis of a lifting on the literal X_j *if $\alpha_{ij} > 0$ and that is* a hypothesis of a lifting on the literal $1 - X_j$ *if $\beta_{ij} > 0$.*

Definition 2.10. *A Lovász-Schrijver (LS) derivation of $a^T X \geq b$ from a set of linear inequalities I is a sequence of inequalities $g_1, \ldots, g_q$ such that each g_i is either an inequality from I, or follows from previous inequalities by an N-cut as defined above, and such that the final inequality is $a^T X \geq b$. Similarly, a LS_0 derivation uses N_0-cuts and LS_+ uses N_+-cuts.*

An elimination *of a point $x \in \mathbb{R}^n$ from I is a derivation from I of an inequality $c^T X \geq d$ such that $c^T x < d$. A* refutation of *I is a derivation of* $0 \geq 1$ *from I.*

An LS (LS_0, LS_+) tightening *of a polytope P_I is a set of inequalities, $\{c_j^T X \geq d_j \mid j \in J\}$ so that each $c_j^T X \geq d_j$ is a formula in some derivation Γ from the hypotheses I. (Note that it is possible for Γ to have multiple sinks.)*

Definition 2.11. *Let $\mathcal{P}$ be one of the proof systems LS, LS_0 or LS_+. Let Γ be an $\mathcal{P}$-derivation from I, viewed as a directed acyclic graph. The derivation Γ is tree-like if each inequality in the derivation, other than the initial inequalities, is used at most once. In a tree-like derivation the underlying graph, excluding the leaf nodes, is a forest. The inequalities in Γ are represented with all coefficients in binary notation. The size of Γ is the size of the underlying directed acyclic graph; the rank of Γ is the depth of the underlying directed acyclic graph. For a set of boolean inequalities I, the $\mathcal{P}$-size of I is the minimal size over all $\mathcal{P}$ refutations of I. The $\mathcal{P}$-tree-size of I is the minimal size over all tree-like $\mathcal{P}$ refutations of I. The $\mathcal{P}$-rank of I is the minimal rank over all $\mathcal{P}$-refutations of I.*

A few technical points. First, it is entirely possible that some nodes of the derivation-DAG are labeled with the same inequality. For DAG-like derivations, we may assume this is not the case, but for tree-like derivations, it is a common situation. Second, we define tree-size to be the number of nodes in the derivation tree, not the sum of the bit-sizes needed to represent each inequality of the derivation (the bit-size of the derivation). This is because the tree-size trade-offs and lower bounds that we prove apply regardless of the sizes of the coefficients. On the other hand, the upper bounds that we make use of are easily seen to create derivations that are of polynomial bit-size. Third, in our definition of the Lovász-Schrijver systems, we can derive a new inequality from any number of previous inequalities in one step. However, in light of Caratheodory's theorem, we may assume without loss of generality that the fan-in in is at most $n^2 + n + 1$.

Definition 2.12. *Let I be a system of inequalities over the variables $X_1, \dots X_n$ that includes $0 \le X_i \le 1$ for all $i \in [n]$. Define $LS_0^r(I)$ to be the set of all linear inequalities with LS_0 derivations from I of rank at most r, $LS^r(I)$ to be the set of all linear inequalities with LS derivations from I of rank at most r, and $LS_+^r(I)$ to be the set of all linear inequalities with LS_+ derivations from I of rank at most r.*

The following simple fact is repeatedly used in this article. It holds simply because the equalities that define N_0-cuts (N-cuts, N_+-cuts) are preserved under substituting 0 or 1 for a variable.

Lemma 2.13. *Let Γ be an LS_0 (LS, LS_+) derivation of $c^T X \ge d$ from hypotheses I. Let ρ be a restriction to the variables of X. $\Gamma \upharpoonright_\rho$ is an LS_0 (LS, LS_+) derivation of $\left(c^T X \ge d\right) \upharpoonright_\rho$ from the hypotheses $I \upharpoonright_\rho$.*

Corollary 2.14. *Let Γ be an LS_0 (LS, LS_+) elimination of $w \in \mathbb{R}^n$ from hypotheses I. Let ρ be a restriction to the variables of X such that for all $i \in [n]$, $\rho(X_i) \in \{0,1\} \Rightarrow w_i = \rho(X_i)$. Let w' be the vector indexed by variables from $[n] \setminus dom(\rho)$ that agrees with w on $[n] \setminus dom(\rho)$. $\Gamma \upharpoonright_\rho$ is an LS_0 (LS, LS_+) elimination of w' from the hypotheses $I \upharpoonright_\rho$.*

2.3 Protection matrices and protection vectors

When analyzing the rank needed to refute systems of inequalities and to eliminate points from systems of inequalities, a dual perpective (introduced by Lovász and Schrijver [23]) has often been used [5, 9, 2, 27, 25, 24].

Definition 2.15. *Let $y \in \mathbb{R}^{n+1}$ be given, and let $K \subseteq \mathbb{R}^{n+1}$ be a cone. An* LS_0 protection matrix for y with respect to K *is a matrix $Y \in \mathbb{R}^{(n+1)\times(n+1)}$ such that:*

1. *$Ye_0 = diag(Y) = Y^T e_0 = y$,*
2. *For all $i = 0, \dots n$, $Ye_i \in K$ and $Y(e_0 - e_i) \in K$.*
3. *If $x_i = 0$ then $Ye_i = 0$, and if $x_i = y_0$ then $Ye_i = y$.*

If Y is also symmetric, then Y is said to be an LS protection matrix. *If Y is also positive semidefinite, then Y is said to be an* LS_+ protection matrix. *If Y is an LS_0 (LS, LS_+) protection matrix for y with respect to $\mathbb{R}^{n+1}$ (ie. if it is protection matrix for y with respect to some cone $K \subseteq \mathbb{R}^{n+1}$) then we simply say that Y is an LS_0 (LS, LS_+) protection matrix for y.*

Definition 2.16. *Let $K \subseteq \mathbb{R}^{n+1}$ be a cone. Define $N_0(K)$ to be set of $y \in \mathbb{R}^{n+1}$ such that there exists an LS_0 protection matrix for y with respect to K, define $N(K)$ to be set of $y \in \mathbb{R}^{n+1}$ such that there exists an LS protection matrix for y with respect to K, and define $N_+(K)$ to be set of $y \in \mathbb{R}^{n+1}$ such that there exists an LS_+ protection matrix for y with respect to K.*

The sets $N_0(K)$, $N(K)$ and $N_+(K)$ are easily seen to be cones, and therefore the construction can be iterated.

Definition 2.17. *Let $K \subseteq \mathbb{R}^{n+1}$ be a cone. Inductively define $N_0^0(K) = K$ and $N_0^{r+1}(K) = N_0(N_0^r(K))$. Define $N^r(K)$ and $N_+^r(K)$ similarly.*

The connection between the N_0, N and N_+ operators, which work on cones in $\mathbb{R}^{n+1}$, and the syntactic definition of the LS_0, LS and LS_+ deduction systems is summarized in the following fundamental theorem of Lovász and Schrijver.

Theorem 2.18. *[23] Let I be a set of inequalities in $\{X_1,\dots X_n\}$ that includes the inequalities $0 \le X_i \le 1$ for all $i \in [n]$, and let $K_I \subseteq \mathbb{R}^{n+1}$ be the polyhedral cone given by the homogenization of I. $P_{LS_0^r(I)} = N_0^r(K_I) \restriction_{X_0=1}$, $P_{LS^r(I)} = N^r(K_I) \restriction_{X_0=1}$, and $P_{LS_+^r(I)} = N_+^r(K_I) \restriction_{X_0=1}$.*

Corollary 2.19. *Let I be a set of inequalities in $\{X_1,\dots X_n\}$ that includes the inequalities $0 \le X_i \le 1$ for all $i \in [n]$, and let $K_I \subseteq \mathbb{R}^{n+1}$ be the polyhedral cone given by the homogenization of I. There exists a rank $\le r$ LS refutation of I if and only if every point of $N^r(K_I)$ satisfies $0 \ge X_0$, if and only if $N^r(K_I) \restriction_{X_0=1}$ is empty. There exists a LS elimination of $x \in \mathbb{R}^n$ from I of rank at most r if and only if $\binom{1}{x} \notin N^r(K_I)$. The analogous statements relate LS_0 with N_0, and LS_+ with N_+.*

A contrapositive reading of the definition shows that for $y \in \mathbb{R}^{n+1}$ and a protection matrix Y for y, for any cone Q with $y \in Q$, if $y \notin N_+(Q)$ then there exists some $i \in [n]$ with either $Ye_i \notin Q$ or $Y(e_0 - e_i) \notin Q$. That is, if y fails to make it into the next round of LS_+ tightening, it is because column of Y fails to belong to Q. By a variant of Theorem 2.18, we are able to make analogous claims for the syntactic formulation of N_+ cuts.

Definition 2.20. *Let $x \in \mathbb{R}^n$ be given, and let Y be an LS_0 protection matrix for $\binom{1}{x}$. For each $i = 0,\dots n$, let y^i be the bottom n entries of the $n+1$ dimensional column vector Ye_i, so that $Ye_i = \binom{x_i}{y^i}$. For $i \in E(x)$, let $PV_{i,1}(Y)$ denote the vector y^i/x_i and let $PV_{i,0}(Y)$ denote the vector $(x - y^i)/(1 - x_i)$. For $i \in Supp(x)$, let $PV_{i,0}(Y) = PV_{i,1}(Y) = x$. These $2n$ vectors are collectively known as the* protection vectors for x from Y.

Lemma 2.21. *(proof in Appendix) Let $I = \{a_1^T X \ge b_1, \dots a_m^T X \ge b_m\}$ be a system of inequalities. Let $c^T X \ge d$ be an inequality obtained by one one round of LS_+ lift-and-project from I, that is:*

$$d - c^T X \quad = \quad \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{i,j}(b_i - a_i^T X)X_j + \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_{i,j}(b_i - a_i^T X)(1 - X_j) + \sum_{j=1}^{n} \lambda_j (X_j^2 - X_j) + \sum_k (g_k + h_k^T X)^2$$

with each $\alpha_{i,j}, \beta_{i,j} \ge 0$. Let $x \in \mathbb{R}^n$ be given such that $c^T x < d$. If Y is an LS_+ protection matrix for $\binom{1}{x}$, then there exists an $i \in [m]$ and a $j \in [n]$ so that either:

1. *$a_i^T X \ge b_i$ is used as the hypothesis for a lifting inference on X_j, $x_j \ne 0$, and $a_i^T PV_{j,1}(Y) < b_i$.*
2. *$a_i^T X \ge b_i$ is used as the hypothesis for a lifting inference on $1 - X_j$, $x_j \ne 1$, and $a_i^T PV_{j,0}(Y) < b_i$.*

The proof of Lemma 2.21 is immediate from the usual proof of Theorem 2.18. The following lemma is immediate from the definitions:

Lemma 2.22. *Let $x \in \mathbb{R}^n$ be given, and let Y be an LS_0 protection matrix for $\binom{1}{x}$. For all $i \in E(x)$, $\varepsilon \in \{0,1\}$, $(PV(Y)_{i,\varepsilon})_i = \varepsilon$. For all $i \in Supp(x)$, all $\varepsilon \in \{0,1\}$, $(PV(Y)_{i,\varepsilon})_i = x_i$.*

3 Tree-size versus rank

The proof of the tree-size/rank trade-off is based upon constructing a partial assignment that kills all paths that lift on a large number of variables - this should then create a low rank refutation of the system. However, it is not clear what happens to paths that repeatedly lift on a small number of variables. The distinction is between rank and what we dub *variable rank*.

We show that rank and variable rank are equal in Subsection 3.2, and we use this to prove the tree-size/rank trade-off in Subsection 3.3. First, we need some properties of how the Lovász-Schrijver operators behave on the faces of a polyhedral cone.

3.1 Lovász-Schrijver operators and projections

The following lemma and its consequences are crucial for the results of this paper.

Lemma 3.1. *(Lemma 3.6 of [13]) If F is a face of a polyhedral cone K, then $N_0(F) = N_0(K) \cap F$, $N(F) = N(K) \cap F$ and $N_+(F) = N_+(K) \cap F$.*

Proof. We present the argument for the N_0 operator; the other cases are analogous.

Let $y \in N_0(K \cap F)$ be given. By definition, there is an LS_0 protection matrix for y with respect to $K \cap F$. This is clearly also an LS_0 protection matrix for y with respect to K. Therefore, $y \in N_0(K)$ and thus $y \in N_0(K) \cap F$.

For the other direction, choose a system of homogenized inequalities A so that $K = \{y \in \mathbb{R}^{n+1} \mid Ay \geq 0\}$; let $A_1, \ldots A_m$ denote the rows of A. Choose $J \subseteq [m]$ so that $F = \{y \in K \mid A_J y = 0\}$. Let $y \in N_0(K) \cap F$ be given. There is an LS_0 protection matrix Y for y with respect to K. Let $i \in \{0, \ldots n\}$ and $j \in J$ be given. Because Y is an LS_0 protection matrix for y with respect to K, $Ye_i \in K$ and $Y(e_0 - e_i) \in K$. Therefore $A_j(Ye_i) \geq 0$ and $A_j(Ye_0 - Ye_i) \geq 0$. However, because $Ye_0 = y \in F$, $A_j Ye_0 = 0$, and therefore $A_j(-Ye_i) \geq 0$. Because we also have that $A_j(Ye_i) \geq 0$, $A_j(Ye_i) = 0$. Because $j \in J$ was arbitrary, both $Ye_i \in K \cap F$ and $Ye_0 - Ye_i \in K \cap F$. Thus Y is an LS_0 protection matrix for y with respect to $K \cap F$, and therefore $y \in N_0(K \cap F)$. □

Lemma 3.2. *Let I be system of inequalities over the variables $X_1, \ldots X_n$, such that I includes $0 \leq X_i \leq 1$ for each $i \in [n]$. For every $i \in [n]$, and every inequality $c^T X \geq d$, if there is a derivation of $(c^T X \geq d) \restriction_{X_i=0}$ from $I \restriction_{X_i=0}$ of rank r, then there is $\varepsilon \geq 0$ and a derivation of $c^T X + \varepsilon X_i \geq d$ of rank at most r. Similarly, if there is a derivation of $(c^T X \geq d) \restriction_{X_i=1}$ from $I \restriction_{X_i=1}$ of rank r, then there is $\varepsilon \geq 0$ and a derivation of $c^T X + \varepsilon(1 - X_i) \geq d$ of rank at most r.*

Proof. We present the case of $X_i = 0$ for the LS system, the case of $X_i = 1$ and the LS_0 and LS_+ systems are entirely analogous. Let I, $i \in [n]$, and $c^T X \geq d$ be given as in the statement of the Lemma. Suppose that there is a rank r derivation of $(c^T X \geq d) \restriction_{X_i=0}$ from $I \restriction_{X_i=0}$. As a consequence, we have that there is a rank $\leq r$ derivation of $c^T X \geq d$ from $I \cup \{X_i = 0\}$, and therefore, by Theorem 2.18, for all $x \in (N^r(K_I \cap \{X_i = 0\})) \restriction_{X_0=1}$, $c^T x \geq d$. On the other hand:

$$\begin{aligned}(N^r(K_I \cap \{X_i = 0\})) \restriction_{X_0=1} &= (N^r(K_I) \cap \{X_i = 0\}) \restriction_{X_0=1} = (N^r(K_I) \restriction_{X_0=1}) \cap \{X_i = 0\} \\ &= P_{LS^r(I)} \cap \{X_i = 0\} = P_{LS^r(I)} \cap \{X_i \leq 0\}\end{aligned}$$

Therefore, by the affine Farkas lemma, Lemma 2.1, there exist $\alpha_1, \ldots \alpha_m$, with each $\alpha_j \geq 0$, $\varepsilon \geq 0$, and inequalities $a_j^T - b_j \geq 0$, each derivable from I within rank r, so that: $\sum_{j=1}^m \alpha_j(a_j^T - b_j) + \varepsilon(-X_i) = c^T X - d$, and thus $\sum_{j=1}^m \alpha_j(a_j^T - b_j) = c^T X + \varepsilon X_i - d$. Therefore $c^T X + \varepsilon X_i - d$ can be derived in LS rank $\leq r$ from I.

□

Corollary 3.3. *Let I be system of inequalities over variables X_i, $i \in [n]$. For every $i \in [n]$, if there is a refutation of $I \restriction_{X_i=0}$ of rank r, then there is $\varepsilon > 0$ and a derivation of $X_i \geq \varepsilon$ of rank at most r. Similarly, if there is a refutation of $I \restriction_{X_i=1}$ of rank r, then there is $\varepsilon > 0$ and a derivation of $(1 - X_i) \geq \varepsilon$ of rank at most r.*

Proof. Suppose that there is a refutation of $I\restriction_{X_i=0}$ of rank at most r. That is, there is a derivation of $0 \geq 1$ from $I\restriction_{X_i=0}$ of rank at most r. By Lemma 3.2, there exists $a \geq 0$ so that there is a rank at most r derivation of $aX_i \geq 1$ from I. If $a > 0$, we multiply by $1/a$ and have $X_i \geq 1/a > 0$. If $a = 0$, there is a derivation of $0 \geq 1$ from I - we add $X_i \geq 0$ to this to obtain $X_i \geq 1$. The case for $I\restriction_{X_i=1}$ is analogous.

$\square$

Definition 3.4. *Let $y \in \mathbb{R}^{n+1}$ be given with $y_0 = 1$. and let $K \subseteq \mathbb{R}^{n+1}$ be a cone. Let Y be a LS_0 (LS, LS_+) protection matrix for y with respect to K. Y is said to be* support extending *if for all $i \in [n]$, for all $j \in [n]$, $y_j = 1 \Rightarrow (Ye_i)_j = y_i$, and $y_j = 0 \Rightarrow (Ye_i)_j = 0$.*

The designation "support extending" was chosen because of the following lemma:

Lemma 3.5. *Let $x \in \mathbb{R}^n$ be given and let I be a set of inequalities that includes $0 \leq X_i \leq 1$ for all $i \in [n]$. If Y is a support-extending protection matrix for $\binom{1}{x}$ with respect to the cone K_I, then for each $i \in [n]$, $\varepsilon \in \{0,1\}$, $Supp(x) \cup \{i\} \subseteq Supp(PV_{i,\varepsilon}(Y))$.*

Proof. For $i \in \mathrm{Supp}(x)$, $PV_{i,0}(Y) = PV_{i,1}(Y) = x$, so the claim holds. Now consider $i \in E(x)$. For each $\varepsilon \in \{0,1\}$, Lemma 2.22 guarantees that $i \in \mathrm{Supp}(PV_{i,\varepsilon}(Y))$. Now, let $j \in \mathrm{Supp}(y)$ be given.

$$(PV_{i,0}(Y))_j = \frac{x_j - (Ye_i)_j}{1 - x_i} = \begin{cases} \frac{0-0}{1-x_i} = 0 = x_j & \text{if } x_j = 0 \\ \frac{1-x_i}{1-x_i} = 1 = x_j & \text{if } x_j = 1 \end{cases}$$

$$(PV_{i,1}(Y))_j = \frac{(Ye_i)_j}{x_i} = \begin{cases} \frac{0}{x_i} = 0 = x_j & \text{if } x_j = 0 \\ \frac{x_i}{x_i} = 1 = x_j & \text{if } x_j = 1 \end{cases}$$

Thus, $\mathrm{Supp}(x) \subseteq \mathrm{Supp}(PV_{i,\varepsilon}(Y))$. We actually get that the protection vectors also agree with x on the support of x, but we do not need that in any arguments of this paper. $\square$

Lemma 3.6. *Let $K \subseteq \mathbb{R}^{n+1}$ be a polyhedral cone that satisfies the inequalities $0 \leq X_i \leq X_0$ for all $i \in [n]$. For all $y \in K$ with $y_0 = 1$, $y \in N_0(K)$ ($N(K)$, $N_+(K)$) if and only if there exists a support extending LS_0 (LS, LS_+) protection matrix for y with respect to K.*

Proof. We present the proof for LS_0 operator; the other cases are identical. Clearly, if such a protection matrix exists, then $y \in N_0(K)$. Now suppose that $y \in N_0(K)$. Let $F = \{z \in K \mid \forall i \in [n],\ (y_i = 1 \Rightarrow z_i = z_0),\ (y_i = 0 \Rightarrow z_i = 0)\}$, this is a face of K because K satisfies the inequalities $0 \leq X_i \leq X_0$. Of course, $y \in N_0(K) \cap F$, and by Lemma 3.1, $N_0(K) \cap F = N_0(K \cap F)$, so $y \in N_0(K \cap F)$. Therefore, there exists an LS_0 protection matrix Y for y with respect to $K \cap F$. By definition, Y is also a protection matrix for y with respect to K. Furthermore, because Y is a protection matrix for y with respect to $K \cap F$, for each $i \in [n]$, $Ye_i \in K \cap F$. Of course, membership in F guarantees that for all $i \in [n]$ for all $j \in [n]$, if $y_j = 1$ then $(Ye_i)_j = (Ye_i)_0 = y_i$, and if $y_j = 0$, then $(Ye_i)_j = 0$. $\square$

3.2 Variable rank

Variable rank measures how many distinct variables must be lifted upon along some path in a derivation. More precisely: Let I be a set of linear inequalities over the variables $X_1, \ldots, X_n$, and let Γ be a tree-like LS_+ derivation from I. Label the edges of the tree by the literal that is being lifted on in that inference. Let π be a path from an axiom to the final inequality. The *variable rank of* π is the number of distinct variables that appear as lift-variables in the edges of π. The *variable rank* of Γ is the maximum variable rank of any path from an axiom to the final inequality in Γ. For any inequality $c^T X \geq d$, the *variable rank of* $c^T X \geq d$ *with respect to* I, *vrank*$^I(c^T X \geq d)$, is defined to be the minimal variable rank of any derivation of $c^T X \geq d$. If there is no such derivation, then the variable rank is defined to be ∞. The variable rank of I, $\text{vrank}(I)$, is defined to be $\text{vrank}(0 \geq 1)$. The variable rank of a vector $x \in [0,1]^n$ with respect to I, $\text{vrank}^I(x)$, is the minimum variable rank with respect to I of an inequality $c^T X \geq d$ such that $c^T x < d$.

It turns out that rank equals variable rank. This is what allows us to prove a tree-size/rank trade-off in Theorem 3.10 instead of tree-size/variable rank trade-off: The strategy for the proof of Theorem 3.10 is to apply restrictions that kill all paths of high variable rank, possibly leaving some high rank but low variable rank branches.

Theorem 3.7. *Let I be a set of inequalities, then for LS_0, LS and LS_+, for any x, $vrank^I(x) = rank^I(x)$.*

Proof. Let $x \in [0,1]^n$. Clearly $\text{vrank}^I(x) \leq \text{rank}^I(x)$. We will prove the other direction by induction on $\text{rank}^I(x)$. We will show that for any x, if x has rank r, then any elimination of x must have a path that lifts on at least r distinct variables from $E(x)$. (Recall that $E(x)$ are those indices/coordinates of x that take on nonintegral values.) For $r = 0$ the proof is trivial.

For the inductive step, let x be a vector such that $\text{rank}^I(x) \geq r+1$. By Lemma 3.6, there is a support extending protection matrix Y for $\binom{1}{x}$ with respect to $N_+^r(P_I)$. Let Γ be a minimum variable rank elimination of x that is frugal in the sense that x satisfies every inequality of Γ except for the final inequality. Let the final inference of Γ be:

$$d - c^T X \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{i,j}(b_i - a_i^T X)X_j + \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_{i,j}(b_i - a_i^T X)(1-X_j) + \sum_{j=1}^{n} \lambda_j (X_j^2 - X_j) + \sum_k (g_k + h_k^T X)^2$$

By Lemma 2.21, there exists $i \in [m]$ and $j \in [n]$ so that either $a_i^T X \geq b_i$ is the hypothesis of an X_j lifting and $a_i^T PV_{1,j}(Y) < b_i$, or $a_i^T X \geq b_i$ is the hypothesis of an $1 - X_j$ lifting and $a_i^T PV_{0,j}(Y) < b_i$.

Suppose that the lifting is on X_j (the case of $1 - X_j$ is exactly the same). We now want to argue that j is not in $\text{Supp}(x)$. Suppose $j \in \text{Supp}(x)$. Then $PV_{0,j}(Y) = PV_{1,j}(Y) = x$. But this implies that $a_i^T x < b_i$ so Γ is not frugal, as we could have removed this last inference. Thus, we can assume that j is not in $\text{Supp}(x)$. Now let $y = PV_{j,1}(Y)$. Because Y is a protection matrix for $\binom{1}{x}$ with respect to $N_+^r(K_I)$, $y = PV_{j,1}(Y) \in N_+^r(K_I)$. Therefore y has rank r and by the induction hypothesis, this implies that this derivation of $a_i^T X \geq b_i$ must have some long path that lifts on at least r variables from $E(y)$. Consider this long path plus the edge labelled X_j from $a_i^T X \geq b_i$ to $c^T X \geq d$. We want to show that this path lifts on $r+1$ distinct variables from $E(x)$. First, let S be the set of r distinct variables from $E(y)$ that label the long path in the derivation of $a_i^T X \geq b_i$. Because Y is support extending, by Lemma 3.5, these r variables are also in $E(x)$. Now consider the extra variable X_j labelling the edge from $a_i^T X \geq b_i$ to $c^T X \geq d$. We have argued above that j is in $E(x)$ but not in $E(y)$ and therefore X_j is distinct from S. Thus altogether we have $r+1$ distinct variables from $E(x)$ that are mentioned along this long path, completing the inductive step. $\square$

3.3 A trade-off for rank and tree-size

Before we prove the tree-size/rank trade-off, we need a few elementary lemmas.

Lemma 3.8. *(proof in Appendix) Let $\varepsilon > 0$ be given. From the inequality $X_i \geq \varepsilon$ there is a rank one LS_0 derivation of $X_i \geq 1$, and from the inequality $1 - X_i \geq \varepsilon$ there is a rank one LS_0 derivation of $-X_i \geq 0$.*

Lemma 3.9. *(proof in Appendix) For all systems of in equalities I, all positive integers r, and all $\varepsilon, \delta > 0$: If there is a rank $\leq r-1$ derivation from I of $X_i \geq \varepsilon$ and a rank $\leq r$ derivation from I of $1 - X_i \geq \delta$, then there is a rank $\leq r$ refutation of I. If there is a rank $\leq r-1$ derivation from I of $1 - X_i \geq \varepsilon$ and a rank $\leq r$ derivation from I of $X_i \geq \delta$, then there is a rank $\leq r$ refutation of I.*

Theorem 3.10. *For any set of inequalities I with no $0/1$ solution, in each of the systems LS_0, LS, and LS_+, $rank(I) \leq 3\sqrt{n \ln S_T(I)}$.*

The high-level strategy for the proof of Theorem 3.10 is very similar to that used by Clegg, Edmonds and Impagliazzo, showing a relationship between degree and size for the polynomial calculus [11], and that used by Ben-Sasson and Wigderson showing a size/width trade-off for resolution [8]. The primary difference is in how refutations of $I\restriction_{X=0}$ and $I\restriction_{X=1}$ are combined into a refutation of I. To convert a refutation of $I\restriction_{X=0}$ into a derivation of $X > 0$, rather than dragging along a side formula, as in [8], the proof of Theorem 3.10 uses Lemma 3.2.

Proof. (of Theorem 3.10) Let Γ be a minimum tree-size refutation of I, and let $S = |\Gamma|$. Set $d = \sqrt{2n \ln S_T(I)}$, and $a = (1 - d/2n)^{-1}$. Let F be the set of paths in Γ of *variable rank* at least d. Call such paths "long". We show by induction on n and b that if $|F| < a^b$ then $\text{rank}(I) \leq d + b$. Observe that the claim trivially holds when $d \geq n$, because every refutation that uses at most n variables has rank at most n, so we may assume that $d < n$. In the base case, $b = 0$ and there are no paths in Γ of variable rank more than d, and thus by Theorem 3.7, $\text{rank}(I) \leq d$. In the induction step, suppose that $|F| < a^b$. Because there are $2n$ literals making at least $d|F|$ appearances in the $|F|$ many long paths, there is a literal X (here X is X_i or $1 - X_i$ for some $i \in [n]$) that appears in at least $\frac{d}{2n}|F|$ of the long paths. Setting $X = 1$, $\Gamma\restriction_{X=1}$ is a refutation of $I\restriction_{X=1}$ with at most $\left(1 - \frac{d}{2n}\right)|F| < a^{b-1}$ many long paths. By the induction hypothesis, $\text{rank}(I\restriction_{X=1}) \leq d + b - 1$. By Lemma 3.2, there is $\varepsilon \geq 0$ and a derivation of $1 - X \geq \varepsilon$ from I of rank at most $d + b - 1$. On the other hand, $\Gamma\restriction_{X=0}$ is a refutation with at most $|F| < a^b$ many long paths and in $n-1$ many variables. By induction on the number of variables, $\text{rank}(I\restriction_{X=0}) \leq d + b$. By Lemma 3.2, there is $\delta \geq 0$ and a derivation of $X \geq \delta$ from I of rank at most $d + b$. Therefore by Lemma 3.9, $\text{rank}(I) \leq d + b$. This concludes the proof that if $|F| < a^b$, then $\text{rank}(I) \leq d + b$.

Because $|F| < |\Gamma| \leq a^{\log_a(S)}$, we have that $\text{rank}(I) \leq \log_a(S) + d$ so that:

$$
\begin{aligned}
\text{rank}(I) &\leq d + \log_a(S) = d + \log_{\left(\frac{2n}{2n-d}\right)}(S) \\
&= d + \log_{\left(1+\frac{d}{2n-d}\right)} S = d + (\ln S) \log_{\left(1+\frac{d}{2n-d}\right)}(e) \\
&= d + (\ln S)\left(\ln\left(1 + (d/(2n-d))\right)\right)^{-1}
\end{aligned}
$$

Because $0 \leq d < n$, we have that $0 \leq d/(2n-d) < 1$, so we may apply the bound $\ln(1+x) \geq x - x^2/2 \geq x/2$ with $x = d/(2n-d)$. Therefore:

$$
\begin{aligned}
\text{rank}(I) &\leq d + (\ln S)\left(d/2(2n-d)\right)^{-1} \\
&\leq d + (\ln S)(2 \cdot 2n/d) \\
&= 3\sqrt{2n \ln S}
\end{aligned}
$$

$\square$

Corollary 3.11. *For the LS_0, LS and LS_+ systems, we have that for any set of inequalities I in n variables with no $0/1$ solution, $S_T(I) \geq e^{(rank(I))^2/9n}$.*

3.4 Asymptotic tightness for LS and LS$_0$

Up to logarithmic factors, the trade-off for rank and tree-size is asymptotically tight for LS$_0$ and LS refutations. This follows from well-known bounds for the propositional pigeonhole principle: On the one hand, it is shown in [17] that LS refutations of PHP_n^{n+1} require LS rank $\Omega(n)$, but on the other hand, there are tree-like LS$_0$ refutations of PHP_n^{n+1} of size $n^{O(1)}$ (this seems to be a folklore result).

Theorem 3.12. *For each $n \in \mathbb{N}$, there is is a CNF F on $N = \Theta(n^2)$ many variables such that $rank(F) = \Omega\left(\sqrt{(N/\log N) \cdot \ln S_T(F)}\right)$.*

The propositional pigeonhole principle has a LS$_+$ refutation of rank one [17], so that example does not show the trade-off to be asymptotically tight for LS$_+$. Determining whether or not the trade-off is asymptotically tight for LS$_+$ is an interesting open question.

3.5 No trade-off for arbitrary derivations in LS$_0$ and LS, and the cost of deduction

Theorem 3.10 shows that for LS or LS$_+$ *refutations*, strong enough rank lower bounds automatically imply tree-size lower bounds. But what about derivations of arbitrary inequalities? Somewhat counter-intuitively, a similar trade-off does not apply for LS or LS$_0$ derivations of arbitrary inequalities, nor for the elimination of points from a polytope. It is an interesting open problem to determine whether or not such a tree-size/rank tradeoff for arbitrary derivations holds for LS$_+$.

A natural approach for transforming results abut refutations into results about derivations would be to use some form of deduction. *Deduction* is the logical principle that says: If there is a refutation of $\{\psi_1, \dots \psi_n\}$ in some logical system $\mathcal{F}$, then there is an $\mathcal{F}$ derivation of $\neg\psi_n$ from the hypotheses $\{\psi_1, \dots \psi_{n-1}\}$. Many systems of propositional logic enjoy an efficient version of the deduction theorem, in which passing from refutations to derivations does not increase the size (or some other parameter) very much. In the context of the Lovász-Schrijver systems, deduction means transforming a refutation of $\{a_i^T X \geq b_i \mid i \in [m]\}$ into a derivation of $a_m^T X \leq b_m - \varepsilon$ from the hypotheses $\{a_i^T X \geq b_i \mid i \in [m-1]\}$ for some $\varepsilon > 0$.

One hypothetical approach to obtain a tree-size/rank trade-off for arbitrary derivations would proceed as follows: If we know that deriving $a_m^T X < b_m$ from the hypotheses $\{a_i^T X \geq b_i \mid i \in [m-1]\}$ requires high rank, then “by deduction” refuting $\{a_i^T X \geq b_i \mid i \in [m]\}$ requires high rank and thus large tree-size, therefore deriving $a_m^T X < b_m$ from the hypotheses $\{a_i^T X \geq b_i \mid i \in [m-1]\}$ requires large tree-size. Unfortunately, the hypothetical use of the deduction theorem is fallacious: For LS$_0$ and LS systems, deduction can blow up the rank.

Theorem 3.13. *For sufficiently large n, there exists a system of inequalities I over the variables $\{X_1, \dots X_n\}$ and an inequality $a^T X \leq b$ such that:*

1. *Any LS derivation of $a^T X \leq b$ from I requires rank $\Omega(n)$.*

2. *For any* $\varepsilon > 0$, $I \cup \{a^T X \geq b + \varepsilon\}$ *has a rank one* LS_0 *refutation.*

3. *There is a tree-like* LS_0 *derivation of* $a^T X \leq b$ *from* I *of polynomial size.*

Proof. Let I be the following system of inequalities: For each $1 \leq i < j \leq n$, there is $X_i + X_j \leq 1$. Let $a^T X \leq b$ be the inequality $\sum_{i=1}^{n} X_i \leq 1$. We show that deriving $a^T X \leq b$ from I requires rank $\Omega(n)$. This is just a reduction from the well-known rank lower bound for LS refutations of PHP^n_{n-1} [17]. Let r be the minimum rank derivation of $\sum_{i=1}^{n} X_i \leq 1$ from I. In the n to $n-1$ pigeonhole principle, there are clauses $X_{i,j} + X_{i',j} \leq 1$ (for all $i, i' \in [n]$ with $i \neq i'$, and all $j \in [n-1]$), and $\sum_{j=1}^{n-1} X_{i,j} \geq 1$ (for all $i \in [n]$). In rank r we can derive $\sum_{i=1}^{n} X_{i,j} \leq 1$ for each $j \in [n-1]$. Summing up over all j gives $\sum_{j=1}^{n-1} \sum_{i=1}^{n} X_{i,j} \leq n-1$. On the other hand, there is a rank zero derivation of $\sum_{i=1}^{n} \sum_{j=1}^{n-1} X_{i,j} \geq n$ from the inequalities of PHP^n_{n-1}. Thus we have a rank r refutation of PHP^n_{n-1}. Because the LS rank of PHP^n_{n-1} is $\Omega(n)$, it follows that $r = \Omega(n)$.

Next we want to show that for any ε, the system $I \cup \{\sum_{i=1}^{n} X_i \geq 1 + \varepsilon\}$ has a rank one LS_0 refutation: By multiplying $X_i + X_j \leq 1$ by X_i and multilinearizing, we get $X_i + X_j X_i \leq X_i$, equivalently, $X_j X_i \leq 0$. Do this for all $i \neq j$, thus obtaining $X_j X_i \leq 0$ for all $i \neq j$. By multiplying $\sum_{j=1}^{n} X_j \geq (1+\varepsilon)$ by X_i and multilinearizing, we get $\sum_{j \neq i} X_j X_i \geq \varepsilon X_i$. However, adding this with the previously derived $X_j X_i \leq 0$ inequalities, and scaling, we get $0 \geq X_i$, for all $i = 1, \ldots n$. Thus we have $0 \geq \sum_{i=1}^{n} X_i \geq (1+\varepsilon)$, which yields $0 \geq 1$ after scaling. Finally, it is not hard to show by induction on k that there is a polynomial tree-size LS_0 derivation of $\sum_{i=1}^{k} X_i \leq 1$ from I. □

We do not yet know whether or not there is a "rank efficient deduction theorem" for LS_+. Theorem 3.13 does not apply because it relies upon a rank lower bound the propositional pigeonhole principle, and PHP^{n+1}_n has rank one LS_+ refutations [17]. Finally, known bounds for the pigeonhole principle show that for LS_0 and LS, there is no tree-size/rank trade-off for eliminations of points.

Theorem 3.14. *For sufficiently large* $n \in \mathbb{N}$, *there exists a set of inequalities* I_n *over* $X_1, \ldots, X_n$ *and a point* $x \in [0,1]^n$ *such that there is a polynomial size tree-like* LS_0 *derivation of* x *from* I_n, *but any LS elimination of* x *requires rank* $\Omega(n)$.

Proof. As in the proof of Theorem 3.13, let I be the following system of inequalities: For each $1 \leq i < j \leq n$, there is $x_i + x_j \leq 1$. By the argument of the proof of Theorem 3.13, all derivations of $\sum_{i=1}^{n} x_i \leq 1$ from I require rank $r_0 = \Omega(n)$. Therefore, by the affine Farkas Lemma, Lemma 2.1, for all $r < r_0$ there exists $z \in N^r(P_I)$ such that $\sum_{i=1}^{n} z_i > 1$. Let x be such a point belonging to $N^{(r_0-1)}(P_I)$. On the other hand, there is a tree-like LS_0 derivation of $\sum_{i=1}^{n} x_i \leq 1$ from I of size $n^{O(1)}$. Upon deriving $\sum_{i=1}^{n} x_i \leq 1$, the point x is eliminated. □

4 Tree-size bounds based on expanding constraints

The tree-size/rank trade-off of Theorem 3.10 and Corollary 3.11 allows us to quickly deduce tree-size bounds from previously known rank bounds for LS_+ refutations of prominent "sparse and expanding" unsatisfiable formulas. Specifically, we derive exponential tree size lower bounds for the Tseitin principles, random 3CNF formulas, and random mod 2 linear equations.

In this section, let F be a set of mod-2 equations over n variables. That is, each equation in F is of the form $\sum_{i \in S} X_i \equiv a \pmod 2$, where $S \subseteq [n]$ and $a \in \{0,1\}$. Notice that each such equation can be represented by the conjunction of $2^{|S|-1}$ clauses, each of which can be represented as a linear inequality. We denote by P_F the polytope bounded by these inequalities and by the inequalities $0 \le X_i \le 1$.

Let G_F be the bipartite graph from the set F to the set of variables where each equation is connected to the variables it contains.

Definition 4.1. *For $x \in \{0,1,1/2\}^n$, we say an equation $f \in F$ is fixed with respect to x if x sets all the variables set of f to 0/1 and f is satisfied by x. Let $G_F(x)$ be the subgraph of G_F induced by the set of variables $E(x)$ (those variables that are not integral valued) and the set of nonfixed equations.*

Definition 4.2. Random linear equations over $\mathbb{Z}_2$*: There are $2\binom{n}{k}$ linear, mod-2 equations over n variables that contain exactly k different variables; let $\mathcal{M}_m^{k,n}$ be the probability distribution induced by choosing m of these equations uniformly and independently.* Random k-CNFs*: There are $2^k\binom{n}{k}$ clauses over n variables that contain exactly k different variables; let $\mathcal{N}_m^{k,n}$ be the probability distribution induced by choosing m of these clauses uniformly and independently.*

Definition 4.3. *The Tseitin formula for an odd-sized graph $G=(V,E)$ has variables x_e for all edges $e \in E$. For each $v \in V$ there is one equation expressing that the sum of all edges incident with v is odd: $\sum_{e,v\in e} x_e = 1 \; mod \; 2$.*

The following theorem proven by [9] gives a rank lower bound for mod 2 equations as a function of the expansion.

Theorem 4.4. *[9] Let $\varepsilon > 0$ and let $w \in \frac{1}{2}\mathbb{Z}^n$. If $G_F(w)$ is an (r,c)-boundary expander, then it has LS_+ rank at least $r(c-2)$.*

The following results from [9] yield linear rank bounds for instances of Tseitin, 3-CNF, and 3-LIN formulas.

Fact 4.5. *For any constant δ, ε, k, there exists $\alpha > 0$ such that the following holds: Let $F \sim \mathcal{M}_{\Delta n}^{k,n}$. Then G_F is almost always an $(\alpha n, k-1-\varepsilon)$ boundary expander. Likewise for G_C where $C \sim \mathcal{N}_{\Delta n}^{k,n}$.*

Theorem 4.6. *[9]*

1. *The Tseitin tautology on a graph H has LS_+ rank at least $(c-2)n/2$ where c is the edge-expansion of H;*
2. *Let $k \ge 5$. There exists c such that for all constants $\Delta > c$, $F \sim \mathcal{M}_{\Delta n}^{k,n}$ requires LS_+ rank $\Omega(n)$ with high probability;*
3. *Let $k \ge 5$. There exists c such that for all constants $\Delta > c$, $C \sim \mathcal{C}_{\Delta n}^{k,n}$ requires LS_+ rank $\Omega(n)$ with high probability.*

As a consequence of Theorem 4.6 combined with Theorem 3.10, we get exponential tree-size bounds for these formulas.

Theorem 4.7.
1. *Let G be an odd-size graph on n nodes with edge-expansion c such that $c > 4$, and maximum degree Δ. All LS_+ refutations of $P_{TS(G)}$ require tree-size $2^{\Omega(n/\Delta)}$.*
2. *Let $k \ge 5$. There exists c such that for all constants $\Delta > c$, for $F \sim \mathcal{M}_{\Delta n}^{k,n}$, with probability $1-o(1)$, all LS_+ refutations of P_f require tree-size $2^{\Omega(n)}$.*

3. *Let $k \geq 5$. There exists c such that for all constants $\Delta > c$, for $C \sim \mathcal{N}_{\Delta n}^{k,n}$, with probability $1-o(1)$, all LS_+ refutations of P_C require tree-size $2^{\Omega(n)}$.*

The above proofs rely on the fact that for $k \geq 5$, the boundary expansion is greater than 2. In a subsequent paper, Alekhnovich, Arora and Tourlakis prove linear rank for random 3-CNFs [2].

Lemma 4.8. *[2] For a CNF ϕ, let C_ϕ be the bipartite graph between clauses and variables in which there is an edge between each clause and the variables that it contains. If C_ϕ is a $(\delta n, 2-\varepsilon)$ expander, then $(1, 1/2, \ldots 1/2) \in N_+^{\varepsilon\delta n/2}(SAT(\phi))$.*

By a well-known application of Markov's inequality, the probability that a random 3-CNFs with at least $5.2n$ clauses is unsatisfiable is $1-o(1)$ as $n \to \infty$. Furthermore, there exists a constant κ so that the probability that a random 3-CNF on Δn clauses is a $(\kappa n/\Delta^2, 4/3)$ exapnder is $1-o(1)$ as $n \to \infty$ (cf. [7], although a slightly different definition of expansion is used there). Thus we have:

Theorem 4.9. *There exists a constant $\beta > 0$ such that if ϕ is random Δn clause 3-CNF on n variables with $\Delta \geq 5.2$, then with probability $1-o(1)$ as $n \to \infty$, ϕ is unsatisfiable and all LS_+ refutations of ϕ require rank at least $\beta n/\Delta^2$.*

An immediate application of Corollary 3.11 extends this to:

Theorem 4.10. *There exist constants a constant $\gamma > 0$ such that if ϕ is random Δn clause 3-CNF on n variables, with $\Delta \geq 5.2$, then with probability $1-o(1)$ as $n \to \infty$, ϕ is unsatisfiable and all LS_+ refutations of ϕ require tree-size at least $2^{\gamma n/\Delta^2}$.*

5 Tree-size based integrality gaps

In this section, we will prove integrality gaps for small tree-like LS_+ derivations. Suppose we want to get an integrality gap of g for size s tree-like LS_+ derivations for some optimization problem P. Our goal will be the following. Given an arbitrary polytope P' obtained by a size s LS_+ tightening of the original polytope P, we want to exhibit a (nonintegral) point r such that: (i) r is in P'; and (ii) the value of objective function (what we are trying to maximize) on r is off from the optimal integral solution by a factor of g.

In this section, we establish tree-size based LS_+ integrality gaps for three combinatorial problems: Max-k-SAT, max-k-LIN, and vertex cover. As discussed in Subsection 3.5, we cannot always use Theorem 3.10 directly to obtain tree-size based integrality gaps. Nonetheless, we prove integrality gaps for sub-exponential tree-size LS and LS_+ relaxations by using variants of the method. For max-k-SAT and max-k-LIN, the method for establishing a rank-based integrality gap actually establishes a rank bound for refuting the system stating “all constraints are satisfied” and we will apply Theorem 3.10 in that manner. For vertex cover, on the other hand, we apply a random restriction to the derivation so that after applying the restriction, all high variable rank paths are killed, but, on the other hand, the restricted vertex cover instance still requires high variable rank to eliminate all points with a poor integrality gap.

5.1 Max-k-SAT and Max-k-LIN

The problem MAX-k-SAT (MAX-k-LIN) is the following: Given a set of k-clauses (mod-2 equations), determine the maximum number of clauses (equations) that can be satisfied simultaneously. It is known that it cannot be well-approximated in polynomial time if $P \neq NP$. Here we show inapproximation results (that are unconditional) for a restricted class of approximation algorithms that involve LS_+-relaxations of a linear program.

Given a set of k-mod-2 equations $F = \{f_1,\ldots,f_m\}$ over variables $X_1,\ldots,X_n$, add a new set of variables $Y_1,\ldots,Y_m$. For each f_i: $\sum_{j\in I_i} X_j \equiv a \pmod 2$, let f_i' be the equation $Y_i + \sum_{j\in I_i} X_j \equiv a+1 \pmod 2$. Let F' be the set of f_i''s. If Y_i is 1, then f_i' is satisfied if and only if f_i is satisfied. Hence we want to optimize the linear function $\sum_{i=1}^m Y_i$ subject to the constraints F'.

Call this linear program L_F. In the same way, we can obtain a maximization problem, L_C, corresponding to a set of k clauses C. analagous manner. An r-round LS_+ relaxation of L_F (or any linear program) is a linear program with the same optimization function but with any additional constraints that can be generated in depth r from the original constraints using LS_+. Similarly, a size s tree-like LS_+ relaxation of L_F (or any linear program) is a linear program with the same optimization function, but with s additional constraints that are derived from the original ones via a tree-like LS_+ proof.

Theorem 5.1. *Let $k \geq 5$. For any constant $\varepsilon > 0$, there are constants $\Delta,\beta > 0$ such that if $F \sim \mathcal{M}_{\Delta n}^{k,n}$ then the integrality gap of any size $s \leq 2^{\beta n}$ tree-like LS_+ relaxation of L_F is at least $2-\varepsilon$ with high probability. Similarly, for any $k \geq 5$ and any $\varepsilon > 0$, there exists $\Delta,\beta > 0$ such that if $C \sim \mathcal{N}_{\Delta n}^{k,n}$, then the integrality gap of any size $s \leq 2^{\beta n}$-round relaxation of L_C is at least $\frac{2^k}{2^k-1}$ with high probability.*

Proof. We will obtain size based integrality gaps via a reduction to the tree-size lower bounds proven in the previous section for 3-CNF and 3-LIN refutations.

We present the proof for L_F; an analgous argument works for L_C. Given $F \sim \mathcal{M}_{\Delta n}^{k,n}$, we want to show that there is no derivation of $\sum Y_i < m$ (where m is the number of mod 2 equations) via a polynomial-size tree derivation from the original equations F'. Consider a new constraint $g = \sum_{i=1}^m Y_i \geq m$. The set of constraints $F' \cup g$ is unsatisfiable with $F \sim \mathcal{M}_{\Delta n}^{k,n}$. In fact, for $\Delta \geq (8-4\varepsilon+\varepsilon^2)/\varepsilon^2$, a Chernoff bound and a union bound show that with high probability, no boolean assignment satisfies more than a $1/(2-\varepsilon)$ fraction of F''s equations.

First, we show that the unsatisfiable system of inequalities $F' \cup \{g\}$ requires large tree size refutations. We do this by applying the tree-size/rank trade-off of Theorem 3.10 For the rank bound, we will show that the the assignment z where all Y_i's are set to 1 and all X_i's are set to 1/2 survives for $\Omega(n)$ many rounds of LS_+ lift-and-project. This assignment clearly satisfies all inequalities in $F' \cup \{g\}$. Now, when we consider the equations restricted to the nonintegral values, it is just the original equations of F. With probability $1-o(1)$ over $F \sim \mathcal{M}_{\Delta n}^{k,n}$, the associated graph G_F is an $(\alpha n, 2+\delta)$-boundary expander for some $\alpha,\delta > 0$ that depend on Δ. Let $\beta = \alpha\delta$. Hence by Theorem 4.4, the $\mathrm{rank}^{F\cup\{g\}}(z) = \Omega(n)$, and therefore $\mathrm{rank}(F\cup\{g\}) = \Omega(n)$. By Theorem 3.10, we can conclude that the extended system $F' \cup g$ requires tree-size $2^{\Omega(n)}$ to refute in LS_+.

Now, we show that that the above superpolynomial tree-size needed to refute $F' \cup \{g\}$ implies the same tree-size lower bound for deriving $\sum_{i=1}^m Y_i \leq m-\varepsilon$ for all $\varepsilon > 0$: Suppose that we can derive $\sum_{i=1}^m y_i \leq m-\varepsilon$ from the original equations F' for some $\varepsilon > 0$ using tree-size S, can derive the empty polytope from $F' \cup g$ by summing $\sum_{i=1}^m y_i \leq m-\varepsilon$ with g, to yield $0 \geq \varepsilon$. Thus $S = 2^{\Omega(n)}$.

□

5.2 $LS+$ Integrality Gap for Vertex Cover

Given a 3XOR instance F over $\{X_1, \ldots X_n\}$ with $m = \Delta n$ equations, we define the FGLSS graph G_F as follows. G_F has $N = 4m$ vertices, one for each equation of F and for each assignment to the three variables that satisfies the equation. We think of each vertex as being labelled by a partial assignment to three variables. Two vertices u and v are connected if and only if the partial assignments that label u and v are inconsistent. The optimal integral solution for F is equal to the largest independent set in G_F. Note that $N/4$ is the largest possible independent set in G_F, where we choose exactly one node from each 4-clique.

The vertex cover and independent set problems on G_F is encoded in the usual way, with a variable $Y_{C,\eta}$ for each node (C,η) of G_F, where C corresponds to a 3XOR equation in F, and η is a satisfying assignment for C. Its polytopes is denoted $VC(G_F)$.

The following lemma was proven in [24].

Lemma 5.2. *Let F be a $(k, 1.95)$-expanding 3XOR instance such that any two equations of F share at most one variable, and let G_F be the corresponding FGLSS graph. The point $(3/4, \ldots, 3/4)$ is in the polytope generated after $\frac{k-4}{44}$ rounds of LS_+ lift-and-project applied to $VC(G_F)$.*

The following lemma, also proven in [24], shows that there are instances of 3XOR satisfying the hypotheses of Lemma 5.2.

Lemma 5.3. *For every $c < 2$, $\varepsilon > 0$, there exist $\alpha, \Delta > 0$ such that for every $n \in N$ there is a 3XOR instance F of mod 2 equations on n variables with $m = \Delta n$ equations such that: (i) No more than $(1/2+\varepsilon)m$ of equations of F are simultaneously satisfiable; (ii) Any two equations of F share at most one variable; and (iii) F is $(\alpha n, c)$-expanding.*

The above lemmas combine to give the following lower bound.

Theorem 5.4. *[24] For every $\varepsilon > 0$ there exists $c_\varepsilon > 0$ such that for infinitely many n, there exists a graph G with n vertices such that the ratio between the minimum vertex cover of size G and the optimum solution produced by any rank $c_\varepsilon n$ LS_+ tightening of $VC(G)$ is at least $7/6 - \varepsilon$.*

Proof. Let $\varepsilon > 0$ be given. Apply Lemma 5.3 and take $\alpha, \Delta > 0$, t sufficiently large (to demonstrate that the theorem holds for arbitrary large graphs), and a 3XOR instance F over $X_1, \ldots X_t$ with $m = \Delta t$ many equations so that G_F is $(\alpha t, 1.95)$ edge expanding, at most $(1/2+\varepsilon)m$ equations of F are simultaneously satisfiable, and no two equations of F share more than one variable.

Note that for any 3XOR instance F, a minimum size vertex cover of G_F consists of all nodes, less some independent set of maximum size, and an independent set in G_F that contains m_0 nodes corresponds to a an assignment that satisfies m_0 equations of F. Therefore, the minimum vertex cover size for G_F is $\geq 4m - m(1/2+\varepsilon)$. On the other hand, by Lemma 5.2, the all $3/4$ point remains after $\frac{\alpha t-4}{44}$ rounds of LS_+ lift-and-project from $VC(G_F)$. Thus, the integrality gap for $N_+^{\frac{\alpha t-4}{4}}(VC(G_F))$ is at least $\frac{4m - m(1/2+\varepsilon)}{(3/4)4m} = \frac{7}{6} - \frac{\varepsilon}{3} \geq \frac{7}{6} - \varepsilon$. The number of vertices in G_F is $4\Delta t$, so $c_\varepsilon \leq \frac{\alpha t - 4}{44(4\Delta t)}$ suffices for the Theorem statement. $\square$

We will improve Lemma 5.2 by proving a $7/6-\varepsilon$ integrality gap not only for small rank LS_+ tightenings of vertex cover but also for small tree LS_+ tightenings of vertex cover. The basic idea is to apply a random restriction $\rho = \rho_X \cup \rho_Y$, with ρ_X to the X variables of the 3XOR instance and ρ_Y to the Y variables of the independent set instance, so that: (i) The independent set constraints for G_F become the independent set constraints of $G_{F\restriction_{\rho_X}}$ after applying ρ_Y, ie. $VC(G_F)\restriction_{\rho_Y}=VC(G_{F\restriction_{\rho_X}})$. (ii) $F\restriction_{\rho_X}$ retains the expansion properties needed to apply Lemma 5.2. (iii) In an LS_+ derivation from $VC(G_F)$, any path that lifts on $\Omega(n)$ variables will have some lifting-literal falsified by ρ_Y with probability at least $1-2^{-\Omega(n)}$.

Regarding the issue of relating the ρ_X and ρ_Y assignments: Given a partial assignment ρ_X to the X's, we simply define ρ_Y via:

$$\rho_Y(Y_{C,\eta}) = \begin{cases} 1 & \text{if } \eta \text{ is a sub-assignment of } \rho_X \\ 0 & \text{if } \eta \text{ is inconsistent with } \rho_X \\ Y_{C,\eta} & \text{otherwise} \end{cases}$$

It is immediate upon inspection that for any ρ_X that does not falsify any equation of F, with ρ_Y defined as above, $VC(G_F)\restriction_{\rho_Y}=VC(G_{F\restriction_{\rho_X}})$ (up to renaming variables $Y_{C,\eta}$ in which ρ_X and η are consistent, but ρ_X sets at most two variables of C).

We now take an alternative view to point (iii), in which we replace the goal of "falsifying some literal of a long path" with the goal of satisfying a 3-DNF *in the X variables*. We construct the 3-DNF on a literal-by-literal basis: For a negative literal literal $1-Y_{C,\eta}$ let $\phi^-_{C,\eta}$ be the 3-DNF stating that "ρ_X satisfies η", that is, let x_i, x_j, x_k denote the variables of equation C, and set $\phi^-_{C,\eta}$ to be $x_i^{\eta(i)}\wedge x_j^{\eta(j)}\wedge x_k^{\eta(k)}$. For a positive literal $Y_{C,\eta}$, let $\phi^+_{C,\eta}$ be the 3-DNF stating "ρ_X satisfies C by satisfying some $\eta'\neq\eta$", that is, let x_i, x_j, x_k denote the variables of equation C, let β_1,β_2,β_3 the three assignments that satisfy C but are not η, and set $\phi^+_{C,\eta}$ to be $\bigvee_{l=1}^3 x_i^{\beta_l(x_i)}\wedge x_j^{\beta_l(x_j)}\wedge x_k^{\beta_l(x_k)}$. For a path π in an LS_+ derivation, let ϕ_π denote the 3-DNF obtained by taking the disjunction of $\phi^+_{C,\eta}$, for each $Y_{C,\eta}$ that is used positively in some lift of π, and of $\phi^-_{C,\eta}$ for each $Y_{C,\eta}$ that is used negatively in some lift of π. We clearly have that: If $\phi_\pi\restriction_{\rho_X}=1$ then ρ_Y falsifies some lift-literal of π.

We are now faced with the task of constructing a restriction to the X variables that will preserve the expansion properties of the 3XOR instance, but will satisfy the 3-DNF ϕ_π with overwhelming probability when π is a long a path. This was solved by Misha Alekhnovich in his analysis of $\text{Res}(k)$ refutations of random $3XOR$ instances [1]. We now revisit the definitions and results of [1], and show why they may be applied. The primary difference between our restriction and that of [1] is that we focus on the preservation of *edge expansion*, as opposed to *boundary expansion*. All that is needed about these closure operators is that they guarantee expansion after their application, and that the number of equations eliminated is bounded by a constant times the number of variables set. The correctness of the random restriction lemma of [1] does require that the initial system of equations have constant-rate boundary expansion. This applies in our use because by Fact 2.5, a (r,η) edge expander is an $(r,2\eta-d)$ boundary expander, and we apply the restriction lemma to an $(\alpha n,1.98)$ edge expander with 3 variables per equation.

Definition 5.5. *(after [3, 1]) Let $A\in\{0,1\}^{m\times n}$ be an (r,η) edge expander, let $\delta\in(0,1)$ be given, and let $J\subseteq[n]$ be given. Define the relation $\vdash^e_J$ on subsets of $[m]$ as:*

$$I_1\vdash^e_J I_2 \iff |I_2|\leq(r/2)\wedge\left|N_A(I_2)\setminus\left(\bigcup_{i\in I_1}A_i\cup J\right)\right|<\delta\cdot\eta|I_2| \tag{2}$$

Define the δ expansion closure of J, *$ecl^\delta_A(J)$, via the following iterative procedure: Initially let $I=\emptyset$. So long as there exists I_1 so that $I\vdash^e_J I_1$, let I_1 be the lexicographically first such set, replace I by $I\cup I_1$ and remove all rows in I_1 from the matrix A. Set $ecl^\delta_A(J)$ to be the value of I after this process stops. When the matrix A is clear from the context, we drop the subscript. Let* the δ-cleanup of A after removing J, $\text{CL}^\delta_J(A)$, *be the matrix that results by removing all rows of $ecl^\delta(J)$ and all columns of $J\cup\bigcup_{i\in ecl^\delta_A(J)}A_i$ from A.*

Lemma 5.6. *[3, 1] Let $A \in \{0,1\}^{m\times n}$, $\delta \in (0,1)$, and $J \subseteq [n]$ be given. If $CL_J^\delta(A)$ is non-empty, then $CL_J^\delta(A)$ is an $(r/2, \delta\cdot\eta)$ edge expander.*

Lemma 5.7. *(after [3, 1], proof of in Appendix) Let $A \in \{0,1\}^{m\times n}$ be an (r,η)-edge expander, let $\delta \in (0,1)$ be given, and let $J \subseteq [n]$ be given. If $|J| < \frac{r(1-\delta)\eta}{2}$, then $|ecl_A^\delta(J)| < \frac{|J|}{(1-\delta)\eta}$.*

Lemma 5.8. *[1] Let $A \in \{0,1\}^{m\times n}$ be an (r,η) edge expander, and let $J \subseteq [n]$ be given. For all $I_0 \subseteq [m]$, if $N_A(I_0) \subseteq J$ then $I_0 \subseteq ecl_A(J)$.*

Lemma 5.9. *(folklore, cf. [1]) Let $Ax = b$ be a system of equations so that A is an (r,β) boundary expander with $\beta > 0$. For every $I \subseteq [m]$ with $|I| \leq r$, $A_I x = b_I$ is satisfiable.*

Definition 5.10. *Fix $\delta, \gamma \in (0,1)$. Let $A \in \{0,1\}^{m\times n}$ be an (r,β)-boundary expander, and let $b \in \{0,1\}^m$ be given. Let $\mathcal{D}\ (A,r,\beta,\delta,\gamma)$ be the distribution on partial assignments to the variables $X_1, \ldots X_n$ generated by the following experiment: Uniformly select a subset $S_0 \subseteq \{X_1, \ldots X_n\}$ of size $\frac{r\beta(1-\delta)\gamma}{2}$. Let $I = ecl_A^\delta(S_0)$. Let $S = S_0 \cup \{X_j \mid \exists i \in I,\ A_{i,j} = 1\}$. The restriction ρ is a uniformly selected assignment to the variables of S that satisfies $A_I X = b_I$.*

In the above definition, take note that $|S_0| \leq \frac{r\beta(1-\delta)\gamma}{2} \leq \frac{r}{2}$, so that by Lemma 5.7, $|I| = |ecl_A^\delta(S_0)| < \frac{|S_0|}{\eta(1-\delta)} \leq \frac{r(1-\delta)\beta\gamma}{2\eta(1-\delta)} \leq \frac{r(1-\delta)\eta\gamma}{2\eta(1-\delta)} = \gamma r/2 < r/2$. Therefore, by Lemma 5.9, the system of equations $A_I X = B_I$ is satisfiable. Below is the random restriction lemma of [1]. We defer the definition of "normal form" until after the statement.

Definition 5.11. *Let F be a DNF, and let S be a set of variables. If every term of F contains a variable from S, then we say that S* is a cover of *F. The* covering number of *F, $c(F)$, is the minimum cardinality of a cover of F.*

Lemma 5.12. *[1] Let $A \in \{0,1\}^{m\times n}$ be an (r,β)-boundary expander such that each column of A contains at most d ones. Let $b \in \{0,1\}^m$ be arbitrary. There exists $a > 0$ (dependent upon only on β, γ and δ, and decreasing in β) such that for any k-DNF F so that F is in normal form:*

$$Pr_{\rho\in\mathcal{D}\ (A,r,\beta,\delta,\gamma)}[F\restriction_\rho \neq 1] < 2^{-c(F)/d^{ak}}$$

The notion of normal form used in [1] depends upon another definition of "closure".

Definition 5.13. *(after [4, 1]) Let $A \in \{0,1\}^{m\times n}$ and $J \subseteq [n]$ be given. Define the* closure of *J, $cl_A(J)$, via the following iterative procedure: Initially let $I = \emptyset$. So long as there exists I_1 so that $\partial_A(I_1) \subseteq J \cup I$, let I_1 be the lexicographically first such set, replace I by $I \cup I_1$ and remove all rows in I_1 from the matrix A. Set $cl_A(J)$ to be the value of I after this process stops. When the matrix A is clear from the context, and we drop the subscript. Let t be a term. We define $cl(t)$ to be $cl(Vars(t))$. We say that t is* locally consistent *if the formula $t \wedge [A_{cl(t)}X = b_{cl(t)}]$ is satisfiable. A DNF F is said to be in* normal form *if every term $t \in F$ is locally consistent.*

Lemma 5.14. *Let F be an instance of 3XOR, written as $AX = b$, where A is an (r,η) edge expander with $r \geq 2$ and $\eta > 1.5$. Let π a set of literals over the variables $\{Y_{C,\eta} \mid (C,\eta) \in V(G_F)\}$. The formula ϕ_π is in normal form.*

Proof. Let t be a term of ϕ_π. By definition, t is of the form $x_i^{\eta(x_i)} \wedge x_j^{\eta(x_j)} \wedge x_k^{\eta(x_k)}$ where C is an equation of F, whose variables are x_i, x_j, and x_k, and η is an assignment to these three variables satisfying C. By Definition 5.13, we clearly have that equation C belongs to $cl_A(t) = cl_A(vars(C))$. However, the closure process cannot proceed past the second step, because the edge expansion of A guarantees all other equations C' contain at least one variable not in $vars(C)$, so that $N(C') \not\subseteq vars(t) \cup vars(C) = vars(C)$. Therefore, $cl_A(t) = \{C\}$. Because η is assignment to $\{x_i, x_j, x_k\}$ that satisfies C, we have that $t = x_i^{\eta(x_i)} \wedge x_j^{\eta(x_j)} \wedge x_k^{\eta(x_k)}$ and the equation C can be simultaneously satisfied. $\square$

We now address how to bound the maximum number of equations in which each variable can occur.

Lemma 5.15. *(after [1]) Let $\varepsilon, \alpha, \Delta > 0$ and $n \in \mathbb{N}$ be given. Let F be a system of $m = \Delta n$ many 3XOR equations that satisfies: (i) No more than $(1/2 + \varepsilon)m$ of the equations of F are simultaneously satisfiable; (ii) No two equations of F share more than one variable; (iii) F is $(\alpha n, 1.99)$ edge-expanding.*

There is a 3XOR instance F' in the X variables satisfying: (i) No more than a $(1/2 + \varepsilon)$ fraction of the equations of F' are simultaneously satisfiable; (ii) No two equations of F' share more than one variable; (iii) F' is $(\alpha n/2, 1.98)$ edge-expanding; (iv) No variable appears in more than $\frac{3000\Delta}{\alpha}$ equations. (v) F' has at most Δn many equations.

Proof. Let A be equation/variable incidence matrix for F. Define J to be the set of $\frac{\alpha n}{1000}$ columns of the largest hamming weight in A, by Lemma 5.7 $|ecl_A^{\frac{199}{200}}(J)| < 200|J| \leq 200(.001r) \leq r/5 = \alpha n/5$. Therefore, $\mathrm{CL}_J^{\delta}(A)$ has at least $\Delta n - \alpha n/5$ many rows, and at least $n - 3\alpha n/5$ many columns. Furthermore, by Lemma 5.6, $\mathrm{CL}_J^{\delta}(A)$ is an $(\alpha n/2, \frac{199}{200} \cdot \frac{199}{1000})$ edge expander, which implies that it is an $(\alpha n/2, 1.98)$ edge expander.

By Lemma 5.9, we may choose an assignment ρ to the variables of $ecl_A^{\frac{199}{200}}(J)$ that satisfies every equation of $ecl_A^{\frac{199}{200}}(J)$. Let $F' = F \restriction_\rho$. F' is non-empty because F is unsatisfiable, and F' is not falsified because any falsified equation would belong to $ecl_A^{\frac{199}{200}}(J)$. The equation/variable incidence matrix of F' is a submatrix of $\mathrm{CL}_J^{\delta}(A)$, and as such is an $(\alpha n/2, 1.98)$ edge expander. Furthermore, as restriction of F, no two equations of F' share more than one variable, and at most a $(1/2 + \varepsilon)$ fraction of the equations of F' are simultaneously satisfiable.

Finally, every variable of F' can appear in at most 3000 equations of F'. If more than $\frac{\alpha n}{1000}$ of the variables occurred in more than $\frac{3000\Delta}{\alpha}$ equations, the total number of variable occurrences would exceed $\frac{3000\Delta}{\alpha} \cdot \frac{\alpha n}{1000} = 3\Delta n$, but this cannot happen since every equation one of the Δn equations contains three variables. $\square$

Lemma 5.16. *Let F be a 3XOR instance over the X variables such that every X variable appears in at most d equations of F. Let π be a set of literals in the Y variables, such that each literal is over a distinct variable. Then $c(\phi_\pi) \geq \frac{|\pi|}{4d}$.*

Proof. Each term of ϕ_π has the form $x_i^{\eta(x_i)} \wedge x_j^{\eta(x_j)} \wedge x_k^{\eta(x_k)}$ where some equation C of F is in the variable x_i, x_j, x_k and η is one of the four assignments to those three variables that satisfies C. Because each X variable can belong to at most d many equation, each X variable can belong to at most $4d$ terms of ϕ_π. Thus $c(\phi_\pi) \geq \frac{|\pi|}{4d}$. $\square$

Theorem 5.17. *For all $\varepsilon > 0$, there exists $\Delta, c > 0$ so that for sufficiently large n, there exists F, a system of at most Δn many 3XOR equations over $\{X_1,\dots X_n\}$, such that any tree-like LS_+ tightening $VC(G_F)$ with integrality gap $\le \frac{7}{6} - \varepsilon$ has size at least 2^{cn}.*

Proof. Choose $\varepsilon_0, \gamma > 0$ so that $\varepsilon_0 + \gamma/2 = 3\varepsilon$. Apply Lemma 5.3, and choose $\Delta, \alpha > 0$, and then, taking n sufficiently large to show that the claim holds for arbitrarily large instances, let F' be a system of Δn many 3XOR equations on n variables such that $G_{F'}$ is an $(\alpha n, 1.99)$ edge expander, no two equations of F' share more than one variable, and at most $\Delta n(1/2 + \varepsilon_0)$ equations of F' are simultaneously satisfiable.

Apply Lemma 5.15 to obtain F so that: (i) No more than a $(1/2 + \varepsilon_0)$ fraction of the equations of F are simultaneously satisfiable (ii) No two equations of F share more than one variable (iii) F is $(\alpha n/2, 1.98)$ edge-expanding (iv) No variable appears in more than $\frac{3000\Delta}{\alpha}$ equations. (v) The number of equations in F is at most Δn. Set $d = \frac{3000\Delta}{\alpha}$, set $\delta = \frac{195}{198}$, and let a be the parameter of Lemma 5.12 with $\delta = \frac{195}{198}$, γ as defined previously, and β equal to the boundary expansion of G_F (and thus $\beta \ge 0.96$).

For each ρ in the support of $\mathcal{D}(A, (\alpha/2)n, \beta, \delta, \gamma)$, as per Definition 5.10, let the point w^ρ be defined by:

$$w^{\rho}_{C,\eta} = \begin{cases} 1 & \text{if } \rho_Y(Y_{C,\eta}) = 1 \\ 0 & \text{if } \rho_Y(Y_{C,\eta}) = 0 \\ 3/4 & \text{otherwise} \end{cases}$$

For each ρ, if $\rho_Y(Y_{C,\eta}) = 1$ then $\rho(Y_{C,\eta'}) = 0$ for all $\eta \neq \eta'$, so $\sum_{(C,\eta)\in V(G_F)} w^{\rho}_{C,\eta} \le 3m$. On the other hand, each such ρ satisfies at most $\gamma(\alpha/2)n/2 \le \gamma m/2$ many equations of F, so the minimum size vertex cover in $G_{F\restriction_\rho}$ has size at least $\left(\frac{7}{2} - \varepsilon_0\right)m - \gamma m/2$. Therefore, the integrality gap of each w^ρ is at least $\frac{\left(\frac{7}{2}-\varepsilon_0\right)m - \gamma m/2}{3m} = \frac{\frac{7}{2} - \varepsilon_0 - \gamma/2}{3} = \frac{7}{6} - \varepsilon$.

Set $R = \frac{(\alpha/4)n-4}{44}$. Assume for sake of contradiction that there is a tree-like LS_+ tightening of $VC(G_F)$ with integrality at most $\frac{7}{6} - \varepsilon$ and tree-size at most $S = \sqrt{2^{R/4d^{3a+1}} - 1}$. Call this forest of derivations Γ. Choose a restriction ρ according to the distribution $\mathcal{D}(A, (\alpha/2)n, \beta, \delta, \gamma)$.

Let π be a path in the derivation Γ from a formula to one of its ancestors that contains at least R many distinct variables as lift variables. By Lemma 5.14, ϕ_π is in normal form, and by Lemma 5.16, $c(\phi_\pi) \ge \frac{R}{4d}$. Therefore, we may apply Lemma 5.12: $Pr_\rho[\phi_\pi \restriction_\rho \neq 1] < 2^{-R/4d^{3a+1}}$. There are at most $S^2 = 2^{R/4d^{3a+1}} - 1$ such paths in Γ, so by the union bound, there exists a ρ in the support of $\mathcal{D}(A, (\alpha/2)n, \beta, \delta, \gamma)$, so that ρ_Y falsifies a literal on every path of Γ of variable rank $\ge R$.

Because the integrality gap of w^ρ is at least $7/6 - \varepsilon$ and the tightening Γ has integrality gap at most $7/6 - \varepsilon$, we may choose an inequality $c^T X \ge d$ that is derived in Γ such that that $c^T w^\rho < d$. Because every path in Γ of variable rank at least R has one of its lifting literals falsified, there is a variable rank $< R$ derivation of $\left(c^T Y \ge d\right) \restriction_{\rho_Y}$ from $VC(G_F) \restriction_{\rho_Y} = VC(G_{F\restriction_\rho})$. Because $c^T w^\rho < d$ and w^ρ agrees with ρ_Y on the variables set by ρ_Y, w^ρ also falsifies $\left(c^T Y \ge d\right) \restriction_{\rho_Y}$. So the variable rank needed to eliminate w^ρ from $VC(G_F) \restriction_{\rho_Y}$ is $< R = \frac{(\alpha/4)n-4}{44}$. Thus by Theorem 3.7, w^ρ can be eliminated from $VC(G_F) \restriction_{\rho_Y}$ with rank $< \frac{(\alpha/4)n-4}{44}$. Let u be the all 3/4's vector indexed by the variables of $VC(G_F) \restriction_{\rho_Y}$. Because $VC(G_F) \restriction_{\rho_Y} = VC(G_{F\restriction_\rho})$, the elimination of w^ρ from $VC(G_F) \restriction_{\rho_Y}$ with rank $< \frac{(\alpha/4)n-4}{44}$ can be transformed into a elimination of u from $VC(G_{F\restriction_\rho})$ with rank $< \frac{(\alpha/4)n-4}{44}$. However, by Lemma 5.6, $F \restriction_\rho$ is an $(\alpha n/4, 1.95)$ expander. Furthermore, any two of its equations share at most one variable. So by Lemma 5.2, u requires rank at least $\frac{(\alpha/4)n-4}{44}$ to eliminate from $VC(G_{F\restriction_\rho})$. Contradiction.

We have shown that any tree-like LS_+ tightening of $VC(G_F)$ with integrality at most $\frac{7}{6} - \varepsilon$ has tree-size $> S = \sqrt{2^{R/4d^{3a+1}} - 1} = \sqrt{2^{\left(\frac{(\alpha/4)n-4}{44}\right)/4d^{3a+1}} - 1} = 2^{\Omega(n)}$.

$\square$

6 Separations between proof systems

In this section, we compare the tree-like LS_+ proof system for proving CNFs unsatisfiable with other methods for proving CNFs unsatisfiable- the method of Gomory-Chvatal cuts, and resolution. We show that tree-like LS_+ refutations can require an exponential increase in size to simulate these systems.

6.1 Tree LS_+ cannot p-simulate tree GC cutting planes

Another method of solving zero-one programs by adding new inequalities to the linear program is the *Gomory-Chvatal cutting planes* (GC) method.

Definition 6.1. *Let a_i be a real vector of dimension n and let x be a vector of n boolean variables. The rules of GC cutting planes are as follows: (1) (Linear combinations) From $a_1^T x - b_1 \geq 0, \ldots, a_n^T x - b_n \geq 0$, derive $\sum_{i=1}^k (\lambda_i a_i^T x - \lambda_i b_i) \geq 0$, where λ_i are positive rational constants; (2) (Rounding) From $a^T x - \lambda \geq 0$ derive $a^T x - \lceil \lambda \rceil \geq 0$, provided that the coordinates of a are integers. Without loss of generality, we can assume that a rounding operation is always applied after every application of rule (1), and thus we can merge (1) and (2) into a single rule, called a Chvatal-Gomory (GC) cut. A GC cutting planes refutation for a system of inequalities, $f = f_1, \ldots f_m$, is a sequence of linear inequalities $g_1, \ldots, g_q$, such that each g_i is either an inequality from f, or an axiom ($x \geq 0$ or $1 - x \geq 0$), or follows from previous inequalities by a GC cut, and the final inequality g_q is $0 \geq 1$. The size of a refutation is the sum of the sizes of all g_i, where the coefficients are written in binary notation.*

In this subsection, we show that tree-like LS_+ cannot p-simulate tree-like GC cutting planes. This is done by establishing a tree-size lower bound for LS_+ refutations of certain counting modulo two principles. The counting principles that we use are a more complicated version of the ordinary count two principle stating that there can be no partition of a universe of size $2n+1$ into pieces of size exactly two, defined below.

Definition 6.2. *For each $n \in \mathbb{N}$, $Count_2^{2n+1}$ is the CNF consisting of the following clauses over the variables $\{x_e \mid e \in \binom{[2n+1]}{2}\}$: For each $v \in [2n+1]$, $\bigvee_{e \ni v} x_e$. For each $e, f \in \binom{[2n+1]}{2}$ with $e \cap f \neq \emptyset$, $\neg x_e \vee \neg x_f$.*

Unfortunately, the rank bounds for the $Count_2^{2n+1}$ principles are of the form $\Omega(n)$, but the number of variables is $\Theta(n^2)$, so we cannot directly apply the tree-size rank trade-off to $Count_2^{2n+1}$ to obtain superpolynomial tree-size lower bounds. Instead we will consider a more complicated version of the count two principle, that we will call $T_G - Count$, and our plan is as follows. We will begin with the well-known Tseitin principle on a sparse graph G; it is good for us because it is similar in proof complexity to the mod 2 counting principle, but it has only linearly many variables.

Linear rank bounds for LS_+ can be proven for the Tseitin principle on a sparse expander graph by observing that this principle has linear degree bounds in the stronger static positivestellensatz proof system, which imply linear rank bounds for LS_+. We then use a reduction from Tseitin to the count two principle from [10], which shows that from a low degree static positivestellensatz refutation of $T_G - Count$, we can obtain a low degree static positivestellensatz refutation of the Tseitin principle. Thus it follows that $T_G - Count$ requires linear rank in LS_+. Now using our rank-treesize tradeoff for LS_+, it follows that $T_G - Count$ requires exponential-size tree-like LS_+ proofs. Finally, it is not hard to show that $T_G - Count$ has polynomial-size tree-like GC cutting planes proofs, thus establishing that tree-like LS_+ cannot polynomially simulate GC cutting planes. We formalize this argument below.

Definition 6.3. *Let $\{f_1, \ldots, f_m\}$ be a system of polynomials over $\mathbb{R}$. A static positivestellensatz refutation of $\{f_1, \ldots, f_m\}$ is a set of polynomials $\{g_1, \ldots, g_m\}$ and $\{h_1, \ldots, h_l\}$ such that $\sum_{i=1}^{m} f_i g_i = 1 + \sum_{i=1}^{l} h_i^2$. The degree of the refutation is the maximum degree of any $f_i g_i$ or h_i^2.*

Definition 6.4. *The Tseitin principle on a graph $G = (V, E)$ is specified as follows. The underlying variables are x_e for all $e \in E$. For each vertex v there is a corresponding constraint that specifies that the mod 2 sum of all variables x_e, where e ranges over all edges incident with v, is 1. We will specify the constraints by a set of inequalities if we are interested in LS_+ proofs, or by a set of polynomial equations if we are interested in static positivestellensatz proofs. (In either case, each constraint is specified with $2^{O(d)}$ inequalities or polynomial equations, where d is the degree of the graph.)*

Theorem 6.5. *[21] For all n sufficiently large, there is a 6-regular graph, G_n, on $2n+1$ vertices such that any static positivestellensatz refutation of the Tseitin principle on G_n requires degree $\Omega(n)$.*

There is a natural reduction from the the Tseitin principle to the count two principle [10]: Start with an instance of the Tseitin principle on a d-regular graph $G = (V, E)$ with $2n+1$ vertices. Let the underlying variables of the Tseitin principle be x_e for all edges $e \in E$. The associated count two principle will be defined on a universe U as follows. The underlying elements of U will consist of one element corresponding to each vertex i in V, and two elements corresponding to each edge $e = (i, j)$ in E. We will denote the element corresponding to vertex i by (i) and the elements corresponding to the edge $e = (i, j)$ by $(i, j, 1)$ and $(i, j, 2)$.

The idea behind the reduction is as follows. Suppose that there is an assignment to the Tseitin variables so as to satisfy all of the underlying mod 2 equations. Then we will define an associated matching on U. Consider a node i in G and the r labelled edges $(i, j_1), (i, j_2), \ldots, (i, j_r)$ leading out of i, where $j_1 < j_2 < \ldots < j_r$. Suppose that the values of these edges are $a_1, a_2, \ldots, a_r$, $a_i \in \{0, 1\}$. Then for each l, $1 \le l \le r$, we take the first a_l elements in U from $(i, j_l, *)$ and group them with the first $(2 - a_l)$ elements in U from $(j_l, i, *)$. This gives us r 2-partitions so far. Note that the number of remaining, ungrouped elements associated with node i is $(2 - a_1) + (2 - a_2) + \ldots + (2 - a_r) + 1$, which is congruent to 0 mod 2 since $(a_1 + \ldots + a_r) mod 2 = 1$. We then group these remaining, ungrouped elements associated with i, two at a time, in accordance with the following ordering. Ungrouped elements from $(i, j_1, *)$ are first, followed by ungrouped elements from $(i, j_2, *)$ and so on, and lastly the element (i). It should be intuitively clear that if we started with an assignment satisfying all of the mod 2 Tseitin constraints, then the associated matching described above will be a partition of U into groups of size 2.

Given a graph G, the formula $T_G - Count$ denotes the mod 2 counting principle defined over the universe U as given by the reduction just described. When G has degree d, the degree of the polynomial equations expressing $T_G - Count$ will be d, and the number of variables is at most $2dn + dn + n\binom{d}{2}$. (See [10] for a formal description of $T_G - Count$.) [10] prove the following theorem, which shows that the above reduction can be formalized with low degree static positivestellensatz refutations. This is not too surprising since the reduction itself, as well as the underlying reasoning behind the correctness of the reduction, is all local.

Theorem 6.6. *[10] Let G be a graph of degree d. If there is no degree max(dr,d) static positivestellensatz refutation of the Tseitin principle, then there is no degree r static positivestellensatz refuatation of $T_G - Count$.*

The theorem below shows that degree lower bounds for static positivestellensatz refutations implies rank lower bounds for LS_+.

Theorem 6.7. *[21] Let G be a degree d graph. If there is no degree $2r+3d$ static positivestellensatz refutation of $T_G - Count$, then there is no rank r LS_+ refutation of $T_G - Count$.*

From Theorems 6.5, 6.7, 6.6 we see that rank of $TG - Count$ is $\Omega(n)$, and because $TG - Count$ has $O(n)$ many variables, we may apply Theorem 3.10 to conclude:

Corollary 6.8. *For all n sufficiently large, there is a graph G_n on $2n+1$ vertices and degree 6 such that any tree-like LS_+ refutation of $T_G - Count$ requires size $2^{\Omega(n)}$.*

On the other hand, it is not hard to show that $T_G - Count$ has GC cutting planes refutations of polynomial size.

Lemma 6.9. *Let G_n be a family of graphs on $2n+1$ vertices, with constant degree d. Then $T_G - Count$ has polynomial-size tree-like GC cutting planes refutations.*

Proof. There is a standard cutting planes derivation of $\sum_{e \ni v} x_e \leq 1$ using the inequalities $x_e + x_f \leq 1$. It has rank $\Theta(n)$ and tree-size polynomial in n. Summing over all of these gives $\sum_{e \in \binom{[2n+1]}{2}} 2x_e = \sum_{v \in [2n+1]} \sum_{e \ni v} x_e \leq 2n+1$. Apply a single GC cut to this and we have $sum_{v \in [2n+1]} \sum_{e \ni v} x_e \leq 2n$. On the other hand, summing over all of the inequalities $\sum_{e \ni v} x_e \geq 1$ yields $\sum_{v \in [2n+1]} \sum_{e \ni v} x_e \geq 2n+1$. ☐

Theorem 6.10. *Tree-like LS_+ does not polynomially simulate GC cutting planes.*

6.2 Tree LS_+ cannot p-simulate DAG-like resolution

It is known that unrestricted (DAG-like) LS_0 p-simulates resolution, but that simulation constructs Lovász-Schrijver derivations that are are also DAG-like. In this section we show that this is necessary: Tree-like LS_+ cannot p-simulate DAG-like resolution. The family of CNFs that we show to be hard for tree-like LS_+ is the "GT_n principle". It is one of the canonical examples for showing that a system cannot p-simulate DAG-like resolution, and it says that in any total order on a finite set, there exists a minimal element.

Definition 6.11. *For $n \geq 1$, the CNF GT_n is a CNF on the variables $X_{i,j}$, for $i, j \in n$, $i \neq j$. The clauses of GT_n include:*

1. *For each $1 \leq i < j \leq n$, $X_{i,j} \vee X_{j,i}$.*
2. *For each $1 \leq i < j \leq n$, $\neg X_{i,j} \vee \neg X_{j,i}$.*
3. *For each i, j, k, $\neg X_{i,j} \vee \neg X_{j,k} \vee X_{i,k}$*
4. *For each i, $\bigvee_{j \neq i} X_{j,i}$.*

Let $E = \{(i,j) \in [n]^2 \mid i \neq j\}$, so we can think of the variables as $X_{u,v}$ indexed by $(u,v) \in E$. The CNF GT_n is translated into a system of linear inequalities in the usual manner.

It was shown by Buresh-Oppenheim et al that LS_0 refutions of GT_n have rank $\Omega(n)$ [9]. Our tree-size lower bound is modeled after the basic ingredients of their argument.

6.3 Protection matrices for GT_n

The first thing we do is strengthen the rank bound of [9] to apply to LS_+, not just LS_0. As in that work, the rank bound is based upon protecting vectors that correspond to so-called *scaled partial orders*.

Definition 6.12. *A partial order $\prec$ on $[n]$ is said to be t-scaled if there is a partition of $[n]$ into sets $A_1, \dots A_t$ such that $\prec$ is a total ordering within each A_i, but elements from different A_i's are incomparable. For each $u \in A_i$, we say that A_i* is the class of u with respect to $\prec$. *We say that $\prec$ is* at least t-scaled *if $\prec$ is t'-scaled for some $t' \geq t$, and that $\prec$ is* at most t-scaled *if $\prec$ is t'-scaled for some $t' \leq t$.*

We say that (i,j) and *(l,k)* are equivalent with respect to $\prec$, written $(i,j) \equiv (l,k)$, *if $i \prec j$ and $l \prec k$, or if $j \prec i$ and $k \prec l$, or if there exist r,s such that $r \neq s$, $i,l \in A_r$ and $j,k \in A_s$. We say that (i,j)* and *(l,k)* are opposing with respect to $\prec$, written $(i,j) \perp (l,k)$, *if $i \prec j$ and $k \prec l$, or if $j \prec i$ and $l \prec k$, or if there exist r,s such that $r \neq s$, $i,l \in A_r$ and $j,k \in A_s$.*

For a partial order $\prec$, let $x^{\prec} \in \mathbb{R}^E$ be defined by:

$$x^{\prec}_{(i,j)} = \begin{cases} 1 & \text{if } i \prec j \\ 0 & \text{if } j \prec i \\ \frac{1}{2} & \text{if } i \text{ and } j \text{ are incomparable with respect to } \prec \end{cases}$$

For $i,j \in [n]$ such that i and j incomparable with respect to $\prec$, let $\prec^{(i,j)}$ denote the scaled partial order that refines $\prec$ by placing every element from the class of i before every element of the class of j. If $i \prec j$, then $\prec^{(i,j)} = \prec$, and if $j \prec i$, then $\prec^{(i,j)} = \prec_R$, where $\prec_R$ denotes the reversal of $\prec$.

Here is an easy fact about assignments from scaled partial orders:

Lemma 6.13. *Let $\prec$ be a scaled partial order on $[n]$. For all $(i,j) \equiv (l,k)$, $x^{\prec}_{(i,j)} = x^{\prec}_{(l,k)}$. For all $(i,j) \perp (l,k)$, $x^{\prec}_{(i,j)} = 1 - x^{\prec}_{(l,k)}$.*

Here are some easy facts about scaled partial orders:

Definition 6.14. *Let P_s denote least polytope containing $\{x^{\prec} \mid \prec$ is at least s-scaled $\}$.*

Lemma 6.15. *(cf. [9]) When $s \geq 3$, $P_s \subseteq P_{GT_n}$.*

Definition 6.16. *Let $\prec$ be a scaled partial order on $[n]$. Define the matrix $Y^{\prec} \in \mathbb{R}^{\{0\} \cup E \times \{0\} \cup E}$ as follows: $Y_{0,0} = 1$, and for all $(i,j) \in E$, $Y_{(i,j),0} = Y_{0,(i,j)} = x_{(i,j)}$. For $(i,j),(l,k) \in E$:*

$$Y^{\prec}_{(i,j),(l,k)} = \begin{cases} x^{\prec}_{(i,j)} & \text{if } (i,j) \equiv (l,k) \\ 0 & \text{if } (i,j) \perp (l,k) \\ x^{\prec}_{(i,j)} x^{\prec}_{(l,k)} & \text{otherwise} \end{cases}$$

The following two lemmas are proved in the Appendix.

Lemma 6.17. *Let $\prec$ be a scaled partial order, let $x = x^{\prec}$, and let $Y = Y^{\prec}$. For each $(i, j) \in E$, if $0 < x_{(i,j)} < 1$ then $PV_{(i,j),1}(Y) = x^{\prec^{(i,j)}}$ and $PV_{(i,j),0}(Y) = x^{\prec^{(j,i)}}$, otherwise $PV_{(i,j),0}(Y) = PV_{(i,j),1}(Y) = x$.*

Lemma 6.18. *For all at least $(s+1)$-scaled partial orders $\prec$, the matrix $Y^{\prec}$ is an LS_+ protection matrix for $x^{\prec}$ with respect to P_s.*

Lemma 6.19. *Let $s \in \mathbb{N}+3$ be given. For every $n \geq s$, if $\prec$ is an at least s-scaled partial order on $[n]$, then $rank^{GT_n}(x^{\prec}) \geq s-3$.*

Proof. We show by induction on $s \in \mathbb{N}+3$ that $P_s \subseteq N_+^{s-3}(P_{GT_n})$. For $s = 3$, this is a consequence of Lemma 6.15, which tells us $P_3 \subseteq P_{GT_n}$. Assume that the claim holds for s. Let $n \geq s+1$ be given, and let $\prec$ be an at least $(s+1)$-scaled partial order. Consider the matrix $Y^{\prec}$: By Lemma 6.18, this is a protection matrix for $x^{\prec}$ with respect to P_s. However, by the induction hypothesis, $P_s \subseteq N_+^{s-3}(P_{GT_n})$, so $Y^{\prec}$ is also a protection matrix for $x^{\prec}$ with respect to $N_+^{s-3}(P_{GT_n})$. Therefore, $x^{\prec} \in N_+^{s-2}(P_{GT_n})$. Because $\prec$ was an arbitary at least $(s+1)$-scaled partial order, $P_{s+1} \subseteq N_+^{s-2}(P_{GT_n})$. $\square$

Corollary 6.20. *For all $n \geq 3$, the LS_+ rank of GT_n is at least $n-3$.*

Because there are $n^2 - n$ variables in GT_n and the rank bound is only $n-3$, the lower bound obtained from the tree-size/rank trade-off is a trivial constant bound. The tree-size bound for LS_+ refutations of GT_n requires more work than that, but the machinery developed to prove Corollary 6.20 is used.

6.4 A measure of rank that corresponds to scaled partial orders

An obvious approach to proving a tree-size lower bound for LS_+ refutations of GT_n would be to apply a random restriction to the refutation and eliminate all paths of high variable rank. A natural choice for such a restriction is to randomly choose $S \subseteq [n]$ of size $n/2$ and place a random total order on those elements, thus creating an $(n/2+1)$-scaled partial order $\prec$. The restricted refutation of GT_n eliminates $x^{\prec}$, yet we would hope that the restriction kills all paths of high variable rank. It turns out that this is not the case. Suppose that the lift-variables of a path are $X_{1,2}, X_{1,3}, X_{1,4}, \ldots$: This path will not be killed unless 1 is placed into the set S, and that happens with probability exactly one $1/2$.

The idea behind the random restriction approach can be salvaged: It suffices to kill the scaled partial order generated by a path. The path of the example actually generates the scaled partial order $1, 2, 3, 4 \ldots$, and this can be killed by simply placing some $j \prec i$ where $i < j$, and this happens with overwhelming probability. A notationally cumbersome issue that arises is that we are now dealing with the scaled partial order generated by a path, which depends not just the set of literals lifted upon, but on the order in which the literals are lifted upon.

Definition 6.21. *Let n be given. All refutations and inequalities in what follows are over the variables of GT_n.*

Let Γ be an LS_+ derivation of $c^T X \geq d$. Let $\prec$ be a scaled partial order on $[n]$. Let π be a path in Γ from an inequality to one of its ancestors (the ancestor is not necessarily a hypothesis of the derivation).

The partial order of π extending $\prec$, $\prec^{\pi}$, *is either a scaled partial order on $[n]$, or a special null value corresponding to "inconsistency". It is defined recursively as follows: If π has length 0 (eg. π begins and* *ends at the same inequality), then $\prec^{\pi} = \prec$. Otherwise, let $X_{u,v}$ (or $1 - X_{v,u}$) be the lifting variable for the inference of the first step in π, and let π_0 be the remainder of π. If $v \prec u$, then we say that π* and *$\prec$* are inconsistent. *Otherwise, $\prec^{\pi} = (\prec^{(u,v)})^{\pi_0}$.*

We make a simple observation that follows by induction:

Lemma 6.22. *Let Γ be an LS_+ derivation of $c^T X \geq d$. Let $\prec$ be a scaled partial order on $[n]$. Let π be a path in Γ from an inequality to one of its ancestors. If $\prec$ and π are consistent, then $\prec^{\pi}$ refines $\prec$.*

Definition 6.23. *Let $\prec$ be a scaled partial order on $[n]$. For any single-step LS_+ derivation: A lift on $X_{u,v}$ or $1 - X_{v,u}$ is said to* have cost 0 with respect to $\prec$ *if $u \prec v$, a lift on $X_{u,v}$ or $1 - X_{v,u}$ is said to be* inconsistent with respect to $\prec$ *if $v \prec u$, otherwise, a lift on $X_{u,v}$ or $1 - X_{v,u}$ is said to* have cost 1 with respect to $\prec$.

Let π be a path in Γ from an inequality to one of its ancestors such that π is consistent with $\prec$. The cost of π with respect to $\prec$, *$cost_{\prec}(\pi)$, is defined recursively as follows: If π has length 0, then $cost_{\prec}(\pi) = 0$. Otherwise, let l be the lifting literal for the inference of the first step in π, chose $u, v \in [n]$ so that $l = X_{u,v}$ or $l = 1 - X_{v,u}$, and let π_0 be the remainder of π. $cost_{\prec}(\pi) = cost_{\prec}(l) + cost_{\prec^{(u,v)}}(\pi_0)$.*

The following lemma is the analog of a rank lower bound, and shows in particular that any derivation of GT_n requires a path of high cost.

Lemma 6.24. *Let $n \in \mathbb{N}$ be given, and let $\prec$ be an s-scaled partial order on $[n]$. Let Γ be an elimination of $x^{\prec}$ from GT_n. Let t be such that every branch of Γ either is inconsistent with $\prec$, or has cost at most t with respect to $\prec$. We have that $s - t \leq 2$.*

Proof. We induct on the size of Γ. The induction hypothesis is: "For every Γ of size at most S, for all $s, t \in \mathbb{N}$, if Γ that is an elimination of an $x^{\prec}$ from GT_n where $\prec$ is an s-scaled partial order and every branch of Γ either is inconsistent with $\prec$, or has cost at most t with respect to $\prec$, then there exists $\prec^*$ which refines $\prec$, such that $\prec^*$ is at least $s - t$ scaled and $x^{\prec^*} \notin P_{GT_n}$." Lemma 6.24 then follows from Lemma 6.15, because that guarantees that $\prec^*$ is at most 2-scaled and thus $s - t \leq 2$.

For the base case, $|\Gamma| = 1$, so Γ consists of a single inequality $a^T X \geq b$ from GT_n such that $a^T x^{\prec} < b$. It immediately follows that $x^{\prec} \notin P_{GT_n}$, moreover, because $\prec$ is s-scaled, for all $t \geq 0$, $\prec$ is at least $(s - t)$-scaled.

Let $S \in \mathbb{N}$ be given and assume that the lemma holds for all eliminations of size at most S. Let $s \in \mathbb{N}$ be given, and let $\prec$ be an s-scaled partial order on $[n]$. Let Γ be an elimination of $x = x^{\prec}$ from GT_n such that the size of Γ is $S + 1$, and let t be an upper bound on the cost of every branch in Γ with respect to $\prec$. Let $d^T X \geq c$ be the final inequality of Γ, and consider its derivation:

$$c - d^T X \quad = \quad \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{i,j} (b_i - a_i^T X) X_j + \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{i,j} (b_i - a_i^T X)(1 - X_j) + \sum_{j=1}^{n} \lambda_j (X_j^2 - X_j) + \sum_k (g_k + h_k^T X)^2$$

with each $\alpha_{i,j}, \beta_{i,j} \geq 0$.

Let $Y = Y^{\prec}$, as per Definition 6.16. By Lemma 2.21, there exists an $i \in [m]$ and a $(u,v) \in E$ such that:

1. $a_i^T X \geq b_i$ is used as the hypothesis for a lifting inference on $X_{(u,v)}$ and $a_i^T PV_{(u,v),1}(Y) < b_i$ and $x_{u,v} \neq 0$.
2. $a_i^T X \geq b_i$ is used as the hypothesis for a lifting inference on $1 - X_{(u,v)}$ and $a_i^T PV_{(u,v),0}(Y) < b_i$ and $x_{u,v} \neq 1$.

Suppose that Case 1 holds; the analysis under Case 2 is essentially the same. Let Γ^* be the sub-derivation of $a_i^T X \geq b_i$. The size of Γ^* is at most S, so the induction hypothesis applies to Γ^*.

If $x_{u,v} = 1$, then $PV_{(u,v),1}(Y) = x$, so that Γ^* is an elimination of $x = x^{\prec}$. Notice that in this situation we have that $u \prec v$, so that $\prec^{(u,v)} = \prec$. Every path in Γ^* from $a_i^T X \geq b_i$ to one of its ancestors that is consistent with respect to $\prec$ is the suffix of a path in Γ from $d^T X \geq c$ to one of its ancestors that is consistent with $\prec$, and therefore has cost at most t with respect to $\prec$. Therefore, by the induction hypothesis, there is $\prec^*$ refining $\prec$ such that $\prec^*$ is at least $s-t$ scaled and $x^{\prec^*} \notin P_{GT_n}$.

Now consider the case when $x_{u,v} \neq 1$. Because Case 1 guarantees that $x_{u,v} \neq 0$, we have that $x_{u,v} = 1/2$, so that u and v are incomparable with respect to $\prec$. Set $y = PV_{(u,v),1}(Y) = x^{\prec^{(u,v)}}$. Note that $\prec^{(u,v)}$ is $s-1$ scaled and that it refines $\prec$. Furthermore, u and v are in different components of $\prec$, so that the lift upon $X_{u,v}$ has cost one with respect to $\prec$. Every path in Γ^* from $a_i^T X \geq b_i$ to one of its ancestors that is consistent with respect $\prec^{(u,v)}$ is the suffix of a path in Γ from $d^T X \geq c$ to one of its ancestors that is consistent with $\prec$, so every path in Γ^* that is consistent with respect to $\prec^{(u,v)}$ has cost at most $t-1$ with respect to $\prec^{(u,v)}$. Therefore, by the induction hypothesis, there is $\prec^*$ refining $\prec^{(u,v)}$ such that $\prec^*$ is at least $(s-1)-(t-1) = s-t$ scaled and $x^{\prec^*} \notin P_{GT_n}$. By the transitivity of refinement, $\prec^*$ also refines $\prec$. ☐

The following lemma is the random restriction lemma. It shows that for any subexponential-sized proof Γ, there exists a restriction that is not too large and such that all relevant paths in Γ under the restriction have low cost.

Lemma 6.25. *There exists $c > 0$ so that for all $n \geq 6$, if Γ is a refutation of GT_n and the size of Γ is at most $\frac{1}{4}2^{cn}$, then there exists a partial order $\prec$ on $[n]$ that is at least $n/4$ scaled, and such that all paths in Γ that are consistent with respect to $\prec$ have cost at most $n/4-3$ with respect to $\prec$.*

Proof. We generate $\prec$ at random as follows: Randomly generate $V \subseteq [n]$ by placing $i \in [n]$ into V with with independent probability $1/2$. Select a total order for the elements of V uniformly at random. All $i \in [n] \setminus V$ are incomparable with the elements of V and with each other.

We reckon the cost of paths with respect to "the degenerate partial order" $\prec_D$, that satisfies for all $x, y \in [n]$, $x \not\prec_D y$. This suffices to prove the lemma, because the cost of π with respect to $\prec$ can only exceed the cost of π with respect to the degenerate partial order.

Let π be a path in Γ such that the cost of π with respect to the degenerate partial order exceeds $n/2-3$. Let $A_1, \ldots A_t$ be the classes of $\prec_\pi$, and note that $t \leq n/2+3$. Let $a_i = |A_i|$. List out the elements of A_i according to $\prec_\pi$, $u_{i,1}, \ldots u_{i,a_i}$. For each $j = 1, \ldots \lfloor a_i/2 \rfloor$, the probability that $\prec$ places $a_{i,2j}$ before $a_{i,2j-1}$ is clearly $\frac{1}{8}$. For distinct j's, these events are independent. Therefore the probability that for all $j = 1, \ldots \lfloor a_i/2 \rfloor$, that $\prec$ and $\prec_\pi$ do not disagree on the relative order of $a_{i,2j-1}$ and $a_{i,2j}$ is at most $(7/8)^{\lfloor a_i/2 \rfloor}$. Because the sets $A_1, \ldots A_t$ are disjoint, the probability that for all $i = 1, \ldots t$, $\prec$ and $\prec_\pi$ do not disagree on the relative order of any $a_{i,2j-1}$ and $a_{i,2j}$ with $j \in \{1, \ldots \lfloor a_i/2 \rfloor\}$ is at most $\prod_{i=1}^t (7/8)^{\lfloor a_i/2 \rfloor}$.

Let n_2 be the number of $u \in [n]$ such that u appears in a class A_i of $\prec_\pi$ with $|A_i| = 2$. Let $n_{\geq 3}$ be the number of $u \in [n]$ such that u appears in a class A_i of $\prec_\pi$ with $|A_i| \geq 3$. We immediately have that $\prod_{i=1}^t \left(\frac{7}{8}\right)^{\lfloor a_i/2 \rfloor} \leq \left(\frac{7}{8}\right)^{(1/2)n_2 + (2/3)n_{\geq 3}}$

At most $t-1$ elements of $[n]$ can appear in singleton classes, and therefore at least $n/2-3$ items appear in classes of size two or more. Thus, $n_2 + n_{\geq 3} \geq n/2-3$. It follows that: $\left(\frac{7}{8}\right)^{(1/2)n_2+(2/3)n_{\geq 3}} \leq \left(\frac{7}{8}\right)^{(1/2)(n/2-3)}$.

Because the event that $\prec_\pi$ and $\prec$ are consistent implies that for all $i = 1, \dots t$, $\prec$ and $\prec_\pi$ do not disagree on the relative order of any $a_{i,2j-1}$ and $a_{i,2j}$ with $j \in \{1, \dots \lfloor a_i/2 \rfloor\}$, the probability that π is consistent with respect to $\prec$ is at most $\left(\frac{7}{8}\right)^{(1/2)(n/2-3)}$. Choose $c > 0$ so that $\left(\frac{7}{8}\right)^{(1/2)(n/2-3)} < 2^{-cn}$ for all $n \geq 6$.

Let Γ be a refutation of GT_n such that the size of Γ is at most $\frac{1}{4}2^{cn}$. Choose $\prec$ by the distribution described above. By the union bound, the probability that there exists a path π in Γ that has cost $\geq (n/4) - 3$ with respect to the degenerate partial order and is also consistent with respect to $\prec$ is at most $1/4$. Because the expected size of $|V|$ is $n/2$, the probability that $|V| \geq (3/4)n$ is at most $2/3$ by Markov's inequality. Therefore, there exists $\prec$ which is at least $n/4$ scaled such that for all π in Γ, if the cost of π with respect to the empty partial order $\geq (n/4) - 3$, then π is inconsistent with respect to $\prec$.

$\square$

Theorem 6.26. *There exists $c > 0$ so that for all $n \in \mathbb{N}$, every tree-like LS_+ refutation of GT_n has size at least 2^{cn}.*

Proof. Suppose for the sake of contradiction that there is an LS_+ refutation of GT_n of size $< 2^{cn}$. By Lemma 6.25, there is partial order $\prec$ on $[n]$ such that $\prec$ is at least $n/4$ scaled, and all paths in Γ that are consistent with $\prec$ have cost at most $n/4 - 3$ with respect to Γ. However, by Lemma 6.24, we must have that $3 = (n/4) - ((n/4) - 3) \leq 2$, which is false. $\square$

It is well-known that the GT_n principle possesses unrestricted resolution refutations of size $O(n^3)$. Thefore we have as a corollary to Theorem 6.26:

Theorem 6.27. *Tree-like LS_+ refutations cannot p-simulate DAG-like resolution.*

Because DAG-like LS_+ can *p*-simulate DAG-like resolution, we have:

Corollary 6.28. *Tree-like LS_+ refutations cannot p-simulate DAG-like LS_+ refutations.*

7 Discussion

Our results bound the size of the derivation tree needed for LS_+ tightening of linear relaxations to obtain strong integrality gaps or to refute an unsatisfiable CNF. Another way to measure the size of an LS_+ derivation is to arrange the formulas as directed acyclic graph. Derivations in this model are called "DAG-like" or simply "unrestricted". The most urgent, burning question left open by this paper is to prove size lower bounds for LS_+ derivations in the *DAG-like* model.

At present, only one bound on DAG-like refutation size is known for LS_0 [14], and no non-trivial bounds are known for any DAG-like LS or LS_+ derivations. Moreover, no bounds are known on the DAG-sizes necessary to obtain good integrality gaps for any natural optimization problem (such as vertex cover or max-*k*-SAT) using any of the Lovász-Schrijver operators.

A natural question to ask is whether or not the techniques of this paper can be extended to the DAG-like model: Is it possible to acheive a general size/rank tradeoff for DAG-like LS? In particular, can we prove that small DAG-like LS proofs imply small rank? We suspect that the answer is negative.

An interesting loose-end to address is whether or not the tree-size/rank tradeoff for LS_+ holds for derivations as well as refutations. A positive answer would simplify the task of proving tree-size based integrality gaps for LS_+. However, we suspect that the answer is negative and that one simply needs to find the right counterexamples. It would also be nice to resolve the issue of whether or not deduction requires an increase in the rank for the LS_+ system, and to determine if Theorem 3.10 is asymptotically tight for LS_+ refutations.

There are some integrality gaps known for low-rank LS_+ and LS tightenings for which we have not yet obtained tree-size based integrality gaps, for example, set cover [2] and max-cut [25]. We suspect that rank-based integrality gaps such as these can be used to obtain tree-size-based integrality gaps in these cases as well.

Finally, there is the question of whether or not a tree-size/rank trade-off holds for other zero-one programming derivation systems, such as the Sherali-Adams system or Lassier proofs. This seems likely and interesting, but stronger (ie. super-logarithmic) rank bounds for those systems are needed before such a trade-off would be of any use.

References

[1] M. Alekhnovich. Lower bounds for *k*-DNF resolution on random 3-CNFs. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 251–256, 2005.

[2] M. Alekhnovich, S. Arora, and I. Tourlakis. Towards strong nonapproximability results in the lovasz-schrijver hierarchy. In *STOC*, pages 294–303, 2005.

[3] M. Alekhnovich, E. Hirsch, and D. Itsykson. Exponential lower bounds for the running times of DPLL algorithms on satisfiable formulas. *Journal of Automated Reasoning*, 35(1-3):51–72, 2005.

[4] M. Alekhnovich and A. Razborov. Lower bounds for the polynomial calculus: Non-binomial case. In *Proceedings of the Forty-Second Annual IEEE Symposium on Foundations of Computer Science*, pages 190–199, 2001.

[5] S. Arora, B. Bollobas, L. Lovasz, and I. Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006.

[6] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitioning. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, pages 222–231, 2004.

[7] E. Ben-Sasson. *Expansion in Proof Complexity*. PhD thesis, Department of Computer Science, Hebrew University of Jerusalem, 2001.

[8] E. Ben Sasson and A. Wigderson. Short proofs are narrow - resolution made simple. *Journal of the ACM*, 48(2), 2001.

[9] J. Buresh-Oppenheim, N. Galesi, S. Hoory, A. Magen, and T. Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory of Computing*, 2:65–90, 2006.

[10] S. Buss, D. Grigoriev, R. Impagliazzo, and T. Pitassi. Linear gaps between degrees for the polynomial calculus modulo distinct primes. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 547–556, Atlanta, GA, May 1999.

[11] M. Clegg, J. Edmonds, and R. Impagliazzo. Using the Gröbner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pages 174–183, Philadelphia, PA, May 1996.

[12] Stephen A. Cook and Robert A. Reckhow. Time bounded random access machines. *Journal of Computer and System Sciences*, 7(4):354–375, 1973.

[13] S. Dash. *On the matrix cuts of Lovász and Schrijver and their use in Integer Programming*. PhD thesis, Department of Computer Science, Rice University, March 2001.

[14] S. Dash. An exponential lower bound on the length of some classes of branch-and-cut proofs. In *IPCO*, 2002.

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

[16] D. Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259:613–622, 2001.

[17] D. Grigoriev, E. Hirsch, and D. Pasechnik. Complexity of semialgebraic proofs. *Moscow Mathematical Journal*, 2:647–679, 2002.

[18] D. Grigoriev and E. Hirsh. Algebraic proof systems over formulas. *Theoretical Computer Science*, 303:83–102, 2003.

[19] D. Grigoriev and N. Vorobjov. Complexity of null- and positivestellensatz proofs. *Annals of Pure and Applied Logic*, 113:153–160, 2001.

[20] A. Haken. The intractability of resolution. *Theoretical Computer Science*, 39:297–305, 1985.

[21] A. Kojevnikov and D. Itsykson. Lower bounds of statis lovasz-schrijver calculus proofs for tseitin tautologies. In *ICALP*, pages 323–334, 2006.

[22] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[23] L. Lovasz and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optimization*, 1(2):166–190, 1991.

[24] G. Schoenebeck, L. Trevisan, and M. Tulsiani. A linear round lower bound for lovasz-schrijver sdp relaxations of vertex cover. In *ECCC*, number 98, 2006.

[25] G. Schoenebeck, L. Trevisan, and M. Tulsiani. Tight integrality gaps for lovasz-schrijver lp relaxations of vertex cover and max cut. In *ECCC*, 2006.

[26] I. Tourlakis. *New lower bounds for approximation algorithms in the Lovász-Schrijver hierarchy*. PhD thesis.

[27] I. Tourlakis. New lower bounds for vertex cover in the lovasz-schrijver hierarchy. In *IEEE Conference on Computational Complexity*, 2006.

A Proof of Lemma 2.21

Proof. (of Lemma 2.21) Express all inequalities in homogenized form: Each $a_i^T X \geq b_i$ becomes $\binom{-b_i}{a_i}^T \binom{1}{X}$, with $u_i = \binom{-b_i}{a_i}$, and $c^T X \geq d$ becomes $h^T \binom{1}{X} \geq 0$ with $h = \binom{-d}{c}$.

Because the coefficients of the non-linear monomials all cancel, there is a skew-symmetric matrix $A \in \mathbb{R}^{(n+1)\times(n+1)}$ and a positive semidefinite matrix $B \in \mathbb{R}^{(n+1)\times(n+1)}$ so that:

$$he_0^T = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{i,j} u_i e_j^T + \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_{i,j} u_i (e_0 - e_j)^T + \sum_{j=1}^{n} \lambda_j e_j (e_0 - e_j)^T + A + B$$

Take the entry-wise product of this matrix with Y we have that $he_0^T \bullet Y = c^T x - d < 0$. Therefore:

$$\begin{aligned}
0 > he_0^T \bullet Y &= \sum_{i,j} \alpha_{i,j} u_i e_j^T \bullet Y + \sum_{i,j} \beta_{i,j} u_i (e_0 - e_j)^T \bullet Y + \sum_j \lambda_j e_j (e_0 - e_j)^T \bullet Y + A \bullet Y + B \bullet Y \\
&\geq \sum_{i,j} \alpha_{i,j} u_i^T Y_j + \sum_{i,j} \beta_{i,j} \left(u_i^T Y_0 - u_i^T Y_j\right) + \sum_j \lambda_j (Y_{0,j} - Y_{j,j}) + 0 + 0 \\
&= \sum_{i,j} \alpha_{i,j} u_i^T Y_j + \sum_{i,j} \beta_{i,j} u_i^T \left(Y_0 - Y_j\right)
\end{aligned}$$

Therefore, there exists some $i \in [m]$ and $j \in [n]$ so that $\alpha_{i,j} u_i^T Y_j + \beta_{i,j} u_i^T (Y_0 - Y_j) < 0$.

In the case that $x_j = 0$, by Definition 2.15, $Y_j = 0$ and $Y_0 - Y_j = \binom{1}{x}$. $0 > \alpha_{i,j} u_i^T Y_j + \beta_{i,j} u_i^T (Y_0 - Y_j) = \beta_{i,j} u_i^T \binom{1}{x}$. Therefore, $\beta_{i,j} > 0$ (so there is some lift upon $1 - X_j$) and $0 > -b_i + a_i^T x = -b_i + a_i^T (PV_{j,0}(Y))$.

In the case that $x_j = 1$, by Definition 2.15, $Y_j = \binom{1}{x}$ and $Y_0 - Y_j = 0$. $0 > \alpha_{i,j} u_i^T Y_j + \beta_{i,j} u_i^T (Y_0 - Y_j) = \alpha_{i,j} u_i^T \binom{1}{x}$. Therefore, $\alpha_{i,j} > 0$ (so there is some lift upon X_j) and $0 > -b_i + a_i^T x = -b_i + a_i^T (PV_{j,1}(Y))$.

Now consider the case with $0 < x_j < 1$. By Definition 2.15, we may choose $y \in \mathbb{R}^n$ so that $Y_j = \binom{x_j}{y}$. Substituting $\binom{x_j}{y}$ for Y_j yields $\alpha_{i,j}(-b_i x_j + a^T y) + \beta_{i,j}(-b_i(1 - x_j) + a_i^T (x - y)) < 0$. If $0 > \alpha_{i,j}(-b_i x_j + a_i^T y)$, then $\alpha_{i,j} > 0$ (so $-b_i + a_i^T X \geq 0$ is used as the hypothesis for some lift on X_j), and also $0 > -b_i + a_i^T (y/x_j) = -b_i + a_i^T (PV_{j,1}(Y))$. Similarly, if $0 > \beta_{i,j}(-b_i(1 - x_j) + a_i^T (x - y))$, then $\beta_{i,j} > 0$ (so $-b_i + a_i^T X \geq 0$ is used as the hypothesis for some lift on $(1 - X_j)$), and $0 > -b_i + a_i^T ((x - y)/(1 - x_j)) = -b_i + a_i^T (PV_{j,0}(Y))$.

□

B Lemmas for the tree-size/rank trade-off

Proof. (of Lemma 3.8) From the hypothesis $X_i \geq \varepsilon$, we may infer $(1 - X_i)X_i \geq \varepsilon(1 - X_i)$, multilinearize by adding a multiple of $X_i^2 - X_i = 0$ and we have $0 \geq \varepsilon(1 - X_i)$. Multiply through by $1/\varepsilon$ and we have $X_i \geq 1$. Clearly this derivation has LS_0 rank one.

From the hypothesis $(1-X_i) \geq \varepsilon$, we may infer $X_i(1-X_i) \geq \varepsilon X_i$, multilinearize by adding a multiple of $X_i^2 - X_i = 0$ and we have $0 \geq \varepsilon X_i$. Multiply through by $1/\varepsilon$ and we have $-X_i \geq 0$. Clearly this derivation has LS_0 rank one. $\square$

Proof. (of Lemma 3.9) The two cases are nearly identical, for brevity we do the first case only. By hypothesis, there is a rank $\leq r-1$ derivation of $X_i \geq \varepsilon$; combine this with Lemma 3.8, and we have a rank $\leq r$ derivation of $X_i \geq 1$ from I. By hypothesis, there is a rank $\leq r$ derivation of $1 - X_i \geq \delta$. Adding these two formulas we have $1 \geq 1+\delta$, which yields $0 \geq 1$ after multiplying by the positive scalar $1/\delta$. $\square$

C Edge expansion closure calculation

Proof. (of Lemma 5.7) Suppose for the sake of contradiction that $|ecl^{\delta}(J)| \geq \frac{1}{(1-\delta)\eta}|J|$. Let $I_1, \ldots I_t$ be the sequence of subsets of $[m]$ that are taken in cleaning procedure, with each $|I_i| \leq r/2$.

First we inductively show that for each $s \leq t$, $|N_A(\bigcup_{i=1}^{s} I_i) \setminus J| \leq \delta \cdot \eta |\bigcup_{i=1}^{s} I_i|$. For the base case, Equation 2 yields $|N_A(I_1) \setminus J| \leq \delta \cdot \eta |I_1|$. For the induction step, assume that $|N_A(\bigcup_{i=1}^{s} I_i) \setminus J| \leq \delta \cdot \eta |\bigcup_{i=1}^{s} I_i|$ for an arbitrary $s < t$. By Equation 2, $|N_A(I_{s+1}) \setminus \left(J \cup \bigcup_{i \in \bigcup_{i=1}^{s} I_i} A_i\right)| \leq \delta \cdot \eta |I_{s+1}|$. Because rows added to $ecl^{\delta}(J)$ are removed from the matrix after each stage of cleaning, the sets $I_1, \ldots I_t$ are pairwise disjoint, thus:

$$\begin{aligned} \left|N_A\left(\bigcup_{i=1}^{s+1} I_i\right) \setminus J\right| &\leq \left|N_A\left(\bigcup_{i=1}^{s} I_i\right) \setminus J\right| + \left|N_A(I_{s+1}) \setminus \left(J \cup \bigcup_{i \in \bigcup_{i=1}^{s} I_i} A_i\right)\right| \\ &\leq \delta \cdot \eta \left|\bigcup_{i=1}^{s} I_i\right| + \delta \cdot \eta \, |I_{s+1}| = \delta \cdot \eta \left|\bigcup_{i=1}^{s+1} I_i\right| \end{aligned}$$

Now, let i_0 be the first index with $|\bigcup_{i=1}^{i_0} I_i| > \frac{1}{(1-\delta)\eta}|J|$. Note that $|\bigcup_{i=1}^{i_0} I_i| \leq |\bigcup_{i=1}^{i_0-1} I_i| + |I_{i_0}| \leq \frac{1}{(1-\delta)\eta}|J| + r/2 \leq \frac{1}{(1-\delta)\eta}\frac{r(1-\delta)\eta}{2} + r/2 = r$. Therefore by edge expansion, $|N_A\left(\bigcup_{i=1}^{i_0} I_i\right)| > \eta |\bigcup_{i=1}^{i_0} I_i|$. Therefore: $|N_A\left(\bigcup_{i=1}^{i_0} I_i\right) \setminus J| \geq \eta |\bigcup_{i=1}^{i_0} I_i| - |J| > \eta |\bigcup_{i=1}^{i_0} I_i| - \eta(1-\delta)|\bigcup_{i=1}^{i_0} I_i| = \delta \cdot \eta |\bigcup_{i=1}^{i_0} I_i|$. This contradicts the previously established fact that $|N_A\left(\bigcup_{i=1}^{i_0} I_i\right) \setminus J| \leq \delta \cdot \eta |\bigcup_{i=1}^{i_0} I_i|$.

$\square$

D Protection matrices for GT_n

Proof. (of Lemma 6.17) The cases for $x_{(i,j)} \in \{0,1\}$ follow from the definition of protection vectors, so consider (i,j) with $x_{(i,j)} = 1/2$.

By definition:

$$(PV_{(i,j),1}(Y))_{(l,k)} = Y_{(l,k),(i,j)}/x^{\prec}_{(i,j)} = \begin{cases} x^{\prec}_{(i,j)}/x^{\prec}_{(i,j)} = 1 = x^{\prec(i,j)}_{(l,k)} & \text{if } (i,j) \equiv (l,k) \\ 0/x^{\prec}_{(i,j)} = 0 = x^{\prec(i,j)}_{(l,k)} & \text{if } (i,j) \perp (l,k) \\ x^{\prec}_{(l,k)}x^{\prec}_{(i,j)}/x^{\prec}_{(i,j)} = x^{\prec}_{l,k} = x^{\prec(i,j)}_{(l,k)} & \text{otherwise} \end{cases}$$

$$(PV_{(i,j),0}(Y))_{(l,k)} = \frac{Y_{(l,k),0} - Y_{(l,k),(i,j)}}{1 - x^{\prec}_{(i,j)}} = \frac{x^{\prec}_{(l,k)} - Y_{(l,k),(i,j)}}{1 - x^{\prec}_{(i,j)}} = \begin{cases} \frac{x^{\prec}_{(l,k)} - x^{\prec}_{(l,k)}}{1 - x^{\prec}_{(i,j)}} = 0 = x^{\prec^{(j,i)}}_{(l,k)} & \text{if } (i,j) \equiv (l,k) \\ \frac{x^{\prec}_{(l,k)} - 0}{1 - x^{\prec}_{(i,j)}} = \frac{1/2}{1/2} = 1 = x^{\prec^{(j,i)}}_{(l,k)} & \text{if } (i,j) \perp (l,k) \\ \frac{x^{\prec}_{l,k} - x^{\prec}_{(i,j)} x^{\prec}_{(l,k)}}{1 - x^{\prec}_{(i,j)}} = x^{\prec}_{(l,k)} = x^{\prec^{(j,i)}}_{(l,k)} & \text{otherwise} \end{cases}$$

$\square$

Proof. (of Lemma 6.18) Let $Y = Y^{\prec}$. Let $y = \binom{1}{x^{\prec}}$.

We just check that the properties of Definition 2.15 hold:

1. That $x^{\prec} \in P_s$: By hypothesis, $\prec$ is $(s+1)$-scaled, so $x^{\prec} \in P_s$.
2. $Ye_0 = diag(Y) = \binom{1}{x^{\prec}}$. By definition, $Y_{0,0} = 1$, $Y_{0,(i,j)} = y_0 y_{(i,j)} = 1 \cdot x^{\prec}_{(i,j)} = x^{\prec}_{(i,j)}$, and $Y_{(i,j),(i,j)} = x^{\prec}_{(i,j)}$.
3. For all $(i,j) \in E$, if $x^{\prec}_{(i,j)} = 1$, then $Ye_{(i,j)} = \binom{1}{x^{\prec}}$.

   By definition, $(Ye_{(i,j)})_0 = x^{\prec}_{(i,j)} = 1$. For $(l,k) \in E(x^{\prec})$, we have:

   $$(Ye_{(i,j)})_{(l,k)} = Y_{(l,k),(i,j)} = \begin{cases} x^{\prec}_{(l,k)} = x^{\prec}_{(i,j)} & \text{if } (i,j) \equiv (l,k) \\ 0 = x^{\prec}_{(l,k)} & \text{if } (i,j) \perp (l,k) \\ x^{\prec}_{(l,k)} x^{\prec}_{(i,j)} = x^{\prec}_{(l,k)} \cdot 1 = x^{\prec}_{l,k} & \text{otherwise} \end{cases}$$

4. For all $(i,j) \in E$, if $x^{\prec}_{(i,j)} = 0$, $Ye_{(i,j)} = 0$. By definition, $(Ye_{(i,j)})_0 = x^{\prec}_{(i,j)} = 0$. For $(l,k) \in E$, we have:

   $$(Ye_{(i,j)})_{(l,k)} = Y_{(l,k),(i,j)} = \begin{cases} x^{\prec}_{(l,k)} = x^{\prec}_{(i,j)} = 0 & \text{if } (i,j) \equiv (l,k) \\ 0 & \text{if } (i,j) \perp (l,k) \\ x^{\prec}_{(l,k)} x^{\prec}_{(i,j)} = x^{\prec}_{(l,k)} \cdot 0 = 0 & \text{otherwise} \end{cases}$$

5. That $PV_{(i,j),0}(Y), PV_{(i,j),1}(Y) \in P_s$ for all othe $(i,j) \in E$. This follows immediately from Lemma 6.17, and the fact that both $\prec^{(i,j)}$ and $\prec^{(j,i)}$ are s-scaled.
6. The matrix Y is positive semidefinite.

   Let $y = \binom{1}{x^{\prec}}$. We define a disjoint family of subsets of E as follows: For each $r, s \in [t]$ with $r \neq s$, there is a set $C_{r,s} = \{(i,j) \mid i \in A_r,\ j \in A_s\}$. For each $1 \leq r < s \leq t$ let $\mathrm{z}^{(r,s)} \in [-1,1]^n$ be defined via: $\mathrm{z}^{(r,s)}_0 = 0$, and for $(i,j) \in E$:

   $$\mathrm{z}^{(r,s)}_{(i,j)} = \begin{cases} \sqrt{y_{(i,j)} - y^2_{(i,j)}} & \text{if } (i,j) \in C_{r,s} \\ -\sqrt{y_{(i,j)} - y^2_{(i,j)}} & \text{if } (i,j) \in C_{s,r} \\ 0 & \text{otherwise} \end{cases}$$

The calculation below reveals that:

$$Y = y^T y + \sum_{1 \le r < s \le t}^{m} (z^{(r,s)})^T z^{(r,s)}$$

This suffices to finish the proof of the claim, because a sum of positive semidefinite matrices is also positive semidefinite.

Checking the calculations: Let $Z = y^T y + \sum_{1 \le r < s \le t}^{m} (z^{(r,s)})^T z^{(r,s)}$.

Let (i,j) and (l,k) with $(i,j) \equiv (l,k)$ be given. First consider the case when $x^{\prec}_{(i,j)} \in \{0,1\}$. This forces that the arcs (i,j) and (l,k) do not cross two pieces of the partition, and that $x^{\prec}_{(l,k)} \in \{0,1\}$. Moreover, $z^{(r,s)}_{(i,j)} = z^{(r,s)}_{(l,k)} = 0$ for all r,s.

$$Z_{(i,j),(l,k)} \;=\; Z_{(i,j),(i,j)} = y_{(i,j)} y_{(l,k)} = x^{\prec}_{(i,j)} \cdot x^{\prec}_{(l,k)} = x^{\prec}_{(i,j)} = Y_{(i,j),(l,k)}$$

Now consider the case when $(i,j) \equiv (l,k)$ and $x^{\prec}_{(i,j)} = 1/2$ (so that both (i,j) and (l,k) cross from some A_r to some A_s, WLOG $r < s$):

$$\begin{aligned}
Z_{(i,j),(l,k)} &= y_{(i,j)} y_{(l,k)} + z^{(r,s)}_{(i,j)} z^{(r,s)}_{(l,k)} \\
&= y_{(i,j)} y_{(l,k)} + \sqrt{y_{(i,j)} - y^2_{(i,j)}} \sqrt{y_{(l,k)} - y^2_{(l,k)}} \\
&= 1/4 + \sqrt{1/2 - 1/4} \sqrt{1/2 - 1/4} = 1/2 = x^{\prec}_{(i,j)} = Y_{(i,j),(l,k)}
\end{aligned}$$

Let (i,j) and (l,k) with $(i,j) \perp (l,k)$ be given. When $x^{\prec}_{(i,j)} \in \{0,1\}$, (i,j) and (l,k) do not cross two pieces of the partition, and that $x^{\prec}_{(l,k)} = 1 - x^{\prec}_{(i,j)}$. Moreover, $z^{(r,s)}_{(i,j)} = z^{(r,s)}_{(l,k)} = 0$ for all r,s. So we have:

$$Z_{(i,j),(l,k)} \;=\; y_{(i,j)} y_{(l,k)} = x^{\prec}_{(i,j)} (1 - x^{\prec}_{(i,j)}) = 0 = Y_{(i,j),(l,k)}$$

Now consider the case when (i,j) crosses from A_r to A_s and (l,k) crosses from A_s to A_r and both $x^{\prec}_{(i,j)} = x^{\prec}_{(l,k)} = 1/2$.

$$\begin{aligned}
Z_{(i,j),(l,k)} &= y_{(i,j)} y_{(l,k)} + z^{(r,s)}_{(i,j)} z^{(r,s)}_{(l,k)} \\
&= y_{(i,j)} y_{(l,k)} - \sqrt{y_{(i,j)} - y^2_{(i,j)}} \sqrt{y_{(l,k)} - y^2_{(l,k)}} \\
&= x^{\prec}_{(i,j)} x^{\prec}_{(l,k)} - \sqrt{x^{\prec}_{(i,j)} - (x^{\prec}_{(i,j)})^2} \sqrt{x^{\prec}_{(l,k)} - (x^{\prec}_{(l,k)})^2} \\
&= 1/4 - \sqrt{1/2 - 1/4} \sqrt{1/2 - 1/4} = 0 = Y_{(i,j),(l,k)}
\end{aligned}$$

For all other (i,j), (l,k), we have that for all $1 \le r < s \le t$, either $z^{(r,s)}_{(i,j)} = 0$ or $z^{(r,s)}_{(l,k)} = 0$, so that $Z_{(i,j),(l,k)} = y_{(i,j)} y_{(l,k)} = x^{\prec}_{(i,j)} x^{\prec}_{(l,k)} = Y_{(i,j),(l,k)}$.

$\square$

ECCC ISSN 1433-8092
http://eccc.hpi-web.de/