

# Approximate Constraint Satisfaction Requires Large LP Relaxations

Siu On Chan<sup>\*</sup> James R. Lee<sup>†</sup> Prasad Raghavendra<sup>‡</sup> David Steurer<sup>§</sup>

## 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  $\frac{1}{2}$  and any such linear program for MAX 3-SAT has an integrality gap of  $\frac{7}{8}$ .

## 1 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 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

---

<sup>\*</sup>Microsoft Research New England

<sup>†</sup>University of Washington

<sup>‡</sup>U. C. Berkeley

<sup>§</sup>Cornell University

still hope to complete this program for LPs that use the natural encoding 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. This offers a potential framework for understanding the power of linear programs for many problems by relating their expressive power to that of the very explicit Sherali-Adams hierarchy.

In Section 1.2, we discuss our approach for the specific example of MAX CUT, including the class of LPs to which our results apply. Section 2 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 3, we present the technical components of our approach, as well as the proof of our main theorem. Finally, Section 4 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.

## 1.1 History and context

**Extended formulations.** 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<sup>+</sup>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  $O(n^{1/2-\epsilon})$  requires LPs of size  $2^{\Omega(n^\epsilon)}$ . Building on that work, Braverman and Moitra [BM13] show that approximating MAX CLIQUE within  $O(n^{1-\epsilon})$  requires LPs of size  $2^{\Omega(n^\epsilon)}$ .

These three latter papers all use Yannakakis' connection between extension complexity and non-negative rank (see, e.g., [FMP<sup>+</sup>12], for a detailed discussion). They are based on increasingly more sophisticated analyses of a single family of slack matrices first defined in [FMP<sup>+</sup>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, ABT06], 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 [FdVKM07] 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].

## 1.2 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  $x = (x_1, x_2, \dots, x_n) \in \{\pm 1\}^n$  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* the objective function.

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  $\{0 \leq y_{i,j} \leq 1 : i, j \in V\}$  and

$$\left\{ y_{i,j} \leq y_{i,k} + y_{k,j}, \quad y_{i,j} + y_{i,k} + y_{k,j} \leq 2 : i, j, k \in V \right\}.$$

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 1.1.* 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 3, 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 [FdVVKM07, CMM09], it is shown that for every  $\varepsilon > 0$  and every  $d \in \mathbb{N}$ , there are graphs  $G$  with  $\text{opt}(G) \leq s$ , but  $\text{SA}_d(G) \geq c$ , where  $s \leq \frac{1}{2} + \varepsilon$ , and  $c \geq 1 - \varepsilon$ , and  $\text{SA}_d$  denotes the LP-value of the  $d$ -round Sherali-Adams relaxation (see Section 2 for the definition).

**Proving a lower bound.** In Theorem 2.2, we recall that if there is an LP relaxation  $\mathcal{L}$  of size  $R$  such that  $\mathcal{L}(G) \leq c$  whenever  $\text{opt}(G) \leq s$ , then a simple application of Farkas' Lemma shows that there are non-negative functions  $q_1, \dots, q_R : \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  such that for every graph  $G$  with  $\text{opt}(G) \leq s$ , there are coefficients  $\lambda_1, \dots, \lambda_R \geq 0$  satisfying

$$c - G = \lambda_1 q_1 + \dots + \lambda_R q_R. \tag{1.1}$$

(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

$q_i : \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  encode the slack of each constraint of the LP. Thus if the  $i$ th LP constraint is of the form  $\langle A_i, z \rangle \leq b_i$ , then  $q_i(x) = b_i - \langle A_i, y_{S_x} \rangle$  where  $y_{S_x}$  is the cut vector corresponding to  $x \in \{\pm 1\}^n$ .

The  $d$ -round Sherali-Adams relaxation for a graph  $G_0$  has  $\text{SA}_d(G_0) \leq c$  if and only if there exist a family of non-negative  $d$ -juntas  $\{f_i : \{-1, 1\}^n \rightarrow \mathbb{R}_{\geq 0}\}$  such that

$$c - G_0 = \sum_i \lambda_i f_i, \quad (1.2)$$

where  $\lambda_i \geq 0$  for each  $i$ . See Observation 2.4 and the surrounding discussion.

We recall that a  $d$ -junta is a function whose value depends on at most  $d$  of its inputs. In particular, if  $G_0$  is a  $(c, s)$  gap instance, i.e. such that  $\text{opt}(G_0) \leq s$  but  $\text{SA}_d(G_0) > c$ , then no such representation (1.2) with  $d$ -juntas can exist. Our goal is to derive an instance  $G$  from  $G_0$  such that no representation of the form (1.1) can exist.

This proceeds in three steps: First, we argue that, by a truncation argument, it suffices to consider functions  $\{q_i\}$  that are sufficiently smooth. Then in Section 3.1, we show that any sufficiently smooth  $q_i$  can be approximated (in a certain weak sense) by a  $K$ -junta  $q'_i$  for  $K$  which may be quite large (e.g.,  $K = n^{0.2}$ ).

In Section 3.2, we employ a random restriction argument: By planting the instance  $G_0$  at random inside a larger graph  $G$ , we can ensure that for every  $q'_i$ , the set of significant coordinates *when restricted to  $G_0$*  is much smaller; in fact, we show that with high probability over the random planting, every such  $q'_i$  has only  $d$  significant coordinates in the support of  $G_0$ . Here we use crucially the fact that we have only  $R$  functions  $\{q_i\}$ .

This means that, restricted to  $G_0$ , every  $q_i$  is weakly approximated by a  $d$ -junta  $q'_i$ . More specifically, all the low-degree Fourier coefficients of  $q_i - q'_i$  are small. Now the fact that no representation of the form (1.2) exists should disprove a representation of the form (1.1) as long as the “approximation” does not hurt us too much. Here we use the fact that the  $d$ -round Sherali-Adams relaxation is only capable of “perceiving” low-degree functions (more technically, the  $d$ -round Sherali-Adams functional introduced in (2.1) is a degree- $d$  multilinear polynomial). In particular, it suffices that the low-degree parts of  $q_i$  and  $q'_i$  are close.

The ingredients are all put together in Section 3.3, where one can find the proof of our main theorem for general CSPs.

## 2 Background

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

Throughout the paper, for a function  $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ , we write  $\mathbb{E} f = 2^{-n} \sum_{x \in \{\pm 1\}^n} f(x)$ . If  $g : \{\pm 1\}^n \rightarrow \mathbb{R}$ , we denote the inner product  $\langle f, g \rangle = \mathbb{E}[fg]$  on the Hilbert space  $L^2(\{-1, 1\}^n)$ . Recall that any  $f : \{\pm 1\}^n \rightarrow \mathbb{R}$  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  $|S| \leq d$ . 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  $\mu_f$  denote the corresponding probability measure on  $\{\pm 1\}^n$ . Observe that for any  $g : \{\pm 1\}^n \rightarrow \mathbb{R}$ , we have  $\mathbb{E}_{x \sim \mu_f}[g(x)] = \langle f, g \rangle$ .

**Constraint Satisfaction Problems.** Constraint satisfaction problems 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  $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ , we let MAX- $\Pi$  denote the following optimization problem: An instance  $\mathfrak{I}$  consists of boolean variables  $X_1, \dots, X_n$  and a collection of  $\Pi$ -constraints  $P_1(X) = 1, \dots, P_m(X) = 1$  over these variables. A  $\Pi$ -constraint is a predicate  $P_0 : \{\pm 1\}^n \rightarrow \{0, 1\}$  such that  $P_0(X) = P(X_{i_1}, \dots, X_{i_k})$  for some  $P \in \Pi$  and *distinct* indices  $i_1, \dots, i_k \in [n]$ . 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

$$\mathfrak{I}(x) \stackrel{\text{def}}{=} \frac{1}{m} \sum_i P_i(x).$$

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

*Examples:* MAX CUT corresponds to the case where  $\Pi$  consists of the binary inequality predicate. For MAX 3-SAT,  $\Pi$  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- $\Pi_n$  be the set of MAX- $\Pi$  instances on  $n$  variables. An *LP-relaxation of size R* for MAX- $\Pi_n$  consists of the following.

**Linearization:** For every  $\mathfrak{I} \in \text{MAX-}\Pi_n$ , we associate a vector  $\tilde{\mathfrak{I}} \in \mathbb{R}^m$  and for every assignment  $x \in \{\pm 1\}^n$ , we associate a point  $\tilde{x} \in \mathbb{R}^m$ , such that  $\mathfrak{I}(x) = \langle \tilde{\mathfrak{I}}, \tilde{x} \rangle$  for all  $\mathfrak{I} \in \text{MAX-}\Pi_n$  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  $\mathfrak{I}$  of MAX- $\Pi_n$ .

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

*Remark 2.1.* 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  $x \mapsto \mathfrak{I}(x)$  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  $\mathcal{L}(\mathfrak{I}) \leq c$  for all instances  $\mathfrak{I} \in \text{MAX-}\Pi_n$  with  $\text{opt}(\mathfrak{I}) \leq s$ . We also say that  $\mathcal{L}$  achieves an  $\alpha$ -factor approximation if  $\mathcal{L}(\mathfrak{I}) \leq \alpha \text{opt}(\mathfrak{I})$  for all  $\mathfrak{I} \in \text{MAX-}\Pi_n$ .

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 2.2.** *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  $q_1, \dots, q_R : \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  such that for every instance  $\mathfrak{I} \in \text{MAX-}\Pi_n$  with  $\text{opt}(\mathfrak{I}) \leq s$ , the function  $c - \mathfrak{I}$  is a nonnegative combination of  $q_1, \dots, q_R$ , i.e.*

$$c - \mathfrak{I} \in \left\{ \sum_i \lambda_i q_i \mid \lambda_1, \dots, \lambda_R \geq 0 \right\}.$$

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

**A communication model.** The characterization in [Theorem 2.2](#) has an illustrative interpretation as a two-party, one-way communication complexity problem: Alice's input is a  $\text{MAX-}\Pi$  instance  $\mathfrak{I}$  with  $\text{opt}(\mathfrak{I}) \leq s$ . Bob's input is an assignment  $x \in \{\pm 1\}^n$ . Their goal is to compute the value  $\mathfrak{I}(x)$  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  $v \leq c$ . The protocol is correct if for every input pair  $(\mathfrak{I}, x)$ , the expected output satisfies  $\mathbb{E}v = \mathfrak{I}(x)$  (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 2.2](#).

On the other hand, if there exist  $R = 2^L$  functions  $\{q_1, q_2, \dots, q_R\}$  as in [Theorem 2.2](#), 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  $\mathfrak{I}$  (specified by the coefficients  $\{\lambda_i\}$ ), and then Bob outputs  $c - q_i(x) \leq c$ .

*Example:* Suppose the optimization problem is  $\text{MAX CUT}$ . 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  $|\mathbb{1}_S(u) - \mathbb{1}_S(v)|$ , 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 allow 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  $c - \mathfrak{I}$  and the cone generated by  $q_1, \dots, q_R$ . The presence of  $\varepsilon$  will allow us to tolerate some noise in the arguments that follow.

**Lemma 2.3.** *In order to show that  $(c, s)$ -Max- $\Pi_n$  requires LP relaxations of size greater than  $R$ , it is sufficient to prove the following: For every collection of densities  $q_1, \dots, q_R: \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$ , there exists  $\varepsilon > 0$ , a function  $H: \{\pm 1\}^n \rightarrow \mathbb{R}$  and a Max- $\Pi_n$  instance  $\mathfrak{I}$  such that*

1.  $\langle H, c - \mathfrak{I} \rangle < -\varepsilon$
2.  $\langle H, q_i \rangle \geq -\varepsilon$ .

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

$$\langle H, c - \mathfrak{I} \rangle = \langle H, \sum_i \lambda_i q_i \rangle = \sum_i \lambda_i \langle H, q_i \rangle \geq -\varepsilon \sum_i \lambda_i = -\varepsilon \gamma,$$

which contradicts the condition  $\langle H, c - \mathfrak{I} \rangle < -\varepsilon$  since  $\gamma \leq c \leq 1$ .  $\square$

**Sherali-Adams LP relaxations for CSPs.** A primary component of our approach involves leveraging known integrality gaps for the Sherali-Adams (SA) 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].

A  $d$ -round Sherali-Adams LP relaxation for a Max- $\Pi_n$  instance will consist of variables  $\{X_S : S \subseteq [n], |S| \leq d\}$  for all products of up to degree  $d$  on the  $n$ -variables. These variables  $\{X_S : |S| \leq d\}$  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  $\{X_S : |S| \leq d\}$  is that these moments agree with a set of local marginal distributions. In particular, for every set  $S \subseteq [n]$  with  $|S| \leq d$  there exists a distribution  $\mu_S$  over  $\{\pm 1\}^S$  such that,

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

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  $\tilde{\mathbb{E}} P \geq 0$  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

$$\tilde{\mathbb{E}} \chi_\alpha = \mathbb{E}_{x \sim \mu_S} \chi_S(x) \quad \forall |S| \leq d, S \subseteq [n].$$

The  $d$ -round Sherali-Adams value of a MAX- $\Pi_n$  instance  $\mathfrak{I}$  is defined as

$$\text{SA}_d(\mathfrak{I}) \stackrel{\text{def}}{=} \max_{d\text{-l.e.f. } \tilde{\mathbb{E}}} \tilde{\mathbb{E}} \mathfrak{I}.$$

This optimization problem can be implemented by an  $n^{O(d)}$ -sized linear programming relaxation for MAX- $\Pi_n$ . (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 MAX- $\Pi_n$ , 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  $\{\chi_\alpha : |\alpha| \leq d\}$ . We can represent  $\tilde{\mathbb{E}}$  by a function  $H : \{\pm 1\}^R \rightarrow \mathbb{R}$  such that  $\tilde{\mathbb{E}} f = \langle H, f \rangle$  for all  $f : \{\pm 1\}^R \rightarrow \mathbb{R}$ . Concretely,

$$H = \sum_{\alpha \subseteq [n]: |\alpha| \leq d} c_\alpha \chi_\alpha \quad \text{for } c_\alpha = \tilde{\mathbb{E}} \chi_\alpha, \tag{2.1}$$

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

In Section 3, We will use a modification of  $H$  as a test function in the sense of Lemma 2.3. The crucial property of the test function  $H$  that makes it useful for our purposes is the following.

*Observation 2.4.* Suppose that  $f : \{\pm 1\}^n \rightarrow \mathbb{R}$  depends only on a subset of at most  $d$  coordinates  $S \subseteq [n]$ , then

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

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

One should consider this observation in light of Lemma 2.3. 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  $\langle H, q_i \rangle \geq 0$ .

*Remark 2.5.* Some work on Sherali-Adams relaxations for Max Cut focus on edge variables instead of vertex variables. This includes [FdIVKM07, CMM09]. In those papers, the  $d$ -round Sherali-Adams relaxation consists of variables  $\{X_S : S \subseteq \binom{[n]}{2}, |S| \leq d\}$  for every subset of  $d$  edges in the complete graph. Since their base polytope also includes triangle inequalities, any  $\binom{d}{2}$ -round Sherali-Adams solution with edge variables can be converted to a  $d$ -round solution for

vertex variables. One should observe that the  $d$ -round vertex relaxation is at least as strong as the  $d$ -round edge relaxation.

Moreover, both papers [FdVVKM07, CMM09] actually prove a lower bound against the  $d$ -round vertex version and then argue that this yields a lower bound for the weaker edge relaxation. For general max-CSPs, the vertex version is arguably the canonical relaxation, and it is perhaps misguided to consider the edge version even for MAX CUT. In [Sch08] (which studies general CSPs), the more natural vertex version is considered.

A major benefit of the “extended formulation” model to which our results apply is that the edge/vertex relaxation distinctions are not relevant; in fact no specific meaning is ascribed to the variables of the LP. All that matters is the number of defining inequalities.

### 3 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 3.1** (Main). *Fix a positive number  $d \in \mathbb{N}$ . Suppose that the  $d$ -round Sherali-Adams relaxation cannot achieve a  $(c, s)$ -approximation for  $\text{MAX-}\Pi_n$  for every  $n$ . Then no sequence of LP relaxations of size at most  $n^{d/2}$  can achieve a  $(c, s)$ -approximation for  $\text{MAX-}\Pi_n$  for every  $n$ .*

We prove the following more general result in Section 3.3.

**Theorem 3.2.** *Consider a function  $f : \mathbb{N} \rightarrow \mathbb{N}$ . Suppose that the  $f(n)$ -round Sherali-Adams relaxation cannot achieve a  $(c, s)$ -approximation for  $\text{MAX-}\Pi_n$ . Then for all sufficiently large  $n$ , no LP relaxation of size at most  $n^{f(n)^2}$  can achieve a  $(c, s)$ -approximation for  $\text{MAX-}\Pi_N$ , where  $N \leq n^{10f(n)}$ .*

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

#### 3.1 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  $(X_1, \dots, X_n) \in \{\pm 1\}^n$  be correlated random bits with distribution  $\mu$ . 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  $\mu$ . Here, and in what follows,  $H(\cdot)$  denotes the Shannon entropy measured in bits.

**Lemma 3.3.** *For all  $1 \leq d, t \leq n$  and  $\beta > 0$ , the following holds. If  $\mu$  has entropy  $\geq n - t$ , there exists a set  $J \subseteq [n]$  of at most  $\frac{td}{\beta}$  coordinates such that for all subsets  $A \not\subseteq J$  with  $|A| \leq d$ , we have*

$$\max_{v \in A} H(X_v | X_{A \setminus v}) \geq 1 - \beta. \quad (3.1)$$

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

$$H(X_v | X_{e \setminus v}) \leq 1 - \beta. \quad (3.2)$$

Lemma 3.3 follows directly from the next claim.

**Proposition 3.4.** *If  $\mu$  has entropy  $n - t$ , then  $|\bigcup_{e \in E(G_\beta)} e| \leq \frac{td}{\beta}$ .*

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

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

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

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

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

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

We now record a Fourier-theoretic consequence of Lemma 3.3. To this end, recall that for two probability measures  $\mu$  and  $\nu$  over  $\{\pm 1\}^n$ , one defines the *KL-divergence* as the quantity

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

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

$$D(\mu \| \nu) \geq \frac{1}{\ln 4} \|\mu - \nu\|_1^2. \quad (3.3)$$

**Lemma 3.5.** Let  $\mu$  be a distribution as in the statement of Lemma 3.3, and let  $J \subseteq [n]$  be the corresponding set of coordinates. If  $A \subseteq [n]$  satisfies  $|A| \leq d$  and  $A \not\subseteq J$ , then

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

*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  $|A| \leq d$  and  $A \not\subseteq J$ , Lemma 3.3 implies that some  $v \in A$  satisfies (3.1). 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

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

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

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

by (3.3).

Since

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

we get

$$\left| \mathbb{E}_{\mu}[\chi_A(x)] \right| \leq \mathbb{E}_{y \sim \mu} \sqrt{(\ln 4)(1 - H(\mu_{|y}))} \leq \sqrt{(\ln 4)(1 - \mathbb{E}_y[H(\mu_{|y})])},$$

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

Finally, we arrive the primary goal of this section.

**Lemma 3.6** (High-Entropy Distributions vs Juntas). *Let  $q: \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  be a density and let  $\mu_q$  denote the corresponding measure on  $\{\pm 1\}^n$ . If  $\mu_q$  has entropy least  $n - t$  for some  $t \leq n$ , then for every  $1 \leq d \leq n$  and  $\gamma > 0$ , there exists a set  $J \subseteq [n]$  with*

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

such that for all subsets  $\alpha \not\subseteq J$  with  $|\alpha| \leq d$ , we have  $|\hat{q}(\alpha)| \leq \gamma$ .

*Proof.* Set  $\beta = \frac{\gamma^2}{4}$ . Now apply Lemma 3.5 and use the fact that  $\hat{q}(\alpha) = \mathbb{E}_{x \sim \mu_q}[\chi_{\alpha}(x)]$ .  $\square$

## 3.2 Random Restrictions

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

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq \frac{pn}{2}\right) \geq 1 - e^{-pn/8}. \quad (3.4)$$

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

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) = \sum_{i=t}^n \binom{n}{i} p^i (1-p)^{n-i} \leq (pn)^t \sum_{i=0}^{\infty} (pn/(1-p))^i \leq 2(pn)^t. \quad (3.5)$$

**Lemma 3.7.** *For any  $d \in \mathbb{N}$ , the following holds. Let  $Q$  be a collection of densities  $q: \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  such that the corresponding measures  $\mu_q$  have entropy at least  $n - t$ . If  $|Q| \leq n^{d/2}$ , then for all integers  $m$  with  $3 \leq m \leq n/4$ , 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  $\mu_q$ , all  $d$ -wise correlations in  $S - J(q)$  are small. Quantitatively, we have

$$|\hat{q}(\alpha)| \leq \left(\frac{32mtd}{\sqrt{n}}\right)^{1/2} \quad \forall \alpha \subseteq S, \alpha \not\subseteq J(q), |\alpha| \leq d$$

*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 (3.4), we have  $|S| \geq m$  with probability at least  $1 - e^{-m/4} > 1/2$ .

Fix  $\gamma = \left(\frac{32mtd}{\sqrt{n}}\right)^{1/2}$ . By Lemma 3.6, for each  $q \in Q$  there exists a set  $J'(q)$  of at most  $\frac{4td}{\gamma^2} \leq \frac{\sqrt{n}}{8m}$  coordinates such that for all subsets  $\alpha \not\subseteq J'(q)$  with  $|\alpha| \leq d$ , we have  $|\hat{q}(\alpha)| \leq \gamma$ .

The set  $J(q)$  for a distribution  $q$  is given by  $J(q) = J'(q) \cap S$ . Clearly,  $\mathbb{E}[|J'(q) \cap S|] = \frac{2m}{n} |J'(q)| \leq \frac{2m}{n} \cdot \frac{\sqrt{n}}{8m} \leq 1/4$ . Thus by (3.5), we can write

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

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

### 3.3 Proof of Main Theorem – Theorem 3.1

In this subsection, we will prove Theorems 3.2 and 3.1. Let  $m \leq n$  be parameters  $m, n \in \mathbb{N}$  to be chosen later. Let  $\mathfrak{I}_0$  be a MAX- $\Pi_m$  instance with  $\text{SA}_d(\mathfrak{I}_0) \geq c$  and  $\text{opt}(\mathfrak{I}_0) \leq s$ . 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- $\Pi_n$  cannot certify  $\text{opt}(\mathfrak{I}) < c - \varepsilon$  for all “shifts”  $\mathfrak{I}$  of  $\mathfrak{I}_0$ . (Here, “shift” means that we plant  $\mathfrak{I}_0$  on some subset of the variables  $\{1, \dots, n\}$ .)

Let  $Q$  denote a collection of densities  $q: \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  on  $\{\pm 1\}^n$  with  $|Q| \leq n^{d/2}$ . We will show that there exists a shift  $\mathfrak{I}$  of  $\mathfrak{I}_0$  and a test functional  $H \in L^2(\{-1, 1\}^n)$  that satisfies the conditions of Lemma 2.3 (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

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

By Markov's inequality, the set  $B$  has measure at most  $\mathbb{E} \mathbb{1}_B \leq |Q| \cdot n^{-d} \leq n^{-d/2}$ .

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

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

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

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

Since  $q(x) < n^d$  for every point  $x \notin B$ , the function  $q'$  satisfies  $q' \leq n^d + 1$ , 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 3.7](#), there exists a set  $S \subseteq [n]$  of size  $m$  such that every  $q \in Q$ , every Fourier coefficient  $\hat{q}'(\alpha)$  with degree  $|\alpha| \leq d$  and  $\alpha \notin J(q)$  satisfies  $|\hat{q}'(\alpha)| \leq K$ , where

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

For a subset  $T \subseteq [n]$ , let  $q_{|T}$  denote the marginal of  $q$  on  $T$ , so that  $q_{|T}(x) = \mathbb{E}_{y \in \{\pm 1\}^{[n] \setminus T}} q(x_T, y_{[n] \setminus T})$ . (Equivalently,  $q_{|T} = \sum_{\alpha \subseteq T} \hat{q}(\alpha) \chi_\alpha$ .) In this notation, the  $\ell_1$ -norm of the degree- $\leq d$  Fourier coefficients of  $q'_{|J(q)} - q'_{|S}$  is bounded by

$$\sum_{\substack{\alpha \subseteq S \\ \alpha \not\subseteq J(q) \\ |\alpha| \leq d}} |\hat{q}'(\alpha)| \leq Km^d. \quad (3.6)$$

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  $\mathfrak{I}$  be the Max- $\Pi_n$  instance obtained by planting the Max- $\Pi_m$  instance  $\mathfrak{I}_0$  on the variable set  $S$ . Let  $H$  be a  $d$ -round Sherali-Adams solution (in the sense of [\(2.1\)](#)) for  $\mathfrak{I}$  with value  $c$ . Further, we will choose a Sherali-Adams solution  $H$  that satisfies  $\hat{H}(\alpha) = 0$  for  $\alpha \notin S$ . This is possible since all the constraints in  $\mathfrak{I}$  are contained within  $S$ . (Since  $\mathfrak{I}$  is a shift of  $\mathfrak{I}_0$ , it has the same Sherali-Adams value,  $\text{SA}_d(\mathfrak{I}) = \text{SA}_d(\mathfrak{I}_0)$ .)

We claim that  $H' = (1 - \mathbb{1}_B) \cdot H$  is a test function in the sense of [Lemma 2.3](#). Recall that  $\|\mathfrak{I}\|_\infty \leq 1$ . Furthermore, by [\(2.1\)](#), we have  $\|H\|_\infty \leq m^d$ . Using these,

$$\begin{aligned} \langle H', c - 2\eta - \mathfrak{I} \rangle &= \langle H, c - 2\eta - \mathfrak{I} \rangle - \langle H \cdot (c - 2\eta - \mathfrak{I}), \mathbb{1}_B \rangle \\ &\leq -2\eta - \langle H \cdot (c - 2\eta - \mathfrak{I}), \mathbb{1}_B \rangle \\ &\quad (\text{because } \langle H, 1 \rangle = 1 \text{ and } \langle H, \mathfrak{I} \rangle = \text{SA}_d(\mathfrak{I})) \end{aligned}$$

$$\begin{aligned}
&\leq -2\eta + (c + 2\eta + \|\mathfrak{I}\|_\infty) \|H\|_\infty \cdot \mathbb{E} \mathbf{1}_B \\
&\leq -2\eta + 3m^d \cdot n^{-d/2} \\
&< -\eta.
\end{aligned}$$

Finally,

$$\begin{aligned}
\langle H', q \rangle &= \langle H', q' - \mathbb{E} q_{\text{bad}} \rangle + \langle H', q_{\text{bad}} \rangle \\
&= \langle H', q' - \mathbb{E} q_{\text{bad}} \rangle + 0 \quad (\text{using } \text{supp}(q_{\text{bad}}) \cap \text{supp}(H') = \emptyset) \\
&= \langle H, q' - \mathbb{E} q_{\text{bad}} \rangle + 0 \quad (\text{using } \text{supp}(q' - \mathbb{E} q_{\text{bad}}) \cap \text{supp}(H - H') = \emptyset) \\
&= \langle H, q' \rangle - \mathbb{E} q_{\text{bad}} \quad (\text{using } \langle H, 1 \rangle = 1) \\
&= \langle H, q'_{|S} \rangle - \mathbb{E} q_{\text{bad}} \quad (\text{using } \hat{H}(\alpha) = 0 \text{ for } \alpha \not\subseteq S) \\
&= \langle H, q'_{|J(q)} \rangle - \langle H, q'_{|J(q)} - q'_{|S} \rangle - \mathbb{E} q_{\text{bad}} \\
&\geq \langle H, q'_{|J(q)} \rangle - \eta - \mathbb{E} q_{\text{bad}} \quad (\text{using } |\hat{H}(\alpha)| \leq 1 \text{ for all } \alpha \subseteq S, \deg(H) \leq d \\
&\quad \text{and our Fourier-}\ell_1\text{-norm bound (3.6) for } q'_{|J(q)} - q'_{|S}) \\
&\geq \mathbb{E} q_{\text{bad}} - \eta - \mathbb{E} q_{\text{bad}} \\
&\geq -\eta.
\end{aligned}$$

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

We conclude that by Lemma 2.3, the cone generated by  $Q$  does not contain the function  $c - 2\eta - \mathfrak{I}$ .

We see that for any collection  $Q$  of at most  $n^{d/2}$  densities, there exists a Max- $\Pi_n$  instance  $\mathfrak{I}$  with  $\text{opt}(\mathfrak{I}) \leq s$  (a shift of  $\mathfrak{I}_0$ ) such that  $c - 2\eta - \mathfrak{I}$  is not in the cone generated by  $Q$ . By Theorem 2.2, it follows that any  $n^{d/2}$ -sized linear programming relaxation cannot achieve a  $(c - 2\eta, s)$ -approximation for Max- $\Pi_n$ . Finally, note that asymptotically, we have

$$\eta = O\left(\frac{m^d \sqrt{md \log n}}{n^{1/4}}\right). \quad (3.7)$$

□

*Proof of Theorem 3.2.* Fix an instance size  $m$  and put  $d = f(m)$ . In the preceding argument, require that  $n$  grows like  $m^{10d} = m^{10f(m)}$  so that  $\eta = o(1)$  (see (3.7)). The lower bound achieved is  $n^{d/2} \geq m^{5f(m)^2}$ . □

## 4 Symmetric Linear Programs

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

**Theorem 4.1.** 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  $\leq \binom{N}{t}$  can achieve a  $(c, s)$ -approximation on  $N$ -vertex graphs, where  $N = 2n$ .

We will require the following lemma of Yannakakis.

**Lemma 4.2** ([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  $J$ .

**Lemma 4.3.** Suppose a family of functions  $\mathcal{F} = \{f_i : \{\pm 1\}^n \rightarrow \mathbb{R} : i = 1, 2, \dots, M\}$  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  $\sum_{a \in [n]} x_a$ .

*Proof.* Let  $\text{Orb}(f)$  denote the orbit of a function  $f$  under permutation of its inputs. Since  $\mathcal{F}$  is closed under permutation of inputs to the functions, it contains the orbits of each of the functions  $f_1, \dots, f_M$ .

This implies that for each of the functions  $f_i$ , we have  $|\text{Orb}(f_i)| < \binom{n}{k}$ . Hence for each  $f_i$  the automorphism group of all permutations that preserve the function  $f_i$  is large, i.e.,  $|\text{Aut}(f_i)| > k!(n-k)!$ . By Lemma 4.2, the automorphism group  $\text{Aut}(f_i)$  contains all even permutations that fix a subset of coordinates  $J_i$  with  $|J_i| \leq k$ .

We claim that  $\text{Aut}(f_i)$  contains all permutations that fix the coordinates in  $J_i$ . We know that for every  $x$ , and every even permutation  $\sigma$  of  $\bar{J}_i$ ,  $f_i(x) = f_i(\sigma x)$ . For every  $x \in \{\pm 1\}^n$ , there will be two coordinates  $a, b \in \bar{J}_i$  such that  $x_a = x_b$ . Let  $\pi_{ab}$  denote the permutation that swaps  $x_a, x_b$ . Since  $\pi_{ab}(x) = x$ , we have  $f_i(\pi_{ab}(x)) = f_i(x)$ . So for all even permutations  $\sigma$ ,

$$f_i(\sigma \pi_{ab} x) = f_i(\pi_{ab} x) = f_i(x)$$

As  $\sigma$  varies over all even permutations,  $\sigma \pi_{ab}$  varies over all odd permutations, leading to the conclusion that for all permutations  $\pi$  fixing the coordinates of  $J_i$ ,  $f_i(x) = f_i(\pi x)$ .

This symmetry of the function  $f_i$  implies that it depends only on the assignment to  $J_i$  and the hamming weight of the assignment to  $\bar{J}_i$ . The hamming weight of the assignment to  $\bar{J}_i$  is equal to  $\sum_a x_a - \sum_{a \in J_i} x_a$ . This shows that  $f_i$  only depends on the coordinates in  $J_i$  and the value  $\sum_{a \in [n]} x_a$ .  $\square$

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

Construct a new graph  $\mathfrak{I}' = \mathfrak{I}_1 \cup \mathfrak{I}_2$  consisting of two disjoint copies of the instance  $\mathfrak{I}$ . Let  $N = 2n$  denote the number of vertices of  $\mathfrak{I}'$ . Let us suppose that  $\mathfrak{I}_1$  is on vertices  $\{1, \dots, n\}$  and  $\mathfrak{I}_2$  is on  $\{n+1, \dots, 2n\}$ .

We will now extend the Sherali-Adams LP solution for  $\mathfrak{I}$  to a Sherali-Adams solution for  $\mathfrak{I}'$ . Roughly speaking, we will copy the SA-solution as is on to  $\mathfrak{I}_1$

and negate all its values on  $\mathfrak{I}_2$ . In other words, the Sherali-Adams solution on  $\mathfrak{I}'$  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  $S_1 = S \cap \{1, \dots, n\}$  and  $S_2 = \{i - n \mid i \in S \setminus S_1\}$ . Then, we can describe the SA solution  $\{Y_S\}_{S \subseteq [2n], |S| \leq t}$  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 4.4.** For a set  $S \subseteq [2n]$  of the form  $S = S_1 \cup (S_2 + n)$  where  $S_1, S_2 \subseteq [n]$ , define  $\text{wt}(S) = |S_1 \oplus S_2|$ .

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

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

*Observation 4.5.* For any set  $S$ ,  $\text{wt}(S \oplus \{i\}) = \text{wt}(S \oplus \{n+i\})$ .

*Proof.* It is easy to see that both values are equal to the hamming weight of the  $\mathbb{F}_2$  vector  $\mathbf{1}_{S_1} \oplus \mathbf{1}_{S_2} \oplus e_i$  where  $\mathbf{1}_{S_1}, \mathbf{1}_{S_2} \in \mathbb{F}_2^n$  are indicators of subsets  $S_1, S_2$  and  $e_i \in \mathbb{F}_2^n$  is the  $i^{\text{th}}$  standard basis vector.  $\square$

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

$$H(x) = \sum_{S \subseteq [N]} Y_S \chi_S(x), \quad (4.1)$$

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

**Lemma 4.6.** 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 4.5,  $\text{wt}(S \oplus \{i\}) = \text{wt}(S \oplus \{n+i\})$ . If  $\text{wt}(S \oplus \{i\}) = \text{wt}(S \oplus \{n+i\}) > t$ , then by definition we will have  $Y_{S \oplus \{i\}} = Y_{S \oplus \{n+i\}} = 0$ .

On the other hand, if  $\text{wt}(S \oplus \{i\}) = \text{wt}(S \oplus \{n+i\}) \leq t$  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  $Y_{S \oplus \{i\}} = -Y_{S \oplus \{n+i\}}$ .  $\square$

**Lemma 4.7.** For any polynomial  $p(x_1, \dots, x_N)$ , 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,

$$\langle H, \left( \sum_i x_i \right) p \rangle = \mathbb{E}_{x \in \{\pm 1\}^N} \left[ H(x) \left( \sum_i x_i \right) p(x) \right]$$

Since  $x$  takes values in  $\{\pm 1\}^N$ , 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,

$$\begin{aligned} \left\langle \sum_{A \subseteq [N]} Y_A \chi_A(x), \chi_S(x) \left( \sum_i x_i \right) \right\rangle &= \mathbb{E}_{x \in \{\pm 1\}^N} \left[ \chi_S(x) \sum_{A \subseteq [N], i \in [2n]} Y_A \chi_{A \oplus \{i\}}(x) \right] \\ &= \mathbb{E}_{x \in \{\pm 1\}^N} \left[ \chi_S(x) \sum_{B \subseteq [N]} \chi_B(x) \left( \sum_{i \in [2n]} Y_{B \oplus \{i\}} \right) \right] \\ &= 0 \quad (\text{because } \sum_{i \in [2n]} Y_{B \oplus \{i\}} = 0 \text{ since } Y_{B \oplus \{i\}} = -Y_{B \oplus \{n+i\}}) \end{aligned}$$

□

**Lemma 4.8.** *If  $f : \{\pm 1\}^N \rightarrow \mathbb{R}_{\geq 0}$  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*

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

*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 4.6, we have

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

Since  $p_0$  depends on at most  $t$  coordinates, by Observation 2.4 we can write,

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

where  $\mu_J$  is some distribution on  $x_J$ .

Define a distribution  $\mu$  on  $\{\pm 1\}^N$  as follows: Sample  $x_J$  from  $\mu_J$  and then sample  $x_J$  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  $\mu$  is supported entirely on the set  $\{x \in \{\pm 1\}^N \mid \sum_i x_i = 0\}$ . Therefore, we have

$$\begin{aligned} \langle H, p_0 \rangle &= \mathbb{E}_{x_J \sim \mu_J} [p_0(x_J)] \\ &= \mathbb{E}_{x \sim \mu} [p_0(x)] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_{x \sim \mu} \left[ p_0(x_J) + \sum_{i=1}^N p_i(x_J) \left( \sum_i x_i \right)^i \right] \quad (\text{because } \sum_i x_i = 0 \text{ on the support of } \mu) \\
&= \mathbb{E}_{x \sim \mu} [f(x)] \geq 0 \quad (\text{because } f \text{ is pointwise non-negative}) \quad \square
\end{aligned}$$

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

*Proof of Theorem 4.1.* Suppose that  $N$  is even. Let  $\mathcal{F} = \{f_1, \dots, f_M\}$  denote the family of functions from  $\{\pm 1\}^N$  to  $\mathbb{R}_{\geq 0}$  associated with some symmetric LP relaxation of MAX CUT, i.e. those coming from an application of Theorem 2.2.

By the symmetry assumption, the family  $\mathcal{F}$  is closed under permutations of its inputs. Hence, by Lemma 4.3 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  $\mathfrak{I}$  on  $n = N/2$  nodes with  $\text{SA}_t(\mathfrak{I}) > c$ , and let  $\mathfrak{I}'$  be the graph obtained by taking two copies of  $\mathfrak{I}$  as discussed before. Then  $\mathfrak{I}'$  has the property that  $\text{SA}_t(\mathfrak{I}') = \text{SA}_t(\mathfrak{I})$  and  $\text{opt}(\mathfrak{I}') = \text{opt}(\mathfrak{I})$ . Let  $H$  be the corresponding functional defined in (4.1).

Let us consider  $\mathfrak{I}'$  as a function on  $\{\pm 1\}^N$  assigning cuts to their MAX CUT value in  $\mathfrak{I}'$ . Suppose we can express

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

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

$$\langle H, c - \mathfrak{I}' \rangle = c - \text{SA}_t(\mathfrak{I}') < 0$$

while,

$$\langle H, f_i \rangle \geq 0 \quad \forall i \quad \text{by Lemma 4.8,}$$

a contradiction.  $\square$

## 5 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  $n^{c_\varepsilon}$  rounds, where  $c_\varepsilon$  is some constant depending on  $\varepsilon$  [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 (3.6).

**Question 5.1.** 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  $\text{poly}(\frac{\log R(n)}{\log n})$  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 5.2.** 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. [1](#), [3](#)
- [ABLT06] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis, *Proving integrality gaps without knowing the linear program*, Theory Comput. **2** (2006), 19–51. [3](#)
- [BFPS12] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer, *Approximation limits of linear programs (beyond hierarchies)*, FOCS, 2012, pp. 480–489. [2](#), [3](#), [7](#)
- [BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani, *SDP gaps from pairwise independence*, Theory of Computing **8** (2012), no. 12, 269–289. [3](#)
- [BM13] Mark Braverman and Ankur Moitra, *An information complexity approach to extended formulations*, STOC, ACM, 2013, pp. 161–170. [2](#)
- [CMM09] M. Charikar, K. Makarychev, and Y. Makarychev, *Integrality gaps for Sherali-Adams relaxations*, Proc. STOC, ACM, 2009, pp. 283–292. [3](#), [4](#), [9](#), [10](#), [19](#)
- [FdIVKM07] Wenceslas Fernández de la Vega and Claire Kenyon-Mathieu, *Linear programming relaxations of Maxcut*, SODA, 2007, pp. 53–61. [3](#), [4](#), [9](#), [10](#)

- [FFGT11] Y. Faenza, S. Fiorini, R. Grappe, and H. R. Tiwary, *Extended formulations, non-negative factorizations and randomized communication protocols*, arXiv:1105.4127, 2011. 8
- [FMP<sup>+</sup>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. 2
- [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. 2
- [Hås01] Johan Håstad, *Some optimal inapproximability results*, J. ACM **48** (2001), no. 4, 798–859. MR 2144931 (2006c:68066) 2
- [Lau03] M. Laurent, *A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. (2003), 470–496. 3, 8
- [LS91] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization*, SIAM J. Optim. **1** (1991), 166–190. 1
- [McD98] Colin McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248. MR 1678578 (2000d:60032) 12
- [Pas12] K. Pashkovich, *Extended formulations for combinatorial polytopes*, Ph.D. thesis, Magdeburg Universität, 2012. 7
- [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. 1
- [Sch08] G. Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, Proc. FOCS, IEEE, 2008, pp. 593–602. 3, 10
- [Tsy09] Alexandre B. Tsybakov, *Introduction to nonparametric estimation*, Springer, 2009. 11
- [Vaz01] Vijay V. Vazirani, *Approximation algorithms*, Springer-Verlag, Berlin, 2001. MR 1851303 (2002h:68001) 1
- [WS11] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*, Cambridge University Press, Cambridge, 2011. 1
- [Yan91] M. Yannakakis, *Expressing combinatorial optimization problems by linear programs*, J. Comput. System Sci. **43** (1991), no. 3, 441–466. 1, 2, 16, 20