Electronic Colloquium on Computational Complexity, Report No. 105 (2012)

# $LS_+$ Lower Bounds from Pairwise Independence

Madhur Tulsiani* Pratik Worah†

August 17, 2012

**Abstract**

We consider the complexity of $\mathrm{LS}_+$ refutations of unsatisfiable instances of Constraint Satisfaction Problems (CSPs) when the underlying predicate supports a pairwise independent distribution on its satisfying assignments. This is the most general condition on the predicates under which the corresponding MAX-CSP problem is known to be approximation resistant.

We show that for random instances of such CSPs on $n$ variables, even after $\Omega(n)$ rounds of the $\mathrm{LS}_+$ hierarchy, the integrality gap remains equal to the approximation ratio achieved by a random assignment. In particular, this also shows that $\mathrm{LS}_+$ refutations for such instances require rank $\Omega(n)$. We also show the stronger result that refutations for such instances in the *static* $\mathrm{LS}_+$ proof system requires size $\exp(\Omega(n))$.

## 1 Introduction

We prove lower bounds for constraint satisfaction problems in the context of semidefinite programming (SDP) hierarchies of Lovász and Schrijver (LS) [15]. The motivation for studying such lower bounds arises from two settings.

First, in the context of 0-1 optimization problems, many of the natural problems are NP-complete and so a natural strategy is to model the problem as an integer program (IP) with the desired objective function. Next one relaxes the integer program to a linear program (LP) by allowing fractional solutions, solves the LP optimally in polynomial time and rounds the fractional solution to a 0-1 solution. Therefore, it is desirable that we obtain theoretical bounds on the quality of the rounded solution with respect to the optimum and in most known cases [10] such a proof crucially depends on the integrality gap of the LP [20, 10]. In this paper the lower bounds show that a large integrality gap remains even when one uses certain (systematically) stronger and stronger LP and SDP relaxations.

Second, in the context of proof complexity, one starts with some propositional formula with no 0-1 solution and hopes to show that proof systems of varying strength, from simple resolution to extended Frege, have no short refutations for the formula (see the book [13] for a survey of this area). The LP/SDP hierarchies we consider have been studied as proof systems (see [11] for a survey). A refutation in the LS type proof systems typically consists of showing that a strong enough relaxation of a convex program encoding the propositional formula contains no integer points. The level in the hierarchy to which one needs to go to obtain this conclusion is called the "rank" of the proof and the number of inequalities needed to deduce that the polytope has no

*TTI Chicago, madhurt@ttic.edu
†University of Chicago, pworah@uchicago.edu



ISSN 1433-8092

integer points is called the "size" of the proof. The results in this paper may be interpreted as lower bounds on the rank and size of refutations in various LS SDP hierarchies.

We now give some brief background on constraint satisfaction problems we study. A $k$-constraint satisfaction problem ($k$-CSP(P)) with an underlying predicate $P$ on $\{0,1\}$ variables is a formula $F : \{0,1\}^n \to \{0,1\}$ which is a conjunction of constraints of the form $c := P(x_{i_1}+b_{i_1,c}, ..., x_{i_k}+b_{i_k,c})$, where $P : \{0,1\}^k \to \{0,1\}$ is the underlying predicate, $x_i$s are the variables, $b_{i,c}$s are $\{0,1\}$ constants and '+' denotes addition modulo 2. Thus CNF-SAT is a constraint satisfaction problem where the predicate is disjunction. The MAX-$k$-CSP(P) problem refers to the problem of finding a maximum set of simultaneously satisfiable constraints given a $k$-CSP with predicate $P$. Clearly, CSPs are very important in computer science as well as discrete optimization [20]. A $k$-CSP predicate $P$ is said to *support a probability distribution* $\mu : \{0,1\}^k \to [0,1]$ if $\mu$ is non-zero only on the set $P^{-1}(1)$. Similarly we may define a distribution supported on a particular instance of the predicate $P$ i.e., a constraint. A probability distribution $\mu_c : \{0,1\}^k \to [0,1]$ supported on a constraint $c = P(x_{i_1} + b_{i_1}, ..., x_{i_k} + b_{i_k})$ is *balanced* if $\forall x_i \in c,\ b \in \{0,1\}\ \mu_c(x_i = b) = \frac{1}{2}$ and *pairwise independent* if $\forall x_i, x_j \in c,\ b, b' \in \{0,1\}^2\ \mu_c(x_i = b, x_j = b') = \frac{1}{4}$. In this paper we investigate lower bounds for LP and SDP relaxations of randomly generated CSPs for any given promise predicate (defined below).

**Definition 1.1 ([6],[3])** *A* promise predicate *on $k$ variables is a predicate $P : \{0,1\}^k \to \{0,1\}$ which supports a balanced pairwise independent distribution $\mu : \{0,1\}^k \to [0,1]$ on $P^{-1}(1)$.*

Such a class of predicates was defined by Austrin and Mossel in the context of studying approximation resistant predicates [3]. If it is hard to approximate the MAX-$k$-CSP(P) problem better than a random assignment i.e $\frac{P^{-1}(1)}{2^k}$ then $P$ is said to be approximation resistant. Austrin and Mossel [3] prove that under the Unique Games Conjecture any promise predicate is approximation resistant. Benabbas et al [6] proved that any MAX-$k$-CSP(P) instance on promise predicates has an integrality gap of $\frac{P^{-1}(1)}{2^k}$ for $\Omega(n)$ rounds of the *mixed hierarchy* i.e. Sherali-Adams (SA) with one round of SDP. Although Schoenebeck [17] has shown a linear rank lower bound on the Lasserre relaxation of sufficiently expanding CSPs with XOR clauses i.e. the predicate supports a $(k-1)$-wise independent distribution, so far no other integrality gaps for MAX-$k$-CSP(P) with promise predicates were known for any other SDP based hierarchy i.e. other than the mixed hierarchy. In this paper we prove the following statement.

**Theorem 1.2** *Let $P$ be a promise predicate on $k$ variables, $\delta > 0$ and $F$ be a random instance of MAX-$k$-CSP(P) on $n$ variables for sufficiently large $n$. Then with probability $exp(-O(k^4 2^{2k}/\delta^4))$ over the choice of $F$, the $LS_+$ hierarchy has an integrality gap of $\frac{2^k}{P^{-1}(1)}(1-\delta)$ on $F$ even after $\Omega_{k,\delta}(n)$ rounds. Also, with the above probability, any $LS_+$ refutation of $F$ requires rank $\Omega_{k,\delta}(n)$.*

Observe that a LP or SDP relaxation of a MAX-$k$-CSP(P) instance has no constraints except that the variables lie in $[0,1]$ and so in order to achieve an integrality gap of $\frac{2^k}{P^{-1}(1)}$ one usually has to find a (fractional) solution that satisfies all the constraints in the objective function. In such cases one can equivalently interpret the problem of proving integrality gaps for a MAX-$k$-CSP(P) instance for many rounds in some LP or SDP hierarchy as the problem of showing that the proof system corresponding to the LP or SDP hierarchy requires large rank.

There is also a connection between such rank bounds and traditional size bounds in proof complexity since Pitassi and Segerlind [16] prove size vs rank trade-offs for refutations in LP and SDP hierarchies (like LS$_+$). Their results imply exponential tree-like size bounds for $LS_+$ from linear rank lower bounds. Pitassi and Segerlind [16] also prove trade-offs between size and rank for SA and Lasserre proof systems[1], where size is defined as the number of monomials in the refutation. Usually the number of monomials is a weaker measure of size for geometric proof systems since even the SA refutation of the pigeon-hole principle has an exponential number of monomials but only a polynomial number of lifted inequalities.

While promise predicates have not been studied specifically in the context of static-$LS_+$[2] or even in propositional proof complexity, there have been several distantly related lower bound results for expanding instances of CSPs for example [5, 4, 2] and [12]. In all such previous lower bounds the following meta-theorem was true

$$\text{High enough expansion} + \text{Property X} \Longrightarrow \text{Lower bound for Proof system Y},$$

where we can replace the pairs $X$ and $Y$ in the above with sensitivity and resolution [5], binomial ideals and Polynomial Calculus [4], and immunity and Polynomial Calculus [2]. In this paper we (roughly) show the following kind of meta-theorem:

$$\text{High enough expansion} + \text{Promise predicates} \Longrightarrow \text{Lower bound for Static-LS}_+,$$

where our measure of size is roughly the number of lifted inequalities and has been used previously in Grigoriev et al. [11] and Itsykson and Kojevnikov [12]. We prove the following statement.

**Theorem 1.3** *Let $P$ be a promise predicate on $k$ variables and $F$ be a random instance of MAX-k-CSP(P) on $n$ variables for sufficiently large $n$. Then with probability $exp(-O(k^4 2^{2k}))$ over the choice of $F$, any static-$LS_+$ refutation of $F$ requires size $exp(\Omega_k(n))$.*

Note that Theorem 1.3 also implies the rank bounds in Theorem 1.2 since the size lower bound from Theorem 1.3 also implies exponential tree-like size lower bounds for $LS_+$ [12]. Furthermore size lower bounds in static-$LS_+$ are incomparable to the, as yet elusive, size lower bounds for $LS_+$.

**Our techniques:** In order to show our lower bounds we need to exhibit candidate fractional solutions which survive many rounds of the $LS_+$ hierarchy.

Proofs of lower bounds in the $LS_+$ hierarchy can often be viewed as strategies for a prover-adversary game, where in the candidate fractional solution, the adversary "fixes" the value of a variable at each step and the prover is required to provide a new fractional solution consistent with this fixing. The solutions for the fixings of all the variables can be viewed as columns of a matrix (called the *protection matrix*) and $LS_+$ lower bounds require proving that such matrices that arise in the proof are positive semidefinite. Our main tool in this regard is to systematically decompose the large $n \times n$ matrix into relatively smaller matrices and show that each such matrix is positive semidefinite. Since our instance is derived from expanders, a common theme in proof complexity, we use expansion correction to construct the systematic decomposition for multiple rounds of $LS_+$ hierarchy. Expansion correction was first used in this context by Alekhnovich et al. [1]. Alekhnovich et al. also proved the matrices were positive semidefinite by using a decomposition in to simpler matrices, but in their case these simpler matrices were positive semidefinite simply by diagonal

[1]In passing, we note that Pitassi and Segerlind [16] seem to view these proof systems as dynamic proof systems as opposed to static proof systems (cf. Grigoriev et al [11]).

[2]A *static* proof system gives the refutation in one-shot as opposed to a *dynamic* proof system which gradually builds the refutation [11]. For example the Nullstellensatz maybe viewed as a static version of the Polynomial Calculus and Sherali-Adams as a static version of the Lovász-Schrijver proof system (cf. [11, 19]).

dominance. In this paper we crucially use properties of the "locally consistent distributions of partial assignments to variables" as defined by Benabbas et al. [6] to show that our fractional solutions satisfy semidefiniteness constraints. This is interesting because positive semidefiniteness is a global property and previous papers used the local consistency property only to show fractional solutions for LP hierarchies without many rounds of positive semidefiniteness constraints [6, 9].

Our size bound for static-$LS_+$ proof systems also requires proving certain related matrices are positive semidefinite and crucially uses the tools developed in the rank lower bound argument. We think that a similar approach may be useful with (as yet elusive) more general lower bounds for the stronger Lasserre hierarchy.

A brief walkthrough of the paper follows. In Section 2 we restate the basic definitions for hierarchies, proof systems and their related parameters. Section 3 is divided into four subsections. In Section 3.1 we revisit the notion of expansion correction from [1] and [6]. In Section 3.2 we define our family of measures, prove they are locally consistent and show some useful properties for these measures. In Section 3.3 we prove the rank lower bound and integrality gap results for $LS_+$ using the prover adversary technique. We extend the lemmata from this section to prove the size lower bound for static-LS$_+$ in Section 3.4 Finally, in Section 4 we sketch how to generalize Theorems 1.2 and 1.3 to the case of almost all random instances of CSPs with promise predicates.

## 2 Basic Formalism

We briefly recap some of the necessary basic definitions. More details about basic convex optimization concepts like cones, polytopes, linear and semidefinite programming are discussed in [7]. We assume some familiarity with linear and semidefinite programming hierarchies especially lift and project hierarchies. Details about these are discussed in the surveys [14, 10].

Lovász and Schrijver [15] introduced the $LS_+$ lift and project hierarchy as a tool to generate successively tighter relaxations of combinatorial optimization problems. Starting with the polytope $K \in \mathbb{R}^n$ (say) of an initial linear relaxation for the problem, the hierarchy generates progressively tighter relaxations. The definition of $LS_+$ [15] uses the homogenized cone $\tilde{K}$ defined as $\tilde{K} := \{(\lambda, \lambda x_1, \ldots, \lambda x_n) \mid \lambda > 0, (x_1, \ldots, x_n) \in K\}$. The polytope $K$ corresponding to the homogenized cone $\tilde{K}$ is simply obtained by intersecting it with the hyperplane $x_0 = 1$.

**Definition 2.1 ([15])** *Given a convex cone $\tilde{K}$ in $\mathbb{R}^{n+1}$ define the cone $M(\tilde{K_1})$ (the lifted $LS_+$ cone) as the cone consisting of all $(n+1) \times (n+1)$ matrices $Y$ in $\mathbb{R}$ satisfying the conditions:*

1. *$Y$ is symmetric and positive semidefinite.*
2. *$\forall i,\ Y_{ii} = Y_{i0}$.*
3. *$Y_i \in \tilde{K}, Y_0 - Y_i \in \tilde{K} \quad \forall 1 \leq i \leq n$.*

*Here $Y_i$ denotes the $i^{th}$ column of the matrix $Y$. Let $N_+(\tilde{K})$ denote the projection $Ye_0$ of $M(\tilde{K})$. Define $N_+(\tilde{K})$ (or simply $N_+(K) = N_+(\tilde{K}) \cap (x_0 = 1)$) as the cone (polytope) obtained after a single $LS_+$ lift and project step.*

In proof complexity one typically starts with a formula which is a negation of some tautology and encodes the formula as a system of inequalities in $[0,1]^n$ [11]. The $LS_+$ rank of an unsatisfiable (in $\{0,1\}^n$) set of inequalities given by the cone $\tilde{K}$ is the minimum value of $r$ such that $N_+^r(\tilde{K})$ is empty. An equivalent characterization of the $LS_+$ hierarchy as a proof system is given below.

**Definition 2.2 ([11])** *Given a set $K$ of linear inequalities on the variables $\{x_1, ..., x_n\}$ and the $\forall i,\ x_i^2 - x_i = 0$, we have the following inference rules for $LS_+$:*

1. $\frac{p \geq 0}{p \cdot q \geq 0}$ *where $deg(pq) \leq 2$ and $q \in \{x_i, 1 - x_i : i \in [n]\}$.*
2. $\frac{p \geq 0 \quad q \geq 0}{\alpha p + \beta q \geq 0}$ *for $\alpha, \beta \in \mathbb{R}^+$.*
3. $l^2 \geq 0$ *for $deg(l) \leq 1$*

*A valid refutation must obtain the contradiction $-1 \geq 0$.*

An application of rule 1 increases the degree of the inequality $p \geq 0$ by 1 (which is called the lift step) and thus one must reduce the degree before applying rule 1 again, by taking positive linear combinations as in rule 2 (which is called the projection step). In this case the $LS_+$ rank simply refers to the maximum number of applications of inference rule 1 in any path leading to a contradiction i.e., the root in the proof DAG. Grigoriev et al [11] also defined and studied size lower bounds for the static-$LS_+$ proof system which is at least as strong as the proof system corresponding to the mixed hierarchy studied in Benabbas et al [6].

**Definition 2.3 ([11])** *The axioms consist of the inequalities $\forall i\ x_i \geq 0$, $\forall i\ 1 - x_i \geq 0$, $\forall i\ x_i^2 - x_i \geq 0$ and a given set of linear inequalities $K$. A valid static-$LS_+$ refutation consists of positive linear combination of the terms $\varphi_{I,J} = s_{I,J} \cdot \Pi_{i \in I} x_i \Pi_{j \in J}(1 - x_j)$ where $I$ and $J$ are multisets of variable indices and $s_{I,J}$ is an axiom or the square of some linear form. A refutation is obtained by deriving*

$$\sum_l \omega_l \cdot \varphi_{I_l, J_l} \ = \ -1 \tag{2.1}$$

*where each $\omega_l \in \mathbb{R}^+$.*

The *size* of a static-$LS_+$ refutation (as in equation 2.1) is defined as the number of summands i.e. the number of distinct $\varphi$s in the static-$LS_+$ refutation. The *multilinear or boolean degree* of a polynomial in $\mathbb{R}[x_1, .., x_n]$ is defined as the degree after multilinearizing the polynomial i.e. the polynomial is viewed to be in the Smolensky ring $S_n(\mathbb{R}) := \mathbb{R}[x_1, .., x_n]/\{x_i^2 - x_i : i \in [n]\}$. The degree of a static-$LS_+$ refutation (as in equation 2.1) is the maximum among the multilinear degree of its summands. Throughout this paper we will use degree to stand for multilinear degree (which only makes the lower bounds stronger). This means that we can assume $I_l$ and $J_l$ in equation 2.1 are just sets and not multisets. This also means that we will essentially work in the ring $S_n(\mathbb{R})$ and so we will ignore degree of the summands with $s_{I,J} = x_i^2 - x_i$ for any $i \in [n]$ in equation 2.1 since their multilinear degree is always 0.

## 3 Rank Lower Bounds and Integrality Gaps

In order to show our lower bounds we use the same family of instances used in Benabbas et al. [6]. Before describing the proofs we first recap some of the already known results and provide some useful extensions of our own.

### 3.1 Expansion Correction

In this subsection we mainly recap some of the concepts from Alekhnovich et al. [1] and Benabbas et al. [6] which will prove useful later.

**Definition 3.1** *Given a $k$-CSP formula $F$ we define its* constraint graph *$G$ as the bipartite graph (bigraph) $G = (L, R, E)$ formed by variables on the right side and constraints on the left side. An edge in $E$ connects variable $x_i$ to constraint $c_i$ if $x_i \in c_i$.*

Let $N(X)$ denote the set of neighbors of vertex set $X$ in the constraint bigraph and let $\partial X := \{v \in G : |N(v) \cap X| = 1\}$ denote the boundary of $X$.

**Definition 3.2** *We say a constraint bigraph $G$ is $(r, e)$* expanding *if for any set $X$ vertices corresponding to constraints with $|X| \leq r$, $|N(X)| \geq e \cdot |X|$. Similarly, we call the graph $(r, e)$* boundary expanding *if for any set $X$ of vertices corresponding to constraints with $|X| \leq r$, $|\partial X| \geq e \cdot |X|$.*

Given a promise predicate $P$, we use the same model for generating a random instance of $k$-CSP(P) as [6]. A constraint $c$ is generated randomly by uniformly selecting a set of $k$ variables from the set of $n$ variables. Each selected variable $x_i \in c$ is then XOR-ed with $b_{i,c}$ which are 0 or 1 with equal probability. The constraint graph of a random $k$-CSP(P) is an expander and the random formula is unsatisfiable as formally witnessed by the following theorem. The satisfiability of the formula is measured by the quantity $OPT(F)$ which the maximum number of constrains satisfied by any assignment to the variables of the formula.

**Theorem 3.3 ([6])** *Given $\varepsilon, \delta > 0$ and a predicate $P$ then there exist $\gamma = O(2^k/\delta^2)$, $\eta = \Omega(\frac{1}{\gamma^{10/\varepsilon}})$ and $N_0 \in \mathbb{N}$ such that with probability at least $exp(-O(k^4\gamma^2))$ a random instance $F$ of $k$-CSP(P) with number of variables $n \geq N_0$ and $m = \gamma n$ constraints is:*

1. *Very unsatisfiable: $OPT(F) \leq \frac{P^{-1}(1)}{2^k}(1+\delta)m$*
2. *A good expander: For any set of constraints $C$ with $|C| \leq \eta n$, we have $|N(C)| \geq (k-1-\varepsilon)|C|$.*
3. *Has large girth: The constraint graph has girth at least 5 i.e. no two constraints share more than one variable.*

Following Alekhnovich et al. [1], Benabbas et al [6] define the closure operation on expanders which allows one to preserve expansion and boundary expansion in the presence of deletion of vertices (and edges). Given a constraint graph $G$ and a set of variables $X$, let $\mathcal{C}(X) := \{c \in G : N(c) \subseteq X\}$ denote the constraints supported on the variables in $X$. Let $G - X$ denote the constraint graph obtained by removing all the variables in $X$ and all the constraints in $\mathcal{C}(X)$ from $G$.

**Definition 3.4 ([6])** *The Closure or Advice set $Cl(X)$ of a set of variables $X$ in the $(r_1, e_1)$ expanding constraint bigraph $G$ is a superset of $X$ obtained by Algorithm Closure, when started with the set $X$.*

Note that $Cl(X)$ is unique upto some arbitrary ordering of the constraints and variables and we say that a set of variables is *closed* if it is a closure of some set of variables $X$. Unlike [6], which uses the operation with an arbitrary ordering, we will have to choose the ordering somewhat carefully. We also note that definition 3.4 is very similar to (but slightly stronger than) that in Benabbas et al [6] with boundary expansion replaced by expansion.

**Algorithm Closure**

The input is an $(r_1, e_1)$ expanding bipartite graph $G = (L, R, E)$, some $e_2 \in (0, e_1)$, and some $S \subseteq R$, $|S| < (e_1 - e_2)r_1$, with some arbitrary order $S = \{x_1, \ldots, x_t\}$. The output is the closure of $S$ i.e. $Cl(S)$.

Initially set $Cl(S) \leftarrow \emptyset$ and $\xi \leftarrow r_1$
**For** $j = 1, \ldots, |S|$ **do**
  $M_j \leftarrow \emptyset$
  $Cl(S) \leftarrow Cl(S) \cup \{x_j\}$
  **If** $G - Cl(S)$ is not $(\xi, e_2)$ expanding **then**
    Enumerate in increasing lexicographic order $M_j \subset L$ to find a maximal $M_j$ in $G-Cl(S)$ such that $|M_j| \le \xi$ and $|N(M_j)| \le e_2|M_j|$ in $G - Cl(S)$
    $Cl(S) \leftarrow Cl(S) \cup N(M_j)$
    $\xi \leftarrow \xi - |M_j|$
**Return** $Cl(S)$

**Lemma 3.5** *If a constraint bigraph $H$ is $(r, k-1-\varepsilon)$ expanding then it is $(r, k-2-2\varepsilon)$ boundary expanding.*

**Proof:** For a set of constraints of size $s \le r$, let $b$ be the number of boundary variables. Since the total number of variable occurrences is $k \cdot s$ and each of the non-boundary variables occur at least twice, we have by expansion $b + (k \cdot s - b)/2 \ \ge \ (k-1-\varepsilon) \cdot s$ which gives $b \ge (k-2-2\varepsilon) \cdot s$. ■

In this paper we will need to perform alternate closure and variable restriction on the constraint bigraph i.e. $(((G - Cl(i_1)) - Cl(i_2)).. - Cl(i_r))$, which we abbreviate as $G - \bigcup_{i \in I} Cl(i)$ for an ordered set $I$. The following lemmata are essentially restatements of Theorem 3.1 a,b from [6] with boundary expansion replaced by expansion.

**Lemma 3.6 ([6])** *If $G$ is $(r_1, e_1)$ expanding then $G - Cl(X)$ is $(r_2, e_2)$ expanding, where $X$ is a set of variables with $|X| \le (e_1 - e_2)r_1$, $r_2 \ge r_1 - \frac{|X|}{e_1 - e_2}$ and $|Cl(X)| \le \frac{k+2e_1-e_2}{e_1-e_2}|X|$.*

Throughout this paper we will set $e_1 = k - 1 - \varepsilon$ and $e_2$ will be set very close to $e_1$ in Algorithm Closure. So we will typically use the following kind of instantiation of the previous lemma.

**Lemma 3.7 ([6])** *Suppose $G$ is a $(s, k-1-\varepsilon)$ expander and $R = \bigcup_{i \in I_R} Cl(i)$ for an ordered set of variables $I_R$ with $|I_R| \le r$, then $G - R$ is a $(s', k-1-2\varepsilon)$ expander for $s' \ge s - \Omega(\frac{r}{\varepsilon})$.*

The proofs of the previous two lemmas are given in the appendix. We end this subsection with an important remark about the definition of closure.

**Remark 3.8** *Algorithm Closure is an online algorithm. Given a set $X_k = \{x_1, \ldots, x_k\}$ where the elements are in increasing order, for each $j < k$, it first computes the closure of $X_j = \{x_1, \ldots, x_j\}$ without looking at $X_k \setminus X_j$. It then extends this to the closure of $X_k$.*

### 3.2 Locally Consistent Measures

One can visualize successive rounds of $LS_+$ as contracting the initial polytope until we reach the empty polytope since our CSP has no 0-1 solution. Therefore it is reasonable to expect that points which survive many rounds of $LS_+$ will be highly symmetric in some way with respect to our initial constraints. In order to capture this symmetry Benabbas et al. [6] define a family of measures $m_S$ for small subsets $S$ of variables in the constraint bigraph by "symmetrizing" the pairwise independent measure $\mu$ supported on $P^{-1}(1)$, where $P$ is underlying predicate for the CSP under consideration.

Let $X$ be a set of variables. Given an assignment $\beta \in \{0,1\}^X$, let $\mu_c(\beta)$ denote the value of $\mu_c$ after restricting $\beta$ to the support of constraint $c$ and summing over all possible values for variables in the support which are not assigned by $\beta$. In particular $\mu_c(\beta) = 1$ if $X$ does not contain any variables in the support of the constraint $c$. On the other hand if $\beta$ is an assignment that violates the constraint $c$, then $\mu_c(\beta) = 0$. Also for $S \subseteq X$ let $\beta_{|S}$ denote the restriction of $\beta$ to the variables in $S$. We now define the family of "local distributions" that we will use.

**Definition 3.9 ([6])** *For a constraint $c := P(x_{i_1}+b_1, ..., x_{i_k}+b_k)$ where $b_j \in \{0,1\}$, let $\mu_c(x_{i_1}, \ldots, x_{i_k})$ be the pairwise independent distribution*

$$\mu_c(x_{i_1}, \ldots, x_{i_k}) \quad := \quad \mu(x_{i_1} + b_1, ..., x_{i_k} + b_k)$$

*supported on the satisfying assignments of $c$. For a subset of variables $S$ let $\alpha \in \{0,1\}^S$ be an assignment to $S$. Define the measure $m_S$ supported on the satisfying assignments of constraints in $\mathcal{C}(Cl(S))$ as*

$$m_S(\alpha) := \frac{1}{Z_S} \cdot \left( \sum_{\beta \in \{0,1\}^{Cl(S)}, \beta_{|S}=\alpha} \prod_{c \in \mathcal{C}(Cl(S))} \mu_c(\beta) \right) \quad \textit{where} \quad Z_S = \sum_{\beta \in \{0,1\}^{Cl(S)}} \prod_{c \in \mathcal{C}(Cl(S))} \mu_c(\beta)$$

Note that the definition of the measure $m_S$ depends on $Cl(S)$, which in turn depends on the ordering the variables according to which the closure is computed. The measures above can formally be defined for all sets $S$ (as long as $Z_S > 0$), but our notion of consistency will only apply to measures $m_S$ from this family where the size of $S$ is bounded by some parameter $t$.

**Definition 3.10** *A family of measures (distributions) $\{m_S\}$ is $t$-*locally consistent *if*

$$\forall T \subseteq S \subseteq [n], |S| \leq t, \ \forall \alpha \in \{0,1\}^T, \ m_S(\alpha) = m_T(\alpha),$$

*where $m_S(\alpha) = \sum_{\beta \in \{0,1\}^S, \beta_{|T}=\alpha} m_S(\beta)$.*

The existence of a $t$-locally consistent measure family implies the existence of feasible Sherali-Adams solutions from the measures [6, 9]. In the case of promise predicates, Benabbas et al [6] show that the measures $m_S$ are well defined i.e., have $Z_S > 0$ (Lemma 3.2) and are $t$-locally consistent for $|S| \leq t = \Omega_{k,\varepsilon,\delta}(n)$.

**Lemma 3.11 (Claim 4.2 from Benabbas et al [6])** *Let $F$ be a random instance of MAX-k-CSP(P) as in Theorem 3.3 chosen for given $\varepsilon, \delta > 0$. Let $S_1 \subseteq S_2 \subseteq [n]$ be sets of variables with $|S_2| \leq \eta\varepsilon n/4k$. Then $m_{S_1}$ and $m_{S_2}$ are equal on $S_1$.*

In this paper we will need to work with conditional probabilities of the form $m_S(\cdot|X = \alpha)$ so for clarity we define them below.

**Definition 3.12** *Let $X, S$ be disjoint sets of variables. Let $\alpha \in \{0,1\}^X$ and $\beta \in \{0,1\}^S$. We then define*

$$m_S(S = \beta | X = \alpha) := \frac{m_{X \cup S}(S = \beta, X = \alpha)}{m_{X \cup S}(X = \alpha)} .$$

Note that we are effectively conditioning each $\mu_c$ involved in defining $m_{X\cup S}(\cdot)$ on the variables that it shares with $X$. Such a conditional measure is well defined for all $\alpha$ such that $m_{X\cup S}(X = \alpha) > 0$.

For the remainder of this subsection let $X$ denote a *closed* set of variables $X := \bigcup_{i\in I_X} Cl(i)$ for some *ordered* set of variables $I_X$. As mentioned before we set $e_1 = k-1-\varepsilon$ and $e_2$ close to $e_1$ (say $k-1-2\varepsilon$) in Algorithm Closure. We now show that if the the family of measures$\{m_S\}$ is $t$-locally consistent, then $\{m_S(\cdot|X=\alpha)\}$ is also $t'$-locally consistent for $t' = t - |X|$.

**Lemma 3.13 (Local Consistency of Conditional Measures)** *Let $X$ be a subset of variable and let $\alpha \in \{0,1\}^X$ be such that $\mu_c(\alpha) > 0$ for all $c \in G$. Let $\{m_S\}$ be a $t$-locally consistent family for $t \geq |X|$. Then the family of measures $\{m_S(\cdot|X=\alpha)\}$ defined for sets $S$ such that $S\cap X = \emptyset$ and $S$ with $|S\cup X| \leq t$ is $(t-|X|)$-locally consistent. More formally, for any $T \subseteq S$, and $|S\cup X| \leq t$*

$$\forall \beta \in \{0,1\}^T \quad m_S(T=\beta|x=\alpha) \;=\; m_T(T=\beta|X=\alpha)\,.$$

**Proof:** By local consistency of the (unconditioned) measures $m_S$, we have that $m_{T\cup X}(X=\alpha) = m_{S\cup X}(X=\alpha)$. Also, both these are positive by the assumption on $\alpha$. Thus, we have for any $\beta \in \{0,1\}^T$

$$m_S(T=\beta|X=\alpha) \;=\; \frac{m_{S\cup X}(T=\beta)}{m_{S\cup X}(X=\alpha)} \;=\; \frac{m_{T\cup X}(T=\beta)}{m_{T\cup X}(X=\alpha)} \;=\; m_T(T=\beta|X=\alpha)$$

where the second equality used the consistency of the (unconditioned) family $\{m_S\}$. ■

In this paper we will have to deal with expanding constraint graphs which arise out of expanding $k$-CSP instances due to conditioning i.e., graphs of the form $G-X$, where $X$ is a set of variables whose values have been fixed (conditioned on).

**Definition 3.14** *Given a set $X$ of variables, we say that a constraint $c$ is* uncompromised *in $G-X$ if it has $k$ neighbors in $G-X$. Otherwise we say that it is* compromised.

The matrices $Y$ that we construct in the proof of our rank lower bound for the $LS_+$ hierarchy will be of the form $Y_{ij} = \mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$ where the set $X$ and the assignment $\alpha$ will change with the number of rounds of the $LS_+$ hierarchy. Here $\mathbb{1}_{\{i\}}(\beta)$ is is the indicator function which is 1 if and only if the variable $x_i$ is set to 1 in the 0-1 assignment $\beta$. The following lemma will be extremely useful as it lets us explicitly calculate (most of) the entries of such matrices.

**Lemma 3.15 (Explicit Evaluation of Conditional Expectations)** *Let $G-X$ be $(r, k-1-\varepsilon)$ expanding for $r > 2$ and $\varepsilon < 0.1$. Let $\alpha \in \{0,1\}^X$ be an assignment to $X$ with $\mu_c(\alpha) > 0$ for all constraints $c$ and Let the family $m_S(\cdot|X=\alpha)$ be defined as before. Then:*

- *for any variable $i \notin X$, we have $\mathbb{E}_{m_i(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\right] = 0.5$.*
- *for any $i,j \notin X$ which do not belong to the same compromised constraint in $G-X$, we have $\mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right] \;=\; 0.25$*

**Proof:** We first prove that the graph remains expanding after removing the variables $i$ and $j$.

**Claim 3.16** *If the constraint bigraph $G-X$ is $(r, k-1-\varepsilon)$ expanding with $\varepsilon \leq 0.1$, girth at least 5, and variables $i$ and $j$ do not belong to the same compromised constraint then $G-X-\{i,j\}$ is $(r-2, k-2-\varepsilon')$ boundary expanding for some $\varepsilon' < 0.9$.*

**Proof:** Let $S$ be a set of constraints in bigraph $G - X - \{i,j\}$. We can check for expansion by considering the following cases based on the size of $S$:

1. If $|S| = 1$ then $|\partial S| = |N(S)| \geq k-2$ since $G - X$ is a $k-1-\varepsilon$ expander and so even a compromised constraint in $G - X - \{i,j\}$ has at least $k-2-\varepsilon$ neighbors as long as $i$ and $j$ both do not belong to that same compromised constraint.
2. If $|S| = 2$ then $|\partial S| \geq 2(k-1-\varepsilon) - 1 - 2 = 2(k-2-(1/2+2\varepsilon))$ since two constraints in $G$ share at most one variable (since $G$ has girth at least 5).
3. If $|S| \geq 3$ then by Lemma 3.5
$$|\partial S| \geq (k-2-2\varepsilon))|S| - 2 = (k-2-(2\varepsilon + \frac{2}{|S|}))|S|$$
and so for $\varepsilon \leq 1/10$ we have $2\varepsilon + \frac{2}{|S|} \leq 9/10$.

■

The previous lemma allows us to decompose the expression for $\mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$ using the following result from [6], the proof of which is by repeated application of the pigeon-hole principle.

**Claim 3.17 ([6])** *Let $H$ be a constraint bigraph and $S_1$ and $S_2$ be sets of variables with $S_1 \subseteq S_2$, $S_2 = Cl(S_2')$ for some $S_2'$ and $|\mathcal{C}(S_2)| \leq r$ such that $H$ and $H - S_1$ are $(r, k-2-\varepsilon')$ boundary expanding for some $\varepsilon' < 1$. Let $\mathcal{C}(S_2) \setminus \mathcal{C}(S_1) := \{c_{i_1}, ..., c_{i_s}\}$ be the set of constraints supported on variable sets $T_{i_1}, ..., T_{i_s}$ respectively. There exists an ordering of the constraints $c_{i_1}, ..., c_{i_s}$ and a partition of $S_2 \setminus S_1$ into $F_1, ..., F_t, F_{s+1}$ such that for all $j \leq t$, $F_j \subseteq T_{i_j}$, $|F_j| \geq k-2$ and $F_j \cap (\cup_{l>j} T_{i_l}) = \emptyset$.*

To summarize, given a sum $\sum_{\gamma \in \{0,1\}^{S_2 \setminus S_1}} \prod_{c \in \mathcal{C}(S_2)} \mu_c(\gamma)$, Claim 3.17 will give an ordered partition of $S_2 \setminus S_1$ into $F_i$s such that in this ordering each successive constraint $c \in \mathcal{C}(S_2) \setminus \mathcal{C}(S_1)$ has at most 2 of its variables shared with the constraints following it in the ordering. This will allow us to use pairwise independence and find a numerical value of such terms.

More precisely, observe that

$$\mathop{\mathbb{E}}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right] = m_{ij}(\mathbb{1}_{\{ij\}}|X=\alpha) = \frac{\sum_{\beta \in \{0,1\}^{Cl(X \cup \{i,j\})}, \beta_{|X}=\alpha, \beta_{|i,j}=1,1} \prod_{c \in \mathcal{C}(Cl(X \cup \{i,j\}))} \mu_c(\alpha \cdot \beta)}{\sum_{\beta \in \{0,1\}^{Cl(X \cup \{i,j\})}} \prod_{c \in \mathcal{C}(Cl(X \cup \{i,j\}))} \mu_c(\alpha \cdot \beta)}. \tag{3.1}$$

Here $\alpha \cdot \beta$ denotes the concatenation of the two assignments $\alpha$ and $\beta$. Note that $m_{X \cup \{i,j\}}(X = \alpha) > 0$ by local consistency of $m_{Cl(X \cup i,j)}(\cdot)$ and the definition of $\alpha$. Furthermore, $X$ is closed, $i,j$ do not belong to the same compromised constraint, and $k \geq 3$ imply that $\mathcal{C}(X) = \mathcal{C}(X \cup \{i,j\})$.

Any terms of the form $\mu_c(\alpha \cdot \beta)$ for $c \in \mathcal{C}(X) = \mathcal{C}(X \cup \{i,j\})$ then appear both in the numerator and denominator. Also, since we are looking at $c \in \mathcal{C}(X)$, these terms depend only on $\alpha$ and can be cancelled from the numerator and denominator. We are then left with a product of terms for $c \in \mathcal{C}(Cl(X \cup \{i,j\})) \setminus \mathcal{C}(X \cup \{i,j\})$. Therefore we can reduce the fraction by cancelling terms of the form $\prod_{c \in \mathcal{C}(X \cup i,j)} \mu_c$. Now we evaluate the numerator and denominator of this reduced fraction separately using Claim 3.17. We set $H = G$, $S_1 = X \cup \{i,j\}$ and $S_2 = Cl(X \cup i,j)$ in Claim 3.17. Observe that Claim 3.17 gives us an ordered partition i.e. $F_1, \ldots, F_{s+1}$s for the variables in $Cl(X \cup \{i,j\}) \setminus (X \cup \{i,j\})$. Let $\overline{\alpha} \in \{0,1\}^{X \cup \{i,j\}} = \alpha \cdot \{1,1\}$. Then the numerator equals

$$\sum_{\beta_{s+1} \in \{0,1\}^{F_{k+1}}} \cdots \sum_{\beta_1 \in \{0,1\}^{F_1}} \mu_{c_{i_1}}(\overline{\alpha} \cdot \beta) \cdots \mu_{c_{i_s}}(\overline{\alpha} \cdot \beta)$$

When we sum over variables in $F_1$, the term $\mu_{c_1}(\overline{\alpha} \cdot \beta)$ disappears since we are summing over at least $k-2$ variables from $c_{i_1}$ and $\mu$ is balanced and pairwise independent (and also none of the other constraints share variables with $F_1$). Summing over variables in $F_1$ reduces the first term to $1/2^{k-|F_1|}$. Proceeding similarly for the rest of the summation, we reduces the expression to $1/2^R$ for some $R \in \mathbb{N}$. The denominator reduces in exactly the same way, except it has two extra variables $i$ and $j$ which we sum over. This gives the denominator as $1/2^{R-2}$ and the ratio as 0.25.

Note that we did not use the fact that we set $j = 1$ in the denominator. Thus, the same proof also shows that $\mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}(1 - \mathbb{1}_{\{j\}})\right] = 0.25$ and hence $\mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\right] = \mathbb{E}_{m_i(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\right] = 0.5$ by doing the above calculation with a $j$ such that $i$ and $j$ are not in the same compromised constraint. ■

Before we begin the actual lower bound proofs in the next two subsections a remark about the proofs in this subsection.

**Remark 3.18** *Lemma 3.13 (Local Consistency Lemma) and Lemma 3.15 (Explicit Evaluation Lemma) are both with respect to* any *ordering of variables as far as computation of closures is concerned. We have not assumed anything about variable ordering thus far, a freedom we will use in the next two subsections.*

The reason we made remark 3.18 is because from the next section onwards we will need to assume that for our restricted formula $F_X$, the variables in $X$ are ordered before other variables when computing the Closure. This is equivalent to saying that for every additional variable that is fixed in $F_X$, we not compute the closure in the graph $G - X$. This was also used in the previous works on the $LS_+$ hierarchy. In absence of this assumption, we may not have $Cl(X) \subseteq Cl(X \cup A)$, which will be needed to bound the number of variables fixed by the closure operations over different rounds.

## 3.3 Rank Bounds for $LS_+$

Given the initial cone specified by the linear constraints derived from our unsatisfiable $k$-CSP(P) instance, we want to show that the cone obtained after many rounds of $LS_+$ lift and project is non-empty i.e. there exists a satisfying (fractional) assignment. To this end we describe a Prover-Adversary game from Schoenebeck et al [18] originally introduced by Buresh-Oppenheim et al [8]. The game has two players a Prover and an Adversary. In round $r$, starting from a point $y^{(r)} \in [0,1]^{n+1}$ the Prover constructs a lifted point / protection matrix $Y^{(r)}$ and a set of vectors $O_r$ with each element in $\mathbb{R}^{n+1}$ such that

1. Each element of $O_1$ satisfies all initial constraints
2. $Y^{(r)}$ satisfies the conditions 1 and 2 in Definition 2.1
3. $\forall i$ the column vectors $Y_i^{(r)}$ and $Y_0^{(r)} - Y_i^{(r)}$ can be expressed as a positive linear combination of elements in $O_r$.

The Adversary chooses a vector $z \in O_r$, sets $y^{(r+1)}$ as $z$ and the game continues as long as the Prover can construct the protection matrices.

**Theorem 3.19 ([18])** *Suppose, starting from a point $y$ that satisfies all our initial constrains, the Prover has a strategy that lasts $r+1$ rounds against any* Adversary*, then $y$ survives $r$ rounds of $LS_+$ lift and project.*

We now describe a Prover strategy which allows us to establish our required integrality gap and rank lower bounds. The Prover initially starts with the point $y = y^{(1)} \in \mathbb{R}^{n+1}$ such that $y_0^{(1)} = 1$ and $\forall i \in [n]$,

$$y^{(1)}(i) \;=\; \mathop{\mathbb{P}}_{m_i(\cdot)}[x_i = 1] \;=\; \mathop{\mathbb{E}}_{m_i(\cdot)}\left[\mathbb{1}_{\{i\}}\right].$$

Note that by Lemma 3.15, we in fact have $y_i^{(1)} = 1/2$ for all $i \in [n]$. The protection matrix $Y^{(1)}$ corresponding to $y^{(1)}$ is given by

$$Y^{(1)}(i,j) = \mathop{\mathbb{E}}_{m_{ij}(\cdot)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right].$$

To describe the set $O_1$, we first observe that the constraints $Y_{ii}^{(1)} = Y_{i0}^{(1)} = y_i^{(1)}$ imply that the vectors $\frac{Y_i^{(1)}}{y_i^{(1)}}$ and $\frac{Y_0^{(1)} - Y_i^{(1)}}{1-y_i^{(1)}}$ correspond to fractional solutions with the $i^{th}$ variable set to 1 and 0 respectively. The vectors in the set $O_1$ are defined by setting additional variables in $Cl(\{i\})$ so that the constraint graph on the variables with fractional values still remains expanding.

Define the set $R_{1,j}$ as the set of variables $Cl(j)$ for $j \in [n]$ and define the set $A_{1,j}$ as possible $\{0,1\}$ assignments to $R_{1,j}$ i.e. $\alpha \in \{0,1\}^{Cl(j)}$. Here we only consider $\alpha \in \{0,1\}^{Cl(j)}$ such that for any constraint $c_i \in \mathcal{C}(Cl(j))$ we have $\mu_i(\alpha) > 0$. Let $R_1$ be the *set of sets* $R_1 = \{R_{1,j} \,:\, j \in [n]\}$. The set $O_1$ consists of vectors $y^{(1)}_{R_{1,j}=\alpha}(i) \;=\; \mathbb{E}_{m_i(\cdot|R_{1,j}=\alpha)}\left[\mathbb{1}_{\{i\}}\right]$, where $m_i(\cdot|R_{1,l} = \alpha)$ stands for the measure $m_i$ conditioned on fixing the variables in $R_{1,j}$ to an assignment $\alpha$ from $A_1$.

Before we describe the Prover strategy for higher number of rounds we make an assumption about the ordering of variables in Algorithm Closure from here onwards for the rest of this section. This is an important subtlety in our application of the closure algorithm, as opposed to [6] where the ordering on the variables could be chosen arbitrarily. In our setting, when we consider a sequence of variables $x_1, \ldots, x_r$ which has been fixed by the adversary, we want these variables to occur first in the ordering, before any other variables for which we are computing the conditional distributions. Thus, at round $r$, we want the distributions $\{m_S\}$ to be defined according to an ordering which has $x_1, \ldots, x_r$ in that order and before any other variables. However, at round $r+1$, when the adversary conditions on an additional variable $x_{r+1}$, we define the measures according to the *new* ordering which puts $x_1, \ldots, x_{r+1}$ first (in that order). We now want this to be consistent with the measure at round $r$, even though we did not know $x_{r+1}$ at round $r$. This is where the online nature of the closure algorithm (see Remark 3.8) is useful for proving consistency.

**Assumption 3.20** *Given an ordered set of variables $X = \{x_{i_1}, .., x_{i_s}\}$ which have been conditioned by the adversary in the same order we assume that these variables are ordered in the same fashion when computing $Cl(X)$. Furthermore, for any set of variables $S$, $S \cap X = \emptyset$ we assume that the variables in $X$ are ordered before those in $S$ when computing closure.*

**Remark 3.21** *The ordering of variables and thus the family of locally consistent measures that we construct after $r$ rounds of adversary moves may differ across various adversary fixings / moves which is allowed by the online nature of Algorithm Closure (cf. Remark 3.8.)*

We now describe the Prover strategy for round $r+1$ for $r \geq 1$. At the end of round $r$, the Adversary would have chosen a vector $z \in O_r$ such that $z = y^{(r)}_{X=\alpha}$ for $X \in R_r$. So in round $r+1$ the Prover has input $y^{(r+1)} = z$ and the Prover outputs a matrix with entries $Y^{(r+1)}(i,j) \;=\; \mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$. We then define $R_{r+1}$ as

$$R_{r+1} \;:=\; \{X \cup Cl(j) : j \notin X\}\,. \tag{3.2}$$

The allowed assignments for $R_{r+1,j}$ in $A_{r+1}$ are $\alpha' \in \{0,1\}^{X'}$ such that for any constraint $c \in \mathcal{C}(X')$ we have $\mu_c(\alpha') > 0$. We define the set

$$O_{r+1} := \{y^{(r+1)}_{R_{r+1,j}=\alpha'} \in \mathbb{R}^{n+1} : R_{r+1,j} \in R_{r+1}, \alpha' \in A_{r+1}\}, \tag{3.3}$$

where $y^{(r+1)}_{R_{r+1,j}=\alpha'}(i) \;=\; \mathbb{E}_{m_i(\cdot|R_{r+1,j}=\alpha')}\left[\mathbb{1}_{\{i\}}\right]$.

Finally, the game continues as long as the measure $m_{ij}(\cdot|R_{r,l}=\alpha)$ is well defined for all allowed assignments $\alpha \in A_r$ to members of $R_r$. Hence we need to prove the following statements based on the Prover strategy. For the statements below, let $F_{X,\alpha_X}$ ( abbreviated to $F_X$ when $\alpha_X$ is irrelevant) denote the formula obtained by a $\{0,1\}$ assignment $\alpha_X$ to variables in $X$.

**Lemma 3.22** *Let $F_{X,\alpha}$ be a formula obtained by fixing variables in $X$, for an instance of MAX-k-CSP(P) for a promise predicate $P$. Let the constraint graph $G-X$ be $(3,k-1-\varepsilon)$-expanding. Then for any 0-1 assignment to a given variable $i \in G-X$ there exists an assignment $\alpha' \in \{0,1\}^{X\cup Cl(i)}$ such that $\alpha'_{|X} = \alpha$ and $\mu_c(\alpha') > 0$ for any constraint $c$ in $F$.*

**Proof:** The proof is an immediate corollary of Lemma 3.15 since we know $\mathbb{E}_{m_i(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\right] = \mathbb{E}_{m_i(\cdot|X=\alpha)}\left[1-\mathbb{1}_{\{i\}}\right] = 0.5 > 0$. Hence, there must exist an extension to the assignment $\alpha$ to $X \cup Cl(i)$ with the required property. ■

**Lemma 3.23** *The measure $m_{ij}(\cdot|X=\alpha)$ for allowed assignments $\alpha \in A_r$ to $X$ for any $X \in R_r$ is well defined, and therefore the protection matrices $Y^{(r)}$ are also well defined i.e. the entries of $Y^{(r)}$ lie in $[0,1]$, for all $r \le r_0$ where $r_0 = \Omega_{k,\varepsilon,\delta}(n)$ rounds.*

**Proof:** There are two scenarios which can lead to $m_{ij}(\cdot|X=\alpha)$ being undefined. First, it is possible that some $X$ itself is undefined since the closure is larger than the parameter $t$ for which we have $t$-consistency. However, we know from Lemma 3.6 that $|X \cup Cl(i)| \le O(kr/\varepsilon)$ (cf. Remark 3.18). Thus, since $t = \Omega_{k,\varepsilon,\delta}(n)$, the sets $X$ are well defined for $r = \Omega_{k,\varepsilon}(n)$ rounds.

Second, it is possible that for some value of variable $i$ i.e. 0 or 1, in round $r$ we may not be able to extend a 0-1 assignment $\alpha_i$ over $Cl(X \cup i)$ such that all $c \in \mathcal{C}(Cl(X\cup i))$ have $\mu_c(\alpha_i) > 0$. However, we can discount this possibility due to Lemma 3.22. Hence the statement follows. ■

Note that we have used the assumption about ordering of variables (Assumption 3.20) in the previous two Lemmas, but the rest of the proofs in this subsection will continue to hold without Assumption 3.20 as long as the conditioning set is closed and of small enough size i.e. $O_{k,\varepsilon,\delta}(n)$.

**Lemma 3.24** *The matrices $Y^{(r)}$ i.e. $Y^{(r)}(i,j) = \mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$, are positive semidefinite for all $X \in R_r$, $\alpha \in A_r$ and for some $r = \Omega_{k,\varepsilon,\delta}(n)$.*

**Proof:** Observe that both $G$ and $G-X$ are $k-1-\varepsilon$ expanding and so any constraint $c \in G-X$ has at least $k$ or $k-1$ neighbors in $G-X$. Given variables $i$ and $j$, recall that a constraint $c$ is *uncompromised* if it has at least $k-2$ neighbors in $G-X-\{i,j\}$. Otherwise, we say that $c$ is *compromised*. If the variables $i$ and $j$ do not belong to the same compromised constraint then by Lemma 3.15:

$$\mathop{\mathbb{E}}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right] = 0.25,$$

where $X \in R_r$, $\alpha \in \{0,1\}^X$. To take care of the remaining case i.e. if both variables $i$ and $j$ belong to some compromised constraint $c$, we decompose $Y^{(r)}$ into a single base matrix $B$ and a series of "correction" matrices $C_c$. The protection matrix $Y^{(r)}$ has the following properties:

- $Y^{(r)}(i,i) = Y^{(r)}(0,i) = Y^{(r)}(i,0) = 0.5$ and $Y^{(r)}(0,0) = 1$ for all $i \in [n]$.
- Else, $Y^{(r)}(i,j) = \mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$ for both $i,j \in c$ and the constraint $c$ is compromised.
- Otherwise $Y^{(r)}(i,j) = 0.25$.

Let $B = bb^T$ where $b \in \mathbb{R}^{n+1}$, $b_0 = 1$, $b_i = 0.5$ for $i \in [n] \setminus X$ and $b_i = \alpha_{|i}$ for $i \in X$. For every compromised constraint $c$ we define the $(k-1)\times(k-1)$ matrix $C_c(i,j) := [\mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}})\right] - 0.25]$ for $i,j \in c$. Since a variable can not belong to two compromised constraints due to high expansion, observe that $Y^{(r)} = B + \sum_c C_c$. Hence it suffices to prove that each $C_c$ is positive semidefinite. Let us consider a specific matrix $C_c$ from the LHS of the decomposition for $Y^{(r)}$. As long as $|Cl(X)|$ is small enough so that $G - X$ is $(k-1-\varepsilon)$ expanding and Lemma 3.15 applies, we get

$$\mathop{\mathbb{E}}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right] - 0.25 = \mathop{\mathbb{E}}_{m_{ij}(\cdot|X=\alpha)}\left[(\mathbb{1}_{\{i\}} - 0.5)(\mathbb{1}_{\{j\}} - 0.5)\right].$$

Let $S$ denote the union over all $i,j \in c$ of $Cl(X \cup i,j)$. As long as $|S|$ is small enough to meet the local consistency requirement of Lemma 3.13, which is true for $|X| = O_{k,\varepsilon,\delta}(n)$, we can rewrite the matrix $C_c$ as an expectation over measures so that the underlying measure no longer depends on $i$ and $j$, i.e.

$$C_c(i,j) = \mathop{\mathbb{E}}_{m_S(\cdot|X=\alpha)}\left[(\mathbb{1}_{\{i\}} - 0.5)(\mathbb{1}_{\{j\}} - 0.5)\right].$$

Therefore we have expressed $C_c$ as a positive linear combination of rank-one matrices of the form $(\overrightarrow{\mathbb{1}_{\{i\}} - 0.5})(\overrightarrow{\mathbb{1}_{\{i\}} - 0.5})^T$. Hence each $C_c$ is positive semidefinite and so is $Y^{(r)}$. ■

**Lemma 3.25** *$Y_i^{(r)}$ and $Y_0^{(r)} - Y_i^{(r)}$ can be expressed as a convex combination of elements in $O_r$ for all $r \le r_0$ where $r_0 = O_{k,\varepsilon,\delta}(n)$.*

**Proof:** We give the argument for $Y_i^{(r)}$ below. The argument for the column $Y_0^{(r)} - Y_i^{(r)}$. is identical. Let $y^{(r)}_{X=\alpha}$ (or simply $y^{(r)}$) be obtained by fixing $X = \alpha$ for some $X \in R_{r-1}$. Also, note that $X$ is a closed set and hence $y_i^{(r)} = 0.5$ for any $i \notin X$ by Lemma 3.15. Consider the column vector $Y_i^{(r)}/y_i^{(r)}$. For $j \in X$, we have $Y_{ij}^{(r)} = y_i^{(r)} \cdot \alpha_j$ and hence the corresponding entry in $Y_i^{(r)}/y_i^{(r)}$ is just $\alpha_j$. For $j \notin X \cup Cl(i)$, $i$ and $j$ can not be in the same compromised constraint and hence $Y_{ij}^{(r)} = 0.25$ by Lemma 3.15 which gives $Y_{ij}^{(r)}/y_i^{(r)} = 0.5$.

Thus, we only need to bother about $j \in Cl(i) \setminus X$. For such a $j$, we have

$$\frac{Y_{ij}^{(r)}}{y_i^{(r)}} = \frac{\mathbb{E}_{m_{ij}(\cdot|\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]}{\mathbb{E}_{m_i(\cdot|\alpha)}\left[\mathbb{1}_{\{i\}}\right]} = \frac{\sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_{i,j}=1,1}\prod_{c\in\mathcal{C}(X\cup Cl(i))}\mu_c(\beta)}{\sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_i=1}\prod_{c\in\mathcal{C}(X\cup Cl(i))}\mu_c(\beta)}$$

Let the product in the above terms be denoted by $p_\beta$. Then we have

$$\frac{Y_{ij}^{(r)}}{y_i^{(r)}} = \frac{\sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_i=1} p_\beta \cdot \beta_j}{\sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_i=1} p_\beta} = \frac{\sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_i=1} p_\beta \cdot \beta_j}{Z}$$

where $Z = \sum p_\beta$. Since this holds for all $j \in Cl(i) \setminus X$, we can write

$$\frac{Y_i^{(r)}}{y_i^{(r)}} = \frac{1}{Z} \cdot \sum_{\beta\in\{0,1\}^{X\cup Cl(i)},\beta|_X=\alpha,\beta_i=1} p_\beta z_\beta$$

where $(z_\beta)_j$ is $\beta_j$ for $j \in X \cup Cl(i)$ and 0.5 otherwise. Since all the vectors $z_\beta$ for which $p_\beta$ is non-zero correspond to elements of $O_r$, this completes the proof. ■

Finally, we can fix the expansion parameter $\varepsilon$ to a suitable constant (say 0.1) and conclude the main theorem below.

**Theorem 3.26** *Let $P$ be a promise predicate on $k$ variables, $\delta > 0$ and $F$ be a random instance of MAX-k-CSP(P) on $n$ variables for sufficiently large $n$. Then with probability $exp(-O(k^4 2^{2k}/\delta^4))$ over the choice of $F$, the $LS_+$ hierarchy has an integrality gap of $\frac{2^k}{P^{-1}(1)}(1-\delta)$ on $F$ even after $\Omega_{k,\delta}(n)$ rounds. Also, with the above probability, any $LS_+$ refutation of $F$ requires rank $\Omega_{k,\delta}(n)$.*

## 3.4 Size Lower Bound

In this subsection we use the techniques developed in the previous section, esp. Lemma 3.24, for proving rank lower bounds and ideas similar to [12] to prove a size lower bound for promise predicates in static-$LS_+$. We continue to assume that the variables in the ordered set $X$ are ordered before all other variables when computing closure i.e. Assumption 3.20 holds. Given a formula $F$ for a MAX-$k$-CSP(P) with promise predicate $P$ we let $F_X$ denote the formula obtained by fixing variables in a set $X$ and since the actual assignment to variables in $X$ will not be important we do not include it in the notation. Note that the constraint graph $G - X$ naturally corresponds to $F_X$.

**Theorem 3.27 (Degree Lower Bound)** *Let $F$ be an unsatisfiable instance of MAX-k-CSP(P) on $n$ variables, for a promise predicate $P$. Let $X$ be a subset of variables such that the formula $F_X$ with the constraint graph $G - X$ is $(r, k-1-\varepsilon)$ expanding for some $r = \Omega_{k,\varepsilon,\delta}(n)$. Then $F_X$ has (multilinear) degree $\Omega_{k,\varepsilon,\delta}(n)$ for any valid static-$LS_+$ refutation.*

**Proof:** Observe that any static-$LS_+$ refutation of $F$, with degree $\le d$, has the form (equation 2.1):

$$\sum_l \omega_l \cdot \varphi_{I_l,J_l} = -1,$$

where $\varphi_{I_l,J_l} = s_l \cdot \prod_{i\in I_l} x_i \cdot \prod_{j\in J_l}(1-x_j)$ for an axiom $s_l \ge 0$ and $\omega_l \in \mathbb{R}^+$. Let $Var(s_l)$ denote the set of variables in the axiom $s_l \ge 0$ and let $S_l := I_l \cup J_l \cup Var(s_l)$. We evaluate $\varphi_{I_l,J_l}$ at $\vec{y} \in \mathbb{R}^{\binom{n}{\le d}}$, where

$$y_S = \mathop{\mathbb{E}}_{m_S(\cdot|X=\alpha)} \left[ \left(\prod_{i\in S} \mathbb{1}_{\{i\}}\right) \right],$$

so that

$$\varphi_{I_l,J_l}(y) = \mathop{\mathbb{E}}_{m_{S_l}(\cdot|X=\alpha)}\left[(s_l(\mathbb{1})\prod_{i\in I_l}\mathbb{1}_{\{i\}}\prod_{j\in J_l}(1-\mathbb{1}_{\{j\}}))\right], \tag{3.4}$$

where $s_l(\mathbb{1})$ denotes the axiom $s_l$ evaluated by setting $x_i \in s_l$ to $\mathbb{1}_{\{i\}}$. Note that this evaluation at $\vec{y}$ defines a $\mathbb{R}$-valued linear map on the set of multilinear polynomials of degree at most $d$. The idea of this proof is to show that under this evaluation, which is well defined for $d = O_{k,\varepsilon,\delta}(n)$, the LHS of any refutation maps to some non-negative quantity while the RHS maps to $-1$ - a contradiction.

If $s_l$ is just an initial linear constraint then to show $\varphi_{I_l,J_l}(y) \geq 0$ we simply check that $m_{S_l}(\cdot|X=\alpha)$ is $\Omega_{k,\varepsilon}(n)$ locally consistent and supported only on satisfying assignments to $\mathcal{C}(S_l \cup X)$, both of which are true as long as $|S_l \cup X| \leq O_{k,\varepsilon,\delta}(n)$ (Lemma 3.13).

In the remaining case $s_l$ is the square of a linear form, say $s_l = (c_l^T x)^2$. Let $A_l := I_l \cup J_l$. Observe that the evaluation for $\varphi_{I_l,J_l} = s_l \cdot \prod_{i\in I_l} x_i \cdot \prod_{j\in J_l}(1-x_j)$ can be viewed as an expectation where all variables in $A_l$ are fixed to some 0-1 values i.e. $A_l = \beta$ for $\beta \in \{0,1\}^{A_l}$ such that each variable in $I_l$ is 1 and each variable in $J_l$ is 0. So to prove that $\varphi_{I_l,J_l}(y) \geq 0$ it suffices to prove that the following $n \times n$ matrix is positive semidefinite:

$$M_l(i,j) := m_{\{i,j\}\cup A_l}(x_i = 1, x_j = 1, A_l = \beta | X = \alpha),\ \forall i,j \in [n], \tag{3.5}$$

since $\varphi_{I_l,J_l}(y) = c_l^T M_l c_l$. However by local consistency (Lemma 3.13) we have,

$$m_{\{i,j\}\cup A_l}(x_i = 1, x_j = 1, A_l = \beta | X = \alpha) = \frac{m_{\{i,j\}\cup A_l \cup X}(x_i = 1, x_j = 1, A_l = \beta, X = \alpha)}{m_X(X=\alpha)}. \tag{3.6}$$

Local consistency also allows us to write

$$\begin{aligned} &m_{\{i,j\}\cup A_l\cup X}(x_i = 1, x_j = 1, A_l = \beta, X = \alpha) \\ &= \sum_{\substack{\overline{\beta}\in\{0,1\}^{Cl(A_l)} \\ \overline{\beta}_{|A_l} = \beta}} m_{\{i,j\}\cup Cl(A_l)\cup X}(x_i = 1, x_j = 1, Cl(A_l) = \overline{\beta}, X = \alpha)\,. \end{aligned} \tag{3.7}$$

Consider each term in the RHS of equation 3.7, they represent the following matrix as $i$ and $j$ vary over $[n]$:

$$M_{l,\overline{\beta}}(i,j) := m_{\{i,j\}\cup Cl(A_l)\cup X}(x_i = 1, x_j = 1, Cl(A_l) = \overline{\beta}, X = \alpha).$$

It suffices to show that all $M_{l,\overline{\beta}}$ are positive semidefinite. If $m_{Cl(A_l)\cup X}(Cl(A_l) = \overline{\beta}, X = \alpha)$ is positive i.e. matrix $M_{l,\overline{\beta}}$ is not all-zero, then we divide $M_{l,\overline{\beta}}$ by $m_{Cl(A_l)\cup X}(Cl(A_l) = \overline{\beta}, X = \alpha)$ and use local consistency to obtain:

$$M_{l,\overline{\beta}}(i,j) := m_{\{i,j\}}(x_i = 1, x_j = 1 | Cl(A_l) = \overline{\beta}, X = \alpha).$$

Two cases arise in this process:

1. $i \in Cl(A_l)$ or $j \in Cl(A_l)$. Assume without loss of generality that $i \in Cl(A_l)$. This case can be divided into two further cases:
   (a) $\overline{\beta}_{|i} = 0$ then the entire row and column for $i$ in $M_{l,\overline{\beta}}$ is 0 and we simply ignore the variable as far as the positive semidefiniteness of $M_{l,\overline{\beta}}$ is concerned.
   (b) $\overline{\beta}_{|i} = 1$ then two cases arise:

i. $j \in Cl(A_l)$ then $\overline{\beta}_j = 1$ otherwise we can ignore $j$. Moreover $M_{l,\overline{\beta}} = 1$.

ii. $j \notin Cl(A_l)$ then $M_{l,\overline{\beta}} = 0.5$ by Lemma 3.15.

2. $i, j \notin A_l$ then matrix $M_{l,\overline{\beta}}$ is the same as the matrix $Y$, restricted to unfixed variables, in Lemma 3.24.

So the matrix $M_{l,\overline{\beta}}$ looks exactly like the matrix $Y^{(r)}$ in Lemma 3.24, where the fixed variables in $Y^{(r)}$ correspond to the variables in $X \cup Cl(A_l)$. Therefore an argument exactly as in Lemma 3.24 can be used to show $M_{l,\overline{\beta}}$ is positive semidefinite and finish the proof. ■

Given the degree lower bound it is now easy to prove the size lower bound (cf. Grigoriev et al [11], Itsykson and Kojevnikov [12]) which implies Theorem 1.3.

**Theorem 3.28** *Let $P$ be a promise predicate on $k$ variables and $F$ be a random instance of MAX-k-CSP(P) on $n$ variables for sufficiently large $n$. Then with probability $exp(-O(k^4 2^{2k}))$ over the choice of $F$, any static-$LS_+$ refutation of $F$ requires size $exp(\Omega_k(n))$.*

**Proof:** By Theorem 3.3 the underlying constraint graph of $F$ is $\Omega_{k,\varepsilon,\delta}(n)$-expanding with pairwise independent support with the required probability. So it suffices to consider only such instances for this proof. Moreover, we assume that $\varepsilon$ and $\delta$ are fixed to some small constants for the purposes of this proof.

Let $M_G$ be the number of degree $\geq d$ terms of the form $\varphi$ in the static-$LS_+$ refutation of $G$. By pigeon-hole principle there exists a variable $x_i$ and a $\{0,1\}$ assignment $b_i$ to $x_i$ which sets to 0 i.e. "kills", at least $\frac{M_G d}{2n}$ of the $\varphi$s with degree $d$.

Ideally, we would like to use the previous idea repeatedly and "kill" all high degree inequalities and derive a low degree refutation of a smaller instance, where the constraint graph $G$ is expanding with pairwise independent support and then use Theorem 3.27 to prove our statement. However, we may end up with a $G$ which is not expanding or we may end up with a $G$ where some constraints do not support distributions that meet our conditions. We may even end up falsify a constraint when fixing the variables. In all the previously mentioned cases we will not obtain a valid static-$LS_+$ refutation of some instance of a formula which is $\Omega_{k,\varepsilon,\delta}(n)$ expanding with promise predicates. We can remedy the above problem by repeated expansion correction as used in the previous subsection (or see [11, 12]).

Starting with constraint graph $G_{new} := G$ we alternately fix a variable $x_i$ to "kill" $\geq \frac{M_G d}{2n}$ constraints of degree $\geq d$ and then fix variables in $Cl(i)$ to ensure that $G_{new} \leftarrow G - Cl(i)$ is expanding with pairwise independent support. The variables in $Cl(i) \setminus \{i\}$ can be fixed according to Lemma 3.22 to obtain a valid static-$LS_+$ refutation of the formula $F_{X_t}$, where $X_t$ is the set of fixed variables in $t$ steps of the form $\bigcup_{j \leq t} Cl(i_j)$ ($i_j$ is the variable fixed at step $j$).

From Theorem 3.27, we know that there exists constants $\delta_1$ and $\delta_2$, such that for $t \leq \delta_2 n$, any static $LS_+$ refutation of $F_{X_t}$ must require degree at least $d = \delta_1 n$. We continue the above process for $\delta_2 n$ steps. At each step we kill at least a $d/2(n - |X_t|) \geq d/2n$ fraction of terms of degree at least $d$. After $t_0 = \delta_2 n$ steps, we are left with a static $LS_+$ refutation of $F_{X_{t_0}}$ which has at most $M_G \cdot \left(1 - \frac{d}{2n}\right)^{\delta_2 n}$ terms of degree at least $d$. But a refutation of $F_{X_{t_0}}$ must still require degree $d$. Hence, there should be at least one term of degree at least $d$ remaining and we should have

$$M_G \cdot \left(1 - \frac{d}{2n}\right)^{\delta_2 n} \;=\; M_G \cdot \left(1 - \frac{\delta_1}{2}\right)^{\delta_2 n} \;\geq\; 1$$

which gives the required bound.

■

# 4 Extension to Almost All Formulae

Thus far all our claims hold for random instances of CSPs with promise predicates with a constant probability. In this section we sketch an approach which will allow us to generalize our main theorems to almost all random instances of CSPs with promise predicates. This section avoids details and is meant to mainly illustrate the idea of how local consistency could potentially be used to prove integrality gaps for more stronger SDP hierarchies. In order to accomplish our goal all we need to do is remove the assumption on the girth of our constraint graph in Theorem 3.3, since for a random instance the first two points of Theorem 3.3 hold with probability $1 - o(1)$ [6]. We will show that it is sufficient to replace the assumption of large girth by the assumption that there are no large sunflowers in our constraint graph.

**Definition 4.1** *Given a hypergraph $H$, a $(s,t)$-*sunflower *is a set of $t, (t \geq 2)$, hyperedges $E := \{h_i\}$ such that $\forall i, j \in E,\ h_i \cap h_j = S$ for a set $S$ of vertices of size $s$.*

Given a formula $F$ on a $k$-CSP(P), where $P$ is a promise predicate, its constraint bigraph $G$ naturally defines a hypergraph with variables as vertices and constraints as hyperedges. Observe that for any $(s,t)$-sunflower in such a hypergraph $s \leq 2$ (due to $k-2-\varepsilon$ expansion) and $t \leq O_{k,\varepsilon,\delta}(poly \log n)$ with high probability. This follows by a simple union bound. So from now onwards we replace the girth condition in Theorem 3.3 with: "$F$ contains no $(s, \Omega_{k,\varepsilon,\delta}(poly \log n))$-sunflower".

Now if we inspect closely the main tools of our lower bounds then we have essentially two main themes:

1. Local consistency (cf. Lemmas 3.13 and 3.11) and

2. Explicit evaluation i.e. construction of our positive semidefinite matrices (cf. Lemma 3.15).

A close inspection reveals that neither Lemma 3.13 or 3.11 depend on the girth of $G$ being large. Only case 2 in Claim 3.16 needed in Lemma 3.15 requires the girth to be large to ensure $k-2-\varepsilon'$ boundary expansion of $G - X - \{i,j\}$. Due to high boundary expansion a compromised constraint can not share two variables with an uncompromised constraint. Therefore, in absence of the girth assumption, Claim 3.16 may fail only when variables $i$ and $j$ belong to the same $(2,t)$-sunflower $\sigma$, where each hyperedge in $\sigma$ corresponds to an uncompromised constraint in $F_X$. Let us denote such sunflowers as *uncompromised sunflowers*. Therefore we can prove a (weaker) version of Lemma 3.15 without making the large girth assumption but as long as we additionally assume that variables $i$ and $j$ do not simultaneously belong to the same uncompromised $(2,t)$-sunflower. We take care of the two outlier cases i.e. variables $i$ and $j$ belong to the same compromised constraint or the same uncompromised $(2,t)$-sunflower, by using a few more correction matrices in Lemma 3.24. Note that the definition of our (protection) matrices will remain the same i.e. $Y^{(r)}(i,j) = \mathbb{E}_{m_{ij}(\cdot|X=\alpha)}\left[\mathbb{1}_{\{i\}}\mathbb{1}_{\{j\}}\right]$. We briefly sketch why the proof of Lemma 3.24 should continue to hold even with the weaker version of Lemma 3.15.

Each of the following is a straightforward result of $k-2-\varepsilon$, for small enough $\varepsilon > 0$, boundary expansion:

1. A variable in $F_X$ can not simultaneously belong to two compromised constraints. We already used this fact before.
2. A variable in $F_X$ can not simultaneously belong to a compromised constraint and any $(2, t)$-sunflower.
3. A variables in $F_X$ can not simultaneously belong to two uncompromised $(2, t)$-sunflowers.

Therefore we have a local structure (similar to a compromised constraint) which induces a partition among all the effected variables. So we can write small (polylogarithmic) sized correction matrices as in Lemma 3.24, and show using the same methods (essentially local consistency of measures) that each such correction matrix will be positive semidefinite. Thereby proving Lemma 3.24, which is sufficient to ensure that both the rank and size lower bounds extend to the case of almost all instances of random CSPs.

# 5 Open Problems

A couple of interesting problems remain for promise predicates. It would be interesting to prove rank lower bounds for many consecutive rounds of positive semidefinite steps without intermediate projection steps. Like prior work on resolution [5] and PC [2] we need high expansion (in fact higher) and it would be interesting to get the same bounds with moderately expanding CSPs.

# References


[1] Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards Strong Non-Approximability Results in the Lovász-Schrijver Hierarchy. In *STOC*, pages 294–303, 2005.

[2] Michael Alekhnovich and Alexander Razborov. Lower Bounds for Polynomial Calculus: Non-Binomial Case. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pages 190–199, Washington, DC, USA, 2001. IEEE Computer Society.

[3] Per Austrin and Elchanan Mossel. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity*, 18:249–271, 2009. 10.1007/s00037-009-0272-6.

[4] Eli Ben-Sasson and Russell Impagliazzo. Random CNFs are Hard for the Polynomial Calculus. *Comput. Complex.*, 19(4):501–519, December 2010.

[5] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow – resolution made simple. *J. ACM*, 48(2):149–169, March 2001.

[6] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP Gaps from Pairwise Independence. In *Theory of Computing*, To appear.

[7] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[8] Joshua Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank Bounds and Integrality Gaps for Cutting Planes Procedures. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 318–, Washington, DC, USA, 2003. IEEE Computer Society.

[9] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality Gaps for Sherali-Adams Relaxations. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 283–292, New York, NY, USA, 2009. ACM.

[10] Eden Chlamatac and Madhur Tulsiani. Convex Relaxations and Integrality Gaps. In *Handbook on Semidefinite, Cone and Polynomial Optimization*, 2010.

[11] Dima Grigoriev, Edward Hirsch, and Dmitrii Pasechnik. Complexity of Semi-algebraic Proofs. In *STACS 2002*, volume 2285 of *Lecture Notes in Computer Science*, pages 419–430, 2002.

[12] Arist Kojevnikov and Dmitry Itsykson. Lower Bounds of Static Lovász-Schrijver Calculus Proofs for Tseitin Tautologies. In *ICALP*, pages 323–334, 2006.

[13] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory.* Cambridge University Press, New York, NY, USA, 1995.

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

[15] László Lovász and Alexander Schrijver. Cones of Matrices and Set-Functions and 0-1 Optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[16] Toniann Pitassi and Nathan Segerlind. Exponential Lower Bounds and Integrality Gaps for Tree-like Lovász-Schrijver Procedures. In *SODA*, pages 355–364, 2009.

[17] Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, pages 593–602, Washington, DC, USA, 2008. IEEE Computer Society.

[18] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. A Linear Round Lower Bound for Lovász-Schrijver SDP Relaxations of Vertex Cover. In *IEEE Conference on Computational Complexity*, pages 205–216, 2007.

[19] Stefan Dantchev and Barnaby Martin and Mark Rhodes. Tight Rank Lower Bounds for the Sherali-Adams Proof System. *Theoretical Computer Science*, 410(21-23):2054 – 2063, 2009.

[20] Vijay V. Vazirani. *Approximation algorithms.* Springer, 2001.

# A Missing Proofs in Section 3

**Lemma A.1 ([6])** *If $G$ is $(r_1, e_1)$ expanding then $G - Cl(X)$ is $(r_2, e_2)$ expanding, where $X$ is a set of variables with $|X| \leq (e_1 - e_2)r_1$, $r_2 \geq r_1 - \frac{|X|}{e_1 - e_2}$ and $|Cl(X)| \leq \frac{k+2e_1-e_2}{e_1-e_2}|X|$.*

**Proof:** The proof is similar to the one in [6] except that "boundary neighbors" is replaced by "neighbors". We essentially repeat the argument below.

In algorithm Closure let $\xi_S$ be the value of $\xi$ when iteration $|S| = t$ terminates. We prove the loop invariant: $G - Cl(S)$ is $(\xi_S, e_2)$ expanding. Clearly the invariant holds in the beginning as $\xi > \xi_S$ and $e_1 > e_2$. Suppose $G - Cl(S) - \{x_j\}$ is not expanding after iteration $j$. Therefore there must exist a $M' \subseteq L$, disjoint from $M_j$, such that $|N(M')| \leq e_2|M'|$ and $|M'| \leq \xi - |M_j|$, so $|N(M_j \cup M')| \leq e_2|M_j \cup M'|$. But this contradicts the maximality of $M_j$ in the algorithm. Hence the invariant holds.

Next let $M = \bigcup_{j=1}^{t} M_j$ so $|M| = \sum_{j=1}^{t} |M_j| = r_1 - \xi_S$. Then by expansion of $G$: $e_1(r_1 - \xi_S) \leq |N(M)|$. Also since the $M_j$ are disjoint and $\forall j \leq t,\ |N(M_j)| \leq e_2|M_j|$, we have $|N(M)| \leq |S| + e_2(r_1 - \xi_S)$. Combining the two previous bounds for $|N(M)|$ we get our bound on $\xi_S \geq r_1 - \frac{|S|}{e_1 - e_2}$.
Also,

$$|Cl(S)| \leq |S| + \sum_{j=1}^{t} |N(M_j)| \leq |S| + \frac{e_2 + k}{2} \sum_{j} |M_j| \leq |S| + \frac{(e_2 + k)|S|}{2(e_1 - e_2)} = \frac{(2e_1 + k - e_2)|S|}{2(e_1 - e_2)}$$

as required.

Finally, placing $S = X$ and $\xi_S = r_2$ in the above we get the desired results in the required form. ■



ECCC ISSN 1433-8092

http://eccc.hpi-web.de