Approximate Constraint Satisfaction Requires Large LP Relaxations

Noah Fleming

April 19, 2018

Linear programming is a very powerful tool for attacking optimization problems. Techniques such as the ellipsoid method have shown that linear programs are solvable in polynomial time. Furthermore, it is known linear programming is P-complete. Therefore, if one was to show that some NP-hard problem admitted a polynomial-size linear program, then $\mathsf{P} = \mathsf{NP}$. In an attempt to rule out this approach, Yannakakis [4] gave a framework for proving lower bounds on a large class of linear programs known as *extended formulations*.

Consider the $\mathsf{3XOR}_n$ problem on n variables. It's not NP-hard, but it will serve as a good running example. An instance of $\Pi \in \mathsf{3XOR}_n$ consists of m parity constraints $\{P_1, \ldots, P_m\}$, $P_\ell : \{\pm 1\}^n \to \{0, 1\}$ where

$$P_\ell(x) := x_i \oplus x_j \oplus x_k = a_\ell, \quad \text{for } i, j, k \in [n] \text{ and } a_\ell \in \{\pm 1\}^n;$$

the goal is to maximize the number of constraints satisfied. Note that Π can also be represented uniquely as a multilinear polynomial over $\{\pm 1\}^n$ by taking the Fourier expansion. We can rewrite each $P_\ell(x) = x_i \oplus x_j \oplus x_k = a_\ell$ as

$$P_\ell(x) := \frac{1}{2} + \frac{1}{2}(-1)^{\frac{1-a_\ell}{2}} x_i x_j x_k.$$

The value of Π on some assignment $x \in \{-1, 1\}^n$ is given by

$$\Pi(x) = \frac{1}{m} \sum_{i \in [m]} P_i(x),$$

which is the fraction of constraints satisfied by assignment x. We will denote by

$$\mathsf{opt}(\Pi) = \max_{x \in \{-1,1\}^n} \Pi(x),$$

the largest fraction of constraints of Π satisfiable by any assignment $x \in \{-1, 1\}^n$.

If we want to express this as a linear program, then we need to linearize this function. To do this, we can associate some ordering to the $2\binom{n}{3}$ possible $\mathsf{3XOR}_n$ constraints, $P_1, \ldots, P_{2\binom{n}{3}}$.

A natural way of linearizing such a function is to associate with each $\mathsf{3XOR}_n$ instance Π on m vertices, a vector $\tilde{\Pi} \in \mathbb{R}^{2\binom{n}{3}}$, where the ith entry is $1/m$ if Π contains contains constraint P_i and 0 otherwise. Similarly, we can associate with each assignment $x \in \{-1,1\}^n$ a vector $\tilde{x} \in \mathbb{R}^{2\binom{n}{3}}$ in which $\tilde{x}_i = 1$ if $P_i(x) = 1$ and 0 otherwise. This satisfies, for every $\mathsf{3XOR}_n$ instance Π and assignment $x \in \{-1,1\}^n$, that

$$\langle \tilde{\Pi}, \tilde{x} \rangle = \Pi(x).$$

This lends itself to a natural linear program: let $\mathcal{P} \subseteq \mathbb{R}^{2\binom{n}{3}}$ be the convex hull of all $\tilde{x}$ for $x \in \{-1,1\}^n$; the linear program is given by

$$\mathcal{L}(\Pi) = \max_{y \in \mathcal{P}} \langle y, \tilde{\Pi} \rangle.$$

This polytope $\mathcal{P}$ has vertices corresponding to the points $\tilde{x}$ for $x \in \{-1,1\}^n$ and facets corresponding to the encodings $\tilde{\Pi}$ of all $\mathsf{3XOR}_n$ instances Π. Therefore, the value returned by optimizing over $\mathcal{P}$ will be $\mathsf{opt}(\Pi)$.

Unfortunately, the polytope $\mathcal{P}$ has an exponential number of facets and therefore cannot be optimized over efficiently. One possible way to overcome this issue is to find some new polytope $\mathcal{P}'$ in a higher dimensional space $\mathbb{R}^{d \geq n}$ with much fewer facets and such that there is a linear projection from $\mathcal{P}'$ down to $\mathcal{P}$. We could then optimize over the new polytope $\mathcal{P}'$ instead of optimizing over $\mathcal{P}$. Such a polytope $\mathcal{P}'$ is known as an *extended formulation* of the polytope $\mathcal{P}$.

The *size* of an extended formulation is the number of facets of the polytope, while the *extension complexity* of the base polytope $\mathcal{P}$, denoted $\mathsf{xc}(\mathcal{P})$ is the smallest extended formulation of $\mathcal{P}$. We stress that an extended formulation $\mathcal{P}'$ depends only on the input size and not the particular instance $\Pi \in \mathsf{3XOR}_n$ that we want to compute; the instance Π is defined only in he objective function.

Yannakakis gave a beautiful characterization of the extension complexity of a polytope in terms of the non-negative rank of its slack matrix. Consider a linear program $\mathcal{P}$ computing $\mathsf{3XOR}_n$. The *slack matrix* M^S has rows corresponding to the instances $\tilde{\Pi} \in \mathsf{3XOR}_n$, and columns corresponding to the vertices $\tilde{x}$ of $\mathcal{P}$. The entry at some row, column $(\tilde{\Pi}, \tilde{x})$ is the slack between that vertex and that instance,

$$M^S_{\tilde{\Pi},\tilde{x}} := \mathcal{L}(\Pi) - \langle \tilde{x}, \tilde{\Pi} \rangle,$$

where $\mathcal{L}(\Pi) = \max_{y \in \mathcal{P}} \langle y, \tilde{\Pi} \rangle$. The *non-negative rank* of a matrix M, denoted $\mathsf{rk}^+(M)$ is the smallest dimension r such that M can be written as a product of two non-negative matrices F and V with inner-dimension r.

Theorem 1. *(Yannakakis [4]) For any polytope $\mathcal{P}$, $\mathsf{xc}(\mathcal{P}) + 1 = \mathsf{rk}^+(\mathcal{P})$*

The Proof relies on Farkas' Lemma.

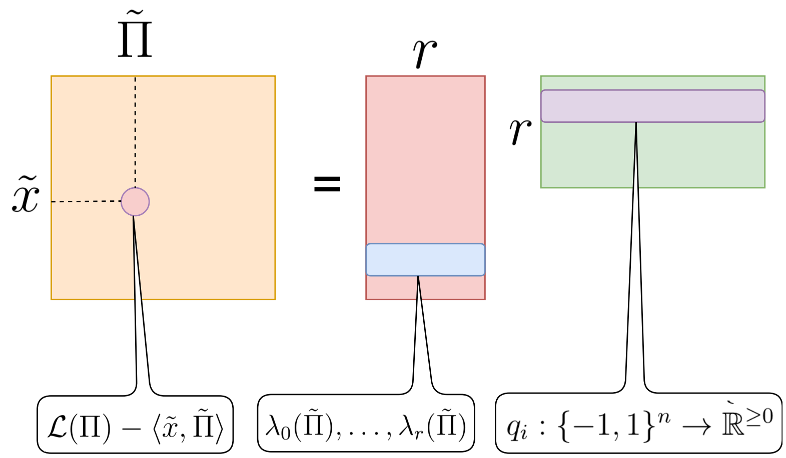

Figure 1: Representation of Theorem 1, the decomposition of the slack matrix into two non-negative matrices with inner dimension r.

Lemma 1. *(Farkas' Lemma) Let $\mathcal{P}$ be a with facets defined by inequalities $\{A_1x \leq b_1, \ldots, A_mx \leq b_r\}$ and let $Cx \leq d$ be an inequality that is valid for $\mathcal{P}$ (that is, every point $\alpha \in \mathcal{P}$ satisfies $C\alpha \leq d$) then there exists $\lambda_0, \ldots, \lambda_r \in \mathbb{R}^{\geq 0}$ such that*

$$d - Cx = \lambda_0 + \sum_{i=1}^{r} \lambda(b_i - A_ix)$$

We will only prove the forward direction, since it is all that we will need.

Proof. (of Theorem 1) Let $\mathcal{P}'$ be an extended formulation of $\mathcal{P}$ such that $\mathcal{P}'$ has r facets, defined by inequalities $A_1x \leq b_1, \ldots, A_rx \leq b_r$. Observe that for every $\Pi \in \mathsf{3XOR}_n$, the inequality $\mathsf{opt}(\Pi) - \langle \tilde{\Pi}, y \rangle \geq 0$ is valid for the polytope $\mathcal{P}$, for every $y \in \mathcal{P}$ and furthermore, that $\mathsf{opt}(\Pi) = \mathcal{L}(\Pi)$ because $\mathcal{P}$ computes $\mathsf{3XOR}_n$ exactly. Applying Farkas' Lemma, we can write

$$\mathcal{L}(\Pi) - \langle \tilde{\Pi}, y \rangle = \lambda_0(\tilde{\Pi}) + \sum_{i=1}^{r} \lambda_i(\tilde{\Pi}) \cdot (b_i - \langle A_i, y \rangle), \tag{1}$$

for some $\lambda_0(\tilde{\Pi}), \ldots, \lambda_r(\tilde{\Pi}) \in \mathbb{R}^{\geq 0}$. Now, because there is a linear projection from $\mathcal{P}$ to $\mathcal{P}'$, there is a vertex v of $\mathcal{P}'$ that projects to each vertex $\tilde{x}$ of $\mathcal{P}$. We will restrict to these vertices,

$$\mathcal{L}(\Pi) - \langle \tilde{\Pi}, \tilde{x} \rangle = \lambda_0(\tilde{\Pi}) + \sum_{i=1}^{r} \lambda_i(\tilde{\Pi}) \cdot (b_i - \langle A_i, v \rangle). \tag{2}$$

Furthermore, the $\tilde{x}$ are in one-to-one correspondence with the $x \in \{-1,1\}^n$, we can rewrite this as each $b_i - \langle A_i, v \rangle$ as a non-negative function $q_i : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$, where

$$q_i(x) = b_i - \langle A_i, v \rangle.$$

Therefore, we can rewrite equation 2 as

$$\mathcal{L}(\Pi) - \langle \tilde{\tilde{\Pi}}, \tilde{x} \rangle = \lambda_0(\tilde{\Pi}) + \sum_{i=1}^{r} \lambda_i(\tilde{\Pi}) \cdot q_i(x);$$

this is the slack between vertex $\tilde{x}$ and instance $\tilde{\Pi}$. We now construct the non-negative matrices V and F with inner dimension $r+1$. Let the rows of V be indexed by the $\tilde{\Pi}$ for $\Pi \in \mathsf{3XOR}_n$ and the columns of F be indexed by the $\tilde{x}$ for $x \in \{-1,1\}^n$. The $\tilde{\Pi}$th row of V will be the vector $[\lambda_0(\tilde{\Pi}), \ldots, \lambda_r(\tilde{\Pi})]$ corresponding to $\tilde{\Pi}$. The ith row of F is the truth table encoding of q_i, where the (i,j)th entry of F is the evaluation of $q_i(x)$. The $(r+1)$st row of F is the all 1 vector. This can be seen in figure 1. Therefore, the inner product between $V_{\tilde{\Pi}}$ and $F_{\tilde{x}}$ is

$$\lambda_0(\tilde{\Pi}) + \sum_{i=1}^{r} \lambda_i(\tilde{\Pi}) \cdot q_i(x) = \mathcal{L}(\Pi) - \langle \tilde{\Pi}, \tilde{x} \rangle.$$

□

Note: Because the extended formulation computes $\Pi \in \mathsf{3XOR}_n$ exactly, $\mathcal{L}(\Pi) = \mathsf{opt}(\Pi)$. Furthermore because the rows and columns of the slack matrix are in one-to-one correspondence between $x \in \{-1,1\}^n$ and $\Pi \in \mathsf{3XOR}_n$, the $(\tilde{x}, \tilde{\Pi})$th entry of the slack matrix is equivalent to

$$M^S_{\Pi,x} = \mathsf{opt}(\Pi) - \Pi(x),$$

because, $\langle \tilde{x}, \tilde{\Pi} \rangle = \Pi(x)$. Therefore, the slack matrix will be the same for *any* base polytope $\mathcal{P}$, the particular linearization is irrelevant. Therefore, more generally, we can define an extended formulation that exactly computes $\mathsf{3XOR}_n$ as a polytope $\mathcal{P} \subseteq \mathbb{R}^{d \geq n}$ such that

1. for every assignment $x \in \{-1,1\}^n$ there is a vector $\tilde{x} \in \mathcal{P}$ and for every instance $\Pi \in \mathsf{3XOR}_n$ there is a vector $\tilde{\Pi} \in \mathbb{R}^d$ such that

$$\Pi(x) = \langle \tilde{x}, \tilde{\Pi} \rangle$$

2. $\mathsf{opt}(\Pi) = \max_{y \in \mathcal{P}} \langle y, \tilde{\Pi} \rangle$ for every $\Pi \in \mathsf{3XOR}_n$.

The extension complexity of $\mathsf{3XOR}_n$, $\mathsf{xc}(\mathsf{3XOR}_n)$ is then the smallest extended formulation for $\mathsf{3XOR}_n$.

The key fact from Theorem 1 that we will use is that if an extended formulation $\mathcal{P}$ of size r computes $\mathsf{3XOR}_n$ then, for every instance Π, there exists a representation

$$\mathcal{L}(\Pi) - \Pi = \lambda_0(\Pi) + \sum_{i=1}^{r} \lambda_i q_i(\Pi),$$

where each q_i is a slack function of $\mathcal{P}$. We will call this representation an *extended formulation witness*, because it witnesses that $\mathcal{P}$ computes Π. From now on, we will write $\lambda_i(\Pi)$ as simply λ_i, where the dependence on Π is implicit.

Recall that the degree-d Sherali-Adams hierarchy computes an instance $\Pi \in \mathsf{3XOR}_n$ if $\mathsf{opt}(\Pi) - \Pi$ can be written as a non-negative linear combination of d-juntas, $\tilde{q}_i$,

$$\mathsf{opt}(\Pi) - \Pi = \sum_{i \in I} \lambda_i \tilde{q}_i.$$

This representations is superficially similar, and one might wonder if there is a way to approximate an extended formulation witness with a Sherali-Adams witness. Obviously, it would be too much to hope for that each of the non-negative functions in the extended formulation witness could be well approximated by a non-negative junta. Surprisingly, Chan, Lee, Raghavendra and Steurer [1] showed that after a specialized random restriction, the resulting q_i can be well approximated by non-negative d-juntas. Using this, they are able to *lift* Sherali-Adams lower bounds to extension lower complexity lower bounds. This transformation works for the class of constraint satisfaction problems (CSP), but we will prove it for the special case of $\mathsf{3XOR}_n$.

Theorem 2. *([1]) Suppose that the $d(n)$-round Sherali-Adams relaxation cannot compute $\mathsf{3XOR}_n$, then for all sufficiently large n, no extended formulation of size at most $n^{d(n)^2}$ can compute $\mathsf{3XOR}_N$ for some $N = n^{10d(n)}$*

We begin with the family of $\mathsf{3XOR}_N$ instances over N variables, and some extended formulation $\mathcal{P}$ of size r. By Yannakakis' Theorem above, each instance $\Pi \in \mathsf{3XOR}_N$, can be written as $\mathcal{L}(\Pi) - \Pi = \sum_{i=1}^{r} \lambda_i q_i$, where each q_i is a function $\{-1, 1\}^N \to \mathbb{R}^{\geq 0}$. Our goal is to write (a restriction of) $\mathcal{L}(\Pi) - \Pi$ as a non-negative linear combination of d-juntas plus some small error term. The proof proceeds in three steps.

1. First, we show that we can restrict our attention to q_i that are sufficiently smooth (the infinity norm of these functions is bounded).

2. Then, we show that each of these q_i can be approximated by an $N^{0.2}$-junta q_i', such that the error on the low degree Fourier coefficients of $q_i - q_i'$ is small. Here we crucially use the fact that degree-d Sherali-Adams can only reason about monomials of degree up to d. This step will incur some error, but we will show that this error goes to 0 as n goes to infinity.

3. Up until now, this proof has worked for any instance $\Pi \in \mathsf{3XOR}_N$. We will now fix a particular instance which will allow us to make the connection to Sherali-Adams lower bounds. Let $\Pi_0 \in \mathsf{3XOR}_n$ be a hard instance for Sherali-Adams on n variables. To obtain the instance $\Pi \in \mathsf{3XOR}_N$, we plant Π_0 at random inside a larger space of N variables by picking a subset of n of the variables and defining the constraints of Π_0 on them; the remaining $N - n$ variables will remain unconstrained. Finally, we argue that with high probability, the set of significant coordinates of q_i' when restricted to the variables on which Π_0 is define is at most d, and so the existence of this extended formulation implies that degree-d Sherali-Adams computes this instance exactly.

Fourier Analysis

We will need several tools from Fourier analysis. We will define the inner product between two n-variable functions f and g as

$$\langle f, g\rangle := \mathbb{E}_{x\in\{-1,1\}^n}[f(x)g(x)],$$

where the expectation is taken over the uniform distribution on $\{-1,1\}^n$. The Fourier representation of a function $f : \{-1,1\}^n \to \mathbb{R}$ is its unique representation over the basis of parity functions $\chi_\alpha := \prod_{i\in\alpha} x_i$ for $\alpha \subseteq [n]$. We can represent f over this basis as

$$f = \sum_{\alpha\subseteq[n]} \hat{f}(\alpha)\chi_\alpha,$$

where the fourier coefficient $\hat{f}(\alpha)$ is defined as f in the χ_α direction,

$$\hat{f}(\alpha) := \langle f, \chi_\alpha\rangle.$$

Intuitively, $\hat{f}(\alpha)$ measures the correlation of the variables $\Pi_{i\in\alpha}x_i$. Throughout this, we will use the functions regular representation and its Fourier representation interchangeably. Furthermore, if f is non-negative and $\mathbb{E}_{x\in\{-1,1\}^n}[f(x)] = 1$, then we can treat the Fourier coefficients of f as a distribution over $\{-1,1\}^n$.

Step 1: Smooth Slack Functions

We will now prove the main theorem by following the three steps laid out previously. Again, suppose that we have an extended formulation $\mathcal{P}$ of size $r \leq N^{d/2}$ which computes $\mathsf{3XOR}_N$ exactly. By Yannakakis' Theorem, for any $\Pi \in \mathsf{3XOR}_N$, we can write Π as a sum of non-negative slack functions,

$$\mathcal{L}(\Pi) - \Pi = \lambda_0 + \sum_{i=1}^{r} \lambda_i q_i,$$

where $\lambda_i \geq 0$ and $q_i : \{-1,1\}^N \to \mathbb{R}^{\geq 0}$. Furthermore, we can normalize each q_i and write it as $q_i(x) = \gamma_i q_i(x)$ for some $\gamma_i \in \mathbb{R}^{\geq 0}$ such that $\mathbb{E}[q_i] = 1$. That is,

$$\mathcal{L}(\Pi) - \Pi = \lambda_0 + \sum_{i=1}^{r}(\lambda_i\gamma_i)\cdot q_i.$$

Define the set

$$Q := \{i : \|q_i\|_\infty \leq N^d\},$$

of the q_i which are fairly smooth. Recall that d is the degree of the Sherali-Adams proof we are trying to obtain. We will show that restricting attention to the set of functions Q will only incur a small additive error. We can decompose the previous sum into

$$\mathcal{L}(\Pi) - \Pi = \lambda_0 + \sum_{i\in Q}(\lambda_i\gamma_i)\cdot q_i + \sum_{j\in[r]\setminus Q}(\lambda_j\gamma_j)\cdot q_j.$$

Because the value of $\mathsf{opt}(\Pi) - \Pi(x) \in [0,1]$ for every $x \in \{-1,1\}^N$ and $\Pi \in \mathsf{3XOR}_N$ and because $\lambda_j, \gamma_j \geq 0$ and q_j is non-negative and $\mathbb{E}[q_j] = 1$, we must have $\lambda_j \gamma_j \leq N^{-d}$ for every instance $\Pi \in \mathsf{3XOR}_N$. Because of this, $\sum_{j \in [r] \setminus Q} (\lambda_j \gamma_j) q_j$ cannot be very large and we will treat it as some small additive error term, which we will denote by $\varepsilon(\Pi)$. Later, we will bound its value,

$$\mathcal{L}(\Pi) - \Pi = \lambda_0 + \sum_{i \in Q} (\lambda_i \gamma_i) \cdot q_i + \varepsilon(\Pi).$$

Step 2: Approximate Functionals by High-Degree Juntas

The aim now is to show that the smooth slack functions q_i for $i \in Q$ can be well approximated by high-degree juntas. For this, we will use a density version of Chang's Lemma. The proof follows from the entropic proof of Chang's Lemma in Impagliazzo, Moore and Russell [2].

Lemma 2. *(Chang's Lemma) Let q be a density with entropy at least $N - t$ for some $t \geq 0$, let $\sigma > 0$ and define $R = \{\alpha : |\hat{f}(\alpha)| \geq \sigma 2^{-t}\}$. Then R spans a space of dimension less than $2t/\sigma^2$*

A consequence of Chang's Lemma is the following.

Lemma 3. *If q_i has entropy at least $N - d \log N$, then for any $\sigma > 0$, there exists a set $J(q_i) \subseteq [N]$ with*

$$|J(q_i)| \leq \frac{2d^2 \log N}{\sigma^2}$$

such that for every $\alpha \not\subseteq J(q_i)$ with $|\alpha| \leq d$, we have $|\hat{q}(\alpha)| \leq \sigma$.

Proof. Consider $S = \{|\alpha| \leq d : |\hat{q}(\alpha)| \geq \sigma\}$ and let S' be the maximal set of linearly independent elements in S. The density version of Chang's Lemma states that, after setting $t = d \log N$, that $|S'| \leq 2\sigma^{-2} d \log N$. Let $J(q_i) = \cup_{\alpha \in S'} \alpha$, then $|J(q_i)| \leq 2d^2 \log N / \sigma^2$ because each α contains at most d elements (it follows by linear independence that for all $\alpha \notin J(q_i)$ with $|\alpha| \leq d$, that $\hat{q}_i(\alpha) \leq sigma$. $\square$

This lemma says that we can decompose any high-entropy q_i into two parts q_i' and e_i, where

$$q_i' = \sum_{\alpha \subseteq J(q_i)} \hat{q}_i(\alpha) \chi_\alpha, \qquad \text{and} \qquad e_i = \sum_{\alpha \subseteq [N] \setminus J(q_i)} \hat{q}_i(\alpha) \chi_\alpha.$$

That is, q_i' depends only on the set of variables in $J(q_i)$ and in e_i, the Fourier coefficients of correlations up to degree-d are very small.

Beyond degree-d we have no control over the magnitude of the Fourier coefficients in e_i. However, recall that the degree-d Sherali-Adams hierarchy can only *perceive* correlations of degree up to d. Therefore, because our end goal is to convert this into a Sherali-Adams proof, this is a non-issue for us.

Therefore, if we could ensure that each of the q_i' were d-junta – that is, that $|J(q_i)| \leq d$, and that the extra error e_i was small, then the proof would be finished. We would have arrived at a representation of $\mathcal{L}(\Pi)-\Pi$ consisting of d-juntas plus some small additive error. Unfortunately, because we need $\sum_{i\in Q} e_i$ to tend to 0 as $n\to\infty$, it turns out that the largest that we will be able to set σ, and still achieve this is $\sigma = \left(\frac{16nd^2\log N}{\sqrt{N}}\right)^{1/2}$. Under Lemma 3, this only guarantees that each q_i' is an $(\sqrt{N}/8n)$-junta, which is approximately $N^{0.2}$ when the final numbers are plugged in.

Finally, we verify that each q_i with $i\in Q$ indeed has high enough entropy to satisfy the hypothesis of Lemma 3:

$$\begin{aligned} H(q_i) = \sum_{x\in\{-1,1\}^N} \frac{q_i(x)}{2^N}\log\left(\frac{2^N}{q_i(x)}\right) &\geq \left(\sum_{x\in\{-1,1\}^N}\frac{q_i(x)}{2^N}\right)\cdot\log\left(\frac{2^N}{\|q_i\|_\infty}\right) \\ &\geq \left(\sum_{x\in\{-1,1\}^N}\frac{q_i(x)}{2^N}\right)\cdot\log\left(\frac{2^N}{N^d}\right) = N - d\log N, \end{aligned}$$

where we used the fact that $\mathbb{E}[q_i]=1$. So far we have achieved a representation of the form

$$\mathcal{L}(\Pi)-\Pi = \lambda_0 + \sum_{i\in Q}(\lambda_i\gamma_i)\cdot(q_i'+e_i)+\varepsilon(\Pi),$$

where e_i are error terms whose Fourier coefficients corresponding to degree-up-to-d correlations are bounded by σ, and q_i are $\approx N^{0.2}$-juntas.

Step 3: Random Restriction to a Hard Instance for Sherali-Adams

The final step is to reduce the $N^{0.2}$-juntas to d-Juntas. To do this, we will employ a special random restriction which will restrict to an instance $\Pi_0 \in \mathsf{3XOR}_n$ for which we have Sherali-Adams lower bounds. Note that until this point, the steps of the proof have not relied on the particular instance of $\mathsf{3XOR}_N$. We will now restrict attention to a particular sub-family of instances. Consider an instance Π_0 of $\mathsf{3XOR}_n$ on n variables, where n is much smaller than N (Π_0 should be thought of as a hard instance for Sherali-Adams). To create our instance Π, we will randomly plant Π_0 inside a larger space of N unconstrained variables by picking a subset S of n variables and defining the constraints of Π_0 on them. The idea is that since the only constraints in Π are those corresponding to Π_0,

$$\mathcal{L}(\Pi) = \mathsf{opt}(\Pi) = \mathsf{opt}(\Pi_0).$$

Now, because each of the junta q_i' depend on at most $N^{0.2}$ variables, then if we restrict to the variables of Π_0, with high probability only a small fraction of the variables on which q_i depends will remain. This can be seen in figure 2. This will be done in the following lemma; recall that in step 2, using Chang's Lemma, we decomposed $q_i = q_i' + e_i$.

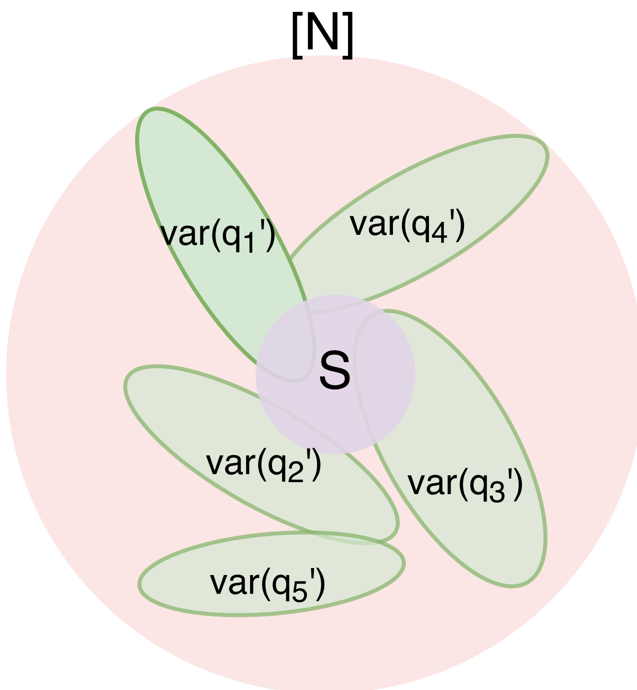

Figure 2: The intersection between the variable space of each $N^{0.2}$-junta q_i' and the restricted set S on which we will plant Π_0.

Lemma 4. *There exists a set $S \subseteq [N]$ of size n such that for each q_i with $i \in Q$, there is a set $J(q) \subseteq S$ with $|J(q)| \leq d$ such that*

$$|\hat{q}(\alpha)| \leq \left(\frac{16nd^2 \log N}{\sqrt{N}} \right)^{1/2},$$

for all $\alpha \subseteq S \setminus J(q)$ with $|\alpha| \leq d$.

For the proof, we will need the following inequality. Let $X_1, \ldots, X_n$ be i.i.d. $\{0,1\}$-random variables, with $\mathbb{E}[X_i] = p$. Then

$$\Pr\left[\sum_{i=1}^{n} X_i \geq t \right] \leq (pn)^t \tag{3}$$

Proof. We will choose the set S as follows:

1. Uniformly at random, pick a partition of $[N]$ into sets $S_1, \ldots, S_n$, each of size N/n.
2. For each variable $i \in [n]$, pick a variable v_i from S_i uniformly at random.
3. Let $S = \{v_i : i \in [n]\}$

In step 2 we argued, using Lemma 3, that we could decompose $q_i = q_i' + e_i$, where each q_i' is an $\sqrt{N}/8n$-junta which depends on a set of coordinates $J(q_i)$, and for every $\alpha \subseteq [N]$ with $|\alpha| \le d$, $|e_i(\alpha)| \le \left(\frac{16nd^2 \log N}{\sqrt{n}}\right)^{1/2}$. We will show that with some positive probability, the intersection of each of the sets $J(q_i)$ with the set S is at most d. For each variable $\ell \in J(q_i)$, let X_ℓ be the event $\ell \in S$. Then, $\mathbb{E}[q_i] = n/N$ because we are choosing each element of S uniformly at random, and so

$$\Pr[|J(q_i') \cap S| \ge d] = \Pr\left[\sum_{\ell \in J(q_i')} X_\ell \ge d\right] \le \left(\frac{n}{N} \cdot |J(q_i')|\right)^d \le \frac{1}{8^d N^{d/2}},$$

where the second inequality follows from inequality 3 above. Finally, because we have assumed that our original extended formulation is of size at most $N^{d/2}$, we have that $|Q| \le N^{d/2}$, and so taking a union bound over all $J(q_i)$ for $i \in Q$ completes the proof. $\square$

Finally, we construct the instance $\Pi \in \mathsf{3XOR}_N$ as follows: Let Π_0 be an instance of $\mathsf{3XOR}_n$ on n variables. Apply Lemma 4 to obtain a subset $S = \{v_1, \dots, v_n\} \subseteq [N]$. Define the constraints of Π as the constrains of Π_0 defined on the variables $\{v_1, \dots, v_n\}$; the remaining $N - n$ variables are left unconstrained.

Let $\mathbb{E}^*$ be the degree-d Sherali-Adams pseudo-expectation which achieves the optimal value on the Π_0,

$$\mathbb{E}^* = \mathsf{SA}_d[\Pi_0] = \max_{\tilde{\mathbb{E}} \sim d-PE} \tilde{\mathbb{E}}[\Pi_0],$$

where we think of $\Pi(x)$ as its representation as a multilinear polynomial so that we can apply $\mathbb{E}^*$ to it. Furthermore, we can represent each of the functions q_i as a multilinear polynomial by taking its Fourier transform. We will think of q_i as having that representation from now on so that we can apply $\mathbb{E}^*$ to them. We now plant $\mathbb{E}^*$ on the set of variables S, that is, we define $\mathbb{E}^*$ on the variables in S and extend it to have Fourier coefficient 0 on all terms outside of S. To do this, we note that Π is unconstrained on variables outside of S and therefore, we define the underlying pseudo-distribution to be uniform on all variables on outside of S. Applying it to both sides of equation **??** we arrive at

$$\mathbb{E}^*[\mathcal{L}(\Pi) - \Pi(x)] = \lambda_0 + \sum_{i \in Q} \lambda_i \gamma_i \cdot (\mathbb{E}^*[q_i'] + \mathbb{E}^*[e_i]) + \mathbb{E}^*\left[\varepsilon(\Pi)\right]$$

$$\mathcal{L}(\Pi) - \mathsf{SA}_d[\Pi_0] = \lambda_0 + \sum_{i \in Q} \lambda_i \gamma_i \cdot (\mathbb{E}^*[q_i'] + \mathbb{E}^*[e_i]) + \mathbb{E}^*\left[\varepsilon(\Pi)\right],$$

Now, because $\mathbb{E}^*$ gives non-zero value only on the variables of S, we have that

$$\mathbb{E}^*[q_i'] = \mathbb{E}^*[q_i' \restriction_S],$$

and so, by Lemma 4, we know that q_i' depends only on at most d variables in S, so it is a non-negative d-junta. Therefore, $\mathbb{E}^*[q_i'] \ge 0$, and so

$$\mathcal{L}(\Pi) \ge \mathsf{SA}_d[\Pi_0] + \sum_{i \in Q} \lambda_i \gamma_i \cdot \mathbb{E}^*[e_i] + \mathbb{E}^*\left[\varepsilon(\Pi)\right].$$

Now, if we can show the error terms, $\sum_{i\in Q} \lambda_i\gamma_i \cdot \mathbb{E}^*[e_i] + \mathbb{E}^*[\varepsilon(\Pi)]$ go to 0 as $n \to \infty$, then we will arrive at a representation of the form

$$\mathcal{L}(\Pi) \geq \mathsf{SA}_d[\Pi_0],$$

and so Sherali-Adams lower bounds will imply extension complexity lower bounds.

Bounding Error Terms

All that is left is to show that the error terms go to 0 as $n \to \infty$. We begin with bounding

$$\mathbb{E}^*[\varepsilon(\Pi)] = \sum_{j\in[r]\setminus Q} (\lambda_j\gamma_j) \cdot \mathbb{E}^*[q_j], \tag{4}$$

the error term that we obtained from Step 1. We will need a simple fact about pseudo-expectations

Claim 1. *For any degree-d pseudo-expectation $\mathbb{E}^*$ in its Fourier representation as a multilinear polynomial over $\{-1,1\}^n$, we have $\|\mathbb{E}^*\|_\infty, \sum_{\alpha\subseteq[n]} |\mathbb{E}^*[\chi_\alpha]| \leq \sum_{i=0}^d \binom{n}{i}$.*

Proof. The Fourier representation of $\mathbb{E}^*$ is

$$\mathbb{E}^* = \sum_{\alpha\leq d} \mathbb{E}^*[\chi_\alpha]\chi_\alpha,$$

where $\chi_\alpha = \prod_{i\in\alpha} x_i$. We know that because $\mathbb{E}^*$ is a pseudo-expectation that $\mathbb{E}^*[\chi_\alpha] \leq 1$ (this follows because a pseudo-expectation is the expectation over a pseudo-distribution). Because it is a degree-d pseudo-expectation, there can be at most $\sum_{i=0}^d \binom{n}{d}$ non-zero Fourier coefficients, each having absolute value at most 1. $\square$

We can apply this claim as follows: Viewing q_j as its Fourier representation as a multilinear polynomial, and noting that $\mathbb{E}^*$ assigns values to monomials of degree at most d we can write

$$\mathbb{E}^*[q_j] \leq \|\mathbb{E}^*\|_\infty \leq \sum_{i=0}^d \binom{n}{d},$$

the first inequality follows because the Fourier representation of $\mathbb{E}^*$ is $\mathbb{E}^* = \sum_{\alpha\leq d} \mathbb{E}^*[\chi_\alpha]$ and so the Fourier coefficient corresponding to the monomial χ_α is $\mathbb{E}^*[\chi_\alpha]$, the value that $\mathbb{E}^*$ assigns to χ_α. Therefore, we can view $\mathbb{E}^*$ as a vector whose αth place is the Fourier coefficient $\mathbb{E}^*[\chi_\alpha]$. Similarly, we view q_j as a vector, where the αth entry is the Fourier coefficient $\hat{f}(\alpha)$, which is the coefficient of χ_α in the representation of q_j as a multilinear polynomial. Therefore, taking the inner product between these two vectors $\langle \mathbb{E}^*, q_j\rangle$ gives the evaluation $\mathbb{E}^*[q_j]$. Because q_i is a density, that is $\mathbb{E}[q_j] = 1$ and $q_j \geq 0$, we can represent it as a distribution over assignments $x \in \{-1,1\}^n$, and associate with each such assignment a set which contains element i if $x_i = -1$. Then, $\langle \mathbb{E}^*, q_j \rangle = \mathbb{E}_{\alpha \sim q_i}[\mathbb{E}^*[\alpha] \leq \|\mathbb{E}^*\|_\infty$

During step 1, we argued that because the $q_j, \lambda_j, \gamma_j \geq 0$, that $\|q_i\|_\infty \geq N^d$ for every $i \in [r] \setminus Q$, and because $\mathcal{L}(\Pi) - \Pi_0 \in [0,1]$, we must have $(\lambda_j \gamma_j) < N^{-d}$ for each $j \in [r] \setminus Q$. Putting all of this together, we have

$$\mathbb{E}^* \left[\varepsilon(\Pi)\right] = \sum_{j \in [r] \setminus Q} (\lambda_j \gamma_j) \cdot \mathbb{E}^*[q_j] \geq -\sum_{i=1}^{d} \binom{n}{i} r N^{-d} \geq -\left(\frac{en}{d}\right)^d N^{-d/2},$$

where the second inequality follows because $|[r] \setminus Q| \leq r$, and the final inequality follows because r, the size of the extended formulation, is at most $N^{d/2}$.

Finally, we bound the error term $\sum_{i \in Q} (\lambda_i \gamma_i) \cdot \mathbb{E}^*[e_i]$ in a similar way, by noting that

$$\begin{aligned} |\mathbb{E}^*[e_i]| &\leq \sum_{\alpha \subseteq S} |\mathbb{E}^*[\chi_\alpha]| \cdot |\hat{e}_i(\alpha)| = \sum_{\alpha \subseteq S : |\alpha| \leq d} |\mathbb{E}^*[\chi_\alpha]| \cdot |\hat{e}_i(\alpha)| \leq \sum_{\alpha \subseteq S : |\alpha| \leq d} |\mathbb{E}^*[\chi_\alpha]| \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2} \\ &\leq \sum_{i=0}^{d} \binom{n}{d} \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2} \leq \left(\frac{en}{d}\right)^d \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2}, \end{aligned}$$

where the equality follows because $\mathbb{E}^*[\chi_\alpha] = 0$ for all $|\alpha| > d$, the third inequality follows from the bound we got on the size of the degree-up-to-d Fourier coefficients of e_i from Lemma 4, and the fourth inequality follows from Claim 1. Therefore, we have

$$\sum_{i \in Q} (\lambda_i \gamma_i) \cdot |\mathbb{E}^*[e_i]| \geq -\left(\frac{en}{d}\right)^d \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2} \sum_{i \in Q} (\lambda_i \gamma_i) \geq -\left(\frac{en}{d}\right)^d \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2},$$

where the final inequality follows from the observation that $\sum_{i=1}^{r} (\lambda_i \gamma_i) \leq 1$. To see this, note that $\mathsf{opt}(\Pi) - \Pi \in [0,1]$, and $\mathbb{E}[q_i] = 1$, and apply an expectation over assignments in $\{-1,1\}^n$ to both sides of $\mathsf{opt} - \Pi(x) = \lambda_0 + \sum_{i=1}^{r} (\lambda_i \gamma_i) \cdot q_i$.

Finishing Up

Finally, putting everything together, we have

$$\begin{aligned} \mathcal{L}(\Pi) - \mathsf{SA}_d[\Pi_0] &= \lambda_0 + \sum_{i \in Q} \lambda_i \gamma_i \cdot (\mathbb{E}^*[q_i'] + \mathbb{E}^*[e_i]) + \mathbb{E}^* \left[\varepsilon(\Pi)\right] \\ &\geq \lambda_0 + \sum_{i \in Q} (\lambda_i \gamma_i) \cdot \mathbb{E}^*[q_i'] - \left(\frac{en}{d}\right)^d \left(\frac{16nd^2 \log N}{\sqrt{N}}\right)^{1/2} - \left(\frac{en}{d}\right)^d N^{-d/2}. \end{aligned}$$

Therefore, we arrive at an expression of the form

$$\mathcal{L}(\Pi) \geq \mathsf{SA}_d[\Pi_0] - \mathsf{err}_n.$$

We now show that $\mathsf{err}_n := \left(\frac{en}{d}\right)^d \left(\frac{4d\sqrt{n\log N}}{N^{1/4}}\right) + \left(\frac{en}{d}\right)^d N^{-d/2}$ goes to 0 as $n \to \infty$. Plugging in our value for $N = n^{10d}$ we have

$$\begin{aligned}\mathsf{err}_n &= \left(\frac{en}{d}\right)^d \left(\frac{4d\sqrt{10dn\log n}}{n^{5d/2}} + n^{-5d}\right), \\ &= \left(\frac{e^d 4d\sqrt{10dn\log n} + 1}{d^d n^{3d/2}}\right), \\ &= o(1).\end{aligned}$$

Therefore, this theorem lifts Sherali-Adams degree lower bounds of up to to extended formulation lower bounds. Unfortunately, because we set need to set $N = n^{10d}$, the best lower bound that we can achieve this way (lifting a Sherali-Adams lower bound of degree $\Omega(n)$ is $N^{o\left(\frac{\log N}{\log\log N}\right)}$. The bottleneck, which limits Chan et al. to only obtaining quasi-polynomial size lower bounds is the application of Chang's Lemma in step 2. Later, Kothari, Meka and Raghavendra [3] overcame this barrier, obtaining truly exponential lower bounds.

References

[1] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. *J. ACM*, 63(4):34:1–34:22, 2016.

[2] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. An entropic proof of chang's inequality. *SIAM J. Discrete Math.*, 28(1):173–176, 2014.

[3] Adam R. Klivans, Pravesh K. Kothari, and Raghu Meka. Efficient algorithms for outlier-robust regression. *CoRR*, abs/1803.03241, 2018.

[4] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 223–228, 1988.