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

Toniann Pitassi*
Computer Science Department
University of Toronto
Toronto, Ontario Canada M5S 1A4
toni@cs.toronto.edu

Nathan Segerlind
Computer Science and Engineering
Portland State University
Portland, OR 97207-0741
nsegerli@cs.pdx.edu

June 18, 2008

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, and to solve certain instances of Boolean satisfiability.

Our first result is a size/rank tradeoff for tree-like Lovász-Schrijver refutations, showing that any refutation that has small size also has small rank. This allows us to immediately derive exponential size lower bounds for tree-like refutations of many unsatisfiable systems of inequalities where prior to our work, only strong rank bounds were known.

Unfortunately, we show that this tradeoff does not hold more generally for *derivations* of arbitrary inequalities. We give a very simple example showing that derivations can be very small but nonetheless require maximal rank. This rules out a *generic* argument for obtaining a size-based integrality gap from the corresponding rank-based integrality gap. Our second contribution is to show that a modified argument can often be used to prove size-based integrality gaps from rank-based integrality gaps. We apply this method to prove size-based integrality gaps for several prominant examples where prior to our work, only rank-based integrality gaps were known.

Our third contribution is to prove new separation results. Using our machinery for converting rank-based lower bounds and integrality gaps into size-based lower bounds, we show that tree-like LS_+ cannot polynomially simulate tree-like Cutting Planes, and that tree-like LS_+ cannot polynomially simulate resolution.

We conclude by examining size/rank tradeoffs beyond the LS systems. We show that for Shirali-Adams and Lasserre systems, size/rank tradeoffs continue to hold, even in the general (non-tree) case.

* Research supported by NSERC.

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 [22] (called LS and LS_+) for tightening a linear relaxation of a zero-one programming problem. For several optimization problems, a small number of applications of the semidefinite LS_+ 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 LS_+, starting from the natural linear program for the independent set problem gives the Lovász Theta functions [21]; one round starting from the natural linear program for the max cut problem gives the famous Goemans-Williamson relaxation for approximating the maximum cut in a graph [14]; and three rounds gives the breakthrough Arora Rao Vazirani relaxation for the sparsest cut problem [6]. Moreover, linear and semi-definite programming methods are widely viewed as a catch-all approach for solving other approximation problems. To back this up, very recent work [7, 23] shows that for a general family of constraint satisfaction problems, the optimal approximation factor (which is actually unknown!) will be equal to the integrality gap obtained after a small number of rounds of matrix cut operators (under the unique games conjecture.)

Due to the importance and seemingly ubiquitous nature of this family of algorithms, there has been a growing body of research aimed at ruling out low-rank LS_+ approximation algorithms for prominent approximation problems. These results prove that a very large family of semidefinite programs (those obtained by optimizing over a low-rank LS_+ polytope) will fail to achieve a good approximation by proving an *integrality gap*. (That is, exhibiting a nonintegral point lying in the polytope, whose value is off from the integral optimal by a certain approximation factor.) Such integrality gaps are important as they show that one of the most promising family of algorithms for solving these problems will not succeed in polynomial time. At present there are rank-based integrality gaps for LS and LS_+ for many important problems, including max-*k*-SAT, max-*k*-LIN, and vertex cover. (For example, see [5, 13, 24, 25, 26, 8, 20, 2, 10].)

While these algorithms rule out a large collection of SDP algorithms, they do not rule out all polynomial-time SDP algorithms. For example, it is certainly conceivable that there are inequalities that one might add that are natural for the problem at hand, but that are not derivable by low rank LS_+ from the initial set of inequalities. Such programs would not be ruled out by rank-based integrality gaps. In this paper we study the *size* of the LS_+ derivation needed to yield good approximations to optimization problems. Exponential (or even superpolynomial) *size*-based integrality gaps are the ultimate negative result as they show that *any* polynomial-time procedure based on LS (or LS_+) will fail to efficiently find an approximate solution (via standard rounding schemes.) In contrast, rank-based lower bounds only rule out algorithms that generate low-rank tightenings of the initial polytope.

We point out that lower bounds for LS_+ are incomparable to PCP-based lower bounds since on the one hand they are unconditional, but on the other hand, they rule out a large class of important algorithms (as opposed to all algorithms). As discussed above, there is an abundance of *rank-based* lower bounds and integrality gaps; however, with respect to the stronger size measure, very little is known. In fact the only unconditional size bound known is due to Kojevnikov and Itsykon [19]. Building on results from [15, 17, 16, 18]) they prove exponential size lower bounds for tree-like LS_+ derivations of certain unsatisfiable formulas (the Tseitin formulas). For integrality gaps, there were no size bounds at all. Our paper is largely inspired by the results of [19]. Can we prove size bounds for other unsatisfiable formula? What about size-based integrality gaps? Finally, what is the connection between size and rank?

1.1 Summary of results

Our first result is a size/rank tradeoff for tree-like LS_0, LS, LS_+ refutations, showing that tree-like refutations can be converted into somewhat balanced refutations. More precisely, we prove the following. Suppose that I is a system of inequalities with a tree-like LS_+ (or LS, LS_0) refutation of size S. Then there is a refutation of I of rank at most $O(\sqrt{n \ln S})$. In particular, if I has a polynomial-size refutation, then it has a refutation of rank $O(\sqrt{n \log n})$. This tradeoff allows us to immediately derive exponential size lower bounds for tree-like refutations for several unsatisfiable systems of inequalities where prior to our work, only rank bounds were known (random 3CNF formulas and random systems of mod 2 equations). In other words, our lower bounds show that a large class of algorithms (those based on constructing tree-like LS_+ proofs) cannot solve SAT exactly in subexponential time. We note that this result is unconditional and rules out a broader class of algorithms than those ruled out by rank bounds.

The main idea behind our size/rank tradeoff is to define a new measure of complexity for a tree-like proof called the *variable rank*. We view a proof derivation as a tree where we label nodes with inequalities and edges with variables that are lifted on in this step. The rank of a proof is thus the longest path in the proof, whereas the variable rank is the largest number of *distinct* variable labels over all paths. Our key insight is to show that for any refutation, the variable rank equals the rank. This allows us to apply well-known methods for balancing the proof by iteratively applying restrictions to kill off long paths. We show that our tradeoff is optimal by exhibiting a family of formulas where our size/rank tradeoff is tight.

Next we try to attack the more interesting problem of proving superpolynomial size bounds for any LS_+ algorithm for approximating an optimization problem. This class of algorithms, say for max-k-SAT, is defined as follows. Begin with the natural polytope corresponding to an instance of max-k-SAT. Apply *any* sequence of LS_+ cuts to the initial polytope to obtain a new refined polytope. The size of the refined polytope is the number of cuts used to derive it from the initial polytope. The tree-size is the number of cuts used where we require that the underlying derivation is a tree. For a maximization problem, the refined polytope has an integrality gap of k if there is a solution with value at least k times OPT; for a minimization problem, the integrality gap is k if there is a solution with value OPT/k. For example, for vertex cover, we would like to show that any subexponential-size tree LS_+ algorithm has an integrality gap of 2. The most natural way to show this is to prove a stronger size/rank tradeoff for LS_+ that holds for *deriations* of arbitrary inequalities (instead of just for refutations, which are derivations of $0 \geq 1$.)

Unfortunately, we prove that this tradeoff does not hold more generally for *derivations* of arbitrary inequalities. We present a very simple example showing that derivations can be very small, but nonetheless require maximal rank. This rules out a *generic* argument for obtaining size-based integrality gaps from the corresponding rank-based integrality gaps. Despite our lack of a general tree-size/rank trade-off for derivations of arbitrary linear inequalities, our second main contribution is to show that a modified argument can often be used to prove size-based integrality gaps from rank-based integrality gaps. We illustrate this method by proving size-based integrality gaps for several optimization problems: We show that for max-k-SAT, every polytope that is obtained by applying an LS_+ tightening of sub-exponential tree-size has integrality gap $1+\frac{1}{2^k-1}$. Similarly we prove a size-based integrality gap of $2-\varepsilon$ for max-k-LIN, and $7/6$ for vertex cover.

Our third main contribution is to prove new separation results in proof complexity. Using our new machinery for converting rank-based lower bounds and integrality gaps into size-based lower bounds (combined with several new ideas), we show that tree-like LS_+ cannot polynomially simulate tree-like Cutting Planes, and that tree-like LS cannot polynomially simulate resolution. This shows in particular that low rank LS_+ cannot polynomially simulate Resolution. We conclude by examining size/rank tradeoffs beyond the LS systems.

We show that for Shirali-Adams and Lasserre systems, size/rank tradeoffs continue to hold, even in the general (non-tree) case.

2 Matrix-cut proof systems

There are several cutting planes proof systems defined by Lovász and Schrijver, collectively referred to as matrix cuts [22]. In these proof systems, we begin with a system of linear inequalities over the variables X. We will present dual definitions for these systems: In the "proof-theoretic" one, we start with a system of linear inequalities and describe precise "cut" rules for obtaining new inequalities from previous ones. In the second "model-theoretic" definition, we will begin with a polytope defined as the set of solutions to the initial system of linear inequalities, and at each round, we will describe a new tightened polytope defined as the set of vectors in the original polytope that have a "protection matrix" associated with them.

2.1 Proof-theoretic View

Definition 2.1. *Given a system of inequalities over $[0,1]^n$ defined by $a_i^T X \geq b_i$ for $i = 1,2,\ldots,m$: $j = 1,\ldots,n$. An inequality $c^T X - d$ is called an N_+-cut if*

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

where $\alpha_{i,j}, \beta_{i,j} \geq 0$, $\lambda_j \in \mathbb{R}$ for $i = 1,\ldots,m,\ j = 1,\ldots,n$ and for each k, $g_k \in \mathbb{R}$, $h_k \in \mathbb{R}^n$. An N-cut is a N_+-cut if $k = 0$. (That is, we cannot use squares of arbitrary linear inequalities.) An N_0-cut is an N-cut if the equality holds when we view X_iX_j as distinct from X_jX_i, $1 \leq i < j \leq n$. 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 (or $1 - X_j$) *if* $\alpha_{ij} > 0$ *(or* $\beta_{ij} > 0$*.)*

Definition 2.2. *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.*

Definition 2.3. *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. The size of Γ is the total bit size of representing all inequalities, with all coefficients in binary notation. The rank of Γ is the depth of the underlying directed acyclic graph. For a set of boolean inequalities I, the $\mathcal{P}$-size ($\mathcal{P}$-tree-size, $\mathcal{P}$-rank) of I is the minimal size (tree-size, rank) over all $\mathcal{P}$ refutations of I. Define $LS_0^r(I)$ ($LS^r(I)$, $LS_+^r(I)$) to be the set of all linear inequalities with LS_0 (LS, LS_+) derivations from I of rank at most r.*

Lemma 2.4. *(Closure under restrictions) Let Γ be an LS_0 (LS, LS_+) derivation of $c^T X \geq d$ from hypotheses I. Let ρ be a restriction to the variables of X. Then $\Gamma\restriction_\rho$ is an LS_0 (LS, LS_+) derivation of $\left(c^T X \geq d\right)\restriction_\rho$ from the hypotheses $I\restriction_\rho$.*

2.2 Model-theoretic view

Definition 2.5. *Let $I = \{a_i^T X \geq b_i \mid i = 1,\ldots,m\}$ be a system of linear equalities in the variables $X_1,\ldots X_n$. Define the polytope of I as $P_I = \{x \in \mathbb{R}^n \mid \forall i \in [m], a_i^T x \geq b_i\}$.*

Following the usual conventions, we will change the setting slightly by working with a convex cone rather than a convex set. Our object of interest is the convex set $P_I \subseteq [0,1]^n$. We first convert it into the *homogenized cone* $K_I \subseteq \mathbb{R}^{n+1}$, defined as $K_I = \{x \in \mathbb{R}^{n+1} \mid \forall i \in [m], a_i^T x - b_i x_0 \geq 0\}$. We will now define the various LS operators, N, N_+ and N_0 such that if K is a cone, then $N_+(K)$, $N(K)$ and $N_0(K)$ are also cones.

Definition 2.6. *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) $Ye_0 = diag(Y) = Y^T e_0 = y$. (The top row, leftmost column, and diagonal of Y are y.); (2) For all $i = 0, \ldots n$, $Ye_i \in K$ and $Y(e_0 - e_i) \in K$. (The i^{th} column and (y minus the i^{th} column) are in K.); (3) If $y_i = 0$ then $Ye_i = 0$, and if $y_i = y_0$ 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.

Definition 2.7. *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. We define $N(K)$ and $N_+(K)$ analogously. The sets $N_0(K)$, $N(K)$ and $N_+(K)$ are easily seen to be cones, and therefore the construction can be iterated. 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. After applying the N operator iteratively to tighten the cone we will then want to project back to $X_0 = 1$ in order to get to the "tightened" polytope: let $K \upharpoonright_{X_0=1} = \{x \in \mathbb{R}^n \mid (1, x_1, \ldots, x_n) \in K\}$.*

2.3 Equivalence between the two views

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, stating that the polytope obtained after r rounds of the cut rule is equal to the polytope obtained after r iterations of the corresponding N operators, projected onto $X_0 = 1$.

Theorem 2.8. *[22] Let I be a set of inequalities in $\{X_1, \ldots X_n\}$ that includes the inequalities $0 \leq X_i \leq 1$ for all $i \in [n]$, and let $K_I \subseteq \mathbb{R}^{n+1}$ be the polyhedral cone given by the homogenization of I. Then $P_{LS_0^r(I)} = N_0^r(K_I) \upharpoonright_{X_0=1}$, $P_{LS^r(I)} = N^r(K_I) \upharpoonright_{X_0=1}$, and $P_{LS_+^r(I)} = N_+^r(K_I) \upharpoonright_{X_0=1}$.*

Corollary 2.9. *Let I be a set of inequalities in $\{X_1, \ldots X_n\}$ that includes the inequalities $0 \leq X_i \leq 1$ for all $i \in [n]$, and let $K_I \subseteq \mathbb{R}^{n+1}$ be the polyhedral cone given by the homogenization of I. The following statements are equivalent: (1) There exists a rank $\leq r$ LS refutation of I; (2) Every point of $N^r(K_I)$ satisfies $0 \geq X_0$; (3) $N^r(K_I) \upharpoonright_{X_0=1}$ is empty. Also, there exists a LS elimination of $x \in \mathbb{R}^n$ from I of rank at most r if and only if $\binom{1}{x} \notin N^r(K_I)$. Analogous statements relate LS_0 with N_0, and LS_+ with N_+.*

Definition 2.10. *Let $x \in [0,1]^n$. $Supp(x)$ are those coordinates i such that x_i is equal to 0 or 1. $E(x)$ are the other coordinates j such that x_j is not integral. Clearly $[n] = Supp(x) \cup E(x)$.*

Definition 2.11. *Let $x \in \mathbb{R}^n$ be given, and let Y be an LS_0 protection matrix for $\binom{1}{x}$. For each $i = 0, \ldots n$, 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.

The following lemma shows that if some $x \in K$ fails to make it into the next round of LS_+ tightening, then any candidate protection matrix Y for x will fail in the sense that one of the $2n$ alleged protection vectors will fail to be in K.

Lemma 2.12. *(proof in Appendix) Let $I = \{a_1^T X \geq b_1, \ldots a_m^T X \geq b_m\}$ be a system of inequalities. Let $c^T X \geq d$ be an inequality obtained by one round of LS_+ lift-and-project from I. Let $x \in \mathbb{R}^n$ be given such that $c^T x < d$. Let Y be a matrix for $\binom{1}{x}$ in the sense that it satisfies the definition of a protection matrix with the possible exception of property (2). Then there exists an $i \in [m]$ and a $j \in [n]$ so that either: (i) $a_i^T X \geq b_i$ is used as the hypothesis for a lifting inference on X_j, $x_j \neq 0$, and $a_i^T PV_{j,1}(Y) < b_i$, or (ii) $a_i^T X \geq b_i$ is used as the hypothesis for a lifting inference on $1 - X_j$, $x_j \neq 1$, and $a_i^T PV_{j,0}(Y) < b_i$.*

We will use the following form of Theorem 2.8, stating that if x is in $N(K)$, then there is a protection matrix Y for x such that all integral bits of x are preserved in all $2n$ protection vectors, and furthermore, the protection vector $PV(Y)_{i,\varepsilon}$ that corresponds to lifting on x_i^{ε} also has its i^{th} bit set to ε.

Lemma 2.13. *(Proof in the Appendix) Let $x \in \mathbb{R}^n$. and let $K \subseteq \mathbb{R}^{n+1}$ be a cone that satisfies $0 \leq X_i \leq X_0$ for all $i \in [n]$. Let $\binom{1}{x} \in N_0(K)$ ($N(K)$,$N_+(K)$). Then there exists a LS_0 (LS, LS_+) protection matrix Y for $\binom{1}{x}$ with respect to K_I such that for each $i \in [n]$, $\varepsilon \in \{0,1\}$, $Supp(x) \cup \{i\} \subseteq Supp(PV_{i,\varepsilon}(Y))$.*

Finally, we will use Farkas lemma which is kind of "completeness theorem" for linear programming:

Lemma 2.14. *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, $c^T x \geq d$. Then there exists $\alpha_1, \ldots, \alpha_m$, each $\alpha_i \geq 0$ such that $c^T X - d = \sum_{i=1}^m \alpha_i (a_i^T X - b_i)$.*

3 Tree-size versus rank

The high-level strategy for the our size/rank tradeoff is very similar to that used by Clegg, Edmonds and Impagliazzo, showing a relationship between degree and size for the polynomial calculus [11]. We first outline this general approach, and then explain the obstacles in using this approach and how we overcome them. As an example, we will outline how to transform a polynomial-size tree refutation into a low rank refutation. Consider the skeleton of the proof tree where nodes are labelled with inequalities and edges are labelled with the literal that is being lifted upon (multiplied by). If we can hit the proof with a restriction such that each long path contains at least one literal set to false, then this will result in a low rank proof, under the restriction. However the low-rank refutation will only be a refutation under the restriction and thus we must continue recursively and argue that there is also a low rank restriction under all other settings to the restricted variables. This will be possible since the size of the restriction will be small. Finally, we will combine all of the low rank refutations (one for each assignment to the restricted variables) in order to obtain a low rank refutation of the entire formula.

In our actual argument, we will select the restriction and recursive somewhat differently than described above, but the intuition is similar. Rather than selecting the whole restriction at once to kill all long paths simultaneously, we will select one variable setting at a time. We will always choose the next variable setting greedily, by picking the variable assignment that kills off the largest number of long paths. We argue that when the variable is set this way (the first case), the number of long paths drops by a large fraction, and when the variable is set the other way (the second case), the total number of variables is reduced by 1. In the first case, we will argue inductively that we can obtain a low rank $r-1$ refutation, and in the second case, a rank r refutation, and finally argue that they can be combined to obtain a rank r refutation.

When applying this argument we run into trouble because a path can be long without mentioning a lot of distinct literals on the edges of the path. A proof is called *regular* if for every path in the proof, a variable occurs in at most one edge labelling along the path. If the proof is regular, then we can apply the above argument. Unfortunately, the proof might be highly *irregular*, potentially making it impossible to apply the restriction argument. An extreme example would be a refutation tree containing two very long paths, one that mentions a literal x_i repeatedly, and another that mentions $\neg x_i$ repeatedly, thereby making it impossible to kill off both long paths simultaneously.

We get around this problem by arguing that in any refutation, if there is a long path, then there must exist another long *regular* path. More precisely, the rank of a tree refutation is the length of the longest path, and we define the *variable rank* of the tree refutation to be the maximum number of variables that are mentioned on a single path. (If the proof is regular, then these two notions of rank are equal.) Theorem 3.1 shows that rank and variable rank are equal. Note that we do not show that for any refutation tree, we can convert it into a regular refutation tree of the same rank. Nonetheless by controlling the irregularities in the proof, we can make the argument outlined above go through. We show that rank and variable rank are equal in Subsection 3.1, and we use this to prove the tree-size/rank trade-off in Subsection 3.2.

3.1 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 $X_1,\dots,X_n$, 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, $\mathrm{vrank}(I)$, is defined to be $\mathrm{vrank}(0 \geq 1)$. The variable rank of a vector $x \in [0,1]^n$ with respect to I, $\mathrm{vrank}^I(x)$, is the minimum variable rank with respect to I of an inequality $c^T X \geq d$ such that $c^T x < d$.

Theorem 3.1. *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 $\mathrm{vrank}^I(x) \leq \mathrm{rank}^I(x)$. We will prove the other direction by induction on $\mathrm{rank}^I(x)$. 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 $\mathrm{rank}^I(x) \geq r+1$. 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 derive the inequality $c^T X - d$. By Lemma 2.13, there is a protection matrix Y for $\binom{1}{x}$ with respect to $N_+^r(P_I)$ satisfying the properties of the lemma. By Lemma 2.12, 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 $\mathrm{Supp}(x)$. Suppose $j \in \mathrm{Supp}(x)$. 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 $\mathrm{Supp}(x)$. 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 $a_i^T X \geq b_i$. By Lemma 2.13, 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. □

3.2 A tight trade-off for rank and tree-size

Theorem 3.2. *For any set of inequalities I with no $0/1$ solution, in each of the systems LS_0, LS, and LS_+, rank$(I) \leq 2\sqrt{2n \ln S_T(I)}$.*

We will need the following two preliminary lemmas.

Lemma 3.3. *(Proof in Appendix) Let I be a 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 $X_i \geq \varepsilon$ from I 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 $(1 - X_i) \geq \varepsilon$ from I of rank at most r.*

Lemma 3.4. *(Proof in Appendix) For all systems of inequalities 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 $\leq r$ deriation from I of $X_i \geq \delta$, then there is a rank $\leq r$ refutation of I.*

Proof. (of Theorem 3.2) Let $S \in \mathbb{N}$ be given. Let $d = \sqrt{2n \ln S}$, and let $a = (1 - d/2n)^{-1} = (1 - \sqrt{\ln S/2n})^{-1}$.

Let I be a set of inequalities in n variables, and let Γ be a refutation of I. Let F be the set of long paths in Γ of *variable rank* at least d. We prove by induction on n and b that if I is a system of inequalities in at most n variables that has a refutation with at most a^b long paths, then rank$(I) \leq d + b$.

The claim trivially holds for all b when $d \geq n$, because every refutation that uses at most n variables has rank at most n. In the base case, $b = 0$ and there are no paths in Γ of variable rank more than d, and thus by Theorem 3.1, rank$(I) \leq d$. For 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 $X = 0$, $\Gamma \restriction_{X=0}$ is a refutation of $I \restriction_{X=0}$ with at most $\left(1 - \frac{d}{2n}\right)|F| < a^{b-1}$ many long paths. By the induction hypothesis, rank$(I \restriction_{X=0}) \leq d + b - 1$. By Lemma 3.3, 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, $\Gamma \restriction_{X=1}$ 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, rank$(I \restriction_{X=1}) \leq d + b$. By Lemma 3.3, there is $\delta \geq 0$ and a derivation of $X \geq \delta$ from I of rank at most $d + b$. Therefore by Lemma 3.4, rank$(I) \leq d + b$. This concludes the proof that if $|F| < a^b$, then rank$(I) \leq d + b$.

Because $|F| < |\Gamma| = a^{\log_a(S)}$, we set $b = \log_a S$ which can be seen to be equal to $\sqrt{2n \ln S}$. Thus rank$(I) \leq 2\sqrt{2n \ln S}$ as desired. □

Corollary 3.5. *For the LS_0, LS and LS_+ systems, for any set of inequalities I in n variables with no $0/1$ solution,* $S_T(I) \geq e^{(rank(I))^2/9n}$.

It is interesting to note that we actually prove a stronger lower bound where size is measured to be the number of inequalities in the proof, and not just the bit size.

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 [16] 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 3.6. *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 [16], 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.

3.3 No trade-off for arbitrary derivations in LS_0 and LS

Theorem 3.2 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_+.

Theorem 3.7. *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 $\Omega(n)$, and (2) 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-1}^n [16]. 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-1}^n. Thus we have a rank r refutation of PHP_{n-1}^n. Because the LS rank of PHP_{n-1}^n is $\Omega(n)$, it follows that $r = \Omega(n)$. Lastly, 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. □

It is interesting to note that for any ε, the system $I \cup \{\sum_{i=1}^n X_i \geq 1 + \varepsilon\}$ has a rank one LS_0 refutation. 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 3.8. *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 3.7, let I be the following system of inequalities: For each $1 \le i < j \le n$, there is $x_i + x_j \le 1$. By the argument of the proof of Theorem 3.7, all derivations of $\sum_{i=1}^n x_i \le 1$ from I require rank $r_0 = \Omega(n)$. Therefore, by the affine Farkas Lemma, Lemma 2.14, 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 $\sum_{i=1}^n x_i \le 1$ from I of size $n^{O(1)}$. Upon deriving $\sum_{i=1}^n x_i \le 1$, the point x is eliminated. $\square$

4 Tree-size Lower Bounds and Integrality Gaps

The tree-size/rank trade-off of Theorem 3.2 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.

Definition 4.1. *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. 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. Finally, the Tseitin formula on an odd-sized graph $G = (V,E)$, $P_{TS(G)}$, has variables x_e for all edges $e \in E$. For each $v \in V$ there is one corresponding equation: $\sum_{e, v \in e} x_e = 1 \bmod 2$.*

Our tree-size tradeoff together with the rank lower bounds from [8] immediately give the following theorem.

Theorem 4.2.

1. *For all odd n sufficiently large, there exists a G on n nodes and degree Δ such that any LS_+ refutation 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 $k \ge 5$. 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 $k \ge 5$, the boundary expansion is greater than 2. In a subsequent paper, Alekhnovich, Arora and Tourlakis prove linear rank for random 3-CNFs [2]. This immediately yields the corresponding exponential tree-size lower bounds for random 3CNF formulas.

As discussed in Subsection 3.3, we cannot appeal to Theorem 3.2 to obtain tree-size based integrality gaps because this theorem holds for refutations but not for more general derivations. Nonetheless, we can obtain integrality gaps for sub-exponential tree-size LS and LS_+ relaxations by using similar ideas.

For max-k-SAT and max-k-LIN, we will actually manage to use Theorem 3.2 directly to prove integrality gaps. For vertex cover, we will demonstrate how to use the ideas behind the proof of 3.2 to obtain size-based integrality gaps based on rank-based integrality gaps using a more hand-tailored approach. This is completely analogous to using a hand-tailored random restriction argument to prove Resolution lower bounds, in cases where the general size-width tradeoff for Resolution cannot be applied.

Recall that the high level idea of the proof of Theorem 3.2 is to hit an alleged small proof with a restriction to kill off all high rank paths, and then figure out how to patch together the low-rank derivations (one where $x_i = 1$ and one where $x_i = 0$.) in a low-rank way. For derivations it is no longer possible to argue that we can patch together the low rank derivations, but we can bypass this step as follows: Begin with an alleged small-size derivation of some inequality g from I. Find a "nice" restriction ρ such that: (i) ρ kills off all high rank paths, and (ii) ρ has the property that $g\restriction_\rho$ still requires high rank.

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. Given a set of k-mod-2 equations $F = \{f_1, \dots, f_m\}$ over variables $X_1, \dots, X_n$, add a new set of variables $Y_1, \dots, Y_m$. 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.

Theorem 4.3. *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.*

$LS+$ **Integrality Gap for Vertex Cover.** Given a 3XOR instance F over $\{X_1, \dots X_n\}$ 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. to three variables. Two vertices u and v are connected if and only if the partial assignments corresponding to 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 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 polytope is denoted $VC(G_F)$. Our final result in this subsection is a generalization of the rank bound of [24] to a tree-size bound.

Theorem 4.4. *For all $\varepsilon > 0$, there exists $\Delta, c > 0$ such 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 of $VC(G_F)$ with integrality gap at most $7/6 - \varepsilon$ has size at least 2^{cn}.*

5 Separations between proof systems

In this section, we show that tree-like LS_+ refutations can require an exponential-size increase to simulate several other proof systems. Our first theorem shows that tree-like LS_+ cannot efficiently simulate *Gomory-Chvatal* (GC) cutting planes, and our second theorem below shows in particular that small rank LS_+ cannot simulate resolution. The proofs of the following theorems (in the Appendix) first prove new rank bounds, and then use the machinery developed in Sections 3 and 4 to obtain size bounds from rank bounds.

Theorem 5.1. *Tree-like LS_+ does not polynomially simulate GC cutting planes.*

Theorem 5.2. *Tree-like LS_+ refutations cannot p-simulate either DAG-like resolution, or DAG-like LS_+.*

6 Tradeoffs beyond LS

We note that rank-size tradeoffs for Lassere and Shirali Adams proof systems are easier to obtain and moreover they are stronger. We will prove that for these systems, linear rank bounds imply exponential size bounds (and not just tree-size bounds as was the case for the LS systems.) For this section, let R be either Lassere or Shirali Adams.

Theorem 6.1. *(Proof in Appendix) Let I be a system of inequalities with n underlying variables, and suppose that I has an R refutation of size S. Then I has a rank $O\sqrt{n \log S}$ R refutation.*

The idea behind the above theorem is very similar to Theorem 3.2. The first step is to show that any R proof can be *multilinearized* without increasing the rank. That is, if I is an system of inequalities with an R refutation, then the refutation can be converted into one where all inequalities are multilinear. This can be accomplished straightforwardly by adding appropriate quantities of $(x_i^2 - x_i)$. The second step is to reprove Lemmas 3.3 and 3.4 for R. These lemmas allow us to combine a rank $r-1$ refutation of $I_{x=0}$ with a rank r refutation of $I_{x=1}$ in order to obtain a rank r R-refutation of I. Finally, with these lemmas at hand, Theorem 6.1 can be proven analogously to the proof of Theorem 3.2.

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] P. Austrin. Towards sharp inapproximability for any 2csp. In *FOCS*, pages 307–317, 2007.

[8] 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.

[9] 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.

[10] M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for sherali-adams relaxations, 2007.

[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] 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.

[13] W. de la Vega and C. Kenyon-Mathiew. Linear programming relaxations of maxcut. In *SODA*, 2007.

[14] 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.

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

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

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

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

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

[20] G. Konstantinos, A. Magen, T. Pitassi, and I. Tourlakis. Optimal integrality gaps for lovasz-schrijver relaxations of vertex cover. In *ECCC*, 2006.

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

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

[23] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *STOC*, pages 245–254, 2008.

[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 vertex cover in the lovasz-schrijver hierarchy. In *IEEE Conference on Computational Complexity*, 2006.

A Proofs from Section 2

Proof. (of Lemma 2.12) 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:

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

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 2.6, $Y_j = 0$ 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 2.6, $Y_j = \binom{1}{x}$ 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 2.6, 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 $0 > -b_i + a_i^T ((x - y)/(1 - x_j)) = -b_i + a_i^T (PV_{j,0}(Y))$.

□

Proof. (of Lemma 2.13)

Let $x \in \mathbb{R}^n$ and let $K \subseteq \mathbb{R}^{n+1}$ be a cone that satisfies $0 \leq X_i \leq X_0$ for all $i \in [n]$. Let $\binom{1}{x} \in N_0(K)$ $(N(K), N_+(K))$. Then we want to show that there exists a LS_0 (LS, LS_+) protection matrix Y for $\binom{1}{x}$ with respect to K_I such that for each $i \in [n]$, $\varepsilon \in \{0,1\}$, $Supp(x) \cup \{i\} \subseteq Supp(PV_{i,\varepsilon}(Y))$.

From the definitions it is clear that for all $i \in [n]$, $\varepsilon \in \{0,1\}$, $i \in Supp(PV_{i,\varepsilon}(Y))$. It is left to show that $Supp(x) \subseteq Supp(PV_{i,\varepsilon}(Y))$.

Let Y be a protection matrix for $\binom{1}{x}$ with respect to K. Y is said to be *support extending* if for all $i \in [n]$ and for all $j \in [n]$, $y_j = 1 \rightarrow (Ye_i)_j = y_i$ and $y_j = 0 \rightarrow (Ye_i)_j = 0$. Note that LS and LS_+ protection matrices are always support extending because of symmetry, but the definition is restrictive for LS_0 protection matrices. We will show that support-extending protection matrices always exist, even for LS_0. That is, we show that if $y = \binom{1}{x} \in N_0(K)$, then there exists a support-extending LS_0 protection matrix for y.

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 3.6 from [12] $N_0(K) \cap F = N_0(K \cap F)$, 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 $(Ye_i)_j = 0$ as desired.

Finally, by definition it is not hard to see that if Y is a support-extending protection matrix for $\binom{1}{x}$ with respect to K, then for each $i \in [n]$, $\varepsilon \in \{0,1\}$, $Supp(x) \subseteq Supp(PV_{i,\varepsilon}(Y))$. This completes the proof.

$\square$

B Proofs from Section 3

Proof. (of Lemma 3.3).

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 $i \in [n]$. We will prove the following stronger statement. 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.

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 2.8, 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 2.14, 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.

Now to complete the proof of Lemma 3.3, 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 the more general statement proven above, this implies 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. $\square$

Proof. (Proof of 3.4)

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 $X_i \geq \varepsilon$. From this 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 $X_i \geq 1$. 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 $1/\delta$. □

C Proofs from Section 4

For this and subsequent sections, we will need the notion of graph expansion.

Definition C.1. *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 C.2. *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 C.3. *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 C.4. *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.*

The following theorem proven by [8] gives a rank lower bound for mod 2 equations as a function of the expansion.

Theorem C.5. *[8] 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 $r(c-2)$.*

Proof. (of Theorem 4.3) 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 analogous 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 3.2 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 C.5, the $\text{rank}^{F\cup\{g\}}(z) = \Omega(n)$, and therefore $\text{rank}(F \cup \{g\}) = \Omega(n)$. By Theorem 3.2, 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 $S = 2^{\Omega(n)}$.

□

In order to prove the final theorem from this section, Theorem 4.4, we will need some preliminary Lemmas and facts.

The following lemma was proven in [24].

Lemma C.6. *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 $(3/4, \dots, 3/4)$ 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 C.6.

Lemma C.7. *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 C.8. *[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 C.7 and take $\alpha, \Delta > 0$, t sufficiently large (to demonstrate that the theorem holds for arbitrary large graphs), and a 3XOR instance F over $X_1, \dots X_t$ 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 C.6, 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. □

We will improve Lemma C.6 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 C.6. (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 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 C.4, 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 C.9. *(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:*

$$I_1 \vdash^e_J I_2 \iff |I_2| \le (r/2) \wedge \left| N_A(I_2) \setminus \left(\bigcup_{i\in I_1} A_i \cup J\right)\right| < \delta\cdot\eta|I_2| \tag{1}$$

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 $ecl_A^\delta(J)$ 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, $\mathrm{CL}_J^\delta(A)$, *be the matrix that results by removing all rows of $ecl^\delta(J)$ and all columns of $J \cup \bigcup_{i \in ecl_A^\delta(J)} A_i$ from A.*

Lemma C.10. *[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 C.11. *(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. If $|J| < \frac{r(1-\delta)\eta}{2}$, then $|ecl_A^\delta(J)| < \frac{|J|}{(1-\delta)\eta}$.*

Lemma C.12. *[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 C.13. *(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 C.14. *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 $X_1,\dots X_n$ generated by the following experiment: Uniformly select a subset $S_0 \subseteq \{X_1,\dots X_n\}$ 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 C.11, $|I| = |ecl_A^\delta(S_0)| < \frac{|S_0|}{\eta(1-\delta)} \leq \frac{r(1-\delta)\beta\gamma}{2\eta(1-\delta)} \leq \frac{r(1-\delta)\eta\gamma}{2\eta(1-\delta)} = \gamma r/2 < r/2$. Therefore, by Lemma C.13, 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 C.15. *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 C.16. *[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 C.17. *(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 C.18. *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 C.17, 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 C.19. *(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 C.11 $|ecl_A^{\frac{199}{200}}(J)| < 200|J| \leq 200(.001r) \leq r/5 = \alpha n/5$. 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 C.10, $\mathrm{CL}_J^{\delta}(A)$ 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 C.13, 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 C.20. *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 $c(\phi_\pi) \geq \frac{|\pi|}{4d}$. $\square$

We are finally ready to prove Theorem 4.4.

Proof. (of Theorem 4.4) Choose $\varepsilon_0, \gamma > 0$ so that $\varepsilon_0 + \gamma/2 = 3\varepsilon$. Apply Lemma C.7, 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 C.19 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 C.16 with $\delta = \frac{195}{198}$, γ as defined previously, and β equal to the boundary expansion of G_F (and thus $\beta \geq 0.96$).

For each ρ in the support of $\mathcal{D}(A, (\alpha/2)n, \beta, \delta, \gamma)$, as per Definition C.14, 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 $\sum_{(C,\eta)\in V(G_F)} w^\rho_{C,\eta} \leq 3m$. On the other hand, each such ρ satisfies at most $\gamma(\alpha/2)n/2 \leq \gamma m/2$ 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 C.18, ϕ_π is in normal form, and by Lemma C.20, $c(\phi_\pi) \geq \frac{R}{4d}$. Therefore, we may apply Lemma C.16: $Pr_\rho[\phi_\pi \restriction_\rho \neq 1] < 2^{-R/4d^{3a+1}}$. 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 $\geq R$.

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 $c^T X \geq d$ 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 $\left(c^T Y \geq d\right) \restriction_{\rho_Y}$ 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 $\left(c^T Y \geq d\right) \restriction_{\rho_Y}$. 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 3.1, 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 C.10, $F \restriction_\rho$ is an $(\alpha n/4, 1.95)$ expander. Furthermore, any two of its equations share at most one variable. So by Lemma C.6, 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 $> S = \sqrt{2^{R/4d^{3a+1}}-1} = \sqrt{2^{\left(\frac{(\alpha/4)n-4}{44}\right)/4d^{3a+1}}-1} = 2^{\Omega(n)}$.

□

D Proofs from Section 5

D.1 Proof of Theorem 5.1

Definition D.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 give the proof of Theorem 5.1, showing 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 D.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 [9], 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 D.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 D.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 D.5. *[19] 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 [9]: 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, $1 \leq l \leq r$, 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 [9] for a formal description of $T_G - Count$.) [9] 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 D.6. *[9] 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 refutation of $T_G - Count$.*

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

Theorem D.7. *[19] 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 D.5, D.7, D.6 we see that rank of $TG - Count$ is $\Omega(n)$, and because $TG - Count$ has $O(n)$ many variables, we may apply Theorem 3.2 to conclude:

Corollary D.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 D.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 $\sum_{v \in [2n+1]} \sum_{e \ni v} x_e \geq 2n+1$. □

Proof. (of Theorem 5.1) The above lemmas easily imply our theorem. □

D.2 Proof of Theorem 5.2

In this section we will prove Theorem 5.2, stating that tree-like LS_+ cannot p-simulate DAG-like resolution.

The family of formulas that we show to be hard for tree-like LS_+ is the GT_n principle. It is one of the canonical exaples for showing that a system cannot p-simulate resolution. Intuitively, it says that in any total ordering of a finite set, there must exist a minimal element.

Definition D.10. *For $n \geq 1$, the formula GT_n is a CNF on the variables $X_{i,j}$ for $i, j \in n$, $i \neq j$. The clauses of GT_n are as follows.*

- *For each $1 \leq i < j \leq n$, $(X_{i,j} \vee X_{j,i})$;*
- *For each $1 \leq i < j \leq n$, $(\neg X_{i,j} \vee \neg X_{j,i})$;*
- *For each i, j, k, $(\neg X_{i,j} \vee \neg X_{j,k} \vee X_{i,k})$;*
- *For each i, $(\vee_{j \neq i} X_{j,i})$.*

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.

The bulk of our work will be to prove the following lower bound showing that tree-like LS_+ refutations must be large. This theorem combined with standard resolution proofs of GT_n will imply Theorem 5.2.

Theorem D.11. *There exists $c > 0$ such that for all n, every tree-like LS_+ refutation of GT_n has size at least 2^{cn}.*

The first thing we do is strengthen the rank bound of [8] 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 D.12. *A partial order $\prec$ on $[n]$ is said to be t-scaled if there is a partition of $[n]$ into sets $A_1, \ldots A_t$ 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 D.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 D.14. *Let P_s denote least polytope containing $\{x^{\prec} \mid \prec$ is at least s-scaled $\}$.*

Lemma D.15. *(cf. [8]) When $s \geq 3$, $P_s \subseteq P_{GT_n}$.*

Definition D.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 $(i,j),(l,k) \in E$:*

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

Lemma D.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 $PV_{(i,j),0}(Y) = PV_{(i,j),1}(Y) = x$.*

Proof. (of Lemma D.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:

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

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

$\square$

Lemma D.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.*

Proof. (of Lemma D.18) Let $Y = Y^{\prec}$. Let $y = \binom{1}{x^{\prec}}$.

We just check that the properties of Definition 2.6 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 D.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 $z^{(r,s)} \in [-1,1]^n$ be defined via: $z_0^{(r,s)} = 0$, and for $(i,j) \in E$:

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

   The calculation below reveals that:

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

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

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

   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, $z_{(i,j)}^{(r,s)} = z_{(l,k)}^{(r,s)} = 0$ 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 $r < s$):

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

   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, $z_{(i,j)}^{(r,s)} = z_{(l,k)}^{(r,s)} = 0$ 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 $x^{\prec}_{(i,j)} = x^{\prec}_{(l,k)} = 1/2$.

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

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

□

Lemma D.19. *Let $s \in \mathbb{N}+3$ be given. For every $n \ge s$, if $\prec$ is an at least s-scaled partial order on $[n]$, then $rank^{GT_n}(x^{\prec}) \ge s-3$.*

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 D.15, which tells us $P_3 \subseteq P_{GT_n}$. Assume that the claim holds for s. Let $n \ge s+1$ be given, and let $\prec$ be an arbitrary partial order that is t-scaled with $t \ge s+1$. Consider the matrix $Y^{\prec}$: By Lemma D.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 arbitrary t-scaled partial order with $t \ge s+1$, $P_{s+1} \subseteq N_+^{s-2}(P_{GT_n})$. □

Corollary D.20. *For all $n \ge 3$, 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 D.20 is used.

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 D.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 $c^T X \ge d$. 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 D.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 D.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 D.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 D.24 then follows from Lemma D.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:

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

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

Let $Y = Y^{\prec}$, as per Definition D.16. By Lemma 2.12, 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 $\prec$. $\square$

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 D.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 $A_1, \dots A_t$ 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 $\prec_\pi$, $u_{i,1}, \dots u_{i,a_i}$. For each $j = 1, \dots \lfloor a_i/2 \rfloor$, 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 $j = 1, \dots \lfloor a_i/2 \rfloor$, 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 $A_1, \dots A_t$ are disjoint, the probability 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 $j \in \{1, \dots \lfloor a_i/2 \rfloor\}$ 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 $\prec$.

□

Theorem D.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 D.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 D.24, we must have that $3 = (n/4) - ((n/4) - 3) \leq 2$, which is false. □

E Proofs from Section 6

Let R be Shirali Adams or Lasserre. Recall that we want to prove that if I is a system of inequalities with n underlying variables, and I has an R refutation of size S, then I has a rank $O\sqrt{n \log S}$ R refutation.

The following two lemmas are the analogs of Lemmas 3.3 and 3.4 and have the same proofs.

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

Lemma E.2. *(Proof in Appendix) For all systems of inequalities I and all integers $r > 0$, and all ε, δ, if there is a rank $\leq r-1$ R-derivation I of $X_i \geq \varepsilon$ and a rank $\leq r$ derivation of from I of $1 - X_i \geq \delta$, then there is a rank $\leq r$ refutation of I. Similarly for $X_i = 0$ and $(1 - X_i) = 0$ reversed.*

The next lemma says that we can multilinearize any R refutation.

Lemma E.3. *Suppose that there exists an R-refutation of g from I. Then there exists a multilinear R refutation of g from I of the same rank. (That is, a refutation where each line consists of a multilinear polynomial inequality.)*

Proof. For any polynomial f, we can derive $f' - f \geq 0$, using the axioms $x^2 - x$ in degree equal to the degree of f, where f' is the multilinearization of f. Now we modify the original refutation so that each line f is replaced by f'. For example, taking linear combinations of multilinear inequalities results in a multilinear inequality. Also, if f is multiplied by x in the original proof, then we first multiply f' by x to get xf', then derive $(xf')' - xf'$ and add to get $(xf')'$ as desired. The rank of the new derivation will be no greater than the old rank. □

Note that this technique will not work for any of the LS systems. We may try to apply it by considering viewing each line in two ways: the first way is as a linear inequality, and the second way is as a formal polynomial. As formal polynomials, notice that any polynomial derivable in the r^{th} round will have degree r (even though its projection will have degree 1). Then we might try to show that these formal polynomials can be assumed to be multilinear as in the argument above. The problem is that we can no longer make this argument because we cannot create the polynomial $f' - f$ at all since we must always project back to degree one.

Theorem E.4. *If I is a set of inequalities and I has an R-refutation containing M nonzero monomials, then I has an R refutation of rank max$(d, 2\lceil (n \log M)^{1/2} \rceil + 1)$.*

Proof. Again the above proof is very similar to our proof of Theorem 3.2. Let $b = \lceil (n \ln M)^{1/2} \rceil$. A monomial is called *wide* if it contains greater than b distinct variables. If $b \geq n$, then the conclusion follows since everything is derivable in rank n. Otherwise for a refutation of I with at least M monomials, let S be the number of monomials containing at least b distinct variables. If $S = 0$ then all the monomials have less than b distinct variables and hence P has a refutation of degree b by Lemma E.3

We now prove that if $S \geq 1$, then there exists a refutation of P of degree at most $b - \log_{(1-b/n)} S + 1$. This completes the proof of the theorem since $S \leq M$ and $ln(1 - b/n) \leq -b/n$ which implies that $b - \log_{(1-b/n)} S + 1 = b - (\ln S)/(ln(1 - b/n)) + 1 \leq b + n/b \ln S \leq b + (n/b) \ln M + 1 \leq b + (n \ln M)/(n \ln M)^{1/3} + 1 = b + (n \ln M)^{1/2} + 1 \leq 2b + 1$.

We prove the statement by induction on S and n. Choose a variable that is in at least $(b/n)S$ monomials of degree larger than b; it exists by an averaging argument. For any $c \in \{0, 1\}$, we can form a refutation of $P \restriction_{x=c}$ by substitution. For $c = 0$ this restriction removes all monomials containing x and thus it removes at least $(b/n)S$ monomials of degree larger than b. Thus it produces a refutation of $P \restriction_{x=0}$ with less than $(1 - b/n)S$ wide monomials. If $(1 - b/n)S \geq 1$ then by induction on S the set $P \restriction_{x=0}$ has a refutation of degree $b - \log_{(1-b/n)}((1 - b/n)S) + 1 = b - \log_{(1-b/n)} S$. Otherwise all monomials have degree less than b and hence by Lemma E.3, $P \restriction_{x=0}$ has a refutation of degree $b \leq b - \log_{(1-b/n)} S$. For $c = 1$ we obtain a refutation of $P \restriction_{x=1}$ with at most S wide monomials. Now using induction on the number of variables, $P \restriction_{x=1}$ has a refutation of degree $b - \log_{(1-b/n)} S + 1$. Thus applying Lemmas E.1 and E.2 there is a refutation of P of degree at most $b - \log(1 - b/n)S + 1$. $\square$

Corollary E.5. *For any set of inequalities I in n variables, if the R rank of I is at least d, then the minimal number of nonzero monomials in such a refutation is at least $2^{\Omega(d^2/n)}$.*