Linear Level Lasserre Lower Bounds for Certain k-CSPs *

Grant Schoenebeck†

Abstract

We show that for $k \geq 3$ even the $\Omega(n)$ level of the Lasserre hierarchy cannot disprove a random k-CSP instance over any predicate type implied by k-XOR constraints, for example k-SAT or k-XOR. (One constant is said to imply another if the latter is true whenever the former is. For example k-XOR constraints imply k-CNF constraints.) As a result the $\Omega(n)$ level Lasserre relaxation fails to approximate such CSPs better than the trivial, random algorithm. As corollaries, we obtain $\Omega(n)$ level integrality gaps for the Lasserre hierarchy of $\frac{7}{6} - \varepsilon$ for VertexCover, $2 - \varepsilon$ for k-UniformHypergraphVertexCover, and any constant for k-UniformHypergraphIndependentSet. This is the first construction of a Lasserre integrality gap.

Our construction is notable for its simplicity. It simplifies, strengthens, and helps to explain several previous results.

1 Introduction

The Lasserre hierarchy [Las01] is a sequence of semidefinite relaxations for certain 0-1 polynomial programs, each one more constrained than the last. The kth level of the Lasserre hierarchy requires that any set of k original vectors be self-consistent in a very strong way. If an integer program has n variables, the nth level of the Lasserre hierarchy is sufficient to obtain a tight relaxation where the only feasible solutions are convex combinations of integral solutions. This is because the nth level requires that the entire set of n vectors are consistent. If one starts from a k-CSP with poly(n) constraints, then it is possible to optimize over the set of solutions defined by the kth level of Lasserre in time $O(n^{O(k)})$, which is sub-exponential for $k = o(n/\log n)$.

The Lasserre hierarchy is similar to the Lovasz-Schrijver hierarchies [LS91], denoted LS and $LS+$ for the linear and semidefinite versions respectively, and the Sherali-Adams [SA90] hierarchy, denoted SA; however, the Lasserre hierarchy is stronger [Lau03]. The region of feasible solutions in ℓth level of the Lasserre hierarchy is always contained in the region of feasible solutions in ℓth level of LS, $LS+$, and SA[1]. A more complete comparison can be found in [Lau03]. While there have been a growing number of integrality gap lower bounds for the LS[ABL02, ABLT06, Tou06, STT], the $LS+$[BOGH+03, AAT05, STT07, GMPT06], and the SA[dlVKM07, CMM07] hierarchies, similar bounds for the Lasserre hierarchy have remained elusive.

*This work first appeared as [Sch08]. This version further simplifies some of the proofs including that of the main lemma, and uses slightly different notation.

†`grant@cs.berkeley.edu`, Computer Science Division, UC Berkeley. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Work for this paper was done while author was visiting Princeton University and Microsoft Research-Silicon Valley

[1]In our definition, for ease of presentation, the ℓth level of Lasserre for a k-CSP is only meaningful if $\ell \geq k$, but this can be modified.

The study of these hierarchies is motivated by the success of semidefinite programs in approximation algorithms. In many interesting cases, for small constant ℓ, the ℓth level of the Lasserre hierarchy provides the best known polynomial-time computable approximation. For example, the first level of the Lasserre hierarchy for the IndependentSet problem implies the Lovasz θ-function and for the MaxCut problem gives the Goemans-Williamson relaxation. The ARV relaxation of the SparsestCut problem is no stronger than the relaxation given in the third level of Lasserre.

In addition, recent work by Eden Chlamtac [Chl07] has shown improved approximation algorithms for coloring and independent set in 3-uniform hypergraphs. In [Chl07] the Lasserre hierarchy was used to find and/or analyze the constraints which led to improved approximations. This work is unlike the aforementioned work, where it was only later realized that the approximation results could be viewed as an application of semidefinite program hierarchies.

Integrality gap results for Lasserre are thus very strong unconditional negative results, as they apply to a "model of computation" that includes the best known algorithms for several problems.

1.1 Previous Lower-Bounds Work

While this is the only work known to us on Lasserre integrality gaps, results are already known about the weaker hierarchical models for several problems, including many problems we study here.

Buresh-Oppenheim, Galesy, Hoory, Magen and Pitassi [BOGH+03], and Alekhnovich, Arora, Tourlakis [AAT05] prove $\Omega(n)$ LS+ round lower bounds for proving the unsatisfiability of random instances of 3-SAT (and, in general, k-SAT with $k \geq 3$) and $\Omega(n)$[2] round lower bounds for achieving approximation factors better than $7/8 - \varepsilon$ for Max 3-SAT, better than $(1-\varepsilon)\ln n$ for Set Cover, and better than $k - 1 - \varepsilon$ for HypergraphVertexCover in k-uniform hypergraphs. They leave open the question of proving LS+ round lower bounds for approximating the Vertex Cover problem.

Much work has been done on Vertex Cover. Schoenebeck, Tulsiani, and Trevisan[STT] show an integrality gap of $2 - \varepsilon$ remains after $\Omega(n)$ rounds of LS, which is optimal. This build on the previous work of Arora, Bollobas, Lovasz, and Tourlakis [ABL02, ABLT06, Tou06] who prove that even after $\Omega(\log n)$ rounds the integrality gap of LS is at least $2-\varepsilon$, and that even after $\Omega((\log n)^2)$ rounds the integrality gap of LS is at least $1.5 - \varepsilon$.

Somewhat weaker results are known for $LS+$. The best known results are incomparable and were show by shown by Georgiou, Magen, Pitassi, and Tourlakis[GMPT06] and Schoenebeck, Tulsiani, and Trevisan [STT07]. The former result [GMPT06] builds on the previous ideas of Goemans and Kleinberg [KG98] and Charikar [Cha02], and shows that an integrality gap of $2 - \varepsilon$ survives $\Omega(\sqrt{\frac{\log n}{\log\log n}})$ rounds of LS+. The later result shows an integrality gap of $\frac{7}{6} - \varepsilon$ survives $\Omega(n)$ rounds. This result builds on past research which we review here as it is relevant for understanding the results of this paper.

The result of Feige and Ofek [FO06] immediately implies a $17/16 - \varepsilon$ integrality gap for one round of LS+, and the way in which they prove their result implies also the stronger $7/6 - \varepsilon$ bound. The standard reduction from Max 3-SAT to VertexCover shows that if one is able to approximate VertexCover within a factor better than 17/16 then one can approximate Max 3-SAT within a factor better than 7/8. This fact, and the $7/8 - \varepsilon$ integrality gap for Max 3-SAT of [AAT05], however do not suffice to derive an $LS+$ integrality gap result for VertexCover. The reason is that reducing an instance of Max 3SAT to a graph, and then applying a VertexCover relaxation to the graph, defines a semidefinite program that is possibly tighter than the one obtained by

[2]In all integrality gap containing an ε, the constant in the Ω depends on ε.

a direct relaxation of the Max 3-SAT problem. Feige and Ofek [FO06] are able to analyze the value of the Lovasz θ-function of the graph obtained by taking a random 3-SAT instance and then reducing it to an instance of IndependentSet (or, equivalently, of VertexCover).

For the Sherali-Adams hierarchy, Charikar, Makarychev, and Makarychev [CMM07] show that, for some ε, after n^ε rounds an integrality gap of $2 - o(1)$ remains.

Other results by Charikar [Cha02] and Hatami, Magen, and Markakis [HMM06] prove a $2 - o(1)$ integrality gap result for semidefinite programming relaxations of Vertex Cover that include additional inequalities. Charikar's relaxation is implied by the relaxation obtained after two rounds of Lasserre. The semidefinite lower bound of Hatami et al is implied after five rounds of Lasserre.

It was compatible with previous results that after a constant number of rounds of Lasserre the integrality gap for Vertex Cover could become $1 + o(1)$.

Our Result

The main result of this paper, is a proof that, for $k \geq 3$, the $\Omega(n)$th level of Lasserre cannot prove that a random k-CSP over any predicate implied by k-XOR is unsatisfiable. From this main results it quickly follows that the $\Omega(n)$th level of Lasserre:

- cannot prove a random k-XOR formula unsatisfiable.
- cannot prove a random k-SAT formula unsatisfiable.
- contains integrality gaps of $1/2 + \varepsilon$ for Max-k-XOR
- contains integrality gaps of $1 - \frac{1}{2^k} + \varepsilon$ for Max k-SAT.
- contains integrality gaps of $\frac{7}{6} - \varepsilon$ for VertexCover.
- contains integrality gaps of any constant for k-UniformHypergraphVertexCover.
- contains integrality gaps of $\Omega(1)$ for k-UniformHypergraphIndependentSet.

In addition to the power of our result, it is also very short and simple. It extends and simplifies results in [STT07] and [AAT05]. To a large extent it also explains the proofs of [FO06] and [STT07], and can be seen as being inspired by these results.

Road Map

In Section 2 we will define notation and provide background to our results. In Section 3 we will prove the main result. In Section 4 we will state and prove the remaining results, which are corollaries of the main result.

2 Background and Notation

We denote the set of Boolean variables $[n] = \{1, \ldots, n\}$. Let the range of variables be denoted $\mathbf{x} = \{x_i\}_{i \in [n]} = \{0,1\}^n$. For $I \subseteq \{1, \ldots, n\}$, let $\mathbf{x}_I = \{x_i\}_{i \in I}$ be the projection of $\mathbf{x}$ to the coordinates of I. We will consider problems where each constraint is local in that it is a k-junta, a function that depends on at most k variables. Formally:

Definition 1 *For $I \subseteq [n]$, let $\mathfrak{F}^I$ be the set of all function that only depend on the variable in I. That is there exists a function $f_{|I} : \mathbf{x}^I \to \{0,1\}$ such that $f(x) \equiv f_{|I}(x_I)$.*

*A k-*junta *f is a function $f : \mathbf{x} \to \{0,1\}$ that depends on at most k variables. Let $\mathfrak{F}^k$ be the set of k-juntas, then*

$$\mathfrak{F}^k = \bigcup_{\substack{|I| \leq k \\ I \subseteq [n]}} \mathfrak{F}^I$$

A *k-constraint* f is a k-junta that appears in the objective function or constraints of an optimization problem.

Sometimes we use $\mathbf{1}_f$ to denote $\mathbf{1}_f = \{x \in \mathbf{x} : f(x) = 1\}$.

Definition 2 *A k-constraint f* implies *another k-constraint g if $\mathbf{1}_f \subseteq \mathbf{1}_g$. We say that a predicate is XOR-implied if it is implied by either parity or its negation.*

For notational convenience, we will denote by $f_{(I)}^{=x_I}$ (or simply $f^{=x_I}$) the constraint where $f_{(I)}^{=x_I}(\bar{x}) = 1$ if $\bar{x}_I = x_I$ and 0 otherwise. We will denote by $\vec{\mathbf{1}}$ and $\vec{\mathbf{0}}$ the functions that are always and never true respectively (which are 0-juntas).

We will look at relaxations for two types of integer programs. In the first, we have a set of constraints, and would like to know if there is any feasible solution. In the second, we have a set of constraints and would like to maximize some objective function subject to satisfying the constraints. We formalize the notions here:

Definition 3 *A k-*constraint satisfiability problem *$\langle \mathbf{x}, \mathbf{C} \rangle$ is a set of n Boolean variables in the domain $\mathbf{x} = \{0,1\}^n$, and a set of k-constraints $\mathbf{C} = \{C_i\}_{i=1}^m$.*

Definition 4 *A k-*constraint maximization (or minimization) problem *$\langle \mathbf{x}, \mathbf{C}, M \rangle$ is a set of n Boolean variables in the domain $\mathbf{x} = \{0,1\}^n$, a set of k-constraints $\mathbf{C} = \{C_i\}_{i=1}^m$, and an objective function $M : \mathbf{x} \to \mathbb{R}$ to be maximized (or minimized) where $M = \sum_{j=1}^{\ell} \lambda_j f_j$ and each $\lambda_j \in \mathbb{R}$ and each f_j is a k-junta.*

Fourier Analysis Let $I \subseteq [n]$, then we define the character $\chi_I : \{0,1\}^n \to \{-1,1\} \subseteq \mathbb{R}$ as

$$\chi_I(x) = \prod_{i \in I} (-1)^{x_i} = (-1)^{\bigoplus_{i \in I} x_i}$$

Note that $\chi_I \cdot \chi_J = \chi_{I \triangle J}$. The *weight* of a character is the number of input variables on which its value depends. We use the following facts:

1. Any function $f : \{0,1\}^n \to \{0,1\} \subseteq \mathbb{R}$ can be written as

$$f(x) = \sum_{I \subseteq [n]} \hat{f}(I) \chi_I(x)$$

   where $\hat{f}(I) = \mathbb{E}_x\, f(x) \chi_I(x)$.

2. For any functions $f, g : \{0,1\}^n \to \{0,1\} \subseteq \mathbb{R}$ we have that $\widehat{f+g}(I) = \hat{f}(I) + \hat{g}(I)$.

3. For any functions $f, g : \{0,1\}^n \to \{0,1\} \subseteq \mathbb{R}$ we have that $\widehat{f \cdot g}(I) = \sum J \subseteq [n]\hat{f}(J)\hat{g}(I \triangle J)$

4. Fix $I \subseteq [n]$ and define $f : \{0,1\}^n \to \{0,1\} \subseteq \mathbb{R}$ as $f(x) = \bigoplus_{i \in I} x_i$ Then $\hat{f}(\emptyset) = \frac{1}{2}$, $\hat{f}(I) = -\frac{1}{2}$ and for $J \subseteq [n], J \notin \{\emptyset, I\}$ then $\hat{f}(J) = 0$.

Lasserre The Lasserre relaxation defined momentarily is designed to progressively restrict the feasible region of a constraint maximization (or minimization) problem $\langle \mathbf{x}, \mathbf{C}, M \rangle$ to be closer and closer to the convex hull of the integer solutions, in such a way that maximizing (or minimizing) over the feasible regions is still trackable.

Definition 5 *The rth round of Lasserre on the k-constraint maximization problem $\langle \mathbf{x}, \mathbf{C}, M \rangle$ is the semidefinite program with a variable v_f for every r-junta $f \in \mathfrak{F}^k$. Let $M = \sum_{i=1}^{\ell} \lambda_i f_i$ be the objective function. For reasons of convention, we will denote by v_0 the vector for the function $\vec{\mathbf{1}}$*

$$\max \sum_{i=1}^{\ell} \lambda_i ||v_{f_i}||^2$$

where

$$||v_0||^2 = 1 \tag{1}$$

$$\forall C \in \mathbf{C} \qquad ||v_C||^2 = 1 \tag{2}$$

$$\begin{array}{r} \forall f, g, f', g' \in \mathfrak{F}^k \text{ where} \\ f \cdot g \equiv f' \cdot g' \end{array} \qquad \langle v_f, v_g \rangle = \langle v_{f'}, v_{g'} \rangle \tag{3}$$

$$\begin{array}{r} \forall f, g, f+g \in \mathfrak{F}^k \text{ where} \\ f \cdot g \equiv \vec{\mathbf{0}} \end{array} \qquad v_f + v_g = v_{f+g} \tag{4}$$

The semidefinite program for the rth Lasserre round of a satisfiability problem is the same, but we only check for the existence of feasibility, we do not try to maximize over any objective function. [3]

First note that this is a relaxation, because any $\{0,1\}$ integer solution can be transformed into a $\{(0),(1)\}$ vector solution.

Now, given a distribution of integer solutions, we know that their exists an equivalent vector solution because each integer solution has an equivalent vector solution and the program in convex. We can easily create *explicit* vectors that satisfy the Lasserre constraints. If $(y_1, \ldots, y_n)$ is from a probability distribution of integral solutions, that is $(y_1, \ldots, y_n) = \sum_{j=1}^{m} p_j (z_1^j, \ldots, z_n^j)$ where $z_i^j \in \{0,1\}$, $z^j = (z_1^j, \ldots, z_n^j)$ are a feasible integral solutions, and $\sum_{j=1}^{m} p_j = 1$ then, for each possible k-junta $f \in \mathfrak{F}^k$ we can produce a vector.

$$v_f(j) = \begin{cases} \sqrt{p_j} & f(z^j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

[3]This definition is slightly different, but equivalent to other definitions of the kth level of the Lasserre hierarchy. The way that it is stated, it would require double exponential time to solve the rth level. This is easily remedied by only defining vectors for the **and** functions of up to r variables and using linear combinations of these vectors to define the remaining vectors. We present it like this for ease of notation.

These vectors will satisfy all the constraints of the Lasserre hierarchy at any level. If the reader is unfamiliar with the definition of the Lasserre hierarchy, then it is a straightforward and useful exercise to verify this fact.

While the Lasserre equations can be confusing, one general intuition is that the vectors define a probability distribution on any set of up to r coordinates (Equations 1, 3, and 4); that the probability distributions always satisfy the constraints (Equation 2); and that the probability distributions properly patch together (Equation 3). While global probability distributions map directly to vectors, vectors only map to local distributions (marginal distributions over r variables).

We will momentarily formalize this intuition, but first note that this intuition is not sufficient In applications, it is usually important that we have vectors and not simply local distributions that patch together. The fact that we have vectors gives some global orientation. The Goemans-Williamson MaxCut algorithm generates a global cut with a hyperplane. It is not clear how to do this with a local distributions alone.

If we define scalar variables p_f so that $p_f = ||v_f||^2$, and think of $(y_1, \ldots, y_n)$ as a probability distribution over integer solutions, then p_f is the probability that a randomly drawn solution satisfies the function f. Also we denote by v_{x_I} (or p_{x_I}) the vector (or "probability") corresponding to $f^{=x_I}$.

Claim 6 *Fix $I \subseteq [n]$ such that $|I| \leq r$. Then we can get probability distribution over the elements of $x_I \in \mathbf{x}_I$ by defining the "probability" of x_I, p_{x_I} to be $||v_{x_I}||^2$, where $v_{x_I} \equiv v_{f^{=x_I}}$ Actually, these vectors are all orthogonal, and if you sum over them, you get v_0.*

Proof: If $x_I, x'_I \in \mathbf{x}_I$, then v_{x_I} and $v_{x'_I}$ are orthogonal because $f^{=x_I} \cdot f^{=x'_I} = \vec{\mathbf{0}}$ and so by Equation 3 $\langle v_{x_I}, v_{x'_I} \rangle = ||\vec{\mathbf{0}}||^2$ and by Equation 4 $||v_{\vec{\mathbf{0}}}||^2 = 0$

Thus, by Equation 1 then Equation 4: $1 = ||v_{\vec{\mathbf{1}}}||^2 = ||\sum_{x_I \in \mathbf{x}_I} v_{x_I}||^2 = \sum_{x_I \in \mathbf{x}_I} ||v_{x_I}||^2$. So indeed we have a probability distribution. □

Claim 7 *If Equations 1, 3 and 4 are satisfied, then Equation 2 is equivalent to requiring that $||v_{x_I}||^2 = 0$ for all x_I where $x_I \notin \mathbf{1}_{C_{|I}}$ for some $C \in \mathbf{C} \cap \mathfrak{F}^I$.*

Proof: We only used Equations 1, 3 and 4 to show Claim 6. So we know that the v_{x_I} are all orthogonal and by Equation 4 additionally know that if $C \in \mathfrak{F}^I$ then $C = \sum_{x_I \in \mathbf{1}_{C_{|I}}} f^{=x_I}$ and so by Equation 4 we have that $v_C = \sum_{x_I \in \mathbf{1}_{C_{|I}}} v_{x_I}$. Putting these facts together we see.

$$\begin{aligned} 1 - ||v_C||^2 &= ||v_0||^2 - ||v_C||^2 = ||\sum_{x_I \in \mathbf{x}_I} v_{x_I}||^2 - ||\sum_{x_I \in \mathbf{1}_{C_{|I}}} v_{x_I}||^2 \\ &= \sum_{x_I \in \mathbf{x}_I} ||v_{x_I}||^2 - \sum_{x_I \in \mathbf{1}_{C_{|I}}} ||v_{x_I}||^2 = \sum_{x_I \notin \mathbf{1}_{C_{|I}}} ||v_{x_I}||^2 \end{aligned}$$

Thus $||v_C||^2 = 1$ if and only if $\sum_{x_I \notin \mathbf{1}_{C_{|I}}} ||v_{x_I}||^2 = 0$ □

Problems Studied Let $\mathcal{P} : \{0,1\}^k \to \{0,1\}$ be a Boolean predicates on k-variables. Let $\mathbf{R}^{n,k}$ be the set of all k tuples of dictators and anti-dictators such no two depend on the same variable. That is

$$\mathbf{R}^{n,k} = \{(f^{=b_1}_{(\{i_1\})}, \ldots, f^{=b_m}_{(\{i_m\})}) : \forall j \in n : i_j \in [n], b_j \in \{0,1\} \text{ and if } j \neq k \text{ then } i_j \neq i_k\}$$

In the language k-CSP-$\mathcal{P}$ input instances are an integer n and a tuple $\{R^1, \ldots, R^m\}$ where $R^j \in \mathbf{R}^{n,k}$. Each function $\mathcal{P} \circ R^J$ is called a clause or constraint. $(n, \{R^1, \ldots, R^m\} \in k$-CSP-$\mathcal{P}$ if there exists $x \in \mathbf{x}$ such that $\mathcal{P} \circ R^J = 1$ for each $j \in [m]$, and otherwise is not in the language. That is if all clauses can be simultaneously be satisfied. In Max-k-CSP-$\mathcal{P}$ we want to find the maximum number of clauses that can be satisfied simultaneously.

To sample a random instance of k-CSP-$\mathcal{P}$ with m clauses, we can uniformly and independently sample m elements of $\mathbf{R}^{n,k}$, to obtain the instance $(n, \{R^1, \ldots, R^m\}$.

k-XOR is just k-CSP-$\mathcal{P}$ where $\mathcal{P} \equiv \oplus_{i=1}^k x_i$. Not that we can always rewrite the constant $\mathcal{P} \circ R^m$ as $\oplus_{j \in I} x_i = b$ where $I \subseteq [n]$, $|I| = k$, $b \in \{0, 1\}$.

k-SAT is just k-CSP-$\mathcal{P}$ where $\mathcal{P} \equiv \vee_{i=1}^k x_i$.

Definition 8 *Given a predicates $\mathcal{P}$ we define $r(\mathcal{P})$ to be the probability that a random assignment satisfies $\mathcal{P}$.*

For example, in k-XOR, $r(k\text{-XOR}) = 1/2$. For example, in k-SAT, $r(k\text{-SAT}) = 1 - (\frac{1}{2})^k$.

In VertexCover we are given a graph $G = (V, E)$. There is a Boolean variable x_i for each vertex $i \in V$. For each edge $(i, j) \in E$ we have a constraint which says that both x_i and x_j cannot be zero. We are asked to minimize $\sum_{i \in V} x_i$.

In k-UniformHypergraphIndependentSet we are given a k-uniform hypergraph $G = (V, E)$. There is a variable x_i for each vertex $v \in V$. For each edge $(i_1, \ldots, i_k) \in E$ we have a constraint which says that not all $x_{i_1}, \ldots, x_{i_k}$ can be one. We are asked to maximize $\sum_{i \in V} x_i$.

k-UniformHypergraphVertexCover is the same as k-UniformHypergraphIndependentSet except that for each edge $(i_1, \ldots, i_k) \in E$ we have a constraint which says that at least one of $x_{i_1}, \ldots, x_{i_k}$ must be one. We are asked to minimize. $\sum_{i \in V} x_i$.

Background Results Sufficiently dense random k-CSP formulae are far from being satisfiable as the next proposition states.

Proposition 9 *For any $\delta > 0$, with probability $1 - o(1)$, if φ is a random k-CSP-$\mathcal{P}$ with Δn clauses where $\Delta \geq \frac{\ln 2}{2\delta^2} + 1$, at most a $r(\mathcal{P}) + \delta$ fraction of the clauses of φ can be simultaneously satisfied.*

Proposition 9 is well known in the literature, we provide a proof in the appendix for completion.

Definition 10 Width-w resolution *on an XOR formula φ, successively builds up new clauses by deriving a new clause $\bigoplus_{i \in I \Delta J} x_i = b \oplus b'$ whenever the symmetric difference $|I \Delta J| \leq w$ and the clauses $\bigoplus_{i \in I} x_i = b$ and $\bigoplus_{i \in I} x_i = b'$ had either already been derived or belong to φ.*

Width-w resolution proves a formula φ unsatisfiable if it derives the clause $0 = 1$. The following theorem shows that for random 3-XOR formula, even for quite large w, width-w resolution fails to produce a contradiction.

Theorem 11 *For any $k \geq 3$, $d > 0$, $\gamma > 0$, and $0 \leq \varepsilon < k/2 - 1$, there exists some constant $\alpha > 0$, such that if φ is a random k-XOR formula with density dn^{ε}, then with probability $1 - o(1)$ φ cannot be disproved by width $\alpha n^{1 - \frac{\varepsilon}{k/2 - \gamma - 1}}$ resolution nor can any variable be resolved to true or false. Furthermore, this is true even if the parity sign (whether the predicate is parity or its negation) of each clause is adversatively chosen.*

Wigderson and Ben-Sasson [BSW01] show that a variant of Theorem 11 holds for k-SAT formula. The proof of [BSW01] extends to show Theorem 11 using standard techniques. We include a proof in the appendix for completeness.

3 k-CSPs over XOR-Implied Predicates

We now present the main theorem of the paper.

Theorem 12 *Let $\mathcal{P}$ be a XOR-implied predicate. Then for every $\delta, \gamma, d > 0$ and $0 \le \varepsilon < k/2 - 1$ (such that if $\varepsilon = 0$, then $d \ge \frac{\ln 2}{2\delta^2} + 1$) there exists some constant $\alpha \ge 0$, such that with probability $1 - o(1)$, if φ is a random k-CSP-$\mathcal{P}$ with Δn clauses where $\Delta = dn^\varepsilon$ both the following are true:*

1. *at most a $r(\mathcal{P}) + \delta$ fraction of the clauses of φ can be simultaneously satisfied.*
2. *The $\alpha n^{1-\frac{\varepsilon}{k/2-1-\gamma}}$ level of the Lasserre hierarchy permits a feasible solution.*

This theorem implies integrality gaps for XOR-implied k-CSPs because the Lasserre relaxation cannot refute that all clauses can be simultaneously satisfied, but, in fact, at most $r(\mathcal{P}) + \delta$ clauses can be simultaneously satisfied. Notice that an algorithm that simply guesses a random assignment would expect to satisfy an $r(\mathcal{P})$ fraction of clauses in expectation. In particular this theorem shows that with high probability a random k-XOR formula cannot be refuted by $\Omega(n)$ rounds of Lasserre which gives an integrality gap of $1/2 + \delta$ for $\Omega(n)$ rounds of Lasserre for Max k-XOR by setting $\delta = \delta$; $d \ge \frac{\ln 2}{2\delta^2} + 1$; $\varepsilon = 0$; and $\gamma = \frac{1}{2}$. Also, this theorem shows that with high probability a random 3-CNF formula cannot be refuted by $\Omega(n)$ rounds of Lasserre which gives an integrality gap of $7/8 + \delta$ for $\Omega(n)$ rounds of Lasserre for Max k-SAT.

Theorem 12 follows almost immediately from Theorem 11, Proposition 9, and the following Lemma.

Lemma 13 (Main Lemma) *If a k-XOR formula φ cannot be disproved by width-w resolution, then the $\frac{w}{2}$th round of the Lasserre hierarchy permits a feasible solution.*

Proof:[of Theorem 12] Fix $\delta, \gamma, d, \varepsilon, \mathcal{P}$ as allowed in theorem statement, and let φ be a random k-CSP-$\mathcal{P}$ formula with Δn clauses where $\Delta = dn^\varepsilon$. By Proposition 9, 1) holds with probability $1 - o(1)$ because for sufficiently large n, $\Delta = dn^\varepsilon > \frac{\ln 2}{2\delta^2} + 1$.

We can write φ as a k-XOR formula φ_{XOR} so that $\varphi_{XOR} \Rightarrow \varphi$. Now the Lasserre relaxation for φ_{XOR} is strictly tighter than that for φ. Let α' be as guaranteed in Theorem 11 using k, d, γ, and ε as inputs so that by Theorem 11 we know that with probability $1 - o(1)$ it is the case that φ_{XOR} cannot be disproved by width-$\alpha' n^{1-\frac{\varepsilon}{k/2-\gamma-1}}$ resolution. Let $\alpha = \frac{\alpha'}{2}$. By Lemma 13, φ_{XOR} cannot be proven unsatisfiable by $\frac{\alpha'}{2} n^{1-\frac{\varepsilon}{k/2-\gamma-1}} = \alpha n^{1-\frac{\varepsilon}{k/2-\gamma-1}}$ rounds of Lasserre. Because the Lasserre relaxation for φ_{XOR} is tighter than that for φ_{XOR} it must be the case that φ cannot be proven unsatisfiable by Lasserre either. $\square$

Lemma 13 is the main original technical contribution of this work. In the rest of this section we first provide some intuition for the proof of Lemma 13 and then provide its proof.

For a first attempt to prove the lemma we can observe that for any particular set I of at most $w/2$ variables, we can construct vectors for all f as follows: 1) Run bounded width resolution to derive a set of constraints that any satisfying assignment must satisfy. 2) Consider the set SAT_I where

$$SAT_I = \left\{ x_I \in \mathbf{x}_I : \begin{array}{l} x_I \text{ satisfies all the constraints derived by} \\ \text{the resolution whose support is contained in } I \end{array} \right\}$$

Randomize over SAT_I and construct the vectors as we saw in Equation 5. That is each coordinate of v_{f_I} will correspond to an element of $x_I \in SAT_I$, and will be $\sqrt{1/|SAT_I|}$ if $f_I(x_I) = 1$ and 0 if $f_I(x_I) = 0$. These vectors will satisfy the Lasserre Equations 1, 2, and 4; however, these vectors will fail miserably to satisfy Equation 3 of the Lasserre constraints. We have set up valid local distributions; however, these distribution do not patch together consistently. The problem is that when the take the dot product of v_{x_I} and v_{x_J}, the values in each coordinate mean something completely different.

To remedy this misalignment we design a space of equivalence classes of characters of weight at most $w/2$ variables which we will use to index the coordinates of each vector. We will say that $\chi_I \sim \chi_J$ if for all assignments that satisfy the derived resolution clauses, χ_I determines χ_J and vice versa. For example, if φ contained the clause $x_1 \oplus x_2 \oplus x_3 = 0$ then $\chi_{\{1,2\}} \sim \chi_{\{3\}}$ because whatever $x_1 \oplus x_2$ is, x_3 must be the opposite. With some $\sim$ equivalent characters, fixing one character automatically fixes the $\sim$ equivalent character to the opposite value (as above). With other $\sim$ equivalent character, fixing one character automatically fixes the $\sim$ equivalent character to the same value. Using this fact, we can split each equivalence class of $\sim$ equivalent character into two parts, so that the $\sim$ equivalent clauses in each part always fix each other to the same value, and $\sim$ equivalent clauses in opposite parts always fix each other to the opposite value. We can arbitrarily label one part $+$ and the other $-$.

The vector corresponding to a function f will have in each coordinate (which corresponds to an an equivalence class of characters) the sum of the fourier coefficients of f of characters corresponding the characters in this equivalence class. (Each coefficient will be multiplied by ± 1 depending on its label). The intuition here is that characters of the same equivalence class are completely dependant on each other, but non-equivalent characters are completely independent. Note that only some of the coordinates are non-zero.

This relates to the aforementioned construction which satisfies equations 1, 2, and 4 because "locally" we have just taken a rotation! If we project onto only the relevant characters, then the mapping of our previously constructed vectors (that failed to satisfy Equation 3) to these new vectors is simply a rotation. This implies that all the Lasserre equations that were previously satisfied will still be satisfied (because all the irrelevant characters are set to 0 and thus will not affect the dot product).

For each $I \subseteq [n]$, $|I| \leq \frac{w}{2}$, there is a bijection between the set SAT_I and the equivalence classes of χ_J where $J \subseteq I$, because, intuitively, each time resolution derives a new relation, the dimension of each of these sets is reduced by 1.

In particular the vectors $v_{f_{(I)}^{=x_I}}$ for each $x_I \in SAT_I$ still form an orthogonal basis. And if you take the preimage of the vector v_0 (I is still fixed) it corresponds to randomizing over the $x_I \in SAT_I$.

One can develop this intuition into a proof by showing that if $f \in \mathfrak{F}^I$ and $g \in \mathfrak{F}^J$ then v_f and v_g behave well by projecting onto the classes containing the characters involving only variables of $I \cup J$ (these are the only possible non-zero coordinates), and rotating back into the basis of $|SAT_{I\cup J}|$. A previous proof follows this intuition (see [Sch08]. Here we present an easier proof of Lemma 13.

Proof:[Lemma 13]

Construction of Vectors

We first define a set $\mathcal{E}$ which later will be used to index the coordinates of the vectors.

Let φ be a k-XOR formula that has no width-w resolution. Let $\mathcal{C}$ be the collection of clauses generated by width-w resolution running on φ. Let $\mathcal{L}^w$ be all the characters of weight at most w. Let $\mathcal{F} \subseteq \mathcal{L}^w$ be the collection of linear functions cooresponding to the clauses of $\mathcal{C}$. That is if $\bigoplus_{i\in I} x_i = b_i \in \mathcal{C}$ then $\chi_I \in \mathcal{F}$.

Now consider the set $\mathcal{E} = \mathcal{L}^{\frac{w}{2}}/\mathcal{F}$. That is we partition $\mathcal{L}^{\frac{w}{2}}$ into equivalence classes where $\chi_I \sim_{\mathcal{F}} \chi_J \Leftrightarrow \chi_{I\Delta J} \in \mathcal{F}$

For each equivalence class $[\chi_I] \in \mathcal{E}$, we arbitrarily choose some $\chi_{I_0} \in [\chi_I]$ (for notational convenience, we always choose $\chi_\emptyset \in [\chi_\emptyset]$). We define a function $\pi : \mathcal{L}^{|w/2|} \cup \mathcal{F} \to \{+1, -1\}$ such that

$$\pi(\chi_I) = \begin{cases} +1 & \bigoplus_{i\in I\Delta I_0} x_i = 0 \in \mathcal{C} \\ -1 & \bigoplus_{i\in I\Delta I_0} x_i = 1 \in \mathcal{C} \end{cases}$$

Claim 14 *$\sim_F$ is an equivalence relations and π is well defined.*

We now define the vectors. Each vector will have a coordinate corresponding to the each element of $\mathcal{E}$. Let $e_{\chi_I} = \pi(\chi_I)e_{[\chi_I]}$ (where $e_{[\chi_I]}$ is the basis vector with a one in the coordinate corresponding to $[\chi_I]$). Let $f \in \mathfrak{F}^{\frac{w}{2}}$

$$v_f = \sum_{\chi\in\mathcal{L}^{\frac{w}{2}}} \hat{f}(\chi)e_{[\chi]}$$

so that

$$v_f([\chi_I]) = \sum_{\chi\in[\chi_I]} \pi(\chi)\hat{f}(\chi)$$

Proof that constructed vectors satisfy Lasserre constraints

We see that Equations 1 is satisfied by the observation that the fourier expansion of $\vec{\mathbf{1}}$ is 1 in the trivial character and 0 everywhere else. Therefore $v_0 = (1, 0, \ldots, 0)$ where the 1 is in the coordinate of $[\emptyset]$. Therefore $||v_0||^2 = 0$.

We show 2 is satisfied. If $C \in \varphi$ then $C \in \mathcal{C}$.

First assume that C is $\bigoplus_{i\in I} x_i = 1$. Then we must show that $||v_f||^2 = 1$ where $f = \bigoplus_{i\in I} x_i$. We note that $\chi_I \in \mathcal{F}$, $\pi(\chi_I) = -1$, and also recall that $f(x) = \frac{1}{2}\chi_\emptyset - \frac{1}{2}\chi_I = \frac{1}{2} - \frac{1}{2}\chi_I$. Thus $v_f = \frac{1}{2}e_{[\chi_\emptyset]} - \pi(\chi_I)\frac{1}{2}e_{[\chi_I]} = e_{[\chi_\emptyset]}$ because $\chi_\emptyset \sim_{\mathcal{F}} \chi_I$.

Similarly, let $C = \bigoplus_{i\in I} x_i = 0$. Then we must show that $||v_f||^2 = 1$ where $f = 1 - \bigoplus_{i\in I} x_i$, because f is a function that is to be always *satisfied*. We note that $\chi_I \in \mathcal{F}$, $\pi(\chi_I) = 1$, and also recall that $f(x) = 1 - (\frac{1}{2}\chi_\emptyset - \frac{1}{2}\chi_I) = \frac{1}{2} + \frac{1}{2}\chi_I$. Thus $v_f = \frac{1}{2}e_{[\chi_\emptyset]} + \pi(\chi_I)\frac{1}{2}e_{[\chi_I]} = e_{[\chi_\emptyset]}$ because $\chi_\emptyset \;{}_{\mathcal{F}}\chi_I$.

Equation 3 is satisfied because we can write $\langle v_f, v_g\rangle$ in terms of only $\widehat{f \cdot g}$, the fourier coefficients of $f \cdot g$. It will follow that if $f \cdot g \equiv f' \cdot g'$ then $\langle v_f, v_g\rangle = \langle v_{f'}, v_{g'}\rangle$ because the fourier coefficients of $f \cdot g$ and $f' \cdot g'$ are the same. For $[I] \in \mathcal{E}$, let $\hat{f}([\chi_I]) = \sum_{\chi \in [\chi_I]} \pi(\chi_I)\hat{f}(\chi)$. Then

$$\begin{aligned}
\langle v_f, v_g \rangle &= \sum_{[\chi_I] \in \mathcal{E}} \langle \hat{f}([\chi_I]), \hat{g}([\chi_I]) \rangle \\
&= \sum_{\chi \in \mathcal{L}^{\frac{w}{2}}} \pi(\chi)\hat{f}(\chi) \sum_{\theta \in [\chi]} \pi(\theta)\hat{g}(\theta) \\
&= \sum_{\chi \in \mathcal{L}^{\frac{w}{2}}} \pi(\chi)\hat{f}(\chi) \sum_{\psi \in \mathcal{F}} \pi(\chi\psi)\hat{g}(\chi\psi) \\
&= \sum_{\psi \in \mathcal{F}} \pi(\psi) \sum_{\chi \in \mathcal{L}^{\frac{w}{2}}} \hat{f}(\chi)\hat{g}(\chi\psi) \\
&= \sum_{\psi \in \mathcal{F}} \pi(\psi)\widehat{fg}(\psi)
\end{aligned}$$

The second line follows from expanding the summands. The third line follows from the fact that $[\chi] \subseteq \chi \cdot \mathcal{F}$ and because g is a $\frac{w}{2}$-junta, $\hat{g}(\chi\psi) = 0$ if the weight of $\chi\psi$ is greater than $\frac{w}{2}$. The fourth line follows because $\pi(\chi)\pi(\chi\psi) = \pi(\psi)$, and the fifth line from the fact that $\widehat{fg}(\psi) = \sum_{\chi \in \mathcal{L}^{\frac{w}{2}}} \hat{f}(\chi)\hat{g}(\chi\psi)$ because the full fourier expansions of f and g are captured by the characters of $\mathcal{L}^{\frac{w}{2}}$.

Equation 4, is satisfied because $\widehat{f+g}(\chi) = \hat{f}(\chi) + \hat{g}(\chi)$ so that $v_f + v_g = v_{f+g}$ for any functions $f, g, f+g \in \mathfrak{F}^{\frac{w}{2}}$. $\square$

Remark 1 *If the width-bounded resolution not only does not refute φ, but also does not fix any variable x_i to either true or false, then for every $i \in [n]$, $||v_{x_i}||^2 = \frac{1}{2}$. This is because $v_{x_i} = 1/2e_{[\chi_\emptyset]} + 1/2e_{[\chi_{\{i\}}]}$ and if x_i is not fixed than $\chi_\emptyset \not\sim \chi_{\{i\}}$.*

4 Extensions

We now mention the corollaries of Theorem 12 and its proof.

Corollary 15 *For every ε, there exists some constants $\alpha \geq 0$, such that the αn level of Lasserre, an integrality gap of $\frac{7}{6} - \varepsilon$ for* VertexCover *persists.*

The idea of the proof is to rewrite a 3-XOR formula φ as a vertex cover problem on a graph G_φ using the standard FGLSS reduction. We will do it in such a way that any vectors that satisfy the Lasserre relaxation for the 3-XOR instance φ will also satisfy the vertex cover Lasserre relaxation for G_φ.

To prove this corollary, we use the following lemma which states that for a certain type of transformations most of the Lasserre constraints continue to be satisfied:

Lemma 16 *Let $\langle \mathbf{x}, \mathbf{C}, M \rangle$ and $\langle \bar{\mathbf{x}}, \bar{\mathbf{C}}, \bar{M} \rangle$ be two constraint maximization or minimization problems. For $i \in [n]$, $\psi_i : \{0,1\}^{\bar{n}} \to \{0,1\}$ be a k-junta on $\bar{\mathbf{x}}$. Define $\psi : \{0,1\}^{\bar{n}} \to \{0,1\}^n$ as $\psi(\bar{x}) = (\psi_1(\bar{x}), \ldots, \psi_n(\bar{x}))$.*

If a collection of vectors $\{\bar{v}_{\bar{f}}\}_{\bar{f}}$ satisfy the Lasserre constraints after r rounds for $\langle \bar{\mathbf{x}}, \bar{\mathbf{C}}, \bar{M} \rangle$, then the collection of vectors $\{v_f\}_f$ where $v_f \equiv \bar{v}_{f \circ \psi}$ satisfy Equations 1, 3, and 4 for $\lfloor r/k \rfloor$ rounds of Lasserre.

Proof: That we only run for $\lfloor r/k \rfloor$ rounds of Lasserre makes all the vectors well-defined. Each constraint for which we define a vector depends on at most $\lfloor r/k \rfloor$, and so the corresponding vector depends on at most r variables.

We use the following standard identities.

- $\vec{\mathbf{1}} \circ \psi = \vec{\mathbf{1}}$
- $f \circ \psi + g \circ \psi = (f + g) \circ \psi$
- $(f \circ \psi) \cdot (g \circ \psi) = (f \cdot g) \circ \psi$

Now Equation 1 is satisfied because $||v_{\vec{\mathbf{1}}}||^2 = ||\bar{v}_{\vec{\mathbf{1}}\circ\psi}||^2 = ||\bar{v}_{\vec{\mathbf{1}}}||^2 = 1$.

Equation 4 is satisfied because $v_f + v_g = \bar{v}_{f\circ\psi} + \bar{v}_{g\circ\psi} = \bar{v}_{(f+g)\circ\psi} = v_{f+g}$

Equation 3 is satisfied because

$$\langle v_f, v_g \rangle = \langle \bar{v}_{f\circ\psi}, \bar{v}_{g\circ\psi} \rangle = \langle \bar{v}_{(f\cdot g)\circ\psi}, \bar{v}_0 \rangle = \langle \bar{v}_{(f'\cdot g')\circ\psi}, \bar{v}_0 \rangle = \langle \bar{v}_{f'\circ\psi}, \bar{v}_{g'\circ\psi} \rangle = \langle v_{f'}, v_{g'} \rangle$$

□

We now prove Corollary 15

Proof: [Corollary 15] Given a 3XOR instance φ with $\Delta n = m$ equation, we define the FGLSS graph G_φ of φ as follows: G_φ has $N = 4m$ vertices, one for each equation of φ and for each assignment to the three variables that satisfies the equation. We think of each vertex i as being labeled by a partial assignment to three variables $L(i)$. Two vertices i and j are connected if and only if $L(i)$ and $L(j)$ are inconsistent. For example, for each equation, the four vertices corresponding to that equation form a clique. It is easy to see that opt(φ) is precisely the size of the largest independent set of G_φ because there is a bijection between maximal independent sets and assignment to the n variables. Note that, in particular, the size of the largest independent set of G_φ is at most $N/4$, where $N = 4m$ is the number of vertices. Thus the smallest vertex cover of G_φ is $3N/4$ (because the complement of any independent set is a vertex cover).

Let $\gamma = 1/2$ and $d \geq \frac{\ln 2}{2\delta^2} + 1$. Then by Theorem 12, there exists an α such, for large enough n, that we can find a 3XOR formula over n variables that is at most $1/2 + \delta$ satisfiable and cannot be disproved by αn rounds of Lasserre. Let φ be such a formula. Now, we using Theorem 12 we construct the Lasserre vectors for the 3XOR problem φ.

The constraints for the Lasserre hierarchy for VertexCover on this graph G_φ, are defined over the vertices of this graph. Formally, let $\mathbf{x} = \{0,1\}^{V(G_\varphi)}$. Let $\mathbf{C}$ contain the constraint $x_i \vee x_j$ for each edge $(i,j) \in E(G_\varphi)$, so that the constraint is satisfied if and only if at least one of the vertices incident to the edge is in the cover. Let $M = \sum_{i \in V(G_\varphi)} x_i$. Then VertexCover is the 2-constraint minimization problem $\langle \mathbf{x}, \mathbf{C}, M \rangle$. We can convert it into a Lasserre instance using Definition 5.

Now defining $\psi_i = \neg L(i)$[4] , we employ Lemma 16 to construct Lasserre vectors. By Lemma 16, these vectors satisfy Equations 1, 3, and 4 for $\Omega(n)$ rounds of the Lasserre VertexCover relaxation.

We still must show that Equation 2 is satisfied. We must show that for each edge $(i,j) \in E(G_\varphi)$ that $||v_{i\wedge j}||^2 + ||v_{i\wedge\neg j}||^2 + ||v_{\neg i\wedge j}||^2 = 1$. By Claim 7 we can simply show that $||v_{\neg i\wedge\neg j}||^2 = 0$. Let $(i,j) \in E(G_\varphi)$, then

$$||v_{\neg i\wedge\neg j}||^2 = ||\bar{v}_{\neg i\wedge\neg j}||^2 = \langle \bar{v}_{L(i)}, \bar{v}_{L(j)} \rangle = 0$$

[4]That is $\psi_i(x) = 1$ if x is consistent with the label $L(i)$ and $\psi_i(x) = 0$ if x is inconsistent with the label $L(i)$

The first equality follows from Equation 3. The last equality is true because $L(i)$ and $L(j)$ contradict each other. We know this because i and j are joined by an edge.

Knowing that the Lasserre constraints are satisfied, we show that the objective function $\sum_{i\in V(G_\varphi)}||v_i||^2 = \frac{3N}{4}$. Four distinct vertices were created for each of the N clauses. We show that the sum of the $||v_i||^2$ over the four vertices in any clause is always 3. Let $C \in \varphi$ be such a clause, let $i_j : 1 \le j \le 4$ be the four vertices corresponding to C, and let $L(i_j)$ be the label corresponding to vertex i_j. Then $\sum_{j=1}^{4} \bar{v}_{L(i)} = v_0$ by Claim 6 and the fact that the vector corresponding to an unsatisfying assignment is $\vec{0}$. And so

$$\sum_{j=1}^{4} ||v_{i_j}||^2 = \sum_{j=1}^{4} ||\bar{v}_{\neg L(i_j)}||^2 = 3\sum_{j=1}^{4} ||\bar{v}_{L(i_j)}||^2 = 3$$

However, at most $(1/2+\varepsilon)n$ of the clauses of φ can be satisfied, and so G_φ has an independent set of at most $(\frac{1}{8}+\varepsilon)N$, and by taking the complement a vertex cover of size at most $\frac{7}{8}-\varepsilon$. We get the integrality gap of $(\frac{7}{8}+\varepsilon)N/(3N/4) = \frac{7}{6}-\varepsilon$ $\square$

Corollary 17 *For any constants k and c, there exists constants $\alpha, \delta \ge 0$, such that if H is a random Uniform Hypergraph of with n vertices and δn edges, then with probability $1-o(1)$, an integrality gap of c remains at the αn level of the k-*UniformHypergraphIndependentSet *Lasserre hierarchy.*

We will use the following well known proposition which is proved in the appendix for completeness:

Proposition 18 *For every $k \ge 3, \varepsilon > 0$, there exists $\delta > 0$, such that if H is a random k-uniform hypergraph with Δn edges, where $\Delta \ge \delta$, then with probability $1-o(1)$, H has no independent set of size εn, and, equivalently, H has no vertex cover of size $(1-\varepsilon)n$.*

Proof: Let $\varepsilon = \frac{1}{2c}$ and let δ be as in Proposition 18. Let H be a random uniform hypergraph with δn edges. By Proposition 18 we know that with high probability H has no independent set of size εn. We now must show that there exists a good solution to the Lasserre relaxation.

We note that the CSP instance is $\langle \mathbf{x}, \mathbf{C}, M\rangle$ where $\mathbf{x} = \{0,1\}^{V(H)}$, $M = \sum_{i\in V} x_i$, and for each edge $(v_1,\ldots,v_k) \in E(H)$ we add the constraint $\vee_{i=1}^{k}\neg x_i$ to $\mathbf{C}$ which we can transform into a Lasserre relaxation according to Definition 5. Note that any constraint of the form $\vee_{i=1}^{k}\neg x_i$ is implied by either $\oplus_{i=1}^{k} x_i = 1$ if k is even or $\oplus_{i=1}^{k} x_i = 0$ if k is odd. Consider then the k-XOR formula φ_H with Δn clauses which implies $\mathbf{C}$. We see that in each clause of φ the K-XOR constraint is random except for the constant. Thus, by Theorem 11 we know that φ cannot be disproved by width $\Omega(n)$ resolution and no single variable can be fixed. By Theorem 12 φ cannot be disproved by $\Omega(n)$ levels of Lasserre. Moreover by Remark 1 we have that $||v_i||^2 = 1/2$ for all i. Thus $M = \sum ||v_i||^2 = n/2$.

So the ratio of the Lasserre optimum to the actual optimum is $\frac{n/2}{\varepsilon n} = c$. $\square$

Corollary 19 *For any constants k and $\varepsilon > 0$, there exists constants $\alpha, \delta \ge 0$, such that if H is a random Uniform Hypergraph of with n vertices and δn edges, then with probability $1-o(1)$, an integrality gap of $2-\varepsilon$ remains at the αn level of k-*UniformHypergraphVertexCover *Lasserre hierarchy.*

The proof of Corollary 17 is very similar to that of Corollary 19

Proof: Let $\varepsilon = \frac{1}{2c}$ and let δ be as in Proposition 18. Let H be a random uniform hypergraph with δn edges. By Proposition 18 we know that with high probability H has no vertex cover of size $(1-\varepsilon)n$. We now must show that there exists a good solution to the Lasserre relaxation.

We note that the CSP instance is $\langle \mathbf{x}, \mathbf{C}, M \rangle$ where $\mathbf{x} = \{0,1\}^{V(H)}$, $M = \sum_{i \in V} x_i$, and for each edge $(v_1, \ldots, v_k) \in E(H)$ we add the constraint $\vee_{i=1}^{k} x_i$ to $\mathbf{C}$ which we can transform into a Lasserre relaxation according to Definition 5. Note that any constraint of the form $\vee_{i=1}^{k} x_i$ is implied by $\oplus_{i=1}^{k} x_i = 1$. Consider then the k-XOR formula φ_H with Δn clauses which implies $\mathbf{C}$. We see that in each clause of φ the K-XOR constraint is random except for the constant. Thus, by Theorem 11 we know that φ cannot be disproved by width $\Omega(n)$ resolution and no single variable can be fixed. By Theorem 12 φ cannot be disproved by $\Omega(n)$ levels of Lasserre. Moreover by Remark 1 we have that $||v_i||^2 = 1/2$ for all i. Thus $M = \sum ||v_i||^2 = n/2$.

So the ratio of the actual optimum to the Lasserre optimum is $\frac{(1-\varepsilon)n}{n/2} = \frac{1-\varepsilon}{2}$. $\square$

5 Acknowledgements

I would like to both acknowledge and thank Alexandra Kolla and Madhur Tulsiani for many fruitful discussions without which this work would not have been completed. Also, I would like to thank Moses Charikar, Eden Chlamtac, Uriel Feige, and Avi Wigderson for helpful thoughts and pointers. Finally, I would like to thank the anonymous FOCS reviewers for helpful comments.

6 Conclusion

We have shown the first known integrality gaps for Lasserre. On the one hand you can see the main theorem (Theorem 12) as showing gaps for problems that are already known or thought to be **NP**-hard. We say that a predicate A is *approximation resistant* if, given a constraint satisfaction problem over A predicates, it is **NP**-hard to approximate the fraction of such predicates which can be simultaneously satisfied better than the trivial algorithm which randomly guesses an assignment and returns the fraction of predicates it satisfies. In [**?**], Zwick shows that the only 3-CPSs which are approximation resistant are exactly those which are implied by parity or its negation. So, for $k = 3$, the main theorem applies exactly to those problems which we already know are **NP**-hard.

On the other hand, the main theorem applies to results that are known to be in **P**. Deciding if a k-XOR formula is satisfiable is equivalent to solving a set of linear equations over $\mathbb{F}_2$, which can be done with Gaussian elimination.

The corollaries show that this technique can be translated into many different settings, especially when there is a local "gadget" reduction from K-XOR.

References

[AAT05] Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovasz-Schrijver hierarchy. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 294–303, 2005. 1, 2, 3

[ABL02] Sanjeev Arora, Béla Bollobás, and László Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 313–322, 2002. 1, 2

[ABLT06] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006. 1, 2

[BOGH+03] Josh 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 IEEE Symposium on Foundations of Computer Science*, pages 318–327, 2003. 1, 2

[BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow-resolution made simple. *J. ACM*, 48(2):149–169, 2001. 8

[Cha02] Moses Charikar. On semidefinite programming relaxations for graph coloring and Vertex Cover. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms*, pages 616–620, 2002. 2, 3

[Chl07] Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS '07: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 691–701, Washington, DC, USA, 2007. IEEE Computer Society. 2

[CMM07] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations, 2007. manuscript. 1, 3

[dlVKM07] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–61, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics. 1

[FO06] Uriel Feige and Eran Ofek. Random 3CNF formulas elude the Lovász theta function. Manuscript, 2006. 2, 3

[GMPT06] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Tight integrality gaps for Vertex Cover SDPs in the Lovasz-Schrijver hierarchy. Technical Report TR06-152, Electronic Colloquium on Computational Complexity, 2006. 1, 2

[HMM06] Hamed Hatami, Avner Magen, and Vangelis Markakis. Integrality gaps of semidefinite programs for Vertex Cover and relations to ℓ_1 embeddability of negative type metrics. Manuscript, 2006. 3

[KG98] Jon M. Kleinberg and Michel X. Goemans. The Lovász Theta function and a semidefinite programming relaxation of Vertex Cover. *SIAM Journal on Discrete Mathematics*, 11:196–204, 1998. 2

[Las01] Jean B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th International IPCO Conference on Integer Programming and Combinatorial Optimization*, pages 293–303, London, UK, 2001. Springer-Verlag. 1

[Lau03] 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. 1

[LS91] L. Lovasz and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optimization*, 1(12):166–190, 1991. 1

[SA90] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxation between the continuous and convex hull representations. *SIAM J. Discret. Math.*, 3(3):411–430, 1990. 1

[Sch08] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS '08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–692. IEEE Computer Society, 2008. 1, 9

[STT] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of Vertex Cover and Max Cut. In *Proceedings of the 39th ACM Symposium on Theory of Computing*. Earlier version appeared as Technical Report TR06-132 on Electronic Colloquium on Computational Complexity. 1, 2

[STT07] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. A linear round lower bound for Lovasz-Schrijver SDP relaxations of Vertex Cover, booktitle = stoc07, note = "earlier version appeared as technical report tr06-098 on electronic colloquium on computational complexity.". 2007. 1, 2, 3

[Tou06] Iannis Tourlakis. New lower bounds for Vertex Cover in the Lovasz-Schrijver hierarchy. In *Proceedings of the 21st IEEE Conference on Computational Complexity*, 2006. 1, 2

7 Appendix

Proposition 20 *For any $\delta > 0$, with probability $1 - o(1)$, if φ is a random k-CSP chosen from the distribution $\mathcal{D}$ with Δn clauses where $\Delta \geq \frac{\ln 2}{2\delta^2} + 1$, at most a $r(\mathcal{D}) + \delta$ fraction of the clauses of φ can be simultaneously satisfied.*

PROOF: Fix an assignment to n variables. Now if we choose, $m = \Delta n$ clauses at random, the probability that more than a $r(\mathcal{D}) + \delta$ fraction of them are satisfied is at most $\exp(-2\delta^2 m) = \exp(-2\delta^2 \Delta n)$. To get this, we use the Chernoff Bound that says

$$\Pr[X \geq \mathbb{E}[X] + \lambda] \leq \exp(-2\lambda^2/m)$$

where X is the number of satisfied clauses, $\mathbb{E}[X] = r(\mathcal{D})m$, $\lambda = \delta m$. Picking a random formula and random assignment, the probability that more than a $r(\mathcal{D}) + \delta$ fraction of the clauses are satisfied is $\exp(-2\delta^2 \Delta n)$. Taking a union bound over all assignments, we get

$$\begin{aligned} \Pr[\text{any assignment satisfies} \geq (1/2 + \delta)m \text{ clauses}] &\leq \exp(-2\delta^2 \Delta n) \cdot 2^n \\ &= \exp(n(\ln 2 - 2\delta^2 \Delta) = \exp(-2\delta^2 n) \end{aligned}$$

because $\Delta \geq \frac{\ln 2}{2\delta^2} + 1$. $\square$

Theorem 21 *For $k \geq 3$, $d > 0$, $\gamma > 0$, and $0 \leq \varepsilon < k/2-1$, if φ is a random k-XOR formula with density dn^{ε}, then with probability $1-o(1)$ φ cannot be disproved by width $\alpha n^{1-\frac{\varepsilon}{k/2-\gamma-1}}$ resolution nor can any variable be resolved to true or false. Furthermore, this is true even if the parity sign (whether the predicate is parity or its negation) of each clause is adversatively chosen.*

We use the following Proposition:

Proposition 22 *For any $k \geq 3$, $d > 0$, $\gamma > 0$, and $0 \leq \varepsilon < k/2-1$, there exists $\beta > 0$ such that if φ is a random k-XOR formula with density dn^{ε} then with probability $1-o(1)$:*

1. *Every subformula $\varphi' \subseteq \varphi$ where $|\varphi'| \leq \beta n^{1-\frac{\varepsilon}{k/2-1}}$ is satisfiable even after fixing one variable.*
2. *For every subformula $\varphi' \subseteq \varphi$ where $|\varphi'| \in [\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}, \frac{2}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}]$, we have that $2V(\varphi')-k|\varphi'| \geq 2\gamma|\varphi'|$ where $V(\varphi')$ is the number of variables in φ'.*

Proof:[Theorem 21] Let φ be a random XOR formula as in the theorem statement and let C be any clause over the variables of φ. We define $\mu(C)$ to be the smallest size of a subformula $\varphi' \subseteq \varphi$ such that we can start from φ' and imply C using resolution. We note that in any resolution tree, if C_1 and C_2 together imply C_3, then $\mu(C_1) + \mu(C_2) \geq \mu(C_3)$.

From the first part of Proposition 22 we know that with high probability $\mu(0=1) \geq \beta n^{1-\frac{\varepsilon}{k/2-1}}$.

Now consider a resolution tree that derives $0 = 1$, that is, a contradiction. We will show that this tree must contain a clause C with many variables. By the subadditivity of μ as we move up the resolution tree, this tree must contain some clause C such that $\mu(C) \in [\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}, \frac{2}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}]$.

We will now show that with high probability C contains $\frac{\gamma\beta}{6}n^{1-\frac{\varepsilon}{k/2-\gamma-1}}$ variables and thus that the width of the resolution is at least as large. Let φ' be a subformula of size $\mu(C)$ which implies C. By the second part of Proposition 22 we know that $2V(\varphi') - k|\varphi'| \geq \gamma|\varphi'|$. Each variable of φ' must appear either in two of the clauses of φ' or in C itself. If a variable appears in one clause, but not in C; then no matter what the value of the other variables of that clause, the clause could still be satisfied by flipping this one variable. Therefore this clause can always be satisfied independently of the rest of φ' and is not required to imply C. This violates minimality of φ'. So

$$|C| + \frac{k}{2}|\varphi'| \geq V(\varphi') \Rightarrow |C| \geq \frac{1}{2}(2V(\varphi') - k|\varphi'|) \geq \gamma|\varphi'| \geq \frac{\gamma\beta}{3}n^{1-\frac{\varepsilon}{k/2-\gamma-1}}$$

so let $\alpha = \frac{\gamma\beta}{3}$.

To show that you cannot fix one variable to true or false the proof is almost exactly the same. Instead of showing that $\mu(0=1)$ is large, we show that for any x_i, $\mu(x_i = 0)$ and $\mu(x_i = 1)$ are large. This also follows from the first part of Proposition 22.

We note that we never used the parity of individual clauses in the proof, only the variables contained in each clause. Therefore the theorem still applies even if the parity of each clause is adversarially chosen. □

Proof:[Proposition 22] First we bound the probability that for a random formula φ, there exists a set of ℓ clauses containing a total of fewer than $c\ell$ variables by $(O(1)\frac{\ell^{k-c-1}}{n^{k-c-1-\varepsilon}})^{\ell}$;

We can upper bound the probability that there is a set of ℓ clauses containing a total of fewer than $c\ell$ variables by

$$\binom{n}{c\ell} \cdot \binom{\binom{c\ell}{k}}{\ell} \cdot l! \cdot \binom{m}{\ell} \cdot \binom{n}{k}^{-\ell}$$

where $\binom{n}{c\ell}$ is the choice of the variables, $\binom{\binom{c\ell}{k}}{\ell}$ is the choice of the ℓ clauses constructed out of such variables, $\ell! \cdot \binom{m}{\ell}$ is a choice of where to put such clauses in our ordered sequence of m clauses, and $\binom{n}{k}^{-\ell}$ is the probability that such clauses were generated as prescribed.

Using $\binom{N}{K} < (eN/K)^K$, $k! < k^k$, and $m = n^{1+\varepsilon}$ we simplify to obtain the upper bound $\left(O\left(\frac{\ell^{k-c-1}}{n^{k-c-1-\varepsilon}}\right)\right)^\ell$.

We first show that the first part of the proposition is true if we do not fix any variables. If $\varphi' \subseteq \varphi$ is a minimal unsatisfiable subformula of φ, then each variable that appears in φ' must occur twice in φ'. Otherwise the clause in which that variable appears is always satisfiable and φ' is not a minimal unsatisfiable subformula. Thus it is sufficient to show that no set of ℓ clauses contains fewer than $\frac{k}{2}\ell$ variables. We will show that if we set $c = k/2$ in the above formula, the sum over ℓ from 1 to $\beta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}$, can be made $o(1)$ with a sufficiently small β.

$$\sum_{\ell=1}^{\beta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}} \left(O\left(\frac{\ell^{\frac{k}{2}-1}}{n^{\frac{k}{2}-1-\varepsilon}}\right)\right)^\ell$$

Let δ be a sufficiently small constant, and let $\omega(n)$ be some function that grows in an unbounded fashion. We break up the above sum into:

$$\sum_{\ell=1}^{\delta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}\omega(n)^{-1}} \left(O\left(\frac{\ell^{\frac{k}{2}-1}}{n^{\frac{k}{2}-1-\varepsilon}}\right)\right)^\ell + \sum_{\ell=\delta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}\omega(n)^{-1}+1}^{\beta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}} \left(O\left(\frac{\ell^{\frac{k}{2}-1}}{n^{\frac{k}{2}-1-\varepsilon}}\right)\right)^\ell$$

We then bound each of these terms:

$$\sum_{\ell=1}^{\delta n^{(1-\frac{\varepsilon}{\frac{k}{2}-1})}\omega(n)^{-1}} \left(O\left(\frac{\ell^{\frac{k}{2}-1}}{n^{\frac{k}{2}-1-\varepsilon}}\right)\right)^\ell \leq \sum_{\ell=1}^{\infty} \left(O(1)(\delta\omega(n)^{-1})^{k-c-1}\right)^\ell = o(1)$$

for sufficiently small δ and sufficiently large n.

$$\begin{aligned}\sum_{\ell=\delta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}\omega(n)^{-1}+1}^{\beta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}} \left(O\left(\frac{\ell^{\frac{k}{2}-1}}{n^{\frac{k}{2}-1-\varepsilon}}\right)\right)^\ell &\leq \sum_{\ell=\delta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}\omega(n)^{-1}+1}^{\infty} \left(O(1)\beta^{k-c-1}\right)^\ell \\ &\leq \beta^{\delta n^{1-\frac{\varepsilon}{\frac{k}{2}-1}}\omega(n)^{-1}} \sum_{\ell=1}^{\infty} \left(O(1)\beta^{k-c-1}\right)^\ell = o(1)\end{aligned}$$

for sufficiently small β and sufficiently slowly growing $\omega(n)$.

Now we note that small subformulas are satisfiable even if we fix one variable. We can use all the above machinery, but now require that every set of ℓ clauses contains $\frac{k}{2}+1$ variables. However, this change is absorbed into the O constant in $\left(O\left(\frac{\ell^{k-c-1}}{n^{k-c-1-\varepsilon}}\right)\right)^\ell$ because in the above analysis when changing to $\binom{n}{c\ell-1} \cdot \binom{\binom{c\ell-1}{k}}{\ell} \cdot l! \cdot \binom{m}{\ell} \cdot \binom{n}{k}^{-\ell}$ we only get an addition factor of $\frac{c\ell-1}{ne}\left(\frac{c\ell}{c\ell-1}\right)^k$ the first factor helps and the second is bounded by 2^k which is a constant.

Now we show the second part of the Proposition.

We saw above that we can bound the probability that there exists a subformula of size ℓ that fails to satisfy $2V(\varphi') - k|\varphi'| \geq 2\gamma|\varphi'|$ by $\left(O\left(\frac{\ell^{\frac{k}{2}-\gamma-1}}{n^{\frac{k}{2}-\gamma-1-\varepsilon}}\right)\right)^{\ell}$. We will fix β later, and now use a union bound to upper bound the probability that there exists a clause φ' such that $|\varphi'| \in [\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}, \frac{2}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}]$ and $|V(\varphi')| \leq (\frac{k}{2}+\gamma)|\varphi'|$.

$$
\begin{aligned}
&\sum_{\ell=\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}}^{\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}} \left(O\left(\frac{\ell^{\frac{k}{2}-\gamma-1}}{n^{\frac{k}{2}-\gamma-1-\varepsilon}}\right)\right)^{\ell} \\
\leq\ & \left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)\left(O\left(\frac{\left(\frac{2}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)^{\frac{k}{2}-\gamma-1}}{n^{\frac{k}{2}-\gamma-1-\varepsilon}}\right)\right)^{\left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)} \\
\leq\ & \left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)\left(O(\frac{2}{3}\beta)^{k/2-\gamma-1}\right)^{\left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)} \\
\leq\ & \left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)(\frac{1}{2})^{\left(\frac{1}{3}\beta n^{1-\frac{\varepsilon}{k/2-\gamma-1}}\right)} = o(1)
\end{aligned}
$$

for a sufficiently small choice of β

□

Proposition 23 *For every $k \geq 3, \varepsilon > 0$, there exists $\delta > 0$, such that if H is a random k-uniform hypergraph with Δn edges, where $\Delta \geq \delta$, then with probability $1 - o(1)$, H has no independent set of size εn, and, equivalently, H has no vertex cover of size $(1-\varepsilon)n$.*

PROOF:

Let δ be such that $(1-\varepsilon)^{\delta} < \frac{\varepsilon}{e}$. Then the probability that H has an independent set of size εn (or has a vertex cover of size $(1-\varepsilon)n$) is bounded by the probability that there is a set of size εn such that no edge contains only vertices from this set:

$$
\binom{n}{\varepsilon n}(1-\varepsilon)^{\Delta n} \leq \left(\frac{e}{\varepsilon}\right)^{\varepsilon n}(1-\varepsilon)^{\delta n} \leq \left(\frac{e}{\varepsilon}\right)^{\varepsilon n}\left(\frac{\varepsilon}{e}\right)^{n} = \left(\frac{\varepsilon}{e}\right)^{(1-\varepsilon)n} = o(1)
$$

□