Approximate Constraint Satisfaction Requires Large LP Relaxations

Siu On Chan
MSR New England
siuon@cs.berkeley.edu

James R. Lee
University of Washington
jrl@cs.washington.edu

Prasad Raghavendra
UC Berkeley
prasad@cs.berkeley.edu

David Steurer
Cornell University
dsteurer@cs.cornell.edu

***Abstract*—We prove super-polynomial lower bounds on the size of linear programming relaxations for approximation versions of constraint satisfaction problems. We show that for these problems, polynomial-sized linear programs are exactly as powerful as programs arising from a constant number of rounds of the Sherali-Adams hierarchy.**

In particular, any polynomial-sized linear program for MAX CUT has an integrality gap of 1/2 and any such linear program for MAX 3-SAT has an integrality gap of 7/8.

***Keywords*-linear programming, extended formulations, lower bounds, LP hierarchies, constraint satisfaction problems, approximation complexity**

I. INTRODUCTION

Linear programming is one of the most powerful tools known for finding approximately optimal solutions to NP-hard problems. We refer to the books [Vaz01], [WS11] which each contain a wealth of examples. If $P \neq NP$, then for many such problems, polynomial-sized linear programs (LPs) that compute arbitrarily good approximations do not exist.

Thus a line of research has sought to prove lower bounds on the efficacy of small linear programs. The construction of integrality gaps for specific LPs has long been a topic of interest in approximation algorithms. Arora, Bollobás, and Lovász [ABL02] initiated a a more systematic study; they explored the limitations of LPs arising from lift-and-project hierarchies like those of Lovász and Schrijver [LS91] and Sherali and Adams [SA90]. There has now been an extensive amount of progress made in this area; one can see a sampling in the section on previous work.

Arguably, the ultimate goal of this study is to prove unconditional lower bounds for *every* sufficiently small LP. Since linear programming is P-complete under various notions of reduction, this would require proving that NP does not have polynomial-size circuits (see, e.g., the discussion in [Yan91]). But one could still hope to complete this program for LPs that use "natural" encodings of the underlying combinatorial problem.

We make progress toward this goal for the class of constraint satisfaction problems (CSPs). For instance, we prove that every polynomial-sized LP for MAX CUT has an integrality gap of $\frac{1}{2}$, answering a question from [BFPS12]. As another example, every such LP for MAX 3-SAT has an integrality gap of $\frac{7}{8}$. In fact, in both cases these integrality gaps hold for LPs of size $n^{\frac{\delta \log n}{\log \log n}}$ for some constant $\delta > 0$.

Corresponding upper bounds for both problems can be achieved by simple polynomial-sized LPs. For MAX 3-SAT, a $\frac{7}{8}$-approximation is best-possible assuming $P \neq NP$ [Hås01]. For MAX CUT, the famous SDP-based algorithm of Goemans and Williamson [GW95] achieves a 0.878-approximation. In this case, our result yields a strict separation between the power of polynomial-sized LPs and SDPs for a natural optimization problem.

To accomplish this, we show that for approximating CSPs, polynomial-sized LPs are exactly as powerful as those programs arising from $O(1)$ rounds of the Sherali-Adams hierarchy. We are then able to employ the powerful Sherali-Adams gaps that appear in prior work.

In Section I-B, we discuss our approach for the specific example of MAX CUT, including the class of LPs to which our results apply. Section II is devoted to a review of CSPs and their linear relaxations. There we explain our basic approach to proving lower bounds by exhibiting an appropriate separating hyperplane. We also review the Sherali-Adams hierarchy for CSPs. In Section III, we present the technical components of our approach, as well as the proof of our main theorem. Finally, Section IV contains an illustrative discussion of how Sherali-Adams gap examples can be used to construct corresponding gaps for *symmetric* LPs. This connection is quantitatively stronger than our result for general LPs.

A. History and context

Extension complexity. In a seminal paper, Yannakakis [Yan91] proved that every symmetric LP (i.e., one whose formulation is invariant under permutations of the variables) for TSP has exponential size. Only recently was a similar lower bound given for general LPs. More precisely, Fiorini, et. al. [FMP$^+$12] show that the extension complexity of the TSP polytope is exponential. Braun, et. al. [BFPS12] expand the notion of extension complexity to include approximation problems and show that approximating MAX CLIQUE within $O(n^{1/2-\varepsilon})$ requires LPs of size $2^{\Omega(n^{\varepsilon})}$. Building on that work, Braverman and Moitra [BM13] show that approximating MAX CLIQUE within $O(n^{1-\varepsilon})$ requires LPs of size $2^{\Omega(n^{\varepsilon})}$.

These three latter papers all use Yannakakis' connection between extension complexity and non-negative rank (see,

DOI 10.1109/FOCS.2013.45

e.g., [FMP+12], for a detailed discussion). They are based on increasingly more sophisticated analyses of a single family of slack matrices first defined in [FMP+12] (and extended to the approximation setting by [BFPS12]). A major contribution of the present work is that the connection between general LPs and the Sherali-Adams hierarchy allows one to employ a much richer family of hard instances.

LP and SDP hierarchies. As mentioned previously, starting with the works [ABL02], [ABLT06], the efficacy of LP and SDP hierarchies for approximation problems has been extensively studied. We refer to the survey of Laurent [Lau03] for a discussion of the various hierarchies and their relationships.

We mention a few results that will be quite useful for us. Fernández de la Vega and Mathieu [FdlVKM07] showed that for any fixed $\varepsilon > 0$ and k, MAX CUT has an integrality gap of $\frac{1}{2} + \varepsilon$ even after k rounds of the Sherali-Adams hierarchy. In a paper of Charikar, Makarychev, and Makarychev [CMM09], it is shown that MAX CUT and VERTEX COVER have integrality gaps of $\frac{1}{2} + \varepsilon$ and $2 - \varepsilon$, respectively, for $n^{\Omega(\varepsilon)}$ rounds of the Sherali-Adams hierarchy. In work of Schoenebeck [Sch08], tight bounds are given on the number of rounds needed to solve approximate k-CSPs in the Lasserre hierarchy (which, in particular, is stronger than the Sherali-Adams hierarchy). For instance, he shows that for every $\varepsilon > 0$, MAX 3-SAT has a $\frac{7}{8} + \varepsilon$ integrality gap even after $\Omega(n)$ rounds. There are also Sherali-Adams integrality gaps for CSPs with a pairwise independent predicate, due to Benabbas et. al. [BGMT12].

B. *Outline:* MAX CUT

We now present the basic details of our approach applied to the MAX CUT problem. To this end, consider a graph $G = (V, E)$ with $|V| = n$. For any $S \subseteq V$, we use

$$G(S) \stackrel{\text{def}}{=} \frac{|E(S, \bar{S})|}{|E|}$$

to denote the fraction of edges of G crossing the cut $(S, \bar{S})$. The maximum cut value of G is $\text{opt}(G) = \max_{S \subseteq V} G(S)$.

The standard LP. To construct an LP for computing (or approximating) $\text{opt}(G)$, it is natural to introduce variables $x = (x_1, x_2, \ldots, x_n) \in \{\pm 1\}^n$ corresponding to the vertices of G. One can then write, for instance,

$$\text{opt}(G) = \max_{x \in \{\pm 1\}^n} \frac{1}{|E|} \sum_{\{i,j\} \in E} \frac{1 - x_i x_j}{2} .$$

To convert this computation into a linear program, we need to *linearize* it.

The usual way is to introduce new LP variables $y = (y_{i,j}) \in \mathbb{R}^{\binom{n}{2}}$ meant to represent the quantities $(1 - x_i x_j)/2$. Now consider the vector $v_G \in \{0,1\}^{\binom{n}{2}}$ such that $(v_G)_{\{i,j\}} = 1$ precisely when $\{i, j\} \in E$. Given that we have linearized both the graph G and the cut variable x, we can consider the LP relaxation

$$\mathcal{L}(G) = \max_{y \in P} \langle v_G, y \rangle ,$$

where P is any polytope containing all the vectors y such that $y_{i,j} = (1 - x_i x_j/2)$ for some $x \in \{\pm 1\}^n$. The standard relaxation corresponds to a polytope P defined by the constraints $\{0 \leqslant y_{i,j} \leqslant 1 : i, j \in V\}$ and

$$\{y_{i,j} \leqslant y_{i,k} + y_{k,j},\ y_{i,j} + y_{i,k} + y_{k,j} \leqslant 2 : \\ i, j, k \in V\} .$$

Clearly P is characterized by $O(n^3)$ inequalities.

All linearizations are equivalent. But it is important to point out that, for our purposes, any linearization of the natural formulation of MAX CUT suffices. In fact, all such linearizations are equivalent after applying an appropriate linear transformation. We only require that there is a number $m \in \mathbb{N}$ such that:

1) For every graph G, we have a vector $v_G \in \mathbb{R}^m$.
2) For every cut $S \subseteq V$, we have a vector $y_S \in \mathbb{R}^m$.
3) For all graphs G and vectors y_S, the condition $G(S) = \langle v_G, y_S \rangle$ holds.

Now any polytope $P \subseteq \mathbb{R}^m$, such that $y_S \in P$ for every $S \subseteq V$, yields a viable LP relaxation: $\mathcal{L}(G) = \max_{y \in P} \langle v_G, y \rangle$. The *size* of this relaxation is simply the number of facets of P, i.e. the number of linear inequalities needed to specify P.

Remark I.1. We stress that the polytope P depends only on the input size. This is akin to lower bounds in non-uniform models of computation like circuits wherein there is a single circuit for all inputs of a certain size. The input graph G is used only to define the objective function being maximized. In other words, the variables and constraints of the linear program are fixed for each input size while the objective function is defined by the input. To the best of our knowledge, all linear and semi-definite programs designed for approximating max-CSP problems are subsumed by relaxations of this nature.

In Section III, we prove that every such relaxation of polynomial size has an integrality gap of $\frac{1}{2}$ for MAX CUT.

Toward this end, we recall a known *Sherali-Adams gap example:* In [FdlVKM07], [CMM09], it is shown that for every $\varepsilon > 0$ and every $d \in \mathbb{N}$, there are graphs G with $\text{opt}(G) \leqslant s$, but $\mathtt{SA}_d(G) \geqslant c$, where $s \leqslant \frac{1}{2} + \varepsilon$, and $c \geqslant 1 - \varepsilon$, and $\mathtt{SA}_d$ denotes the LP-value of the d-round Sherali-Adams relaxation (see Section II for the definition).

Proving a lower bound. In Theorem II.2, we recall that if there is an LP relaxation $\mathcal{L}$ of size R such that $\mathcal{L}(G) \leqslant c$ whenever $\text{opt}(G) \leqslant s$, then a simple application of Farkas' Lemma shows that there are non-negative functions $q_1, \ldots, q_R : \{\pm 1\}^n \rightarrow \mathbb{R}_{\geqslant 0}$ such that for every graph

G with $\mathrm{opt}(G) \leqslant s$, there are coefficients $\lambda_1, \ldots, \lambda_R \geqslant 0$ satisfying

$$c - G = \lambda_1 q_1 + \cdots + \lambda_R q_R \,. \tag{I.1}$$

(Note that we have earlier viewed G as a function on cuts and we now view it as a function on $\{\pm 1\}^n$ by associating these vectors with cuts.) These functions $q_i : \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ encode the slack of each constraint of the LP. Thus if the ith LP constraint is of the form $\langle A_i, z\rangle \leqslant b_i$, then $q_i(x) = b_i - \langle A_i, y_{S_x}\rangle$ where y_{S_x} is the cut vector corresponding to $x \in \{\pm 1\}^n$.

Our first step is to show that we can effectively truncate any q_i that is not sufficiently spread out (more technically that, when viewed as a distribution over $\{\pm 1\}^n$, has entropy far from n). Since we have only a small (say, polynomial) number of functions q_i, the effect of the truncation on (I.1) can be safely ignored. Thus we can assume that the distribution obtained by scaling q_i to be a probability measure has high entropy.

In Section III-A, we argue that any distribution over $\{\pm 1\}^n$ with large entropy is close to uniform off a small set of coordinates. In other words, every q_i corresponding to such a distribution can be thought of, roughly, as having a few significant coordinates such that if we condition on those, the function is close to uniform on the remaining coordinates.

Then in Section III-B, we employ a random restriction argument: By planting a small instance H at random inside the large graph G, we can ensure that for every q_i, the set of significant coordinates *when restricted to* H is much smaller. By a simple concentration inequality and union bound, there is some fixed planting of H such that this reduction happens for all the q_i's simultaneously.

Now is where our Sherali-Adams gap example enters the picture: An expression as in (I.1) is impossible for the $\mathtt{SA}_d$ gap instance whenever each q_i is a d-junta (i.e., depends on only d of its input variables); see Observation II.4 and the surrounding discussion. Intuitively, since we have argued that we only need to consider q_i's which are of the form "junta plus uniform," the Sherali-Adams lower bound can be made to apply to our functions as well. The key technical estimate involves sufficient control on this level of uniformity.

Of course, matters are made more delicate by the fact that these arguments need to take place simultaneously and the various intuitive properties discussed above are actually analytic in nature (e.g., "close to uniform" corresponds to a bound on the Fourier coefficients of q_i). The ingredients are all put together in Section III-C, where one can find the proof of our main theorem for general CSPs.

II. Background

We now review the maximization versions of boolean CSPs, their linear programming relaxations, and related issues.

Throughout the paper, for a function $f : \{\pm 1\}^n \to \mathbb{R}$, we write $\mathbb{E} f = 2^{-n} \sum_{x \in \{\pm 1\}^n} f(x)$. If $g : \{\pm 1\}^n \to \mathbb{R}$, we denote the inner product $\langle f, g\rangle = \mathbb{E}[fg]$ on the Hilbert space $L^2(\{-1,1\}^n)$. Recall that any $f : \{\pm 1\}^n \to \mathbb{R}$ can be written uniquely in the Fourier basis as $f = \sum_{\alpha \subseteq [n]} \hat{f}(\alpha)\chi_\alpha$, where $\chi_\alpha(x) = \prod_{i \in \alpha} x_i$ and $\hat{f}(\alpha) = \langle f, \chi_\alpha\rangle$. A function f is called a *d-junta* for $d \in [n]$ if f depends only on a subset $S \subseteq [n]$ of coordinates with $|S| \leqslant d$. In other words, f can be written as $f = \sum_{\alpha \subseteq S} \hat{f}(\alpha)\chi_\alpha$.

We say that f is a *density* if it is non-negative and satisfies $\mathbb{E} f = 1$. For such an f, we let μ_f denote the corresponding probability measure on $\{\pm 1\}^n$. Observe that for any $g : \{\pm 1\}^n \to \mathbb{R}$, we have $\mathbb{E}_{x \sim \mu_f}[g(x)] = \langle f, g\rangle$.

Constraint Satisfaction Problems. CSPs form a broad class of discrete optimization problems that include, for example, Max Cut and Max 3-Sat. For simplicity of presentation, we will focus on constraint satisfaction problems with a boolean alphabet, though similar ideas extend to larger domains (of constant size).

For a finite collection $\Pi = \{P\}$ of k-ary predicates $P \colon \{\pm 1\}^k \to \{0,1\}$, we let Max-$\Pi$ denote the following optimization problem: An instance $\Im$ consists of boolean variables $X_1, \ldots, X_n$ and a collection of Π-constraints $P_1(X) = 1, \ldots, P_m(X) = 1$ over these variables. A Π-constraint is a predicate $P_0 \colon \{\pm 1\}^n \to \{0,1\}$ such that $P_0(X) = P(X_{i_1}, \ldots, X_{i_k})$ for some $P \in \Pi$ and *distinct* indices $i_1, \ldots, i_k \in [n]$. The goal is to find an assignment $x \in \{\pm 1\}^n$ that satisfies as many of the constraints as possible, that is, which maximizes

$$\Im(x) \stackrel{\mathrm{def}}{=} \tfrac{1}{m} \sum_i P_i(x) \,.$$

We denote the optimal value of an assignment for $\Im$ as $\mathrm{opt}(\Im) = \max_{x \in \{\pm 1\}^n} \Im(x)$.

Examples: Max Cut corresponds to the case where Π consists of the binary inequality predicate. For Max 3-Sat, Π contains all eight 3-literal disjunctions, e.g., $X_1 \vee \bar{X}_2 \vee \bar{X}_3$.

Linear Programming Relaxations for CSPs. In order to write an LP relaxation for such a problem, we need to linearize the objective function. For $n \in \mathbb{N}$, let Max-Π_n be the set of Max-Π instances on n variables. An *LP-relaxation of size R* for Max-Π_n consists of the following.

Linearization:

> For every $\Im \in$ Max-Π_n, we associate a vector $\tilde{\Im} \in \mathbb{R}^m$ and for every assignment $x \in \{\pm 1\}^n$, we associate a point $\tilde{x} \in \mathbb{R}^m$, such that $\Im(x) = \langle \tilde{\Im}, \tilde{x}\rangle$ for all $\Im \in$ Max-Π_n and all $x \in \{\pm 1\}^n$.

Polytope:

> A convex polytope $P \subseteq \mathbb{R}^m$ described by R linear inequalities, such that $\tilde{x} \in P$ for all assignments $x \in \{\pm 1\}^n$. The polytope P is independent of the instance $\Im$ of Max-Π_n.

Given an instance $\Im \in \text{MAX-}\Pi_n$, the LP relaxation $\mathcal{L}$ outputs the value $\mathcal{L}(\Im) = \max_{y \in P} \langle \tilde{\Im}, y \rangle$. Since $\tilde{x} \in P$ for all assignments $x \in \{\pm 1\}^n$ and $\langle \tilde{\Im}, \tilde{x} \rangle = \Im(x)$, we have $\mathcal{L}(\Im) \geqslant \text{opt}(\Im)$ for instances $\Im \in \text{MAX-}\Pi_n$.

Remark II.1. The choice of linearization does not affect the minimal size of an LP relaxation; by applying an appropriate linear transformation, one sees that all linearizations are equivalent. For concreteness, one could view $x \mapsto \Im(x)$ as a degree-k multilinear polynomial. In the Fourier basis $\{\chi_\alpha : \alpha \subseteq [n]\}$, one would have $\tilde{J} = \sum_\alpha \hat{J}(\alpha)\chi_\alpha$ and $\tilde{x} = \sum_\alpha \chi_\alpha(x)\chi_\alpha$.

(c, s)-approximation. We say that a linear programming relaxation $\mathcal{L}$ for $\text{MAX-}\Pi_n$ achieves a (c, s)-*approximation* if $\mathcal{L}(\Im) \leqslant c$ for all instances $\Im \in \text{MAX-}\Pi_n$ with $\text{opt}(\Im) \leqslant s$. We also say that $\mathcal{L}$ achieves an α-*factor approximation* if $\mathcal{L}(\Im) \leqslant \alpha\, \text{opt}(\Im)$ for all $\Im \in \text{MAX-}\Pi_n$.

The following theorem is inspired by Yannakakis's characterization of exact linear programming relaxations. It appears in similar form in previous works [Pas12] and [BFPS12, Thm. 1]. For simplicity, we specialize it here for constraint satisfaction problems.

Theorem II.2. *There exists an LP relaxation of size R that achieves a (c, s)-approximation for $\text{MAX-}\Pi_n$ if and only if there exist non-negative functions $q_1, \ldots, q_R \colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ such that for every instance $\Im \in \text{MAX-}\Pi_n$ with $\text{opt}(\Im) \leqslant s$, the function $c - \Im$ is a nonnegative combination of $q_1, \ldots, q_R$, i.e.*

$$c - \Im \in \Big\{ \sum\nolimits_i \lambda_i q_i \;\Big|\; \lambda_1, \ldots, \lambda_R \geqslant 0 \Big\}.$$

In other words, the kernel $(\Im, x) \mapsto c - \Im(x)$ has a nonnegative factorization of rank R.

A communication model. The characterization in Theorem II.2 has an illustrative interpretation as a two-party, one-way communication complexity problem: Alice's input is a $\text{MAX-}\Pi$ instance $\Im$ with $\text{opt}(\Im) \leqslant s$. Bob's input is an assignment $x \in \{\pm 1\}^n$. Their goal is to compute the value $\Im(x)$ in expectation. To this end, Alice sends Bob a randomized message containing at most L bits. Given the message Bob outputs deterministically a number v such that $v \leqslant c$. The protocol is correct if for every input pair $(\Im, x)$, the expected output satisfies $\mathbb{E}\, v = \Im(x)$ (the expectation is over Alice's randomness).

An L-bit protocol for this communication problem yields an LP relaxation of size 2^L: If Bob outputs a value $v(x, i)$ based on message i from Alice, then define $q_i(x) = c - v(x, i)$. This yields 2^L non-negative functions satisfying the conditions of Theorem II.2.

On the other hand, if there exist $R = 2^L$ functions $\{q_1, q_2, \ldots, q_R\}$ as in Theorem II.2, then by adding an additional non-negative function q_{R+1}, we may assume that $\sum_{i=1}^{R+1} \lambda_i = 1$, i.e. that we have a convex combination instead of a non-negative combination. This yields a strategy for Alice and Bob: Alice sends an index $i \in [R+1]$, drawn from a distribution depending on $\Im$ (specified by the coefficients $\{\lambda_i\}$), and then Bob outputs $c - q_i(x) \leqslant c$.

Example: Suppose the optimization problem is MAX CUT. In this case, Alice receives a graph $G = (V, E)$ and Bob a cut $S \subseteq V$. If Alice sends Bob a uniformly random edge $\{u, v\} \in E$ and Bob outputs the value $|\mathbb{I}_S(u) - \mathbb{I}_S(v)|$, the result is a communication (in expectation) protocol using at most $\log_2 \binom{n}{2}$ bits of communication. In any protocol achieving a less trivial approximation, Bob would have to always output numbers strictly less than 1. A similar communication in expectation model is considered in [FFGT11], where they show a strong connection to non-negative rank.

LP Lower Bounds from Test Functions. The following lemma will allows us to prove lower bounds for general LP relaxations. For $\varepsilon = 0$, the function H in the statement of the lemma would correspond to a hyperplane separating $c - \Im$ and the cone generated by $q_1, \ldots, q_R$. The presence of ε will allow us to tolerate some noise in the arguments that follow.

Lemma II.3. *In order to show that (c, s)-$\text{MAX-}\Pi_n$ requires LP relaxations of size greater than R, it is sufficient to prove the following: For every collection of densities $q_1, \ldots, q_R \colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$, there exists $\varepsilon > 0$, a function $H \colon \{\pm 1\}^n \to \mathbb{R}$ and a $\text{MAX-}\Pi_n$ instance $\Im$ such that*

1) $\langle H, c - \Im \rangle < -\varepsilon$
2) $\langle H, q_i \rangle \geqslant -\varepsilon\,.$

Proof: We will argue that H certifies that $c - \Im$ is not a nonnegative combination of $q_1, \ldots, q_R$. Let $\gamma = c - \mathbb{E}\, \Im$. For the sake of contradiction, suppose that $c - \Im = \sum_i \lambda_i q_i$ for $\lambda_1, \ldots, \lambda_R \geqslant 0$. Then, $\gamma = c - \mathbb{E}\, \Im = \sum_i \lambda_i \mathbb{E}\, q_i = \sum_i \lambda_i$. Also,

$$\begin{aligned} \langle H, c - \Im \rangle = \langle H, \sum_i \lambda_i q_i \rangle &= \sum_i \lambda_i \langle H, q_i \rangle \\ &\geqslant -\varepsilon \sum_i \lambda_i = -\varepsilon\gamma\,, \end{aligned}$$

which contradicts the condition $\langle H, c - \Im \rangle < -\varepsilon$ since $\gamma \leqslant c \leqslant 1$. ∎

Sherali-Adams LP relaxations for CSPs. A primary component of our approach involves leveraging known integrality gaps for the Sherali-Adams hierarchy. To that end, we now give a brief overview of Sherali-Adams LP relaxations. For a more detailed account, we refer the reader to [Lau03], [CMM09].

A d-round Sherali-Adams LP relaxation for a $\text{MAX-}\Pi_n$ instance will consist of variables $\{X_S : S \subseteq [n], |S| \leqslant d\}$ for all products of up to degree d on the n-variables. These variables $\{X_S : |S| \leqslant d\}$ are to be thought of as the moments up to degree d of the variables, under a purported distribution.

An important property of an SA solution $\{X_S : |S| \leqslant d\}$ is that these moments agree with a set of local marginal distributions. In particular, for every set $S \subseteq [n]$ with $|S| \leqslant d$ there exists a distribution μ_S over $\{\pm 1\}^S$ such that,

$$\mathop{\mathbb{E}}_{x \sim \mu_S} \chi_A(x) = X_A \qquad \forall A \subseteq S\,.$$

In an alternate but equivalent terminology, a d-round SA instance can be thought of as *d-local expectation functional (d-ℓ.e.f.)*. Specifically, a *d-local expectation functional* is a linear functional $\tilde{\mathbb{E}}$ on degree-d multilinear polynomials such that $\tilde{\mathbb{E}}\,1 = 1$ and $\tilde{\mathbb{E}}\,P \geqslant 0$ for every degree-d multilinear polynomial P that is nonnegative over $\{\pm 1\}^n$ and depends only on d variables. In terms of the local marginal distributions, $\tilde{\mathbb{E}}$ is the unique linear functional on degree d polynomials satisfying

$$\tilde{\mathbb{E}}\,\chi_A = \mathop{\mathbb{E}}_{x \sim \mu_S} \chi_A(x) \qquad \forall |S| \leqslant d, A \subseteq S \subseteq [n]\,.$$

The *d-round Sherali-Adams value* of a $\textsc{Max}\text{-}\Pi_n$ instance $\Im$ is defined as

$$\mathtt{SA}_d(\Im) \stackrel{\text{def}}{=} \max_{d\text{-}\ell\text{.e.f. } \tilde{\mathbb{E}}} \tilde{\mathbb{E}}\,\Im\,.$$

This optimization problem can be implemented by an $n^{O(d)}$-sized linear programming relaxation for $\textsc{Max}\text{-}\Pi_n$. (Notice that $\tilde{\mathbb{E}}$ is only an n^d-dimensional object.) In particular, if d-rounds of Sherali-Adams achieve a (c, s)-approximation for $\textsc{Max}\text{-}\Pi_n$, then so do general $n^{O(d)}$-sized LP relaxations.

Given such a d-ℓ.e.f. $\tilde{\mathbb{E}}$, we can extend it linearly to all functions on $\{\pm 1\}^n$ (as opposed to just degree-d polynomials). Concretely, we set $\tilde{\mathbb{E}}\,f = 0$ for all functions f that are orthogonal to the subspace of degree-d functions, i.e., orthogonal to the span of $\{\chi_\alpha : |\alpha| \leqslant d\}$. We can represent $\tilde{\mathbb{E}}$ by a function $H \colon \{\pm 1\}^n \to \mathbb{R}$ such that $\tilde{\mathbb{E}}\,f = \langle H, f\rangle$ for all $f \colon \{\pm 1\}^n \to \mathbb{R}$. Concretely,

$$H = \sum_{\alpha \subseteq [n] \colon |\alpha| \leqslant d} c_\alpha \chi_\alpha \quad \text{ for } c_\alpha = \tilde{\mathbb{E}}\,\chi_\alpha\,, \tag{II.1}$$

where $\{\chi_\alpha\}$ is the Fourier basis.

In Section III, We will use a modification of H as a test function in the sense of Lemma II.3. The crucial property of the test function H that makes it useful for our purposes is the following.

Observation II.4. Suppose that $f : \{\pm 1\}^n \to \mathbb{R}$ depends only on a subset of at most d coordinates $S \subseteq [n]$, then

$$\langle H, f\rangle = \mathop{\mathbb{E}}_{x \sim \mu_S} [f(x)]$$

for some probability measure μ_S on $\{\pm 1\}^n$.

One should consider this observation in light of Lemma II.3. If H is the linear functional corresponding to d rounds of the Sherali-Adams hierarchy and q_i is a non-negative d-junta, then $\langle H, q_i\rangle \geqslant 0$.

III. Sherali-Adams and General LPs

Our main theorem is that general LP relaxations are no more powerful than Sherali-Adams relaxations (in the polynomial-size regime).

Theorem III.1 (Main)**.** *Fix a positive number $d \in \mathbb{N}$. Suppose that the d-round Sherali-Adams relaxation cannot achieve a (c, s)-approximation for $\textsc{Max}\text{-}\Pi_n$ for every n. Then no sequence of LP relaxations of size at most $n^{d/2}$ can achieve a (c, s)-approximation for $\textsc{Max}\text{-}\Pi_n$ for every n.*

We prove the following more general result in Section III-C.

Theorem III.2. *Consider a function $f : \mathbb{N} \to \mathbb{N}$. Suppose that the $f(n)$-round Sherali-Adams relaxation cannot achieve a (c, s)-approximation for $\textsc{Max}\text{-}\Pi_n$. Then for all sufficiently large n, no LP relaxation of size at most $n^{f(n)^2}$ can achieve a (c, s)-approximation for $\textsc{Max}\text{-}\Pi_N$, where $N \leqslant n^{10 f(n)}$.*

In particular, by choosing $f(n) \asymp \frac{\log n}{\log \log n}$, known Sherali-Adams gaps for $\textsc{Max Cut}$ [CMM09] and $\textsc{Max 3-Sat}$ [Sch08] imply the same integrality gaps for LPs of size $n^{\Omega(\frac{\log n}{\log \log n})}$.

A. High-Entropy Distributions vs. Juntas

Fix some $n \in \mathbb{N}$. We now prove that every distribution on n bits with very high entropy has its low-degree part "close to uniform" off a small set of coordinates. For brevity, we write $S \setminus v$ to denote $S \setminus \{v\}$.

Let $(X_1, \ldots, X_n) \in \{\pm 1\}^n$ be correlated random bits with distribution μ. For a subset $S \subseteq [n]$, we use the notation $X_S = \{X_i : i \in S\}$. We first prove the following lemma. Afterward, we use it to bound the Fourier coefficients of μ. Here, and in what follows, $H(\cdot)$ denotes the Shannon entropy measured in bits.

Lemma III.3. *For all $1 \leqslant d, t \leqslant n$ and $\beta > 0$, the following holds. If μ has entropy $\geqslant n - t$, there exists a set $J \subseteq [n]$ of at most $\frac{td}{\beta}$ coordinates such that for all subsets $A \not\subseteq J$ with $|A| \leqslant d$, we have*

$$\max_{v \in A} H(X_v \mid X_{A \setminus v}) \geqslant 1 - \beta\,. \tag{III.1}$$

For $0 < \beta < 1$, consider the hypergraph G_β on vertex set $[n]$ that contains a hyperedge e of size at most d whenever, for all $v \in e$, we have

$$H(X_v \mid X_{e \setminus v}) \leqslant 1 - \beta\,. \tag{III.2}$$

Lemma III.3 follows directly from the next claim.

Proposition III.4. *If μ has entropy $n - t$, then $|\bigcup_{e \in E(G_\beta)} e| \leqslant \frac{td}{\beta}$.*

Proof: Let $J = \bigcup_{e \in E(G_\beta)} e$ denote the set of vertices participating in an edge of G_β. Since each hyperedge contains at most d vertices, we can find a sequence $e_1, e_2, \ldots, e_r$ of at least $r \geqslant |J|/d$ hyperedges such that for each $i = 1, 2, \ldots, r$, the set $e_i \setminus (e_{i-1} \cup \cdots \cup e_1)$ contains at least one vertex $u_i \in J$.

Observe that the vertices $u_1, \ldots, u_r$ are distinct.

Writing $U = \{u_1, \ldots, u_r\}$ and $W = [n] \setminus U$, we can upper bound the entropy of $X_1, \ldots, X_n$ using the chain rule:

$$\begin{aligned} & H(X_1, \ldots, X_n) && \text{(III.3)} \\ & = H(X_W) + \sum_{i=1}^{r} H(X_{u_i} \mid X_{W \cup \{u_1, \ldots, u_{i-1}\}}) \\ & \leqslant H(X_W) + \sum_{i=1}^{r} H(X_{u_i} \mid X_{e_i \setminus u_i}) \\ & \leqslant |W| + |U| \cdot (1 - \beta) \\ & \leqslant n - |J|\beta/d . \end{aligned}$$

The first inequality uses the fact that $e_i \setminus u_i \subseteq W \cup \{u_1, \ldots, u_{i-1}\}$ (because $e_i \setminus u_i$ does not contain any of the vertices $u_i, \ldots, u_r$). The second inequality follows directly from the definition of the hyperedges in G_β.

Since $X_1, \ldots, X_n$ has entropy at least $n-t$ by assumption, it follows that $|J| \leqslant td/\beta$. ■

We now record a Fourier-theoretic consequence of Lemma III.3. To this end, recall that for two probability measures μ and ν over $\{\pm 1\}^n$, one defines the *KL-divergence* as the quantity

$$D(\mu \,\|\, \nu) = \mathop{\mathbb{E}}_{\mu} \left[\log_2 \frac{\mu(x)}{\nu(x)} \right] .$$

In this case, Pinsker's inequality (and its sharp form due to Kullback, Csiszár and Kemperman, see e.g. [Tsy09, Lemma 2.5]) states that

$$D(\mu \,\|\, \nu) \geqslant \frac{1}{\ln 4} \|\mu - \nu\|_1^2. \tag{III.4}$$

Lemma III.5. *Let μ be a distribution as in the statement of Lemma III.3, and let $J \subseteq [n]$ be the corresponding set of coordinates. If $A \subseteq [n]$ satisfies $|A| \leqslant d$ and $A \nsubseteq J$, then*

$$\left| \mathop{\mathbb{E}}_{\mu}[\chi_A(x)] \right| \leqslant \sqrt{(\ln 4)\beta}.$$

Proof: For $x \in \{\pm 1\}^n$ and $S \subseteq [n]$, we use x_S to denote x restricted to the bits in S. Since $|A| \leqslant d$ and $A \nsubseteq J$, Lemma III.3 implies that some $v \in A$ satisfies (III.1). For $y \in \{\pm 1\}^{A \setminus v}$, denote by $\mu_{|y}$ the distribution of x_v conditioned on the event $x_{A \setminus v} = y$. We bound

$$\left| \mathop{\mathbb{E}}_{\mu}[\chi_A(x)] \right| \leqslant \mathop{\mathbb{E}}_{y \sim \mu} \left| \mathop{\mathbb{E}}_{\mu}[\chi_A(x) \mid x_{A \setminus v} = y] \right| .$$

Denote by ν the uniform distribution on $\{\pm 1\}$. For any $y \in \{\pm 1\}^{A \setminus v}$,

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{\mu}[\chi_A(x) \mid x_{A \setminus v} = y] \right| &= \|\mu_{|y} - \nu\|_1 \\ &\leqslant \sqrt{(\ln 4) D(\mu_{|y} \,\|\, \nu)} , \end{aligned}$$

by (III.4).

Since

$$D(\mu_{|y} \,\|\, \nu) = 1 - H(\mu_{|y}) ,$$

we get

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{\mu}[\chi_A(x)] \right| &\leqslant \mathop{\mathbb{E}}_{y \sim \mu} \sqrt{(\ln 4)(1 - H(\mu_{|y}))} \\ &\leqslant \sqrt{(\ln 4)(1 - \mathop{\mathbb{E}}_{y}[H(\mu_{|y})])} , \end{aligned}$$

where the last inequality is Cauchy–Schwarz. The desired bound follows because $\mathbb{E}_y[H(\mu_{|y})] = H(X_v \mid X_{A \setminus v}) \geqslant 1 - \beta$. ■

Finally, we arrive the primary goal of this section.

Lemma III.6 (High-Entropy Distributions vs Juntas)**.** *Let $q \colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ be a density and let μ_q denote the corresponding measure on $\{\pm 1\}^n$. If μ_q has entropy least $n - t$ for some $t \leqslant n$, then for every $1 \leqslant d \leqslant n$ and $\gamma > 0$, there exists a set $J \subseteq [n]$ with*

$$|J| \leqslant \frac{4td}{\gamma^2} .$$

such that for all subsets $\alpha \nsubseteq J$ with $|\alpha| \leqslant d$, we have $|\hat{q}(\alpha)| \leqslant \gamma$.

Proof: Set $\beta = \frac{\gamma^2}{4}$. Now apply Lemma III.5 and use the fact that $\hat{q}(\alpha) = \mathbb{E}_{x \sim \mu_q}[\chi_\alpha(x)]$. ■

B. Random Restrictions

We first recall the following standard estimates (see, e.g., [McD98]). Suppose $X_1, \ldots, X_n$ are i.i.d $\{0, 1\}$ random variables with $\mathbb{E}[X_i] = p$. Then,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant \frac{pn}{2} \right) \geqslant 1 - e^{-pn/8} . \tag{III.5}$$

Furthermore, if $\frac{pn}{1-p} \leqslant 1/2$, then

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant t \right) &= \sum_{i=t}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &\leqslant (pn)^t \sum_{i=0}^{\infty} (pn/(1-p))^i \leqslant 2(pn)^t . \quad \text{(III.6)} \end{aligned}$$

Lemma III.7. *For any $d \in \mathbb{N}$, the following holds. Let Q be a collection of densities $q \colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ such that the corresponding measures μ_q have entropy at least $n - t$. If $|Q| \leqslant n^{d/2}$, then for all integers m with $3 \leqslant m \leqslant n/4$, there exists a set $S \subseteq [n]$ such that*

- *$|S| = m$*
- *For each $q \in Q$, there exists a set of at most d coordinates $J(q) \subseteq S$ such that under the distribution μ_q, all d-wise correlations in $S - J(q)$ are small. Quantitatively, we have*

$$|\hat{q}(\alpha)| \leqslant \left(\frac{32mtd}{\sqrt{n}} \right)^{1/2}$$

for all $\alpha \subseteq S, \alpha \nsubseteq J(q), |\alpha| \leqslant d$.

Proof: We will sample the set $S \subseteq [n]$ by including each element independently with probability $2m/n$, then argue that with non-zero probability, both the conditions on S hold.

First, by (III.5), we have $|S| \geqslant m$ with probability at least $1 - e^{-m/4} > 1/2$.

Fix $\gamma = \left(\frac{32mtd}{\sqrt{n}}\right)^{1/2}$. By Lemma III.6, for each $q \in Q$ there exists a set $J'(q)$ of at most $\frac{4td}{\gamma^2} \leqslant \frac{\sqrt{n}}{8m}$ coordinates such that for all subsets $\alpha \nsubseteq J'(q)$ with $|\alpha| \leqslant d$, we have $|\hat{q}(\alpha)| \leqslant \gamma$.

The set $J(q)$ for a distribution q is given by $J(q) = J'(q) \cap S$. Clearly, $\mathbb{E}[|J'(q) \cap S|] = \frac{2m}{n}|J'(q)| \leqslant \frac{2m}{n} \cdot \frac{\sqrt{n}}{8m} \leqslant 1/4$. Thus by (III.6), we can write

$$\mathbb{P}\left[|J(q) \cap S| \geqslant d\right] \leqslant 2\left(\frac{2m}{n} \cdot |J'(q)|\right)^d \leqslant 2 \cdot \left(\frac{2m}{n} \cdot \frac{\sqrt{n}}{8m}\right)^d \leqslant \frac{2}{4^d n^{d/2}}\,.$$

The existence of the set S follows by taking a union bound over all the $|Q| \leqslant n^{d/2}$ densities in the family Q. Note that we have concluded with $|S| \geqslant m$, but we can remove some elements from S to achieve $|S| = m$. ■

C. Proof of Main Theorem — Theorem III.1

In this subsection, we will prove Theorems III.2 and III.1. Let $m \leqslant n$ be parameters $m, n \in \mathbb{N}$ to be chosen later. Let $\Im_0$ be a MAX-Π_m instance with $\mathtt{SA}_d(\Im_0) \geqslant c$ and $\operatorname{opt}(\Im_0) \leqslant s$. Our goal is to show that for all $\varepsilon > 0$ and all large enough $n \in \mathbb{N}$, any $n^{d/2}$-sized LP relaxation for MAX-Π_n cannot certify $\operatorname{opt}(\Im) < c - \varepsilon$ for all "shifts" $\Im$ of $\Im_0$. (Here, "shift" means that we plant $\Im_0$ on some subset of the variables $\{1, \ldots, n\}$.)

Let Q denote a collection of densities $q\colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ on $\{\pm 1\}^n$ with $|Q| \leqslant n^{d/2}$. We will show that there exists a shift $\Im$ of $\Im_0$ and a test functional $H \in L^2(\{-1,1\}^n)$ that satisfies the conditions of Lemma II.3 (lower bounds from test functions).

Let $B \subseteq \{\pm 1\}^n$ be the subset of points where one of the densities q is exceptionally large. Formally, we define

$$B = \{x \in \{\pm 1\}^n \mid \exists q \in Q.\ q(x) \geqslant n^d\}\,.$$

By Markov's inequality, the set B has measure at most $\mathbb{E}\,\mathbb{I}_B \leqslant |Q| \cdot n^{-d} \leqslant n^{-d/2}$.

We decompose each $q \in Q$ into three parts

$$q = q' + q_{\text{bad}} - \mathbb{E}\, q_{\text{bad}}$$

with $q_{\text{bad}} = q \cdot \mathbb{I}_B$ (and therefore $q' = q \cdot (1 - \mathbb{I}_B) + \mathbb{E}\, q_{\text{bad}}$). The function $q'\colon \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$ is a density because

$$\mathbb{E}\, q' = \mathbb{E}\, q \cdot (1 - \mathbb{I}_B) + \mathbb{E}\, q \cdot \mathbb{I}_B = 1\,.$$

Since $q(x) < n^d$ for every point $x \notin B$, the function q' satisfies $q' \leqslant n^d + 1$, which implies that the min-entropy (and thus Shannon entropy) of the distribution corresponding to q' is at least $n - t$ where $t = d \log_2 n + 1$.

By Lemma III.7, there exists a set $S \subseteq [n]$ of size m such that every $q \in Q$, every Fourier coefficient $\hat{q}'(\alpha)$ with degree $|\alpha| \leqslant d$ and $\alpha \nsubseteq J(q)$ satisfies $|\hat{q}'(\alpha)| \leqslant K$, where

$$K \leqslant 8\left(\frac{md \log_2 n}{\sqrt{n}}\right)^{1/2}.$$

For a subset $T \subseteq [n]$, let $q_{|T}$ denote the marginal of q on T, so that $q_{|T}(x) = \mathbb{E}_{y \in \{\pm 1\}^n}\, q(x_T, y_{[n]\setminus T})$. (Equivalently, $q_{|T} = \sum_{\alpha \subseteq T} \hat{q}(\alpha)\chi_\alpha$.) In this notation, the ℓ_1-norm of the degree-$\leqslant d$ Fourier coefficients of $q'_{|J(q)} - q'_{|S}$ is bounded by

$$\sum_{\substack{\alpha \subseteq S \\ \alpha \nsubseteq J(q) \\ |\alpha| \leqslant d}} |\hat{q}'(\alpha)| \leqslant K m^d. \tag{III.7}$$

Let $\eta = m^d \cdot \max(3n^{-d/2}, K)$, and assume that n is chosen large enough so that $\eta < \frac{1}{2}$.

Let $\Im$ be the MAX-Π_n instance obtained by planting the MAX-Π_m instance $\Im_0$ on the variable set S. Let H be a d-round Sherali-Adams solution (in the sense of (II.1)) for $\Im$ with value c. Further, we will choose a Sherali-Adams solution H that satisfies $\hat{H}(\alpha) = 0$ for $\alpha \nsubseteq S$. This is possible since all the constraints in $\Im$ are contained within S. (Since $\Im$ is a shift of $\Im_0$, it has the same Sherali-Adams value, $\mathtt{SA}_d(\Im) = \mathtt{SA}_d(\Im_0)$)

We claim that $H' = (1 - \mathbb{I}_B) \cdot H$ is a test function in the sense of Lemma II.3. Recall that $\|\Im\|_\infty \leqslant 1$. Furthermore, by (II.1), we have $\|H\|_\infty \leqslant m^d$. Using these,

$$\begin{aligned}
&\langle H', c - 2\eta - \Im\rangle \\
&\quad = \langle H, c - 2\eta - \Im\rangle - \langle H \cdot (c - 2\eta - \Im), \mathbb{I}_B\rangle \\
&\quad \leqslant -2\eta - \langle H \cdot (c - 2\eta - \Im), \mathbb{I}_B\rangle \\
&\qquad\quad \text{(because } \langle H, 1\rangle = 1 \\
&\qquad\qquad\quad \text{and } \langle H, \Im\rangle = \mathtt{SA}_d(\Im)) \\
&\quad \leqslant -2\eta + (c + 2\eta + \|\Im\|_\infty)\|H\|_\infty \cdot \mathbb{E}\,\mathbb{I}_B \\
&\quad \leqslant -2\eta + 3m^d \cdot n^{-d/2} \\
&\quad < -\eta\,.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\langle H', q\rangle \\
&= \langle H', q' - \mathbb{E}\, q_{\text{bad}}\rangle + \langle H', q_{\text{bad}}\rangle \\
&= \langle H', q' - \mathbb{E}\, q_{\text{bad}}\rangle + 0 \\
&\qquad \text{(using } \operatorname{supp}(q_{\text{bad}}) \cap \operatorname{supp}(H') = \emptyset) \\
&= \langle H, q' - \mathbb{E}\, q_{\text{bad}}\rangle + 0 \\
&\qquad \text{(using } \operatorname{supp}(q' - \mathbb{E}\, q_{\text{bad}}) \cap \operatorname{supp}(H - H') = \emptyset) \\
&= \langle H, q'\rangle - \mathbb{E}\, q_{\text{bad}} \qquad \text{(using } \langle H, 1\rangle = 1)
\end{aligned}$$

$= \langle H, q'_{|S}\rangle - \mathbb{E}\, q_{\mathrm{bad}}$ (using $\hat{H}(\alpha) = 0$ for $\alpha \nsubseteq S$)

$= \langle H, q'_{|J(q)}\rangle - \langle H, q'_{|J(q)} - q'_{|S}\rangle - \mathbb{E}\, q_{\mathrm{bad}}$

$\geqslant \langle H, q'_{|J(q)}\rangle - \eta - \mathbb{E}\, q_{\mathrm{bad}}$ (using $|\hat{H}(\alpha)| \leqslant 1$ for all $\alpha \subseteq S$, $\deg(H) \leqslant d$ and our Fourier-ℓ_1-norm bound (III.7) for $q'_{|J(q)} - q'_{|S}$)

$\geqslant \mathbb{E}\, q_{\mathrm{bad}} - \eta - \mathbb{E}\, q_{\mathrm{bad}}$

$\geqslant -\eta.$

The last step uses that $q'_{|J(q)}$ is a d-junta with $q'_{|J(q)} \geqslant \mathbb{E}\, q_{\mathrm{bad}}$ (because $q' \geqslant \mathbb{E}\, q_{\mathrm{bad}}$ by construction). Since H is a d-round Sherali-Adams functional, it satisfies $\langle H, q'_{|J(q)}\rangle \geqslant \mathbb{E}\, q_{\mathrm{bad}}$ (by the d-ℓ.e.f. property).

We conclude that by Lemma II.3, the cone generated by Q does not contain the function $c - 2\eta - \Im$.

We see that for any collection Q of at most $n^{d/2}$ densities, there exists a MAX-Π_n instance $\Im$ with $\mathrm{opt}(\Im) \leqslant s$ (a shift of $\Im_0$) such that $c - 2\eta - \Im$ is not in the cone generated by Q. By Theorem II.2, it follows that any $n^{d/2}$-sized linear programming relaxation cannot achieve a $(c - 2\eta, s)$-approximation for MAX-Π_n. Finally, note that asymptotically, we have

$$\eta = O\left(\frac{m^d \sqrt{md \log n}}{n^{1/4}}\right). \tag{III.8}$$

$\square$

Proof of Theorem III.2: Fix an instance size m and put $d = f(m)$. In the preceding argument, require that n grows like $m^{10d} = m^{10 f(m)}$ so that $\eta = o(1)$) (see (III.8)). The lower bound achieved is $n^{d/2} \geqslant m^{5 f(m)^2}$. $\blacksquare$

IV. Symmetric Linear Programs

We will now prove the following theorem relating Sherali-Adams gaps to those for symmetric LPs for MAX CUT.

Theorem IV.1. *Suppose that the t-round Sherali-Adams relaxation for* MAX CUT *cannot achieve a (c, s)-approximation on graphs with n vertices. Then no symmetric LP of size $\leqslant \binom{N}{t}$ can achieve a (c, s)-approximation on N-vertex graphs, where $N = 2n$.*

We will require the following lemma of Yannakakis.

Lemma IV.2 ([Yan91, Claim 2]). *Let H be a group of permutations whose index in S_n is at most $\binom{n}{k}$ for some $k < n/4$. Then there exists a set J of size at most k such that H contains all even permutations that fix the elements of J.*

The next claim is an elementary consequence of Lemma IV.2; we omit its proof for space considerations in this extended abstract.

Lemma IV.3. *Suppose a family of functions $\mathcal{F} = \{f_i : \{\pm 1\}^n \to \mathbb{R} : i = 1, 2, \ldots, M\}$ is closed under permutation of its inputs, and $M < \binom{n}{k}$ for $k < n/4$, then each function f_i depends only on a subset $S_i \subseteq [n]$ of at most k coordinates and possibly the value $\sum_{a \in [n]} x_a$.*

Let $\Im$ be a Sherali-Adams integrality gap instance of MAX CUT. Suppose $\{X_S\}_{S \in [n], |S| \leqslant t}$ is the t-round Sherali-Adams solution on $\Im$.

Construct a new graph $\Im' = \Im_1 \cup \Im_2$ consisting of two disjoint copies of the instance $\Im$. Let $N = 2n$ denote the number of vertices of $\Im'$. Let us suppose that $\Im_1$ is on vertices $\{1, \ldots, n\}$ and $\Im_2$ is on $\{n+1, \ldots, 2n\}$.

We will now extend the Sherali-Adams LP solution for $\Im$ to a Sherali-Adams solution for $\Im'$. Roughly speaking, we will copy the SA-solution as is on to $\Im_1$ and negate all its values on $\Im_2$. In other words, the Sherali-Adams solution on $\Im'$ is so designed that every pair of vertices x_i, x_{n+i} always have opposite values.

Formally, for any subset $S \subseteq [2n]$ define $S_1 = S \cap \{1, \ldots, n\}$ and $S_2 = \{i - n \mid i \in S \setminus S_1\}$. Then, we can describe the SA solution $\{Y_S\}_{S \subseteq [2n], |S| \leqslant t}$ as follows,

$$Y_S = (-1)^{|S_2|} X_{S_1 \oplus S_2}$$

Here $S_1 \oplus S_2$ is the symmetric difference between the two sets.

Definition IV.4. For a set $S \subseteq [2n]$ of the form $S = S_1 \cup (S_2 + n)$ where $S_1, S_2 \subseteq [n]$, define $\mathrm{wt}(S) = |S_1 \oplus S_2|$.

More generally, we will define $\{Y_S\}_{S \subseteq [2n]}$ for all subsets $S \subseteq [2n]$.

$$Y_S = \begin{cases} (-1)^{|S_2|} X_{S_1 \oplus S_2} & \text{if } \mathrm{wt}(S) \leqslant t \\ 0 & \text{otherwise} \end{cases}$$

Observation IV.5. For any set S, $\mathrm{wt}(S \oplus \{i\}) = \mathrm{wt}(S \oplus \{n+i\})$.

Proof: It is easy to see that both values are equal to the hamming weight of the $\mathbb{F}_2$ vector $\mathbb{I}_{S_1} \oplus \mathbb{I}_{S_2} \oplus e_i$ where $\mathbb{I}_{S_1}, \mathbb{I}_{S_2} \in \mathbb{F}_2^n$ are indicators of subsets S_1, S_2 and $e_i \in \mathbb{F}_2^n$ is the i^{th} standard basis vector. $\blacksquare$

Define a function $H : \{\pm 1\}^N \to \mathbb{R}$ as

$$H(x) = \sum_{S \subseteq [N]} Y_S \chi_S(x), \tag{IV.1}$$

where $\{\chi_S\}$ is the Fourier basis over $\{\pm 1\}^N$.

Lemma IV.6. *For every $S \subseteq [2n]$ and all $i \in [n]$, $Y_{S \oplus \{i\}} = -Y_{S \oplus \{n+i\}}$.*

Proof: Let $S = S_1 \cup (S_2 + n)$ for some $S_1, S_2 \subseteq [n]$. By Observation IV.5, $\mathrm{wt}(S \oplus \{i\}) = \mathrm{wt}(S \oplus \{n+i\})$. If $\mathrm{wt}(S \oplus \{i\}) = \mathrm{wt}(S \oplus \{n+i\}) > t$, then by definition we will have $Y_{S \oplus \{i\}} = Y_{S \oplus \{n+i\}} = 0$.

On the other hand, if $\mathrm{wt}(S \oplus \{i\}) = \mathrm{wt}(S \oplus \{n+i\}) \leqslant t$ then,

$$Y_{S \oplus \{i\}} = (-1)^{|S_2|} X_{(S_1 \oplus \{i\}) \oplus S_2}$$

and

$$Y_{S\oplus\{n+i\}} = (-1)^{|S_2\oplus\{i\}|} X_{S_1\oplus(\{i\}\oplus S_2)}$$

Therefore, also in this case $Y_{S\oplus\{i\}} = -Y_{S\oplus\{n+i\}}$. ■

Lemma IV.7. *For any polynomial* $p(x_1,\ldots,x_N)$*, we have* $\langle H, (\sum_{i=1}^N x_i)p\rangle = 0$.

Proof: By linearity of the inner product, it is sufficient to prove the above claim when p is a monomial. Recall that,

$$\langle H, \left(\sum_i x_i\right) p\rangle = \mathop{\mathbb{E}}_{x\in\{\pm1\}^N}\left[H(x)\left(\sum_i x_i\right)p(x)\right]$$

Since x takes values in $\{\pm1\}^N$, it is sufficient to show the above claim for the elements of the Fourier basis $\{\chi_S\}$.

Fix a monomial $\chi_S(x)$. We can write the above inner product as,

$$\begin{aligned}
&\left\langle \sum_{A\subseteq[N]} Y_A\chi_A(x), \chi_S(x)\left(\sum_i x_i\right)\right\rangle \\
&= \mathop{\mathbb{E}}_{x\in\{\pm1\}^N}\left[\chi_S(x)\sum_{A\subseteq[N], i\in[2n]} Y_A\chi_{A\oplus\{i\}}(x)\right] \\
&= \mathop{\mathbb{E}}_{x\in\{\pm1\}^N}\left[\chi_S(x)\sum_{B\subseteq[N]}\chi_B(x)\left(\sum_{i\in[2n]} Y_{B\oplus\{i\}}\right)\right] \\
&= 0,
\end{aligned}$$

because $\sum_{i\in[2n]} Y_{B\oplus\{i\}} = 0$ since $Y_{B\oplus\{i\}} = -Y_{B\oplus\{n+i\}}$. ■

Lemma IV.8. *If* $f : \{\pm1\}^N \to \mathbb{R}_{\geqslant 0}$ *is a function that depends on a subset* $J \subseteq [N]$ *of at most* t *coordinates and possibly the value* $\sum_{i=1}^N x_i$*, then*

$$\langle H, f\rangle \geqslant 0\,.$$

Proof: Write the function f as a polynomial in $x_J = \{x_i \mid i \in J\}$ and $\sum_i x_i$ as follows,

$$f = p_0(x_J) + \sum_{i=1}^N p_i(x_J)\left(\sum_i x_i\right)^i.$$

Using Lemma IV.6, we have

$$\langle H, f\rangle = \langle H, p_0\rangle\,.$$

Since p_0 depends on at most t coordinates, by Observation II.4 we can write,

$$\langle H, p_0\rangle = \mathop{\mathbb{E}}_{x_J\sim\mu_J}[p_0(x_J)]\,,$$

where μ_J is some distribution on x_J.

Define a distribution μ on $\{\pm1\}^N$ as follows: Sample x_J from μ_J and then sample $x_{\bar{J}}$ uniformly randomly from among all assignments that satisfy $\sum_{i=1}^N x_i = 0$. This is feasible since $|J| = t < n/2$. Note that the distribution μ is supported entirely on the set $\{x \in \{\pm1\}^N \mid \sum_i x_i = 0\}$. Therefore, we have

$$\begin{aligned}
\langle H, p_0\rangle &= \mathop{\mathbb{E}}_{x_J\sim\mu_J}[p_0(x_J)] = \mathop{\mathbb{E}}_{x\sim\mu}[p_0(x_J)] \\
&= \mathop{\mathbb{E}}_{x\sim\mu}\left[p_0(x_J) + \sum_{i=1}^N p_i(x_J)\left(\sum_i x_i\right)^i\right] \\
&\qquad \text{(because } \textstyle\sum_i x_i = 0 \text{ on the} \\
&\qquad\quad \text{support of } \mu) \\
&= \mathop{\mathbb{E}}_{x\sim\mu}[f(x)] \geqslant 0 \\
&\qquad \text{(because } f \geqslant 0 \text{ pointwise)}
\end{aligned}$$

■

We are now in position to prove the main theorem of this section.

Proof of Theorem IV.1: Suppose that N is even. Let $\mathcal{F} = \{f_1,\ldots,f_M\}$ denote the family of functions from $\{\pm1\}^N$ to $\mathbb{R}_{\geqslant 0}$ associated with some symmetric LP relaxation of MAX CUT, i.e. those coming from an application of Theorem II.2.

By the symmetry assumption, the family $\mathcal{F}$ is closed under permutations of its inputs. Hence, by Lemma IV.3 each of its functions f_i depend on a set J_i of at most t coordinates and possibly the value $\sum_{i=1}^n x_i$.

Consider a graph $\Im$ on $n = N/2$ nodes with $\mathtt{SA}_t(\Im) > c$, and let $\Im'$ be the graph obtained by taking two copies of $\Im$ as discussed before. Then $\Im'$ has the property that $\mathtt{SA}_t(\Im') = \mathtt{SA}_t(\Im)$ and $\mathrm{opt}(\Im') = \mathrm{opt}(\Im)$. Let H be the corresponding functional defined in (IV.1).

Let us consider $\Im'$ as a function on $\{\pm1\}^N$ assigning cuts to their MAX CUT value in $\Im'$. Suppose we can express express,

$$c - \Im' = \sum_{i=1}^M \lambda_i f_i\,,$$

wherein $\lambda_i \geqslant 0$. Taking inner product with the functional H on both sides yields

$$\langle H, c - \Im'\rangle = c - \mathtt{SA}_t(\Im') < 0$$

while,

$$\langle H, f_i\rangle \geqslant 0 \quad \forall i \quad \text{by Lemma IV.8}\,,$$

a contradiction. ■

V. Conclusion

We have shown that for constraint satisfaction problems, there is an intimate relationship between general polynomial-sized linear programs and those arising from $O(1)$ rounds of the Sherali-Adams hierarchy. There are a few natural questions that readily suggest themselves.

Firstly, our quantitative bounds are far from optimal. For instance, it is known that the integrality gap of $1/2 + \varepsilon$ for MAX CUT persists for n^{c_ε} rounds, where c_ε is some constant depending on ε [CMM09], while we are only able to prove an integrality gap for LPs of size $n^{\Omega(\frac{\log n}{\log \log n})}$. This is due to the factor of m^d appearing in our Fourier estimate (III.7).

Question V.1. Is it the case that for approximating (boolean) max-CSP problems on n variables, linear programs of size $R(n)$ are only as powerful as those arising from $\mathrm{poly}(\frac{\log R(n)}{\log n})$ rounds of the Sherali-Adams hierarchy?

Secondly, given the connection for linear programs, it is natural to suspect that a similar phenomenon holds for SDPs.

Question V.2. For max-CSP problems, is there a connection between the efficacy of general SDPs and those from the Lasserre SDP hierarchy?

Finally, our techniques have made very strong use of the product structure on the space of feasible assignments for CSPs. One might hope to extend these connections to other types of problems like finding maximum-weight perfect matchings in general graphs [Yan91] or approximating vertex cover.

Acknowledgements

S.O. Chan was supported by NSF grants CCF-1118083 and CCF-1017403. P. Raghavendra was supported by NSF Career Award CCF-1343104 and an Alfred P. Sloan Fellowship. J. R. Lee was supported by NSF grants CCF-1217256 and CCF-0905626.

References

[ABL02] S. Arora, B. Bollobás, and L. Lovász, *Proving integrality gaps without knowing the linear program*, Proc. FOCS, 2002, pp. 313–322.

[ABLT06] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis, *Proving integrality gaps without knowing the linear program*, Theory Comput. **2** (2006), 19–51.

[BFPS12] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer, *Approximation limits of linear programs (beyond hierarchies)*, FOCS, 2012, pp. 480–489.

[BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani, *SDP gaps from pairwise independence*, Theory of Computing **8** (2012), no. 12, 269–289.

[BM13] Mark Braverman and Ankur Moitra, *An information complexity approach to extended formulations*, STOC, ACM, 2013, pp. 161–170.

[CMM09] M. Charikar, K. Makarychev, and Y. Makarychev, *Integrality gaps for Sherali-Adams relaxations*, Proc. STOC, ACM, 2009, pp. 283–292.

[FdlVKM07] Wenceslas Fernández de la Vega and Claire Kenyon-Mathieu, *Linear programming relaxations of Maxcut*, SODA, 2007, pp. 53–61.

[FFGT11] Y. Faenza, S. Fiorini, R. Grappe, and H. R. Tiwary, *Extended formulations, non-negative factorizations and randomized communication protocols*, arXiv:1105.4127, 2011.

[FMP+12] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf, *Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds*, STOC, 2012, pp. 95–106.

[GW95] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach. **42** (1995), 1115–1145.

[Hås01] Johan Håstad, *Some optimal inapproximability results*, J. ACM **48** (2001), no. 4, 798–859. MR 2144931 (2006c:68066)

[Lau03] M. Laurent, *A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. (2003), 470–496.

[LS91] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and* 0-1 *optimization*, SIAM J. Optim. **1** (1991), 166–190.

[McD98] Colin McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248. MR 1678578 (2000d:60032)

[Pas12] K. Pashkovich, *Extended formulations for combinatorial polytopes*, Ph.D. thesis, Magdeburg Universität, 2012.

[SA90] H. D. Sherali and W. P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Discrete Math. **3** (1990), 411–430.

[Sch08] G. Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, Proc. FOCS, IEEE, 2008, pp. 593–602.

[Tsy09] Alexandre B. Tsybakov, *Introduction to nonparametric estimation*, Springer, 2009.

[Vaz01] Vijay V. Vazirani, *Approximation algorithms*, Springer-Verlag, Berlin, 2001. MR 1851303 (2002h:68001)

[WS11] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*, Cambridge University Press, Cambridge, 2011.

[Yan91] M. Yannakakis, *Expressing combinatorial optimization problems by linear programs*, J. Comput. System Sci. **43** (1991), no. 3, 441–466.