Lower Bounds for CSP Refutation by SDP Hierarchies

Ryuhei Mori[*1] **and David Witmer**[†2]

1 **Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan**
`mori@is.titech.ac.jp`

2 **Computer Science Department, Carnegie Mellon University, Pittsburgh, USA**
`dwitmer@cs.cmu.edu`

Abstract

For a k-ary predicate P, a random instance of CSP(P) with n variables and m constraints is unsatisfiable with high probability when $m \geq O(n)$. The natural algorithmic task in this regime is *refutation*: finding a proof that a given random instance is unsatisfiable. Recent work of Allen et al. suggests that the difficulty of refuting CSP(P) using an SDP is determined by a parameter cmplx(P), the smallest t for which there does not exist a t-wise uniform distribution over satisfying assignments to P. In particular they show that random instances of CSP(P) with $m \gg n^{\mathrm{cmplx(P)}/2}$ can be refuted efficiently using an SDP.

In this work, we give evidence that $n^{\mathrm{cmplx}(P)/2}$ constraints are also *necessary* for refutation using SDPs. Specifically, we show that if P supports a $(t-1)$-wise uniform distribution over satisfying assignments, then the Sherali-Adams$_+$ and Lovász-Schrijver$_+$ SDP hierarchies cannot refute a random instance of CSP(P) in polynomial time for any $m \leq n^{t/2-\varepsilon}$.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases constraint satisfaction problems, LP and SDP relaxations, average-case complexity

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.41

1 Introduction

The average-case complexity of constraint satisfaction problems (CSPs) has been studied in computer science, mathematics, and statistical physics. Despite the vast amount of research that has been done, the hardness of natural algorithmic tasks for random CSPs remains poorly understood. We consider random CSPs with n variables and m constraints chosen independently and uniformly at random. Whether or not a random CSP is satisfiable depends on its clause density $\frac{m}{n}$. It is conjectured that for any nontrivial CSP there is a *satisfiability threshold* $\alpha(P)$ depending on the choice of predicate P: For $m < \alpha(P) \cdot n$, an instance is satisfiable with high probability, and $m > \alpha(P)$, an instance is unsatisfiable with high probability. This conjecture has been proven in the case of k-SAT for large enough k [16]. For an arbitrary predicate P, it is only known that there exist constants $\alpha_{lb}(P)$ and $\alpha_{ub}(P)$ such that random instances with $m < \alpha_{lb}(P) \cdot n$ are satisfiable with high probability and random instances with $m > \alpha_{ub}(P) \cdot n$ are unsatisfiable with high probability. In the

* Supported by MEXT KAKENHI Grant Number 24106008.
† Supported by the National Science Foundation Graduate Research Fellowship Program under grant DGE-1252522, by NSF grants CCF-0747250 and CCF-1116594, and by a CMU Presidential Fellowship.

low density, satisfiable regime, the major research goal is to develop algorithms that find satisfying assignments. In the high density, unsatisfiable regime, the goal is to *refute* an instance, i.e., find a short certificate that there is no solution.

In this paper, we study refutation. A refutation algorithm takes a random instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ and returns either “unsatisfiable" or "don't know". It must satisfy two conditions: (1) it is never wrong, i.e., if $\mathcal{I}$ is satisfiable, it must return “don't know" and (2) it returns “unsatisfiable" with high probability over the choice of the instance. As m increases, refutation becomes easier. The objective, then, is to refute instances with m as small as possible. This problem has been studied extensively and is related to hardness of approximation [17], proof complexity [8], statistical physics [13], cryptography [2], and learning theory [14]. As m increases, refutation becomes easier. The objective, then, is to refute instances with m as small as possible. Much research has focused on finding algorithms for refutation, especially in the special case of SAT; see [1] for references.

Most known refutation algorithms are based on semidefinite programming (SDP). For now, we think of an SDP relaxation of an instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ as a black box that returns a number $\mathrm{SDPOpt} \in [0,1]$ that approximates the maximum fraction of constraints that can be simultaneously satisfied. An SDP-based refutation algorithm takes a random instance $\mathcal{I}$ of $\mathrm{CSP}(P)$, solves some SDP relaxation of $\mathcal{I}$, and return “Unsatisfiable" if and only if $\mathrm{SDPOpt} < 1$. The majority of known polynomial-time algorithms for refutation fit into this framework (e.g., [1, 7, 20, 12, 18]). It is then natural to ask the following question.

> What is the minimum number of constraints needed by an efficient SDP-based refutation algorithm for $\mathrm{CSP}(P)$?

Allen et al. give an upper bound on the number of constraints required to refute an instance of $\mathrm{CSP}(P)$ in terms of a parameter $\mathrm{cmplx}(P)$ [1]. They define $\mathrm{cmplx}(P)$ to be the minimum t such that there is no t-wise uniform distribution over satisfying assignments to P. Clearly $1 \le \mathrm{cmplx}(P) \le k$ for nontrivial predicates and $\mathrm{cmplx}(P) = k$ when P is k-XOR or k-SAT. They give the following upper bound.

▶ **Theorem 1** ([1]). *There is an efficient SDP-based algorithm that refutes a random instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ with high probability when $m \gg n^{\mathrm{cmplx}(P)/2}$.*

For special classes of predicates, we know that $n^{\mathrm{cmplx}(P)/2}$ constraints are also necessary for refutation by SDP-based algorithms. Schoenebeck considered arity-k predicates P whose satisfying assignments are a superset of k-XOR's; these include k-SAT and k-XOR. For such predicates, he showed that polynomial-size sum of squares (SOS) SDP relaxations cannot refute random instances with $m \le n^{k/2-\varepsilon}$ [28] using a proof previously discovered by Grigoriev [21]. Based on work of Lee, Raghavendra, and Steurer [24], this implies that no polynomial-size SDP can be used to refute random instances of k-XOR or k-SAT when $m \le n^{k/2-\varepsilon}$. This leads us to make the following conjecture.

▶ **Conjecture 2.** *Let ε be a constant greater than 0. Given a random instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ with $m \le n^{\mathrm{cmplx}(P)/2-\varepsilon}$, with high probability any polynomial-size SDP relaxation of $\mathcal{I}$ has optimal value 1 and can therefore not be used to refute $\mathcal{I}$.*

Proving this conjecture would essentially complete our understanding of the power of SDP-based refutation algorithms. To do this, it suffices to prove it for SOS SDP relaxations, as the SOS relaxation of $\mathrm{CSP}(P)$ is at least as powerful as an arbitrary SDP relaxation of comparable size [24]. Prior to this work, this SOS version Conjecture 2 appeared in [1]; we know of no other mention of this conjecture in the literature.[1]

Some partial progress has been made toward proving this conjecture. Building on results of Benabbas et al. [9] and Tulsiani and Worah [30], O'Donnell and Witmer proved lower bounds for the Sherali-Adams (SA) linear programming (LP) hierarchy and the Sherali-Adams$_+$ (SA$_+$) and Lovász-Schrijver$_+$ (LS$_+$) SDP hierarchies. All three of these hierarchies are weaker than SOS. The SA$_+$ hierarchy gives an optimal approximation for any CSP in the worst case assuming the Unique Games Conjecture [27]. They showed that Sherali-Adams linear programming (LP) relaxations cannot refute random instances with $m \leq n^{\mathrm{cmplx}(P)/2-\varepsilon}$ [26]; this implies that no polynomial-size LP can refute random instances with with $m \leq n^{\mathrm{cmplx}(P)/2-\varepsilon}$ by work of Chan et al. [11]. They showed that SA$_+$ cannot refute random instances with $m \leq n^{\mathrm{cmplx}(P)/2-\varepsilon}$ when a set of $o(m)$ constraints has been removed [26]. Also, they proved that SA$_+$ and LS$_+$ cannot refute fully random instances with $m \leq \Omega(n^{\mathrm{cmplx}(P)/2-1/3-\varepsilon})$. Much less is known for SOS. For predicates P that support a pairwise uniform distribution over satisfying assignments, Barak, Chan, and Kothari showed that polynomial-size SOS relaxations cannot refute random instances with $m = \Omega(n)$ in which $o(m)$ constraints have been removed [5].

In addition, there is a long history of related work on lower bounds for refutation in proof complexity (e.g., [21, 23, 10, 30]). Specifically, SA, SA$_+$, LS$_+$, and SOS have corresponding static semialgebraic proof systems and proving integrality gaps for these LP and SDP relaxations in equivalent to proving rank lower bounds for refutations in these proof systems.

Results

Our contribution is two-fold: First, we remove the assumption that a small number of constraints are deleted to show that fully-random CSP instances have integrality gaps in SA$_+$ for $m \leq \Omega(n^{t/2-\varepsilon})$.

▶ **Theorem 3.** *Let $P : [q]^k \to \{0,1\}$ be $(t-1)$-wise uniform-supporting and let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with n variables and $m \leq \Omega\left(n^{t/2-\varepsilon}\right)$ constraints. Then with high probability the SA$_+$ relaxation for $\mathcal{I}$ has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

Second, we use this result to show that fully random instances have LS$_+$ integrality gaps for $m \leq \Omega(n^{t/2-\varepsilon})$. Recall that LS$_+$ gives relaxations of 0/1-valued integer programs, so we restrict our attention here to Boolean CSPs with $P : \{0,1\}^k \to \{0,1\}$.

▶ **Theorem 4.** *Let $P : \{0,1\}^k \to \{0,1\}$ be $(t-1)$-wise uniform-supporting and let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with n variables and $m \leq \Omega\left(n^{t/2-\varepsilon}\right)$ constraints. Then with high probability the LS$_+$ relaxation for $\mathcal{I}$ has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

In their strongest form, our results hold for a static variant of the LS$_+$ SDP hierarchy that is at least as strong as both SA$_+$ and LS$_+$. We define this static LS$_+$ hierarchy in Section 2.

▶ **Theorem 5.** *Let $P : \{0,1\}^k \to \{0,1\}$ be $(t-1)$-wise uniform-supporting and let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with n variables and $m \leq \Omega\left(n^{t/2-\varepsilon}\right)$ constraints. Then with high probability the static LS$_+$ relaxation for $\mathcal{I}$ has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

[1] Barak, Kindler, and Steurer [6] made a related but different conjecture that the basic SDP relaxation is optimal for random CSPs.

Tulsiani and Worah proved this theorem in the special case of pairwise independence and $O(n)$ constraints [30, Theorem 3.27].[2]

These results provide further evidence for Conjecture 2 and, in particular, give the first examples of SDP hierarchies that are unable to refute CSPs with $(t-1)$-wise uniform-supporting predicates when $m \leq \Omega\left(n^{t/2-\varepsilon}\right)$.

From a dual point of view, we can think of SA_+, LS_+, and static LS_+ as semialgebraic proof systems and our results can be equivalently stated as rank lower bounds for these proof systems.

▶ **Theorem 6.** *Let $P : \{0,1\}^k \to \{0,1\}$ be $(t-1)$-wise uniform-supporting and let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with n variables and $m \leq \Omega\left(n^{t/2-\varepsilon}\right)$ constraints. Then with high probability any SA_+, LS_+, or static LS_+ refutation of $\mathcal{I}$ requires rank $\Omega(n^{\frac{\varepsilon}{t-2}})$.*

In another line of work, Feldman, Perkins, and Vempala [19] showed that if a predicate P is $(t-1)$-wise uniform supporting, then any statistical algorithm based on an oracle taking L values requires $m \geq \widetilde{O}\left(\frac{n^r}{L}\right)$ to refute. They further show that the dimension of any convex program refuting such a CSP must be at least $\widetilde{O}(n^{t/2})$. These lower bounds are incomparable to the integrality gap results stated above: While statistical algorithms and arbitrary convex relaxations are more general models, standard SDP hierarchy relaxations for k-CSPs, including the SA_+ and LS_+ relaxations we study, have dimension $\Theta(n^k)$ and are therefore not ruled out by this work.

Techniques

For simplicity we consider our CSP to have ± 1-values variables and $P : \{-1,1\}^k \to \{0,1\}$. To solve $\mathrm{CSP}(P)$ exactly, it suffices to optimize over distributions on assignments $\{0,1\}^n$. This, of course, is hard, so relaxations like SA, SA_+, LS_+, and SOS instead optimize over "pseudodistributions" on assignments [4], which are objects that look like distributions to simple enough functions. We can also define "pseudoexpectations" $\widetilde{\mathbb{E}}[\cdot]$ over these pseudodistributions. As the rank or degree of the relaxation increases, these pseudodistributions look more like actual distributions over $\{0,1\}^n$. However, the rank-r relaxations have size $n^{O(r)}$. The r-round SA_+ relaxation requires that $\widetilde{\mathbb{E}}[f(x)] \geq 0$ for $f(x) : \{0,1\}^n \to \mathbb{R}$ such that either (1) f depends on at most r variables or (2) f is the square of some affine function. We know that when $m \leq n^{\mathrm{cmplx}(P)/2-\varepsilon}$, there exist pseudodistributions satisfying (1) [9, 26]. Condition (2) is equivalent to positive semidefiniteness of the matrix of second pseudomoments $\widetilde{\mathbb{E}}[x_i x_j]$ of the pseudodistribution. Recalling that we are now considering ± 1-valued variables, previous work had constructed pseudodistributions and proved that their second moment matrix was positive semidefinite (PSD) by showing that it was diagonal and had nonnegative entries. The second moment matrix is diagonal when there are no correlations between assignments to pairs of variables under the pseudodistribution and the marginals of pseudodistribution for each variable is unbiased. This condition holds for instances with low densities, but correlations between variables arise as the density increases.

We show positive semidefiniteness in the presence of these correlations by showing that they must remain local. Our argument extends a technique of Tulsiani and Worah [30]. We prove that the graph induced by correlations between variables must have small connected components, each of which corresponds to a small block of off-diagonal nonzero

[2] Actually, [30] prove a rank lower bound for the dual static LS_+ proof system, but this is equivalent to a rank lower bound for the static LS_+ SDP hierarchy we consider here.

entries in the second pseudomoment matrix. Since each of these off-diagonal blocks is small, condition (1) guarantees that functions supported on these variables will have nonnegative pseudoexpectation. This means that each of these off-diagonal blocks is a PSD matrix. The pseudomoment matrix must then be PSD: It can be written as a sum of a diagonal matrix with nonnegative elements and PSD “correction matrices", each of which corresponds to a small connected component in the correlation graph.

2 Preliminaries

2.1 Constraint satisfaction problems

Given a predicate $P : [q]^k \to \{0, 1\}$, an instance $\mathcal{I}$ of the $\mathrm{CSP}(P)$ problem with variables $x_1, \ldots, x_n$ is a multiset of P-constraints. Each P-constraint is a tuple (c, S), where $c \in [q]^k$ is the negation pattern and $S \in [n]^k$ is the scope. The corresponding constraint is $P(x_S + c) = 1$, where $x_S = (x_i)_{i\in S}$ and $+$ is component-wise addition mod q.

In the decision version of $\mathrm{CSP}(P)$, we wish to determine whether there exists an assignment satisfying all constraints of a given instance $\mathcal{I}$. In the optimization version, the objective is to maximize the fraction of simultaneously satisfied constraints. That is, we define $\mathrm{Val}_{\mathcal{I}}(x) := \frac{1}{m}\sum_{(c,S)\in\mathcal{I}} P(c + x_S)$ and wish to find $x \in [q]^n$ maximizing $\mathrm{Val}_{\mathcal{I}}(x)$. We will write $\max_x \mathrm{Val}_{\mathcal{I}}(x)$ as $\mathrm{Opt}(\mathcal{I})$.

Next, we define our random model. We consider instances in which m constraints are drawn independently and uniformly at random from among all $q^k n^k$ possible constraints with replacement. We distinguish between different orderings of the scope, as P may not be symmetric. The specific details of this definition are not important; our results hold for any similar model. For example, see [1, Appendix D]. A random instance is likely to be highly unsatisfiable: It is easy to show that $\mathrm{Opt}(\mathcal{I}) = \frac{|P^{-1}(1)|}{q^k} + o(1)$ for $m \geq O(n)$ with high probability.

Given an instance $\mathcal{I}$, we can consider the associated k-uniform hypergraph $H_{\mathcal{I}}$. This is the hypergraph on $V = [n]$ that has a hyperedge S if and only if S is the scope of some constraint of $\mathcal{I}$.

Next, we define the main condition on predicates that we will study.

▶ **Definition 7.** A predicate $P : [q]^k \to \{0, 1\}$ is t-wise uniform supporting if there exists a distribution μ over $[q]^k$ supported on P's satisfying assignments such that $\Pr_{z\sim\mu}[z_T = \alpha] = q^{-|T|}$ for all $\alpha \in [q]^{|T|}$ and for all $T \subseteq [k]$ with $1 \leq |T| \leq t$.

2.2 LP and SDP hierarchies

We can define LP and SDP relaxations of both the decision and optimization versions of CSPs. All of our results will apply to both.

2.2.1 Representing $\mathrm{CSP}(P)$with polynomial inequalities

All of the hierarchies we look at start with an initial set of polynomial inequalities representing an instance of a CSP and then iteratively tighten this relaxation. In this section, we will describe standard ways of constructing these initial relaxations.

To write down LP and SDP relaxations of CSP decision problems, we will need to represent the constraints of an instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ as a set of polynomial inequalities. We will do this in two ways.

In SA, SA_+, and LS_+, we will represent each constraint as a degree-k polynomial inequality. Let $P'(x)$ be the unique degree-k polynomial such that $P'(z) = P(z)$ for all $z \in \{0,1\}^k$. Also, given $a \in [0,1]^k$ and $b \in \{0,1\}$, use $a^{(b)}$ to denote a if $b = 0$ and $1 - a$ if $b = 1$. For $z \in [0,1]^k$ and $c \in \{0,1\}^k$, let $z^{(c)} \in [0,1]^k$ be such that $(z^{(c)})_i = z_i^{(c_i)}$. Then we define the base degree-k relaxation of $\mathcal{I}$ to be the set

$$R_{\mathcal{I}} := \{x \in [0,1]^n \mid P'(x_S^{(c)}) = 1 \ \forall (c,S) \in \mathcal{I}\}. \tag{2.1}$$

It will be most natural to construct SA, SA_+, and static LS_+ relaxations starting from this polytope. In addition, this formulation immediately generalizes to larger alphabets.

For LS_+, on the other hand, we will have to start with a *linear* relaxation. Recall that any nontrivial arity-k predicate P can be represented as a conjunction of at most $2^k - 1$ disjunctions of arity k. In particular, letting $F = \{z \in \{0,1\}^k \mid P(z) = 0\}$, we see that

$$P(z) = \bigwedge_{f \in F} \bigvee_{i=1}^{k} f_i \oplus z_i. \tag{2.2}$$

Using (2.2), we can represent $\mathcal{I}$ as a k-SAT instance with at most $(2^k - 1) \cdot m$ constraints. The linear relaxation we will consider is the standard linear relaxation for this k-SAT instance. For each clause $\bigvee_{i=1}^k c_i \oplus z_i$, we will add the inequality $\sum_{i=1}^k z_i^{(c_i)} \geq 1$. We then obtain the following linear relaxation for $\mathcal{I}$.

$$L_{\mathcal{I}} := \left\{ x \in [0,1]^n \;\middle|\; \sum_{i=1}^{k} x_{S_i}^{(c_i \oplus f_i)} \geq 1 \ \forall (c,S) \in \mathcal{I}, f \in F \right\}. \tag{2.3}$$

To get a maximization version of the LS_+ relaxation, we will need a base relaxation with linear constraints and a linear objective function. To do this, we will start with $L_{\mathcal{I}}$ and add variables $z_{c,S}$ for all constraints $(c,S) \in \mathcal{I}$ and consider the following polytope in $\mathbb{R}^{n+m}$.

$$L'_{\mathcal{I}} := \left\{ (x,z) \in [0,1]^{n+m} \;\middle|\; \sum_{i=1}^{k} x_{S_i}^{(c_i \oplus f_i)} \geq z_{c,S} \ \forall (c,S) \in \mathcal{I}, f \in F \right\}. \tag{2.4}$$

We can then write the following standard LP relaxation of $\mathrm{CSP}(P)$.

$$\begin{aligned} &\max \frac{1}{m} \sum_{(c,S) \in \mathcal{I}} z_{c,S} \\ &\quad \text{s.t. } (x,z) \in L'_{\mathcal{I}} \end{aligned}$$

2.2.2 Sherali-Adams

The Sherali-Adams (SA) linear programming hierarchy gives a family of locally consistent distributions on assignments to sets of variables. As the size of these sets increases, the relaxation becomes tighter.

► **Definition 8.** Let $\{D_S\}$ be a family of distributions D_S over $[q]^S$ for all $S \subseteq [n]$ with $|S| \leq r$. We say that $\{D_S\}$ is r-locally consistent if for all $T \subseteq S \subseteq [n]$ with $|S| \leq r$, the marginal of D_S on T is equal to D_T. That is,

$$D_T(\alpha) = \sum_{\substack{\beta \in [q]^S \\ \beta_T = \alpha}} D_S(\beta).$$

Given polynomial inequalities $a_1(x) \geq 0$, $a_2(x) \geq 0$, ..., $a_m(x) \geq 0$ called axioms such that each a_j depends on at most r variables, we consider the set

$$A := \{x \in \mathbb{R}^n \mid a_1(x) \geq 0, a_2(x) \geq 0, \ldots, a_m(x) \geq 0\}.$$

For a polynomial q, let $\mathrm{supp}(q)$ be the set of all variables on which a depends. We then define the *rank-r* SA *relaxation for A* as the set of all families of distributions $\{D_S\}_{S\subseteq[n],\,|S|\leq r}$ over $[q]^S$ satisfying following two properties.

1. $\{D_S\}_{S\subseteq[n],\,|S|\leq r}$ is r-locally consistent.
2. $\mathbb{E}_{\alpha\sim D_{\mathrm{supp}(a_j)}}[a_j(\alpha)] = 1$ for all $j \in [m]$.

We denote this set of families of distributions as $\mathrm{SA}^r(A)$; note that $\mathrm{SA}^r(A)$ is a polytope.

In the case of an instance $\mathcal{I}$ of $\mathrm{CSP}(P)$, we can write r-round SA relaxations in both feasibility and optimization forms. In the feasibility formulation, we check whether or not the polytope $\mathrm{SA}^r(R_\mathcal{I})$ is feasible; this is a relaxation of the problem of checking whether or not all constraints can be simultaneously satisfied.

In the rank-r SA optimization formulation, we solve the following LP.

$$\begin{aligned} \max \quad & \frac{1}{m}\sum_{(c,S)\in\mathcal{I}} \mathbb{E}_{\alpha\sim D_S}[P(\alpha+c)] \\ \text{s.t.} \quad & \{D_S\}_{S\subseteq[n],\,|S|\leq r} \in \mathrm{SA}^r(\emptyset). \end{aligned} \tag{2.5}$$

This is an LP of size $n^{O(r)}$ that is a relaxation of the problem of maximizing the number of satisfied constraints. As r increases, the number of variables and constraints in this LP increases and the relaxation tightens until $r = n$, when SA has $\Theta(q^n)$ variables and gives the exact solution.

2.2.3 Sherali-Adams$_+$

The Sherali-Adams$_+$ (SA$_+$) SDP hierarchy additionally requires the second moment matrix of these distributions to be PSD. Given a family of local distributions $\{D_S\}$, define $M(D) \in \mathbb{R}^{(nq+1)\times(nq+1)}$ to be the symmetric matrix indexed by $(0, [n] \times [q])$ such that

$$\begin{aligned} M(D)_{0,0} &= 1 \\ M(D)_{0,(i,a)} &= D_{\{i\}}(x_i = a) \\ M(D)_{(i,a),(j,b)} &= D_{\{i,j\}}(x_i = a \wedge x_j = b). \end{aligned}$$

Note that the $((i,a),(i,b))$-element of B is $D_{\{i\}}(x_i = a)$ if $a = b$ and is 0 if $a \neq b$. Given an initial set A of axioms as above, we define $\mathrm{SA}^r_+(A)$ as we did $\mathrm{SA}^r(A)$ with the following additional condition.

3. $M(D)$ is PSD.

We define the optimization version of the rank-r SA$_+$ relaxation analogously to the optimization version of SA in (2.5); this is an SDP with size $n^{O(r)}$.

We can equivalently define SA$_+$ by requiring the covariance matrix of the locally consistent $\{D_S\}$ distributions to be positive semidefinite (PSD).

▶ **Definition 9.** The covariance matrix Σ for r-locally consistent distributions $\{D_S\}$ for $r \geq 2$ is defined as

$$\Sigma_{(i,a),(j,b)} = D_{\{i,j\}}(x_i = a \wedge x_j = b) - D_{\{i\}}(x_i = a) \cdot D_{\{j\}}(x_j = b).$$

These two representations are equivalent [31].

▶ **Lemma 10.** *M is PSD if and only if Σ is PSD.*

We include the proof in Appendix B.

The covariance matrix condition will be more convenient for us to work with. A valid global distribution has a PSD covariance matrix, so SA_+ is a relaxation of $\mathrm{CSP}(P)$ and is exact for $r = n$.

2.2.4 Lovász-Schrijver$_+$

We now define the Lovász-Schrijver$_+$ (LS_+) SDP relaxation for binary CSPs whose variables are 0/1-valued. Given an initial polytope $K \in \mathbb{R}^n$, we would like to generate a sequence of progressively tighter relaxations. To define one LS_+ lift-and-project step, we will use the cone

$$\widetilde{K} = \{(\lambda, \lambda x_1, \dots, \lambda x_n) \mid \lambda > 0, x_1, \dots, x_n \in K\}.$$

K can be recovered by taking the intersection with $x_0 = 1$.

▶ **Definition 11.** Let $\widetilde{K}$ be a convex cone in $\mathbb{R}^{n+1}$. Then the lifted LS_+ cone $N_+(\widetilde{K})$ is the cone of all $y \in \mathbb{R}^{n+1}$ for which there exists an $(n+1) \times (n+1)$ matrix Y satisfying the following:

1. Y is symmetric and positive semidefinite.
2. For all i, $Y_{ii} = Y_{i0} = y_i$.
3. For all i, $Y_i \in \widetilde{K}$ and $Y_0 - Y_i \in \widetilde{K}$

where Y_i is the ith column of Y. Then we define $N_+(K)$ to be $N_+(\widetilde{K}) \cap \{x_0 = 1\}$. The r-round LS_+ relaxation of a polytope K results from applying the N_+ operator r times. That is, we define $N_+^r(K) = N_+(N_+^{r-1}(K))$. Y is called a protection matrix for y.

A solution to the r-round LS_+ relaxation for a polytope $K \in \mathbb{R}^n$ defined by $n^{O(1)}$ linear constraints can be computed in time $n^{O(r)}$ using an SDP.

We can write an LS_+ relaxation for $\mathrm{CSP}(P)$ in two ways. The maximization version of the LS_+ CSP relaxation is

$$\begin{aligned} &\max \frac{1}{m} \sum_{(c,S) \in \mathcal{I}} z_{c,S} \\ &\text{ s.t. } (x,z) \in N_+^r(L'_{\mathcal{I}}). \end{aligned}$$

Alternatively, we can check feasibility of $N_+^r(L_{\mathcal{I}})$.

We note here that though it is more natural to apply SA, SA_+, and static LS_+ to (2.1), applying SA, SA_+, and static LS_+ to (2.3) yields a relaxation that is at least as strong.

▶ **Lemma 12.** *Let $r \geq k$. Then the following statements hold.*

1. $\mathrm{SA}^r(R_{\mathcal{I}}) = \mathrm{SA}^r(L_{\mathcal{I}})$.
2. $\mathrm{SA}_+^r(R_{\mathcal{I}}) = \mathrm{SA}_+^r(L_{\mathcal{I}})$.
3. $\mathrm{StaticLS}_+^r(R_{\mathcal{I}}) = \mathrm{StaticLS}_+^r(L_{\mathcal{I}})$

We include the proof in Appendix D. The $\mathrm{StaticLS}_+^r$ operator is defined in the next section.

2.2.5 Static LS_+

The static LS_+ relaxation strengthens both SA_+ and LS_+. As in the case of SA_+, we start with a family of r-locally consistent distributions and then further require that they satisfy certain positive semidefiniteness constraints. In particular, for all $X \subseteq [n]$ with $|X| \leq r-2$ and all $\alpha \in [q]^X$, define the matrices $M_{X,\alpha}(D) \in \mathbb{R}^{(nq+1)\times(nq+1)}$ as follows.

$$\begin{aligned} M_{X,\alpha}(D)_{0,0} &= 1 \\ M_{X,\alpha}(D)_{0,(i,a)} &= D_{\{i\}\cup X}(x_i = a \wedge X = \alpha) \\ M_{X,\alpha}(D)_{(i,a),(j,b)} &= D_{\{i,j\}\cup X}(x_i = a \wedge x_j = b \wedge X = \alpha). \end{aligned}$$

In addition to the SA constraints, the r-round static LS_+ relaxation $\mathrm{StaticLS}_+^r(A)$ satisfies the following constraint.

3'. $M_{X,\alpha}(D)$ is PSD for all $X \subseteq [n]$ with $|X| \leq r-2$ and all $\alpha \in [q]^X$.

Observe that these positive semidefiniteness constraints can be formulated as a positive semidefiniteness constraint for a single matrix. In particular, let M_{total} be the block diagonal matrix with the $M_{X,\alpha}$'s on the diagonal. Then M_{total} has size at most $(qn)^{O(r)}$ and M_{total} is PSD if and only if all of the $M_{X,\alpha}$'s are PSD. The maximization version is again defined exactly as it was for SA and SA_+. Unlike LS_+, this hierarchy immediately generalizes to non-binary alphabets.

For intuition, one can think of this hierarchy as requiring positive semidefiniteness of the covariance matrices of the conditional distributions formed by conditioning on the events that X is assigned α for all X with $|X| \leq r-2$ and all $\alpha \in [q]^X$. We prefer the definition presented here because it more easily handles the case of X getting assigned α with probability 0 in which the corresponding conditional distribution is not defined.

We note that we have not seen this hierarchy defined in this form in previous work, but it is dual to the static LS_+ proof system defined in [22] and described below in Section 2.3 (see Appendix E).

2.2.6 Pseudodistributions: An alternate point of view

We can equivalently define SA, SA_+, and static LS_+ in terms of pseudodistributions [4, 3]. A pseudodistribution is a map $\sigma : \{0,1\}^n \to \mathbb{R}$ that "looks like" a valid distribution over $\{0,1\}^n$ to simple enough functions. Define the corresponding pseudoexpectation $\widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[\cdot]$ as $\widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[f(\mathbf{x})] = \sum_{x\in\{0,1\}^n} \sigma(x)f(x)$.

Sherali-Adams

A *rank-r* SA *pseudodistribution* satisfies that following two conditions.

1. $\sum_{x\in\{0,1\}^n} \sigma(x) = 1$.
2. $\widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[f(\mathbf{x})] \geq 0$ for all nonnegative functions $f : \{0,1\}^n \to \mathbb{R}$ that depend on at most r variables.

Sherali-Adams$_+$

A *rank-r* SA_+ *pseudodistribution* satisfies Condition 1 and a stronger version of Condition 2:

2'. $\widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[f(\mathbf{x})] \geq 0$ for all nonnegative functions $f : \{0,1\}^n \to \mathbb{R}$ satisfying one of the following.
   1. f depends on at most r variables.
   2. $f = \ell^2$ for some function ℓ with degree at most 1.

Static LS_+

A *rank-r static* LS_+ *pseudodistribution* satisfies Condition 1 and a version of Condition 2 that is stronger still:

2″. $\widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[f(\mathbf{x})] \geq 0$ for all nonnegative functions $f : \{0,1\}^n \to \mathbb{R}$ satisfying one of the following.
  1. f depends on at most r variables.
  2. There exists a set X of $r-2$ variables such that for any assignment α to X, the function $f_{X,\alpha} : [q]^{[n]\setminus X} \to \mathbb{R}$ resulting from setting X to α is equal to ℓ^2 for some function ℓ with degree at most 1 possibly depending on X and α.

All three relaxations maximize the objective function

$$\frac{1}{m}\sum_{(c,S)\in\mathcal{I}} \widetilde{\mathbb{E}}_{\mathbf{x}\sim\sigma}[P(x_S+c)]$$

over their corresponding pseudodistributions σ.

2.3 The dual point of view: Static semialgebraic proof systems

We consider refutation of CSPs via semialgebraic proof systems. Starting from a set of axioms $\{a_i(x) \geq 0\}$ that capture the constraints of the CSP as polynomial inequalities, semialgebraic proof systems derive new inequalities the are implied by the axioms and integrality of the variables. To prove that an instance is unsatisfiable, we wish to derive the contradiction $-1 \geq 0$. We consider the SA, SA_+, LS_+, and static LS_+ proof systems. Here, we will only deal with $\{0,1\}$-valued variables.

The SA proof system

A SA refutation has the form

$$\sum_{\ell} \gamma_\ell a_\ell(x)\phi_{I_\ell,J_\ell}(x) = -1,$$

where $\gamma_\ell \geq 0$, a_ℓ is an axiom, and $\phi_{I_\ell,J_\ell} = \prod_{i\in I_\ell} x_i \prod_{j\in J_\ell}(1-x_j)$. This is a proof of unsatisfiability because under the assumption that the all x_i variables are in $\{0,1\}$, every term of the above sum most be nonnegative and it is therefore a contradiction. The rank of this proof (often called the degree) is the maximum degree of any of the terms. The *size* of the proof is the number of terms in the sum; this follows from Farkas' Lemma. Static SA proofs is automatizable: A rank r SA proof may be found in time $n^{O(r)}$ if it exists by solving an LP. The SA proof system first appeared in [22] with the name static LS^∞; the dual hierarchy of LP relaxations was introduced by [29]. A rank-r SA refutation exists if and only if the corresponding rank-r SA relaxation is infeasible.

The SA_+ proof system

In SA_+, a proof has the form

$$\sum_{\ell} \gamma_\ell a_\ell(x)\phi_{I_\ell,J_\ell}(x) + \sum_{s} \nu_s \lambda_s(x)^2 = -1,$$

where $\gamma_\ell, \nu_s \geq 0$ and the λ_s's are affine functions. The rank of the proof is its degree. The dual SA_+ hierarchy of SDP relaxation first appeared in [27]. Again, a rank-r SA_+ refutation exists if and only if the corresponding rank-r SA_+ relaxation is infeasible. SA_+ proofs of rank r can be found in time $n^{O(r)}$ if they exist.

The LS_+ proof system

The LS_+ proof system [25] is dynamic, meaning that a proof is built up over a series of steps. A proof in LS_+ is a sequence of expressions of the form $P(x) \geq 0$. A new inequality is derived from the inequalities already in the proof using inference rules. When $\deg(P(x)) \leq 1$, we allow the following:

$$\frac{P(x) \geq 0}{x_i \cdot P(x) \geq 0} \qquad \frac{P(x) \geq 0}{(1 - x_i) \cdot P(x) \geq 0} \qquad \frac{}{P(x)^2 \geq 0.}$$

We also allow nonnegative linear combinations:

$$\frac{P(x) \geq 0 \quad Q(x) \geq 0}{\alpha \cdot P(x) + \beta \cdot Q(x) \geq 0}$$

for $\alpha, \beta \geq 0$. An LS_+ proof is therefore a sequence of “lifting" steps in which we multiply by some x_i or $(1 - x_i)$ to get a degree-2 polynomial and “projection" steps in which we take nonnegative linear combinations to reduce the degree back to 1. We can view an LS_+ proof as a DAG with inequalities at each vertex and $-1 \geq 0$ at the root. The rank of an LS_+ proof is the maximum number of lifting steps in any path to the root. The LS_+ proof system is not known to be automatizable; see Section 8 of [10] for details. An rank-r LS_+ refutation exists if and only if the corresponding rank-r LS_+ relaxation is infeasible [15].

The static LS_+ proof system

A static LS_+ proof [22] has the following form.

$$\sum_{\ell} \gamma_\ell b_\ell(x) \phi_{I_\ell, J_\ell}(x) = -1,$$

where $\gamma_\ell \geq 0$, b_ℓ is an axiom or the square of an affine function, and ϕ_{I_ℓ, J_ℓ} is as above. Note that this proof system as at least as powerful as the SA_+ proof system: Terms in the sum may be products of a $\phi_{I,J}$ term and the square of an affine function instead of just the square of an affine function or just an axiom multiplied by a $\phi_{I,J}$ term. We do not know of any results on the automatizability of static LS_+. Once again, there exists a static LS_+ refutation if and only if the corresponding static LS_+ relaxation is infeasible. We do not know of any proof of this statement in the literature, so we include one in Appendix E.

2.4 Expansion

Given a set of constraints T, we define its neighbor set $\Gamma(T)$ as $\Gamma(T) := \{v \in [n] \mid v \in \mathrm{supp}(C) \text{ for some } C \in T\}$. We can then define expansion.

▶ **Definition 13.** An instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ is (s, e)-expanding if for every set of constraints T with $|T| \leq s$, $|\Gamma(T)| \geq e|T|$.

We can also define T's boundary neighbors as $\partial T := \{v \in [n] \mid v \in \mathrm{supp}(C) \text{ for exactly one } C \in T\}$ and give a corresponding notion of boundary expansion.

▶ **Definition 14.** An instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ is (s, e)-boundary expanding if for every set of constraints T with $|T| \leq s$, $|\partial T| \geq e|T|$.

We state a well-known connection between expansion and boundary expansion stated in, e.g., [30].

▶ **Fact 15.** *$(s, k-d)$-expansion implies $(s, k-2d)$-boundary expansion.*

It is also well-known that randomly-chosen sets of constraints have high expansion [9, 26]:

▶ **Lemma 16.** *Fix $\delta > 0$. With high probability, a set of $m \leq \Omega(n^{t/2-\delta})$ constraints chosen uniformly at random is both $\left(n^{\frac{\delta}{t-2}}, k - \frac{t}{2} + \frac{\delta}{2}\right)$-expanding and $\left(n^{\frac{\delta}{t-2}}, k - t + \delta\right)$-boundary expanding.*

We give proofs of both of these statements in Appendix A.

2.5 Constructing consistent local distributions

Here, we recall a construction of consistent local distributions supported on satisfying assignments. This construction was first given in [9] and has been used in many subsequent works (e.g, [30, 26, 5]). In Appendix C, we give proofs of all results mentioned in this section.

We first need to define the notion of a closure of a set of variables. For $S \subseteq [n]$, let $H_{\mathcal{I}} - S$ denote the hypergraph $H_{\mathcal{I}}$ with the vertices of S and all hyperedges contained in S removed. Intuitively, the closure of a set $S \subseteq [n]$ is a superset of S that is not too much larger than S isn't very well-connected to the rest of the instance in the sense that $H_{\mathcal{I}} - S$ has high expansion.

▶ **Lemma 17** ([9, 30])**.** *If $H_{\mathcal{I}}$ is (s_1, e_1)-expanding and S is a set of variables such that $|S| < (e_1 - e_2)s_1$ for some $e_2 \in (0, e_1)$, then there exists a set $\mathsf{Cl}(S) \subseteq [n]$ such that $S \subseteq \mathsf{Cl}(S)$ and $H_{\mathcal{I}} - \mathsf{Cl}(S)$ is (s_2, e_2)-expanding with $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$ and $\mathsf{Cl}(S) \leq \frac{k+2e_1-e_2}{2(e_1-e_2)}|S|$.*

We now use the closure to define consistent local distributions supporting on satisfying assignments. We assume that there exists a $(t-1)$-wise independent distribution μ over satisfying assignments to P. For a constraint $C = (c, S)$, let μ_C be the distribution defined by $\mu_C(z) = \mu(z_1 + c_1, \ldots, z_k + c_k)$ and let $\mathcal{C}(S)$ be the set of constraints whose support is entirely contained within S. For a set of variables $S \subseteq [n]$ and an assignment $\alpha \in [q]^S$, we use the notation $S = \alpha$ to indicate the the variables of S are labeled according to the assignment α. For a constraint $C = (c, S)$ and an assignment α to a superset of S, let $\mu_C(\alpha) = \mu_C(\alpha_S)$.

For $S \subseteq [n]$, we can then define the distribution D'_S over $[q]^S$ as

$$D'_S(S = \alpha) = \frac{1}{Z_S} \sum_{\substack{\beta \in [q]^S \\ \beta_S = \alpha}} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta), \text{ where } Z_S = \sum_{\beta \in [q]^S} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta).$$

Using D', we can then define consistent local distributions D_S for $|S| \leq r$ so that $D_S(S = \alpha) = D'_{\mathsf{Cl}(S)}(S = \alpha)$. [9, 26] proved that these distributions are r-locally consistent for $r = n^{\frac{\varepsilon}{t-2}}$.

▶ **Theorem 18.** *For a random instance $\mathcal{I}$ with $m \leq \Omega(n^{t/2-\varepsilon})$, the family of distributions $\{D_S\}_{|S| \leq r}$ is r-locally consistent for $r = n^{\frac{\varepsilon}{t-2}}$ and is supported on satisfying assignments.*

This theorem shows that the SA cannot efficiently refute random $(t-1)$-wise uniforming supporting instances: the r-round SA LP still has value 1 for some $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ when $m \leq \Omega(n^{t/2-\varepsilon})$. In this paper, we show that even when we add the SA_+ requirement that the covariance matrix is PSD, we still cannot refute when $m \leq \Omega(n^{t/2-\varepsilon})$.

As in [30], we can also construct r-locally consistent *conditional* distributions. We will only need these distributions in the proof of our LS_+ result, so we describe them only in the binary alphabet case. Let $S \subseteq [n]$, let $X \subseteq [n]$ be a subset of the variables such that $X \cap S = \emptyset$, and let $\alpha \in \{0,1\}^X$ be an assignment to X such that $\mu_C(\alpha) > 0$ for all constraints in $\mathcal{C}(X)$. Define $D_{S|X=\alpha}$ to be the distribution D_S conditioned on the event that X is assigned α under the distribution D_X. That is, $D_{S|X=\alpha}(S=\beta) = \frac{D_{S\cup X}(S=\beta \wedge X=\alpha)}{D_{S\cup X}(X=\alpha)}$.

▶ **Lemma 19** ([30, Lemma 3.13])**.** *Let $X \subseteq [n]$ and let $\{D_S\}$ be a family of r-locally consistent distributions for sets $S \subseteq [n]$ such that $S \cap X = \emptyset$ and $|S \cup X| \leq r$. Then the family of conditional distributions $\{D_{S|X=\alpha}\}$ is $(r - |X|)$-locally consistent for any $\alpha \in \{0,1\}^X$ such that $\mu_C(\alpha) > 0$ for all constraints C in $\mathcal{C}(X)$.*

The following is a simple corollary that we will use in Section 7:

▶ **Corollary 20.** *Let $\mathcal{I}$ be a random instance of* CSP(P) *with n variables and let $X \subseteq [n]$ such that $|X| \leq \Omega(n^{\frac{\varepsilon}{t-2}})$ and let $\alpha \in \{0,1\}^X$ be any assignment to X such that $\mu_C(\alpha) > 0$ for all constraints in $C \in \mathcal{C}(X)$. Then the family of conditional distributions $\{D_{S|X=\alpha}\}_{|S|\leq r, S\cap X=\emptyset}$ is r-locally consistent for some $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.*

3 Overview of the proof

Previous work [9, 30] only considers instances with a linear number of constraints and relies on the fact that most pairs of variables are uncorrelated in this regime. For $m \gg n$, however, correlations between pairs of vertices do arise because the underlying hypergraph becomes more dense. The major technical contribution of this work is to deal with these correlations by proving that they remain local. More precisely, we consider the graph induced by correlations between variables: Two variables are connected if they have non-zero correlation. We prove that this graph must have connected components of at most constant size. Each of these connected components can then be covered by a local distribution of constant size and this suffices to ensure PSDness of the covariance matrix.

Showing that a set of local distributions is a valid SA_+ solution requires proving that these distributions are consistent and proving that their covariance matrix is PSD. Local consistency was proven in previous work [9, 26]. To prove Theorem 3, it remains to argue that the covariance matrix is PSD. The proof of this statement has three steps: First, we show in Section 4 that if the correlation graph has small connected components, then the covariance matrix is PSD. Second, we show any non-zero correlation must have been caused by a relatively dense subset of constraints in Section 5. In Section 6, we show that connected components in the correlation graph must be small or they would induce large dense subsets of constraints that would violate expansion properties.

In Section 7, we show that positive semidefiniteness of the covariance matrix implies Theorem 4.

4 The correlation graph

In this section, we define the correlation graph, and show that if the correlation graph only has small connected components, then the covariance matrix is PSD.

▶ **Definition 21.** The correlation graph G_c associated with r-locally consistent distributions $\{p_S\}$ is the graph on $[n]$ with an edge between every pair of variables for which there is a non-zero entry in the covariance matrix for $\{p_S\}$. More formally,

$$E(G_c) = \{(u,v) \in [n] \times [n] \mid u \neq v, \exists (a,b) \in [q] \times [q] \text{ s.t. } \Sigma_{(u,a),(v,b)} \neq 0\}.$$

▶ **Lemma 22.** *Let $\{p_S\}$ be a family of r-locally consistent distributions. If all connected components in the correlation graph associated with $\{p_S\}$ have size at most r, then the covariance matrix for $\{p_S\}$ is PSD.*

Proof. Consider the partition $V_1, V_2, \ldots, V_\ell$ of $[n]$ such that u and v are in the same set if and only if they are connected in the correlation graph. We then have nonzero entries in the covariance matrix only for pairs $((u,a),(v,b))$ such that $u, v \in V_i$ for some i. Ordering the rows and columns of the covariance matrix according to the partition, we see that the covariance matrix is block diagonal with a nonzero block on the diagonal for each connected component of the correlation graph. Each of these blocks is PSD since each is the covariance matrix of the Sherali-Adams distribution p_{V_i} for the corresponding set V_i of the partition with size at most r and the covariance matrix of valid distribution is always PSD. Since each block is PSD, the entire matrix is PSD. ◀

We already know that $\{D_S\}$ defined in Section 2.5 is r-locally consistent with high probability when $m \leq \Omega(n^{t/2-\epsilon})$. In the following sections, we will show that connected components in the correlation graph associated with $\{D_S\}$ is small. Hence, from Lemma 22, $\{D_S\}$ is feasible solution for SA_+ SDP while it gives the trivial objective value 1.

5 Correlations are induced by small, dense structures

In this section, we show that pairwise correlations in $\{D_S\}$ are only generated by small, dense subhypergraphs that we will call “bad structures". Given a set of hyperedges W, call a variable v an W-boundary variable if it is contained in exactly one constraint in W.

▶ **Definition 23.** For variables u and v, a bad structure for u and v is a set of constraints W satisfying the following properties:

1. $u, v \in \Gamma(W)$.
2. The hypergraph induced by W is connected.
3. Every constraint contains at most $k - t$ W-boundary variables except for u and v.

We also say W is a bad structure if W is a bad structure for some u and v.

A bad structure for u and v generates correlation between u and v with respect to $\{D_S\}$.

▶ **Lemma 24.** *If there is no bad structure for u and v of size at most $|\mathcal{C}(\mathsf{Cl}(\{u,v\}))|$, then u and v are not correlated with respect to $D_{\{u,v\}}$.*

We need the following technical claim, which states that the distribution D'_S isn't affected removing a constraint with many boundary variables.

▶ **Claim 25.** *Let $T \subseteq S \subseteq [n]$ be sets of variables. Let $C^* \in \mathcal{C}(S)$ be some constraint covered by S. If $|(\partial\mathcal{C}(S) \cap C^*) \setminus T| \geq k - t + 1$, then for any $\alpha \in \{0,1\}^T$,*

$$D'_S(\alpha) = \frac{1}{q^{|T \cap (\partial\mathcal{C}(S) \cap C^*)|}} \cdot D'_{S \setminus (\partial\mathcal{C}(S) \cap C^*)}(\alpha_{T \setminus (\partial\mathcal{C}(S) \cap C^*)}),$$

Proof. Let $B = \partial\mathcal{C}(S) \cap C^*$ be the boundary variables of $\mathcal{C}(S)$ contributed by C^*, i.e., the variables contained in C^* that don't appear in any other constraint of $\mathcal{C}(S)$. First, note that

$$\begin{aligned}\sum_{\substack{\beta \in \{0,1\}^S \\ \beta_T = \alpha}} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta) &= \sum_{\substack{\beta \in \{0,1\}^{S \setminus B} \\ \beta_{T \setminus B} = \alpha_{T \setminus B}}} \prod_{c \in \mathcal{C}(S) \setminus \{C^*\}} \mu_C(\beta) \sum_{\substack{\gamma \in \{0,1\}^B \\ \gamma_{B \cap T} = \alpha_{B \cap T}}} \mu_{C^*}(\beta, \gamma) \\ &= \frac{1}{q^{k - |B \setminus T|}} \sum_{\substack{\beta \in \{0,1\}^{S \setminus B} \\ \beta_{T \setminus B} = \alpha_{T \setminus B}}} \prod_{C \in \mathcal{C}(S) \setminus \{C^*\}} \mu_C(\beta). \end{aligned} \tag{5.6}$$

The last line holds because $|B \setminus T| \geq k - t + 1$ and μ is $(t-1)$-wise independent.

Similarly,

$$Z_S = \sum_{\beta \in \{0,1\}^S} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta) = \frac{1}{q^{k-|B|}} \sum_{\beta \in \{0,1\}^{S \setminus B}} \prod_{C \in \mathcal{C}(S) \setminus \{C^*\}} \mu_C(\beta). \tag{5.7}$$

Dividing (5.6) by (5.7), we see that $D'_S(\alpha) = \frac{1}{q^{|B \cap T|}} \cdot D'_{S \setminus B}(\alpha_{T \setminus B})$. ◀

Using Claim 25, we prove Lemma 24.

Proof of Lemma 24. Let $S_0 = \mathcal{C}(\mathsf{Cl}(\{u,v\}))$. Say there exists a constraint C_1 such that $|(\partial\mathcal{C}(S_0) \cap C_1) \setminus \{u,v\}| \geq k - t + 1$. Let $S_1 = S_0 \setminus (\partial\mathcal{C}(S_0) \cap C_1)$. If there exists a constraint C_2 such that $|(\partial\mathcal{C}(S_1) \cap C_2) \setminus \{u,v\}| \geq k - t + 1$, remove its boundary variables in the same manner to get S_2. Continue in this way until we obtain a set S_ℓ such that $|(\partial\mathcal{C}(S_\ell) \cap C) \setminus \{u,v\}| \leq k - t$ for every constraint $C \in \mathcal{C}(S_\ell)$ ($\mathcal{C}(S_\ell)$ could be empty). Since $|(\partial\mathcal{C}(S_{i-1}) \cap C_i) \setminus \{u,v\}| \geq k - t + 1$ for $1 \leq i \leq \ell$, we can apply Claim 25 ℓ times to see that

$$D_{\{u,v\}}(u = a \wedge v = b) = \begin{cases} \frac{1}{q} \cdot D'_{S_\ell}(u = a) & \text{if } u \in S_\ell, v \notin S_\ell \\ \frac{1}{q} \cdot D'_{S_\ell}(v = b) & \text{if } v \in S_\ell, u \notin S_\ell \\ \frac{1}{q^2} & \text{if } u, v \notin S_\ell \\ D'_{S_\ell}(u = a \wedge v = b) & \text{if } u, v \in S_\ell. \end{cases}$$

In the first three cases, it is easy to see that the lemma holds. In the last case, since $\mathcal{C}(S_\ell)$ cannot be a bad structure, the hypergraph induced by S_ℓ must be disconnected. Say $U_1, U_2, \ldots, U_t$ are the vertex sets of the connected components of S_ℓ. Remove all connected components that contain neither u nor v to get S'_ℓ. Again, the hypergraph induced by S'_ℓ cannot be connected; otherwise, $\mathcal{C}(S'_\ell)$ would be a bad structure. This means that S'_ℓ must be disconnected with u and v in different components. Say S'_u and S'_v are the vertex sets of the connected components of S_ℓ containing u and v, respectively. Then

$$D_{\{u,v\}}(u = a \wedge v = b) = D'_{S'_\ell}(u = a \wedge v = b) = D'_{S'_u}(u = a) \cdot D'_{S'_v}(v = b).$$

The result then follows. ◀

6 All connected components of the correlation graph are small

In this section, we show that all connected components in the correlation graph associated with $\{D_S\}$ are small, which concludes the proof of Theorem 3.

▶ **Theorem 26.** *Assume that the family of distributions $\{D_S\}$ is r-locally consistent and that the hypergraph $H_\mathcal{I}$ is $(\omega(1), k - t/2 + \delta/2)$-expander. Then all connected components in the correlation graph associated with $\{D_S\}$ have size at most $\frac{2k}{\delta}$.*

We will actually prove a slightly more general theorem that we will use to prove LS_+ lower bounds in Section 7. Given a hypergraph H, let $G_{\text{bad}}(H)$ be the graph on $[n]$ such that there is an edge between i and j if and only if there exists a bad structure for i and j in H.

▶ **Theorem 27.** *If the hypergraph H is $(\omega(1), k - t/2 + \delta/2)$-expander, then all connected components in $G_{\text{bad}}(H)$ have size at most $\frac{2k}{\delta}$.*

Lemma 24 implies that $G_{\text{bad}}(H_\mathcal{I})$ contains the correlation graph associated with $\{D_S\}$ as a subgraph, so Theorem 26 immediately implies Theorem 27.

Proof of Theorem 27. For any edge e of $G_{\text{bad}}(H)$, we can find a corresponding bad structure W_e. We will say that W_e induces e. Any such bad structure W_e in a $(\omega(1), k - t/2 + \delta/2)$-expanding hypergraph satisfies

$$\Gamma(W_e) \leq (k-t)|W_e| + 2 + \frac{k|W_e| - ((k-t)|W_e| + 2)}{2} = \left(k - \frac{t}{2}\right)|W_e| + 1. \tag{6.8}$$

The first term upper bounds the number of boundary vertices, the second term counts the endpoints of e, and the last term upper bounds the number of non-boundary vertices. For a connected component in the correlation graph associated with $\{D_S\}$, let $e_1, e_2, \ldots, e_\ell$ be an ordering of edges in the connected component such that $(\bigcup_{j=1}^{i} e_j) \cap e_{i+1}$ is not empty for $i = 1, \ldots, \ell$. That is, $e_1, e_2, \ldots, e_\ell$ is an ordering of the edges in the connected component such that every edge except for the first one is adjacent to some edge preceding it. Let $W_{e_1}, \ldots, W_{e_\ell}$ be corresponding bad structures inducing these edges. Let $T_i = \bigcup_{j=1}^{i} W_{e_j}$ for $i = 1, \ldots, \ell$. While T_i itself is not necessarily a bad structure, we will show that the inequality (6.8) still holds for T_i, i.e.,

$$\Gamma(T_i) \leq \left(k - \frac{t}{2}\right)|T_i| + 1 \tag{6.9}$$

for any $i = 1, \ldots, \ell$. If (6.9) holds, the number of constraints in T_ℓ is at most $\frac{2}{\delta}$; otherwise, expansion is violated. Hence, at most $\frac{2k}{\delta}$ vertices are included in the connected component of the correlation graph associated with $\{D_S\}$.

In the following, we prove (6.9). First, note that $|\Gamma(T_1)| \leq \left(k - \frac{t}{2}\right)|T_1| + 1$ by (6.8). Let $W'_i = W_{e_i} \setminus T_{i-1}$ be the new constraints added at step i. Call any vertex in $\Gamma(T_i) \setminus \Gamma(T_{i-1})$ a new vertex. We will prove that at most $\left(k - \frac{t}{2}\right)|W'_i|$ new vertices are added and this will imply (6.9).

Let n_i be the number of new W'_i-boundary vertices. Then the total number of new vertices is at most

$$n_i + (k|W'_i| - 1 - n_i)/2.$$

The second term upper bounds the number of non-boundary vertices. The -1 comes from the fact that $\Gamma(W'_i)$ must intersect $\Gamma(T_{i-1})$ since e_i must be adjacent to some preceding edge. If $\Gamma(W'_i)$ and $\Gamma(T_{i-1})$ intersect in a boundary vertex, the resulting bound is stronger.

Hence, we would like to upper bound the number n_i of new W'_i-boundary vertices. The number n_i of new W'_i-boundary vertices is at most $(k-t)|W'_i| + 1$ since any new W'_i-boundary vertex must be a new W_{e_i}-boundary vertex and since all but one constraint in W'_i have at most $k - t$ new W_{e_i}-boundary vertices and one constraint in W'_i has at most $k - t + 1$ new W_{e_i}-boundary vertices. Hence, the number of new vertices is at most $(k-t)|W'_i| + 1 + (k|W'_i| - 1 - ((k-t)|W'_i| + 1))/2 = (k - t/2)|W'_i|$. ◀

From Lemmas 16 and 22 and Theorems 18 and 26, we obtain Theorem 3.

7 LS$_+$ rank lower bounds

In this section, we use the techniques of the previous section to prove positive semidefiniteness of the degree-2 moment matrix of the conditional local distributions $\{D_{S|X=\alpha}\}$. From here, degree lower bounds for the static LS$_+$ proof system andrank lower bounds for LS$_+$ follow easily.

7.1 Positive semidefiniteness of conditional covariance matrices

We define $Y_{X,\alpha}$ to be the conditional covariance matrix of the $\{D_{\{i\}|X=\alpha}\}$ distributions. Formally, define $y_{(X,\alpha)} \in \mathbb{R}^{n+1}$ so that $(y_{X,\alpha})_0 = 1$ and $(y_{X,\alpha})_i = D_{\{i\}|X=\alpha}(i = 1)$ for $i \in [n]$. Let $B_{X,\alpha} \in \mathbb{R}^{n\times n}$ have entries $B_{X,\alpha}(i,j) = D_{\{i,j\}|X=\alpha}(i = 1 \wedge j = 1)$. Then define $Y_{X,\alpha} \in \mathbb{R}^{(n+1)\times(n+1)}$ to be

$$\begin{pmatrix} 1 & y_{X,\alpha}^\top \\ y_{X,\alpha} & B_{X,\alpha} \end{pmatrix}.$$

To obtain LS_+ rank lower bounds, we need to show that $Y_{X,\alpha}$ is PSD.

▶ **Lemma 28.** *Let $X \subseteq [n]$ such that $|X| \le \Omega(n^{\frac{\delta}{t-2}})$. For any $\alpha \in \{0,1\}^X$ such that $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$, $Y_{X,\alpha}$ is positive semidefinite.*

To prove this lemma, we first show that $Y_{X,\alpha}$ is PSD when $H - X$ has high expansion. Then we show that any $Y_{X,\alpha}$ can expressed as a nonnegative combination of $Y_{\mathsf{Cl}(X),\beta}$'s for $\beta \in \{0,1\}^{\mathsf{Cl}(X)}$; the first step implies that each of these terms is PSD.

We start by generalizing Lemma 24 to conditional distributions. Let $\mathsf{Cl}_X(S)$ be $\mathsf{Cl}(S)$ in the hypergraph $H - X$.

▶ **Lemma 29.** *Let $X \subseteq [n]$ and $\alpha \in \{0,1\}^X$ such that $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$. If there is no bad structure for u and v in $H - X$ of size at most $|\mathcal{C}(\mathsf{Cl}_X(\{u,v\}))|$, then u and v are not correlated with respect to $D_{\{u,v\}|X=\alpha}$.*

Proof. First, recall that

$$D_{\{u,v\}|X=\alpha}(u = a \wedge v = b) = \frac{D_{\{u,v\}\cup X}(u = a \wedge v = b \wedge X = \alpha)}{D_X(X = \alpha)}.$$

We will show that $D_{\{u,v\}\cup X}(u = a \wedge v = b \wedge X = \alpha)$ is equal to the product of a term depending on u and a but not v and b and a term depending on v and b but not u and a. From there, the lemma immediately follows.

The proof is essentially the same as that of Lemma 24 above. Starting with $S_0 = \mathcal{C}(\mathsf{Cl}(\{u,v\}))$, we apply the same process except we require that each constraint C_i that we remove satisfies $|(\partial\mathcal{C}(S_{i-1}) \cap C_i) \setminus (\{u,v\} \cup X)| \ge k - t + 1$. At the end of this process, we are left with a set S_ℓ such that $|(\partial\mathcal{C}(S_\ell) \cap C) \setminus (\{u,v\} \cup X)| \le k - t$ for every constraint $C \in \mathcal{C}(S_\ell)$ ($\mathcal{C}(S_\ell)$ could be empty). Let $X_\ell := X \cap S_\ell$ and let $\alpha_\ell = \alpha_{X\cap S_\ell}$. By applying Lemma 25 repeatedly, we see that

$$D_{\{u,v\}\cup X}(u = a \wedge v = b \wedge X = \alpha) = \begin{cases} \frac{1}{q^{|X\setminus S_\ell|+1}} D'_{S_\ell}(u = a \wedge X_\ell = \alpha_\ell) & \text{if } u \in S_\ell, v \notin S_\ell \\ \frac{1}{q^{|X\setminus S_\ell|+1}} D'_{S_\ell}(v = b \wedge X_\ell = \alpha_\ell) & \text{if } v \in S_\ell, u \notin S_\ell \\ \frac{1}{q^{|X\setminus S_\ell|+2}} D'_{S_\ell}(X_\ell = \alpha_\ell) & \text{if } u, v \notin S_\ell \\ \frac{1}{q^{|X\setminus S_\ell|}} D'_{S_\ell}(u = a \wedge v = b \wedge X_\ell = \alpha_\ell) & \text{if } u, v \in S_\ell. \end{cases}$$

In all cases except for the last one, the result follows. In the last case, the assumption that there is no bad structure in $H - X$ implies that the hypergraph induced by $S_\ell - X$ in $H - X$

must be disconnected with u and v in separate connected components with hyperedge sets E_u and E_v just as in the proof of Lemma 24. Each connected component in $H - X$ has a corresponding connected component in H. Let E'_u and E'_v be the sets of hyperedges of the connected components in G corresponding to E_u and E_v in $H - X$. Also, let $S_u = \Gamma(E'_u)$ and $S_v = \Gamma(E'_v)$; note that S_u and S_v might intersect. Let $E_{\text{rest}} = \mathcal{C}(S_\ell) \setminus (E'_u \cup E'_v)$.

The subhypergraph induced by S_ℓ then consists of the hyperedges in E'_u, E'_v, and E_{rest} together with the isolated vertices in $S_{\text{iso}} := \mathcal{S}_\ell \setminus \Gamma(\mathcal{C}(S_\ell))$ not contained in any hyperedge covered by S_ℓ. Define $S_{\text{rest}} := \Gamma(E_{\text{rest}}) \cup S_{\text{iso}}$ to be the vertices in S_ℓ that are either contained in some hyperedge of E_{rest} or are isolated. Let $X_u = X \cap S_u$ and $\alpha_u = \alpha_{X \cap S_u}$. Define X_v, X_{rest}, α_v, and α_{rest} in the same way. We can then write $D'_{S_\ell}(u = a \wedge v = b \wedge X_\ell = \alpha_\ell)$ as

$$D'_{S_u}(u = a \wedge X_u = \alpha_u) \cdot D'_{S_v}(v = b \wedge X_v = \alpha_v) \cdot D'_{S_{\text{rest}}}(X_{\text{rest}} = \alpha_{\text{rest}}).$$

Since $D'_{S_{\text{rest}}}(X_{\text{rest}} = \alpha_{\text{rest}})$ depends only on α, the lemma follows. ◀

▶ Remark. When H does not have high expansion, $\mathsf{Cl}(S)$ is still defined for $S \subseteq [n]$ and Lemma 29 still holds. In this case, it is possible that $|\mathsf{Cl}(S)|$ can no longer be bounded in terms of $|X|$.

Using this lemma, we can prove that $Y_{X,\alpha}$ is PSD when $H - X$ has high enough expansion.

▶ **Lemma 30.** *Let $X \subseteq [n]$ such that $H - X$ is $(\omega(1), k - t/2 + \varepsilon)$-expanding for some constant $\varepsilon > 0$. Then for any $\alpha \in \{0,1\}^X$ with $\mu_X(\alpha) > 0$, $Y_{X,\alpha}$ is positive semidefinite.*

Proof. By Lemma 33, $Y_{X,\alpha}$ is PSD if and only if $Q_{X,\alpha} = B_{X,\alpha} - y_{X,\alpha} y_{X,\alpha}^\top$ is. Note that $Q_{X,\alpha}$ is a principle submatrix of the covariance matrix $\Sigma_{X,\alpha}$ of the $\{D_{S|X=\alpha}\}$ distributions, so it suffices to show that $\Sigma_{X,\alpha}$ is PSD. The conditional distributions $\{D_{S|X=\alpha}\}$ are r-locally consistent for $r = \Omega(n^{\frac{\delta}{t-2}})$ by Corollary 20. Then Lemma 22 implies that $\Sigma_{X,\alpha}$ is PSD if the correlation graph of the $\{D_{S|X=\alpha}\}$ distributions has connected components of size at most r. Lemma 29 implies that correlations under $\{D_{S|X=\alpha}\}$ induce bad structures in $H - X$, and we can apply Theorem 27 to $G_{\text{bad}}(H - X)$ to complete the proof. ◀

Finally, we show that any $Y_{X,\alpha}$ can be expressed as a nonnegative combination of $Y_{\mathsf{Cl}(X),\beta}$'s for $\beta \in \{0,1\}^{\mathsf{Cl}(X)}$.

▶ **Claim 31.**

$$Y_{X,\alpha} = \sum_{\substack{\beta \in \{0,1\}^{\mathsf{Cl}(X)} \\ \beta_X = \alpha}} \frac{D_{\mathsf{Cl}(X)}(\mathsf{Cl}(X) = \beta)}{D_X(X = \alpha)} \cdot Y_{\mathsf{Cl}(X),\beta}.$$

Proof. We will prove that

$$Y_{X,\alpha}(i,j) = \sum_{\substack{\beta \in \{0,1\}^{\mathsf{Cl}(X)} \\ \beta_X = \alpha}} \frac{D_{\mathsf{Cl}(X)}(\mathsf{Cl}(X) = \beta)}{D_X(X = \alpha)} \cdot Y_{\mathsf{Cl}(X),\beta}(i,j).$$

for all $0 \le i, j, \le n$.

Let $i, j \ge 1$. By applying definitions, we see that

$$\begin{aligned} Y_{X,\alpha}(i,j) &= D_{\{i,j\}|X=\alpha}(i = 1 \wedge j = 1) \\ &= \frac{D_{X \cup \{i,j\}}(i = 1 \wedge j = 1 \wedge X = \alpha)}{D_X(X = \alpha)}. \end{aligned}$$

Then, by local consistency, this expression is equal to

$$\frac{1}{D_X(X=\alpha)} \sum_{\substack{\beta\in\{0,1\}^{\mathsf{Cl}(X)}\\ \beta_X=\alpha}} D_{\mathsf{Cl}(X)\cup\{i,j\}}(i=1\wedge j=1\wedge \mathsf{Cl}(X)=\beta).$$

We can rewrite this as

$$\sum_{\substack{\beta\in\{0,1\}^{\mathsf{Cl}(X)}\\ \beta_X=\alpha}} \frac{D_{\mathsf{Cl}(X)}(\mathsf{Cl}(X)=\beta)}{D_X(X=\alpha)} \cdot \frac{D_{\mathsf{Cl}(X)\cup\{i,j\}}(i=1\wedge j=1\wedge \mathsf{Cl}(X)=\beta)}{D_{\mathsf{Cl}(X)}(\mathsf{Cl}(X)=\beta)}.$$

Using the definitions of conditional local distributions and $Y_{X,\alpha}(i,j)$ completes the proof for $i,j\geq 1$: the above expression is equal to

$$\sum_{\substack{\beta\in\{0,1\}^{\mathsf{Cl}(X)}\\ \beta_X=\alpha}} \frac{D_{\mathsf{Cl}(X)}(\mathsf{Cl}(X)=\beta)}{D_X(X=\alpha)} \cdot Y_{\mathsf{Cl}(X),\beta}(i,j).$$

When i or j is equal to 0, an essentially identical argument can be used. ◀

Lemma 28 follows immediately from Lemma 30 and Claim 31.

7.2 Rank lower bounds for static LS_+ and LS_+

We now use the results of the previous section to prove Theorem 5, a lower bound on the degree of any static LS_+ refutation. The proof is essentially the same as that of [30, Theorem 3.27], which is the special case of P being pairwise uniform. This will immediately imply Theorem 4, the rank lower bound for LS_+.

Proof of Theorem 5. Assume that the rank of any standard, non-static LS_+ proof is at least r; by Theorem 4, $r=\Omega(n^{\frac{\varepsilon}{t-2}})$. Assume for a contradiction that there exists a static LS_+ refutation of degree $r-k$. Recall that a static LS_+ refutation has the form

$$\sum_{\ell=1}^{r} w_\ell q_\ell(x)\phi_{I_\ell,J_\ell}(x) = -1, \tag{7.10}$$

where $w_\ell\geq 0$, each $q_\ell(x)$ is either an axiom or the square of a linear form, and

$$\phi_{I,J}(x)=\prod_{i\in I} x_i \prod_{j\in J}(1-x_j).$$

By Theorem 18, we know that there exist r-consistent local distributions $\{D_S\}$; let $\widetilde{\mathbb{E}}[\cdot]$ be the corresponding rank-r SA pseudoexpectation. In addition, Corollary 20 states that there also exist conditional $(r-|X|)$-consistent local distributions $\{D_{S|X=\alpha}\}$ for any α such that $\mu(\alpha_C)>0$ for all $C\in\mathcal{C}(X)$,

We will derive a contradiction by applying the operator $\widetilde{\mathbb{E}}[\cdot]$ to both sides of (7.10). Specifically, we will show that if the degree of each term is at most $r-k$, then $\widetilde{\mathbb{E}}[q_\ell(x)\phi_{I_\ell,J_\ell}(x)]\geq 0$. Applying $\widetilde{\mathbb{E}}[\cdot]$ to the left hand side of (7.10) gives a value at least 0, while the right hand side is -1. To show that $\widetilde{\mathbb{E}}[q_\ell(x)\phi_{I_\ell,J_\ell}(x)]\geq 0$, we consider two cases.

Case 1. q_ℓ is an axiom. If q_ℓ is an axiom, then the number of variables in the expression $q_\ell(x)\phi_{I_\ell,J_\ell}(x)$ is $|I_\ell\cup J_\ell\cup \mathrm{supp}(q_\ell)|\leq r-k+k=r$. We also know that $q_\ell(x)\phi_{I_\ell,J_\ell}(x)\geq 0$ for all $x\in\{0,1\}^n$. The definition of rank-r SA pseudodistributions then implies that $\widetilde{\mathbb{E}}[q_\ell(x)\phi_{I_\ell,J_\ell}(x)]\geq 0$.

Case 2. q_ℓ is the square of a linear form. In this case, the result follows almost immediately from Lemma 28. Write $q_\ell(x)$ as follows:

$$q_\ell(x) = \left(a_0 + \sum_{i\in[n]} a_i x_i\right)^2 = \sum_{i,j\in[n]} a_i a_j x_i x_j + 2a_0 \sum_{i\in[n]} a_i x_i + a_0^2.$$

Also, let $A_\ell = I_\ell \cup J_\ell$ and define $\beta \in \{0,1\}^{|A_\ell|}$ so that $\phi_{I_\ell,J_\ell}(x) = 1$ if and only if $x_{A_\ell} = \beta$. Then we have the following calculation:

$$\begin{aligned}
\widetilde{\mathbb{E}}[q_\ell(x)\phi_{I_\ell,J_\ell}(x)] &= \widetilde{\mathbb{E}}[q_\ell(x)\cdot \mathbf{1}(A_\ell = \beta)] \\
&= \sum_{i,j\in[n]} a_i a_j \widetilde{\mathbb{E}}[x_i x_j \cdot \mathbf{1}(A_\ell = \beta)] + 2a_0 \sum_{i\in[n]} a_i \widetilde{\mathbb{E}}[x_i \cdot \mathbf{1}(A_\ell = \beta)] \\
&\qquad + a_0^2 \widetilde{\mathbb{E}}[\mathbf{1}(A_\ell = \beta)] \\
&= \sum_{i,j\in[n]} a_i a_j D_{A_\ell \cup \{i,j\}}(A_\ell = \beta) Y_{A_\ell,\beta}(i,j) \\
&\qquad + 2a_0 \sum_{i\in[n]} a_i D_{A_\ell\cup\{i,j\}}(A_\ell = \beta) Y_{A_\ell,\beta}(i,0) + a_0^2 D_{A_\ell}(A_\ell = \beta) \\
&= D_{A_\ell}(A_\ell = \beta) a^\top (Y_{A_\ell,\beta}) a \quad \text{by } r\text{-local consistency of } \{D_S\} \\
&\geq 0 \qquad \text{by Lemma 28.}
\end{aligned}$$

◀

Theorem 4, our rank lower bound for LS_+ refutations, follows immediately from Theorem 5 and the following fact.

▶ **Fact 32.** *If there exists a rank-r LS_+ refutation of a set of axioms A, then there exists a static LS_+ refutation of A with rank at most r.*

Proof. Let R be a rank-r LS_+ refutation. We look at R as a DAG in which each node is the application of some inference rule, the root is $-1 \geq 0$, and the leaves are axioms or applications of the rule $P(x)^2 \geq 0$ for P with degree at most 1. Starting from the root $-1 \geq 0$ and working back to the axioms, we can substitute in the premises of each inference to get an expression $Q(x) = -1$. Since R has rank r, each path in r has at most r multiplications by a term of the form x_i or $(1 - x_i)$ and $Q(x) = -1$ must be a valid static LS_+ refutation of rank at most r. ◀

Acknowledgments. The first-named author would like to thank Osamu Watanabe for his encouragement. The second-named author would like to thank Anupam Gupta and Ryan O'Donnell for several helpful discussions.

References

1. Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 689–708, 2015.
2. Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 171–180, 2010.
3. Boaz Barak. Lecture 1 – Introduction. Notes from course "Sum of Squares upper bounds, lower bounds, and open questions".

4 Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, Sum-of-Squares Proofs, and their Applications. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pages 307–326, 2012.

5 Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 97–106, 2015.

6 Boaz Barak, Guy Kindler, and David Steurer. On the Optimality of Semidefinite Relaxations for Average-Case and Generalized Constraint Satisfaction. In *Proceedings of the 4th Innovations in Theoretical Computer Science conference*, 2013.

7 Boaz Barak and Ankur Moitra. Tensor Prediction, Rademacher Complexity and Random 3-XOR. *CoRR*, abs/1501.06521, 2015. URL: `http://arxiv.org/abs/1501.06521`.

8 Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 9(3), 2002.

9 Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8:269–289, 2012.

10 Joshua Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory of Computing*, 2:65–90, 2006. `doi:10.4086/toc.2006.v002a004`.

11 Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 350–359, 2013.

12 Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong Refutation Heuristics for Random k-SAT. In Klaus Jansen, Sanjeev Khanna, José D.P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 3122 of *Lecture Notes in Computer Science*, pages 310–321. Springer Berlin Heidelberg, 2004. `doi:10.1007/978-3-540-27821-4_28`.

13 A Crisanti, L Leuzzi, and G Parisi. The 3-SAT problem with large number of clauses in the ∞-replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002.

14 Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 441–448, 2014.

15 Sanjeeb Dash. *On the Matrix Cuts of Lovász and Schrijver and their use in Integer Programming*. PhD thesis, Rice University, 2001.

16 Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 59–68, 2015.

17 Uriel Feige. Relations Between Average Case Complexity and Approximation Complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002.

18 Uriel Feige and Eran Ofek. Easily refutable subformulas of large random 3CNF formulas. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming*, volume 3142 of *Lecture Notes in Comput. Sci.*, pages 519–530. Springer, Berlin, 2004.

19 Vitaly Feldman, Will Perkins, and Santosh Vempala. On the Complexity of Random Satisfiability Problems with Planted Solutions. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 77–86, 2015.

20 Joel Friedman, Andreas Goerdt, and Michael Krivelevich. Recognizing more unsatisfiable random k-SAT instances efficiently. *SIAM J. Comput.*, 35(2):408–430, 2005. `doi:10.1137/S009753970444096X`.

21 Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.

22 Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Complexity of semi-algebraic proofs. In *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, pages 419–430, 2002.

23 Arist Kojevnikov and Dmitry Itsykson. Lower Bounds of Static Lovász-Schrijver Calculus Proofs for Tseitin Tautologies. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, 2006.

24 James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 567–576, 2015.

25 László Lovász and Alexander Schrijver. Cones of Matrices and Set-Functions and 0-1 Optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

26 Ryan O'Donnell and David Witmer. Goldreich's PRG: Evidence for near-optimal polynomial stretch. In *Proceedings of the 29th Annual Conference on Computational Complexity*, pages 1–12, 2014.

27 Prasad Raghavendra. Optimal Algorithms and Inapproximability Results for Every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 245–254, 2008.

28 Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602, 2008.

29 Hanif Sherali and Warren Adams. A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

30 Madhur Tulsiani and Pratik Worah. LS_+ lower bounds from pairwise independence. In *Proceedings of the 28th Annual Conference on Computational Complexity*, pages 121–132, 2013.

31 Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, volume 1. Now Publishers Inc., Hanover, MA, USA, January 2008. doi:10.1561/2200000001.

A Proofs from Section 2.4

▶ **Fact 15.** *$(s, k-d)$-expansion implies $(s, k-2d)$-boundary expansion.*

Proof. Let S be a set of at most s hyperedges. Each of the vertices in $\Gamma(S)$ is either a boundary vertex that appears in exactly one hyperedge or it appears in two or more hyperedges, so $|\Gamma(S)| \leq |\partial S| + \frac{1}{2}(|k|S| - |\partial S|)$. Therefore, we can write

$$|\partial S| \geq 2|\Gamma(S)| - k|S| \geq (k-2d)|S|,$$

where the second inequality follows the expansion assumption. ◀

▶ **Lemma 16.** *Fix $\delta > 0$. With high probability, a set of $m \leq \Omega(n^{t/2-\delta})$ constraints chosen uniformly at random is both $\left(n^{\frac{\delta}{t-2}}, k - \frac{t}{2} + \frac{\delta}{2}\right)$-expanding and $\left(n^{\frac{\delta}{t-2}}, k - t + \delta\right)$-boundary expanding.*

Proof. By Fact 15, it suffices to show that a random instance is $\left(n^{\frac{\delta}{t-2}}, k - \frac{t}{2} + \frac{\delta}{2}\right)$-expanding. We give the proof of [26], which is essentially the same as that of [9].

We want to upper bound the probability that any set of r hyperdges with $r \leq n^{\frac{\delta}{t-2}}$ contains less than $r(k-\frac{t}{2}+\frac{\delta}{2})$ vertices. Fix an r-tuple of edges T; this is a tuple of indices in $[m]$ representing the indices of the hyperedges in T. We wish to upper bound $\Pr[|\Gamma(T)| \leq v]$; we can do this with the quantity

$$\frac{(\#\text{ sets } S \text{ of } v \text{ vertices}) \cdot (\#\text{ sets of } r \text{ edges contained in } S)}{(\#\text{ of ways of choosing } r \text{ edges})}.$$

Taking a union bound over all tuples of size r, we see that

$$\Pr[|\Gamma(S)| \leq v \;\forall S \text{ s.t. } |S| = r] \leq r!\binom{m}{r} \cdot \frac{\binom{n}{v}\binom{k!\binom{v}{k}}{r}}{(k!\binom{n}{k})^r}.$$

Simplifying and applying standard approximations, we get that

$$\Pr[|\Gamma(S)| \leq v \;\forall S \text{ s.t. } |S| = r] \leq e^{(2+k)r+v} v^{kr-v} r^{-r} n^{v-kr} m^r.$$

Set $v = \lfloor r(k-\frac{t}{2}+\frac{\delta}{2}) \rfloor$ and simplify to get

$$\Pr[|\Gamma(S)| < r(k-\frac{t}{2}+\frac{\delta}{2}) \;\forall S \text{ s.t. } |S| = r] \leq (C(k,t) m n^{-(t/2-\delta/2)} r^{t/2-1-\delta/2})^r$$

Then set $m = n^{t/2-\delta}$ and take a union bound over all choices of r to get that

$$\begin{aligned}
&\Pr[H_{\mathcal{I}} \text{ not } (n^{\frac{\delta}{t-2}}, k-\frac{t}{2}+\frac{\delta}{2})\text{-expanding}] \leq \sum_{r=1}^{\lfloor n^{\delta/(t-2)} \rfloor} (C(k,t) n^{-\delta/2} r^{t/2-1-\delta/2})^r \\
&= \sum_{r=1}^{\lceil \log n \rceil} (C(k,t) n^{-\delta/2} r^{t/2-1-\delta/2})^r + \sum_{r=\lceil \log n \rceil+1}^{\lfloor n^{\delta/(t-2)} \rfloor} (C(k,t) n^{-\delta/2} r^{t/2-1-\delta/2})^r \\
&\leq 2C(k,t) n^{-\delta/2} (\log n)^{t/2-\delta/2} + n^{\frac{\delta}{t-2}} (C(k,t) n^{-\delta/2} (n^{\frac{\delta}{t-2}})^{t/2-1-\delta/2})^{\log n} \\
&= O(n^{-\delta/3}).
\end{aligned}$$

◀

B Equivalence between PSDness of the degree-2 moment matrix and the covariance matrix

▶ **Lemma 33.**

$$\begin{pmatrix} 1 & w^T \\ w & B \end{pmatrix} \text{ is PSD } \iff B - ww^T \text{ is PSD.}$$

Proof.

$$\begin{aligned}
\begin{pmatrix} 1 & w^T \\ w & B \end{pmatrix} \text{ is PSD } &\iff \left((v_0\, v) \begin{pmatrix} 1 & w^T \\ w & B \end{pmatrix} (v_0\, v)^T \geq 0 \;\forall v_0 \in \mathbb{R},\, v \in \mathbb{R}^{nq} \right) \\
&\iff \left(v_0^2 + 2\langle w, v \rangle v_0 + \langle Bv, v \rangle \geq 0 \;\forall v_0 \in \mathbb{R},\, v \in \mathbb{R}^{nq} \right) \\
&\iff \left((v_0 + \langle w, v \rangle)^2 - \langle w, v \rangle^2 + \langle Bv, v \rangle \geq 0 \;\forall v_0 \in \mathbb{R},\, v \in \mathbb{R}^{nq} \right) \\
&\iff \left(-\langle w, v \rangle^2 + \langle Bv, v \rangle \geq 0 \;\forall v \in \mathbb{R}^{nq} \right) \\
&\iff \left(v(B - ww^T)v^T \geq 0 \;\forall v \in \mathbb{R}^{nq} \right) \\
&\iff B - ww^T \text{ is PSD.}
\end{aligned}$$

◀

▶ **Lemma 10.** *M is PSD if and only if Σ is PSD.*

Proof. We rewrite M as

$$\begin{pmatrix} 1 & w^T \\ w & B \end{pmatrix},$$

where w is a vector whose (i,a)-element is $p_{\{i\}}(x_i = a)$ for $i \in [n]$ and $a \in [q]$, where B is a matrix whose $((i,a),(j,b))$-element is $p_{\{i,j\}}(x_i = a \wedge x_j = b)$ for $i, j \in [n]$ and $a, b \in [q]$. From Lemma 33, we know that

$$\begin{pmatrix} 1 & w^T \\ w & B \end{pmatrix} \text{ is PSD if and only if } B - ww^T \text{ is PSD.}$$

Observe that $B - ww^T$ is equal to the covariance matrix Σ. ◀

C Proofs from Section 2.5

▶ **Lemma 17.** *If $H_\mathcal{I}$ is (s_1, e_1)-expanding and S is a set of variables such that $|S| < (e_1 - e_2)s_1$ for some $e_2 \in (0, e_1)$, then there exists a set $\mathsf{Cl}(S) \subseteq [n]$ such that $S \subseteq \mathsf{Cl}(S)$ and $H_\mathcal{I} - \mathsf{Cl}(S)$ is (s_2, e_2)-expanding with $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$ and $\mathsf{Cl}(S) \leq \frac{e_1}{e_1 - e_2}|S|$.*

Proof. We calculate $\mathsf{Cl}(S)$ using the closure algorithm of [9, 30]:

Input: An (s_1, e_1)-expanding instance $\mathcal{I}$, $e_2 \in (0, e_1)$, a tuple $S = (x_1, \ldots, x_u) \in [n]^u$ such that $u < (e_1 - e_2)s_2$.
Output: The closure $\mathsf{Cl}(S)$.

Set $\mathsf{Cl}(S) \leftarrow \emptyset$ and $s_2 \leftarrow s_1$.
for $i = 1, \ldots, u$
    $\mathsf{Cl}(S) \leftarrow \mathsf{Cl}(S) \cup \{x_i\}$
    if $H_\mathcal{I} - \mathsf{Cl}(S)$ is not (s_2, e_2)-expanding, **then**
        Find largest set of constraints M_i in $H_\mathcal{I} - \mathsf{Cl}(S)$ such that
        $|M_i| \leq s_2$ and $|\Gamma(M_i)| \leq e_2|M_i|$. Break ties by lexicographic order.
        $\mathsf{Cl}(S) \leftarrow \mathsf{Cl}(S) \cup \Gamma(M_i)$
        $s_2 \leftarrow s_2 - |M_i|$
return $\mathsf{Cl}(S)$

It is clear from the statement of the algorithm that $S \subseteq \mathsf{Cl}(S)$. We need to show that $H_\mathcal{I} - \mathsf{Cl}(S)$ is (s_2, e_2)-expanding, that $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$, and that $\mathsf{Cl}(S) \leq \frac{e_1}{e_1 - e_2}|S|$. We give the proof of [9].

1. $H_\mathcal{I} - \mathsf{Cl}(S)$ is (s_2, e_2)-expanding
   We will show that at every step of the algorithm $H_\mathcal{I} - \mathsf{Cl}(S)$ is (s_2, e_2)-expanding. Say we are in step i and that $H_\mathcal{I} - (\mathsf{Cl}(S) \cup \{x_i\})$ is not (s_2, e_2)-expanding; if it were (s_2, e_2)-expanding, we would be done. Let M_i be the largest set of hyperedges in $H_\mathcal{I} - \mathsf{Cl}(S)$ such that $|M_i| \leq s_2$ and $|\Gamma(M_i)| \leq e_2|M_i|$. We need to show that $H_\mathcal{I} - (\mathsf{Cl}(S) \cup \{x_i\} \cup \Gamma(M_i))$ is $(\zeta - |M_i|, e_2)$-expanding.
   To see this, assume for a contradiction that there exists a set of hyperedges M' in $H_\mathcal{I} - (\mathsf{Cl}(S) \cup \{x_i\} \cup \Gamma(M_i))$ such that $M' \leq s_2 - |M_i|$ and $|\Gamma(M')| < e_2|M'|$. Consider $M_i \cup M'$. Note that M_i and M' are disjoint, so $|M_i \cup M'| \leq s_2$. Also, $|\Gamma(M_i \cup M')| \leq e_2|M_i| + e_2|M'| = e_2|M_i \cup M'|$, contradicting the maximality of M_i.
2. $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$
   Consider the set $M = \bigcup_{i=1}^{u} M_u$. First, note that $|M| = s_1 - s_2$, so $|\Gamma(M)| \geq e_1(s_1 - s_2)$

by expansion of $H_{\mathcal{I}}$. Second, each element of $\Gamma(M) - S$ occurs in exactly one of the M_i's and each M_i has expansion at most e_2. Using these two observations, we see that

$$e_1(s_1 - s_2) \le |\Gamma(M)| \le |S| + \sum_{i=1}^{u} e_2|M_i| = |S| + e_2(s_1 - s_2).$$

This implies the claim.

3. $\mathsf{Cl}(S) \le \frac{e_1}{e_1 - e_2}|S|$
Observe that $\mathsf{Cl}(S) = S \cup \bigcup_{i=1}^{u} \Gamma(M_i)$. Also, every M_i has expansion at most e_2. Therefore, we have that

$$\begin{aligned} |\mathsf{Cl}(S)| &\le |S| + \sum_{i=1}^{u} |\Gamma(M_i)| \\ &\le |S| + e_2 \sum_{i=1}^{u} |M_i| \\ &\le |S| + \frac{e_2|S|}{e_1 - e_2} \\ &= \left(\frac{e_1}{e_1 - e_2}\right)|S|, \end{aligned}$$

where we used that $\sum_{i=1}^{u} |M_i| = s_1 - s_2$ and $s_2 \ge s_1 - \frac{|S|}{e_1 - e_2}$. ◀

▶ **Theorem 18.** *For a random instance $\mathcal{I}$ with $m \le \Omega(n^{t/2-\varepsilon})$, the family of distributions $\{D_S\}_{|S| \le r}$ is r-locally consistent for $r = n^{\frac{\varepsilon}{t-2}}$ and is supported on satisfying assignments.*

To prove the theorem, we will use the following lemma, which says that the local distributions on D'_S and D'_T with $S \subseteq T$ are consistent if $H_{\mathcal{I}} - T$ has high boundary expansion.

▶ **Lemma 34.** *Let P be a $(t-1)$-wise uniform supporting predicate, let $\mathcal{I}$ be an instance of* CSP(P)*, and let $S \subseteq T$ be sets of variables. If $H_{\mathcal{I}}$ and $H_{\mathcal{I}} - S$ are $(r, k - t + \varepsilon)$-boundary expanding for some $\varepsilon > 0$ and $\mathcal{C}(T) \le r$, then for any $\alpha \in [q]^S$,*

$$D'_S(\alpha) = \sum_{\substack{\beta \in [q]^T \\ \beta_S = \alpha}} D'_T(\beta).$$

First, we will use this lemma to prove Theorem 18.

Proof of Theorem 18. Let $S \subseteq T$ be sets of variables with $|T| \le r$. Consider $U = \mathsf{Cl}(S) \cup \mathsf{Cl}(T)$. We will show that both D_S and D_T are consistent with U and therefore must themselves be consistent. Observe that $|\mathsf{Cl}(S)|$ and $|\mathsf{Cl}(T)|$ are at most $\frac{2kr}{\varepsilon}$, so $|U| \le \frac{4kr}{\varepsilon}$. Towards applying Lemma 34, we will first show that $|\mathcal{C}(U)| \le \frac{8r}{\varepsilon}$. Assume for a contradiction that C is a subset of $\mathcal{C}(U)$ of size $\frac{8r}{\varepsilon}$. Then

$$\frac{|\Gamma(C)|}{|C|} \le \frac{|U|}{|C|} = \frac{4kr/\varepsilon}{8r/\varepsilon} = \frac{k}{2} < k - \frac{t}{2} + \delta,$$

which violates expansion.

We know that $H_{\mathcal{I}} - \mathsf{Cl}(T)$ and $H_{\mathcal{I}} - \mathsf{Cl}(S)$ are $(r, k-t+\varepsilon)$-boundary expanding for some $\varepsilon > 0$. We can then apply Lemma 34 twice with sets $\mathsf{Cl}(S) \subseteq U$ and $\mathsf{Cl}(T) \subseteq U$ to see that

$$\begin{aligned} D_S(\alpha) = \sum_{\substack{\beta \in [q]^{\mathsf{Cl}(S)} \\ \beta_S = \alpha}} D'_{\mathsf{Cl}(S)}(\beta) &= \sum_{\substack{\gamma \in [q]^U \\ \gamma_S = \alpha}} D'_U(\gamma) \\ &= \sum_{\substack{\beta' \in [q]^{\mathsf{Cl}(T)} \\ \beta'_S = \alpha}} D'_{\mathsf{Cl}(T)}(\beta') = \sum_{\substack{\alpha' \in [q]^T \\ \alpha'_S = \alpha}} D_T(\alpha'). \end{aligned}$$ ◀

Now we prove Lemma 34.

Proof of Lemma 34. We follow the proof of Benabbas et al. [9]. Let $\mathcal{C}(T) \setminus \mathcal{C}(S) = \{C_1, \ldots, C_u\}$ and, for a constraint C, let $\sigma(C)$ be the variables in the support of C. First, observe that

$$\begin{aligned} Z'_T \sum_{\substack{\beta \in [q]^T \\ \beta_S = \alpha}} D'_T(\beta) &= \sum_{\gamma \in [q]^{T \setminus S}} \prod_{C \in \mathcal{C}(T)} \mu_C((\alpha, \gamma)) \\ &= \left(\prod_{C \in \mathcal{C}(S)} \mu_C(\alpha) \right) \sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^{u} \mu_{C_i}((\alpha, \gamma)) \\ &= (Z'_S D'_S(\alpha)) \sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^{u} \mu_{C_i}((\alpha, \gamma)) \end{aligned}$$

To finish the proof, we will need the following claim.

▶ **Claim 35.** *There exists an ordering $(C_{i_1}, \ldots, C_{i_u})$ of constraints of $\mathcal{C}(T) \setminus \mathcal{C}(S)$ and a partition $V_1, \cdots, V_u, V_{u+1}$ of variables of $T \setminus S$ such that for all $j \leq u$ the following hold.*
1. *$V_j \subseteq \sigma(C_{i_j})$.*
2. *$|V_j| \geq k - t + 1$*
3. *V_j does not intersect $\sigma(C_{i_l})$ for any $l > j$. That is, $V_j \cap \bigcup_{l > j} \sigma(C_{i_l}) = \emptyset$.*

Proof of Claim 35. We will find the sets V_j by repeatedly using $(r, k-t+\delta)$-boundary expansion of $H_{\mathcal{I}} - S$. Let $Q_1 = \mathcal{C}(T) \setminus \mathcal{C}(S)$. We know that $|Q_1| \leq r$, so boundary expansion of $H_{\mathcal{I}} - S$ implies that $|\partial(Q_1) \setminus S| \geq (k-t+\delta)|Q_1|$. There must exist a constraint $C_j \in Q_1$ with at least $k-t+1$ boundary variables in $H_{\mathcal{I}} - S$; i.e., $|\sigma(C_j) \cap (\partial(Q_1) \setminus S)| \geq k-2$. We then set $V_1 = \sigma(C_j) \cap (\partial(Q_1) \setminus S)$ and $i_1 = j$. Let $Q_2 = Q_1 \setminus C_j$. We apply the same process $u-1$ more times until Q_l is empty and then set $V_{u+1} = (T \setminus S) \setminus (\bigcup_{j=1}^{u} V_j)$. We remove constraint C_{i_l} at every step and $F_l \subseteq \sigma(C_{i_l})$, so it holds that $V_j \cap \bigcup_{l > j} \sigma(C_{i_l}) = \emptyset$. ◀

Using the claim, we can write $\sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^{u} \mu_{C_i}((\alpha, \gamma))$ as

$$\sum_{\gamma_{u+1} \in [q]^{V_{u+1}}} \sum_{\gamma_u \in [q]^{V_u}} \mu_{C_u}(\gamma'_u) \sum_{\gamma_{u-1} \in [q]^{V_{u-1}}} \mu_{C_{u-1}}(\gamma'_{u-1}) \cdots \sum_{\gamma_1 \in [q]^{V_1}} \mu_{C_1}(\gamma'_1),$$

where each γ'_j depends on α and γ_l with $l \geq j$ but does not depend on γ_l with $l < j$. We will evaluate this sum from right to left. We know that each V_j contains at least $k-t+1$ elements, so $(t-1)$-wise uniformity of μ implies that $\sum_{\gamma_j \in [q]^{V_j}} \mu_{C_j}(\gamma'_j) = q^{-(k-|V_j|)}$. Applying this repeatedly, we see that

$$\sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^{u} \mu_{C_i}((\alpha, \gamma)) = q^{-(ku - \sum_{j=1}^{u+1} |V_j|)} = q^{|T \setminus S| - k|\mathcal{C}(T) \setminus \mathcal{C}(S)|}.$$

Plugging this quantity into the above calculation, we obtain

$$Z'_T \sum_{\substack{\beta \in [q]^T \\ \beta|_S = \alpha}} D'_T(\beta) = Z'_S D'_S(\alpha) q^{|T \setminus S| - k|\mathcal{C}(T) \setminus \mathcal{C}(S)|}.$$

Since $H_\mathcal{I}$ has $(r, k - t + \delta)$-boundary expansion for some $\delta > 0$, we can set $S = \emptyset$ to get that $Z'_T = q^{|T| - k|\mathcal{C}(T)|}$. Similarly, $Z'_S = q^{|S| - k|\mathcal{C}(S)|}$. Plugging these two quantities in completes the proof. ◀

▶ **Lemma 19.** *Let $X \subseteq [n]$ and let $\{D_S\}$ be a family of r-locally consistent distributions for sets $S \subseteq [n]$ such that $S \cap X = \emptyset$ and $|S \cup X| \leq r$. Then the family of conditional distributions $\{D_S(\cdot|X = \alpha)\}$ is $(r - |X|)$-locally consistent for any $\alpha \in \{0,1\}^X$ such that $\mu(\alpha|_C) > 0$ for all constraints in $\mathcal{C}(X)$.*

Proof. Tulsiani and Worah proved this lemma and we will use their proof [30]. Let $S \subseteq T$ and $|T \cup X| \leq r$. Let β be any assignment to S. Then local consistency of the $\{D_S\}$ measures implies that $D_{S \cup X}(S = \beta \wedge X = \alpha) = D_{T \cup X}(S = \beta \wedge X = \alpha)$ and $D_{S \cup X}(X = \alpha) = D_{T \cup X}(X = \alpha)$. We therefore have that

$$\begin{aligned} D_{S|X=\alpha}(S = \beta) &= \frac{D_{S \cup X}(S = \beta \wedge X = \alpha)}{D_{S \cup X}(X = \alpha)} \\ &= \frac{D_{T \cup X}(S = \beta \wedge X = \alpha)}{D_{T \cup X}(X = \alpha)} = D_{T|X=\alpha}(S = \beta). \end{aligned}$$ ◀

D Equivalence of SA, SA$_+$, and static LS$_+$ tightenings of linear and degree-k relaxations of CSP(P)

▶ **Lemma 12.** *Let $r \geq k$. Then the following statements hold.*
1. $\mathrm{SA}^r(R_\mathcal{I}) = \mathrm{SA}^r(L_\mathcal{I})$.
2. $\mathrm{SA}^r_+(R_\mathcal{I}) = \mathrm{SA}^r_+(L_\mathcal{I})$.
3. $\mathrm{StaticLS}^r_+(R_\mathcal{I}) = \mathrm{StaticLS}^r_+(L_\mathcal{I})$

Proof. The proof is the same for SA, SA$_+$, and static LS$_+$. We use the notation introduced in Section 2.2.1. For $f \in \{0,1\}^k$ and $z \in [0,1]^k$, let $P'_f(z) = \sum_{i=1}^k z^{(f_i)}$. Let $(c, S) \in \mathcal{I}$ be any constraint. Let $\widetilde{\mathbb{E}}[\cdot]$ be any r-round SA pseudoexpectation. We begin by making a couple of observations. Let q be an arbitrary multilinear polynomial satisfying the following conditions.

1. $q(x) \geq 0$ for all $x \in \{0,1\}^n$.
2. $q(x) \cdot (P'(x_S^{(c)}) - 1)$ depends on at most r variables. Equivalently, $q(x) \cdot (P'_f(x_S^{(c)}) - 1)$ depends on at most r variables for all $f \in F$.

First, note that

$$P'(x_S^{(c)}) - 1 = \sum_{f \in F} 1_{\{x_S = f\}}(x) \cdot (P'_f(x_S^{(c)}) - 1). \tag{D.11}$$

This implies that

$$\widetilde{\mathbb{E}}[p(x) \cdot (P'(x_S^{(c)}) - 1)] = \sum_{f \in F} \widetilde{\mathbb{E}}[p(x) \cdot 1_{\{x_S = f\}}(x) \cdot (P'_f(x_S^{(c)}) - 1)]. \tag{D.12}$$

Second, we see that $-q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)\geq 0$ for all $x\in\{0,1\}^n$ for all $(c,S)\in\mathcal{I}$, and for all $f\in F$. Since Condition 2 implies that $-q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)$ depends on at most r variables and $\widetilde{\mathbb{E}}$ is a degree-r SA pseudexpectation,

$$\widetilde{\mathbb{E}}[q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)]\leq 0. \tag{D.13}$$

Now assume that $\widetilde{\mathbb{E}}[\cdot]$ satisfies

$$\widetilde{\mathbb{E}}[p(x)\cdot(P'_f(x_S^{(c)})-1)]\geq 0 \tag{D.14}$$

for all $f\in F$ and for all multilinear polynomials p satisfying conditions 1 and 2. We want to show that $\widetilde{\mathbb{E}}[q(x)\cdot(P'(x_S^{(c)})-1)]=0$ for all multilinear q satisfying conditions 1 and 2. Since $q(x)\cdot 1_{\{x_S=f\}}(x)$ is nonnegative and $q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)$ depends on at most r variables, (D.14) implies that $\widetilde{\mathbb{E}}[q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)]\geq 0$. Together with (D.13), this implies that $\widetilde{\mathbb{E}}[q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)]=0$ and the result follows from (D.12).

For the other direction, assume that $\widetilde{\mathbb{E}}[\cdot]$ satisfies $\widetilde{\mathbb{E}}[p(x)\cdot(P'(x_S^{(c)})-1)]=0$ for all multilinear polynomials $p(x)$ satisfying conditions 1 and 2. Let q be an arbitrary multilinear polynomial satisfying conditions 1 and 2. From (D.11), we see that

$$\sum_{f\in F}\widetilde{\mathbb{E}}[q(x)\cdot 1_{\{x_S=f\}}(x)\cdot(P'_f(x_S^{(c)})-1)]=0.$$

The result then follows from (D.13). ◀

E Correspondence between static LS$_+$roof system and relaxation

Recall the static LS$_+$ relaxation.

$$\begin{aligned}&\sum_{\alpha\in\{0,1\}^S}p_S(\alpha)P(\alpha+c)\geq 1 \text{ for all } (c,S)\in\mathcal{I}\\&\{p_S\}_{S\subseteq[n],\,|S|\leq r} \text{ are } r\text{-locally consistent distributions}\\&\Sigma_{T=\alpha} \text{ is PSD } \forall T\subseteq[n],\alpha\in[q]^T \text{ such that } |T|\leq r-2, p_T(\alpha)>0.\end{aligned} \tag{E.15}$$

A static LS$_+$ refutation has the following form.

$$\sum_{\ell}\gamma_\ell b_\ell(x)\phi_{I_\ell,J_\ell}(x)=-1, \tag{E.16}$$

where $\gamma_\ell\geq 0$, b_ℓ is an axiom or the square of an affine function, and $\phi_{I_\ell,J_\ell}=\prod_{i\in I_\ell}x_i\prod_{j\in J_\ell}(1-x_j)$.

▶ **Proposition 36.** *The static* LS$_+$ *SDP* (E.15) *is infeasible if and only if a static* LS$_+$ *refutation of the form* (E.16) *exists.*

Proof. Recall the definition of SA$_r$. We require that

$$\widetilde{\mathbb{E}}\left[\prod_{i\in I}x_i\prod_{j\in J}(1-x_j)\right]\geq 0 \tag{E.17}$$

for all $I,J\subseteq[n]$ such that $|I\cup J|\leq r$ and

$$\widetilde{\mathbb{E}}\left[a(x)\prod_{i\in I}x_i\prod_{j\in J}(1-x_j)\right]\geq 0 \tag{E.18}$$

for every axiom $a(x) \geq 0$ and for all $I, J \subseteq [n]$ such that $|I \cup J| \leq r$.

Using linearity of $\widetilde{\mathbb{E}}[\cdot]$, we can write this as a linear program in the variables $X_{I,J} := \prod_{i\in I} x_i \prod_{j\in J}(1-x_j)$. Given a set $T \subseteq [n]$ and some assignment $\alpha : T \to \{0,1\}$, define α_0 to be $\{i \in T \ : \ \alpha(i) = 0\}$ and α_1 to be $\{i \in T \ : \ \alpha(i) = 1\}$. In (E.15), we additionally require that the matrices

$$M^{T,\alpha} = \left(\widetilde{\mathbb{E}}\left[x_i x_j \prod_{a\in\alpha_1} x_a \prod_{b\in\alpha_0}(1-x_b)\right]\right)_{i\in[n],j\in[n]} = (X_{\alpha_1\cup\{i,j\},\alpha_0})_{i\in[n],j\in[n]} \tag{E.19}$$

are PSD for all $T \subseteq [n]$ such that $|T| \leq r$ and all $\alpha \in \{0,1\}^T$. As mentioned above, we can arrange the matrices $M^{T,\alpha}$ into a block diagonal matrix M such that M is PSD if and only if each of the $M^{T,\alpha}$'s are PSD. Let d be the dimension of M. Furthermore, we can think of the r-round SA constraints as being linear constraints on the entries of M. In particular, say these constraints have the form $A \cdot \mathrm{vec}(M) \geq b$, where $\mathrm{vec}(M) \in \mathbb{R}^{d^2}$ is the vector formed by concatenating the columns of M. Let c be the number of rows of A.

First, we show that the existence of a refutation of the form (E.16) implies that (E.15) is infeasible. Assume for a contradiction that there exists a solution $\{p_S\}_{S\subseteq[n],\,|S|\leq r}$ to (E.15). This implies the existence of a pseudoexpectation operator $\widetilde{\mathbb{E}}[\cdot]$ satisfying (E.17), (E.18), and (E.19). Now apply $\widetilde{\mathbb{E}}[\cdot]$ to each term of (E.16). The degree of each term $\gamma_\ell b_\ell(x)\phi_{I_\ell,J_\ell}(x)$ is at most r and we have that

$$\widetilde{\mathbb{E}}[\gamma_\ell b_\ell(x)\phi_{I_\ell,J_\ell}(x)] = \gamma_\ell \widetilde{\mathbb{E}}[b_\ell(x) \cdot \mathbf{1}_{\{x=\alpha\}}(x)],$$

where α is the unique assignment to $I_\ell \cup J_\ell$ such that $\phi_{I_\ell,J_\ell} = 1$. Let $U = \mathrm{supp}(b_\ell) \cup I_\ell \cup J_\ell$. If $b_\ell(x) \geq 0$ is an axiom, we know that $\widetilde{\mathbb{E}}[b_\ell(x) \cdot \mathbf{1}_{x=\alpha}(x)] \geq 0$ since every assignment β to U for which $p_U(\beta) > 0$ satisfies $b_\ell(x) \geq 0$. If b_ℓ is the square of some affine function, then PSDness of $\Sigma_{I_\ell\cup J_\ell=\alpha}$ implies that $\widetilde{\mathbb{E}}[\gamma_\ell b_\ell(x)\phi_{I_\ell,J_\ell}(x)] \geq 0$. Every term on the left hand side must be at nonnegative and we have a contradiction.

Now assume that (E.15) in infeasible. If consistent local distributions $\{p_S\}$ do not exist, then an SA refutation must exist and we are done. Assume, then, that consistent local distributions $\{p_S\}$ exist but the corresponding matrix M cannot be PSD. The sets $\{M \in \mathbb{R}^{d\times d} : A \cdot \mathrm{vec}(M) \geq b\}$ and $\{M \in \mathbb{R}^{d\times d} : M \text{ is PSD}\}$ are both nonempty, but their intersection is empty. We will need the following claim.

▶ **Claim 37.** *Let $S \subseteq \mathbb{R}^{d\times d}$ be convex, closed, and bounded. Suppose that for all $M \in S$, M is not PSD. Then there exists a PSD matrix $C \in \mathbb{R}^{d\times d}$ such that $C \cdot M < 0$ for all $M \in S$.*

Proof of Claim. The claim follows from the following two results.

▶ **Theorem 38** (Separating Hyperplane Theorem)**.** *Let $S, T \subseteq \mathbb{R}^d$ be closed, convex sets such that $S \cap T = \emptyset$ and S is bounded. Then there exists $a \neq 0$ and b such that*

$$a^\top x > b \text{ for all } x \in S \text{ and } a^\top x \leq b \text{ for all } x \in T.$$

▶ **Lemma 39.** *A is PSD if and only if $A \cdot B \geq 0$ for all PSD B.*

Applied to our situation, the Separating Hyperplane Theorem says that there exists C and δ such that $C \cdot M < \delta$ for all $X \in S$ and $C \cdot M \geq \delta$ for all PSD X. We need to show that we can choose $\delta = 0$. Applying Lemma 39 will then complete the proof.

We know $\delta \leq 0$ because the zero matrix is PSD. It remains to show that we can choose $\delta \geq 0$. Assume for a contradiction that there exists PSD M such that $C \cdot M < 0$. We can then scale X by a large enough positive constant to get a PSD matrix M' such that $C \cdot M < \delta$, a contradiction. ◀

The claim implies that there is a PSD matrix C such that the set

$$\{M \in \mathbb{R}^{d\times d} : A \cdot \mathrm{vec}(M) \geq b\} \cap \{M \in \mathbb{R}^{d\times d} : C \cdot M \geq 0\}$$

is empty. As this set is defined by linear inequalities, we can apply Farkas' Lemma.

▶ **Theorem 40** (Farkas' Lemma). *Let $A \in \mathbb{R}^{m\times n}$ and consider a system of linear inequalities $Ax \geq b$. Exactly one of the following is true.*

1. *There is an $x \in \mathbb{R}^n$ such that $Ax \geq b$.*
2. *There is a $y \in \mathbb{R}^m$ such that $y \geq 0$, $y^\top A = 0$, and $y^\top b > 0$.*

In particular, this implies that there exist $y \in \mathbb{R}^c$ and $z \in \mathbb{R}$ such that

$$y^\top (A \cdot \mathrm{vec}(M) - b) + zC \cdot M < 0 \tag{E.20}$$

for all $M \in \mathbb{R}^{d\times d}$. Note that the first term is a nonnegative combination of SA constraints. Since C is PSD, we can write the eigendecomposition its $C = \sum_\ell \lambda_\ell v_\ell v_\ell^\top$ with $\lambda_\ell \geq 0$ for all ℓ. Also, recall that M is block diagonal with blocks $M^{T,\alpha}$. This block structure induces a corresponding partition of $[d]$. We can write the vector $v_\ell \in \mathbb{R}^d$ as $(v_{\ell,T,\alpha})_{T,\alpha}$ using this partition. Then the second term of (E.20) is

$$\begin{aligned}
zC \cdot M &= z \sum_\ell \lambda_\ell (v_\ell v_\ell^\top) \cdot M \\
&= z \sum_\ell \lambda_\ell v_\ell^\top M v_\ell \\
&= z \sum_\ell \lambda_\ell \sum_{\substack{|T|\leq r-2 \\ \alpha\in\{0,1\}^T}} v_{\ell,T,\alpha}^\top M^{T,\alpha} v_{\ell,T,\alpha} \\
&= z \sum_\ell \lambda_\ell \sum_{\substack{|T|\leq r-2 \\ \alpha\in\{0,1\}^T}} \sum_{i,j\in[n]} v_{\ell,T,\alpha}(i) v_{\ell,T,\alpha}(j) M_{ij}^{T,\alpha}.
\end{aligned}$$

Overall, we get

$$y^\top (A \cdot \mathrm{vec}(M) - b) + z \sum_\ell \lambda_\ell \sum_{\substack{|T|\leq r-2 \\ \alpha\in\{0,1\}^T}} \sum_{i,j\in[n]} v_{\ell,T,\alpha}(i) v_{\ell,T,\alpha}(j) M_{ij}^{T,\alpha} < 0.$$

Finally, we substitute in $X_{I,J} = \prod_{i\in I} x_i \prod_{j\in J}(1 - x_j)$ and scale appropriately to get an LS_+ refutation of the form (E.16). ◀