Approximate constraint satisfaction requires large LP relaxations

Technical Report

Neha Gupta

1 Introduction

Linear Programming is one of the most important tools for finding approximate solutions to NP hard problems. Thus, many works have attempted to show lower bounds on linear programs i.e. how well can some problem be approximated by a polynomial sized linear program. The authors in [1] started a systematic study of linear programs and in particular showed integrality gaps for vertex cover for Lovasz Shreiver and Sherali Adams hierarchy. Now, there has been a lot of work in this area on proving integrality gaps for these LP/SDP hierarchies with a certain number of rounds for various problems. This rules out a large class of linear programs which cannot be used to approximate a problem within a certain factor however, still there is a gap since this does not rule out some other polynomial sized LPs. In this paper, the authors try to minimize this gap and try to show lower bounds for "natural" polynomial sized LPs for approximating constraint satisfaction problems.

There has been a long line of work started by authors in [2] who showed certain lower bounds for symmetric LPs for Travelling Salesman problem. The authors used connections between extension complexity and communication complexity. This technique was later used in several papers to show lower bounds for LPs for other problems.

Then, there is this another direction of lower bounds for LPs for problems where people show lower bounds for LPs generated by these systematic lift and project hierarchies like (Lovasz Shreiver and Sherali Adams). Very strong lower bounds are known for these hierarchies for certain problems like approximing constraint satisfaction problems. Particularly, we know that authors in [3] showed that for fixed k and ϵ, max-cut has an integrality gap for $\frac{1}{2}+\epsilon$ for k rounds of Sherali Adams relaxation. Authors in [4] showed that Max-Cut has an integrality gap of $\frac{1}{2}+\epsilon$ for $n^{\Omega(\epsilon)}$ rounds of Sherali Adams relaxation. Authors in [5] show tight integrality gaps for k-CSPs for Lasserre hierarchy (which also apply to SA relaxation since it is weaker than the Lasserre hierarchy). For example, the authors show that MAX 3-SAT has an integrality gap of $\frac{7}{8}+\epsilon$ for $\Omega(n)$ rounds of Lasserre hierarchy.

The authors in this paper [9] use the lower bounds on these systematic lift and project hierarchies to show lower bounds for general "natural" polynomial sized LPs for approximation CSPs.

2 Preliminaries and Notation

2.1 Constraint Satisfaction Problems

In constraint satisfaction problems (CSPs), we are given a set of n boolean variables and we are given m k-ary predicates $P_i : \{-1,1\}^k \rightarrow \{0,1\}$ and our aim is to find an assignment which maximizes the number of satisfied predicates.

We will consider the example of Max-Cut on a graph $G = (V, E)$ where $|V| = n$ as an example of Max-CSP problem. As we know, optimal value of max-cut is the maximum fraction of edges that can be cut by any partition of vertices into two sets. Formally, let $x = (x_1, x_2, ..., x_n)$ where $x_i \in \{-1, 1\}$ and x_i represents the ith vertex and $x_i = 1$ if it is in the first set and -1 otherwise. Then,

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

The "natural" LPs that are considered in this paper have two properties:

1. The LP has a vector v_G for every graph G and a vector y in some higher dimension for very cut where the LP value $L(G) = v_G.y$. For max cut, v_G can represent whether for every pair of vertices, whether there is an edge (1) or not (0), and y can represent whether two vertices are separated or not, 1 if separated, 0 otherwise. We can add appropriate required constraints on y. (Note, y may or may not represent an integral cut).
2. The LP polytope (the feasible region) is not allowed to depend on the problem instance and hence, is same for all problems of same size n. Only the objective function depends on the specific problem instance and weights of the edges. We see this is true for the LPs that we considered above.

We say a LP has a (c, s) approximation if for all instance J of size n graph, if $OPT(J) \leq s$, then $L(J) \leq c$.

2.2 Characterization of LPs

This is a characterization of LPs for CSPs as we described above.

Theorem 1. *There exists an LP relaxation of size atmost R that achieves a (c, s) approximation iff there exists non negative functions $q_1, .., q_R : \{-1, 1\}^n \to R_{\geq 0}$ such that for every instance J of Max-CSP with $OPT(J) \leq s$, the function c - J is a non negative combination of functions $q_1, .., q_R$ and 1. More formally,*

$$c - J \in \{\lambda_0 + \sum_{i=1}^{R} \lambda_i q_i | \lambda_i \geq 0\}$$

Proof. First, we prove that (c, s) LP gives a representation of the above form: Let $\widetilde{J}, \widetilde{x}$ be the linear represenation of the instance J and the cut x as described above. Also, let the polytope P of the LP be defined by R inequalities of the form $A_i.y \leq b_i$ and this should be satisfied for every $\widetilde{x}$ representing an actual integral $\{-1, 1\}$ cut. Let us define the functions $q_i : \{-1, 1\}^n \to R_+$ by $q_i(x) = b_i - A_i.\widetilde{x}$.

Now, if for an instance J, opt(J) $\leq s$, then by assumption, $L(J) \leq c$ implying that $\widetilde{J}.y \leq c$ for all y belonging to the LP polytope P. Now, Farkas lemma says that every valid inequality over a polytope P can be written as a non negative combination of the inequalities defining the polytope and the inequality $1 \geq 0$ and hence, we get that $c - \widetilde{y}.\widetilde{J} = \lambda_0 + \sum_{i=1}^{R} \lambda_i (b_i - A_i.\widetilde{y})$ holds for all $\widetilde{y}$ belonging to P. And since $\widetilde{y}.\widetilde{J} =$ cut value, we get the desired result.

Let us show the other direction now:
For every instance J with opt$(J) \leq s$, we know that $c - J \in \{\lambda_0 + \sum_{i=1}^{R} \lambda_i q_i | \lambda_i \geq 0\}$ and $q_1, .., q_R$: $\{-1, 1\}^n \to R_{\geq 0}$. Now, we can always think of q_is as multilinear polynomials on the boolean domain with co-efficients corresponding to the different subsets $S \subseteq [n]$. Let this representation be called $\hat{q}_i$. For any

vector $x \in \{-1,1\}^n$, consider the linearization $\widetilde{y} \in R^{2^n}$ where each co-ordinate represents the product of x_i where $i \in S, S \subseteq [n]$. We can also consider the CSP instance variable J as a vector $\in R^{2^n}$. For example, for max-cut, $J_S = 1$ if $|S| = 2$ and co-ordintes in S form an edge and 0 otherwise. Now, we know that $\widetilde{y}.J =$ cut value for x where x has linearization $\widetilde{y}$. Now, let us consider the polytope defined by the inequalities $\widetilde{y}.\hat{q}_i \geq 0 \forall i \in \{1,..,R\}$. This polytope has size R because this is defined by R inequalities. All the integral cuts belong to this polytope because we know that $\widetilde{y}.\hat{q}_i = q_i(x) \geq 0$ where $\widetilde{y}$ is the linearization of integral cut x. Now, for any $\widetilde{y}$ belonging to the polytope, we know $\widetilde{y}.\hat{q}_i \geq 0$ and hence, $c - J.\widetilde{y} \geq 0$ and hence maximum LP value $\leq c$. Therefore, this direction is also proved.

□

2.3 Characterization of SAs programs

For Sherali Adams relaxation of rounds d, we know that the Sherali Adams pseudo distribution behaves as a locally consistent real distribution for subsets of variables upto size d. Let the Sherali Adams pseudo expectation be $\widetilde{E} = E_{x \sim \mu}$. Using this fact, let us see a few properties of Sherali Adams pseudo distributions:

1. For any non negative polynomial P that depends on atmost d variables, $\widetilde{E}(P) \geq 0$
2. $\widetilde{E}(1) = 1$
3. For any d degree pseudo distribution, there exists another d degree pseudo distribution which can be represented a d degree multilinear polynomial and has the same moments for multilinear polynomials upto degree d. We will consider μ as the pseudo distribution which can be represented as a degree d multinomial.
4. Now, if $\mu = \sum_{S \subseteq [n], |S| \leq d} c_S X_S$ where $X_S = \prod_{i \in S} x_i$
   $\widetilde{E}X_{S'} = \sum_{x \in \{-1,1\}^n} \mu(x) X_{S'}(x) = \sum_{x \in \{-1,1\}^n} \sum_{S \subseteq [n], |S| \leq d} c_S X_S(x) X_{S'}(x)$, The only term with $S = S'$ remains, and all other terms cancel out, hence we see, that $\widetilde{E}X_{S'} = 0$ if $|S'| > d$, otherwise, $2^{-n}\widetilde{E}X_{S'} = c_{S'}$
   Therefore, $\mu = 2^{-n} \sum_{S \subseteq [n], |S| \leq d} (\widetilde{E}X_S) X_S$
5. $|\widetilde{E}X_S| \leq 1$ since, X_S takes values -1 and 1, and this is a real distribution over sets of size atmost d and has expectation 0 otherwise.

The d round Sherali Adams relaxation for a graph G of size n has value $SA(G) \leq c$ iff there exists a family of non negative d-juntas $f_i : \{-1,1\}^n \to R_{\geq 0}$ such that $c - G = \sum_i \lambda_i f_i$ where $\lambda_i \geq 0 \forall i$. Note that d-junta is function whose output values depends on atmost d input variables.

Proof: One direction is easy to see. If we can represent $c - G$ as descibed above, then clearly for any SA pseudo distribution μ, if we take the expectation wrt μ of the equation described above, we get $c - SA(G) = \sum_i \lambda_i \widetilde{E} f_i$ and now since f_i is a non negative d-junta, right hand side ≥ 0. Therefore, we get that $SA(G) \leq c$. Here, we used that $SA(G) = \max_{\text{SA pseudo distriubtion } \mu} E_{x \sim \mu} G(x)$

For the other direction, we need to show that if $SA(G) \leq c$, then $c - G$ can be written as descibed above.

Let us consider the cone of non-negative d-juntas which is just the representation described above $\sum_i \lambda_i f_i$ where $\lambda_i \geq 0 \forall i$, f_is are non-negative d-juntas on the hypercube. Let us consider the dual cone of this cone which is the set of vectors which have a positive dot product with all vectors in the cone when the functions q_i are represented as vectors with dimension 2^n. Now let us prove that these vectors in the dual cone are valid SA pseudo distribution when represented as probability distribution. This is easy to see because their expectation on positive d-juntas will be non-negative and hence they are valid locally consistent distributions on subsets of variables of atmost size d. We can always scale the vectors appropriately which is sufficient to prove they are valid pseudo distribution for SA of rounds d. Now, consider valid SA pseudo distributions, since they are real distributions on subsets of variables of size atmost d, they will have non-negative dot product with all vectors lying the the non-negative d-junta cone. Hence, we see that the dual cones of both the spaces are the same and hence, their cones should also be the same given these are convex sets.

Theorem 2. *For a positive constant d and for k-ary CSPs with $k \leq d$, if d round Sherali Adams relaxation cannot achieve a (c,s) approximation for Max-CSP, then no "natural" LP relaxations of size atmost $n^{\frac{d}{2}}$, can achieve a (c,s) approximation for Max-CSP where n is the number of variables in the CSP.*

Proof. Now, let us recall our Theorem 1, if that representation for any LP had the functions q_i as d-juntas then that LP could not do better than any SA relaxation of size d. Let us see why: $c - J \in \{\lambda_0 + \sum_{i=1}^{R} \lambda_i q_i | \lambda_i \geq 0\}$, taking expectation with respect to SA pseudo distribution μ, we get that
$c - \widetilde{E}(J) = \lambda_0 + \sum_{i=1}^{R} \lambda_i \widetilde{E}[q_i]$, if q_i are non negative d-juntas, then right hand side ≥ 0 and hence, $c - \widetilde{E}(J) \geq 0$ and hence, $c \geq SA(J)$ and if we consider $c = L(J)$, we get that $L(J) \geq S(J)$ and hence, any LP relaxation is as atleast as good as SA relaxation and hence, the integrality gap instance of Sherali Adams relaxation also applied to any LP. Now, these functions may not need to be d-junta in general and hence, the idea is to approximate these functions by non-negative d-juntas.

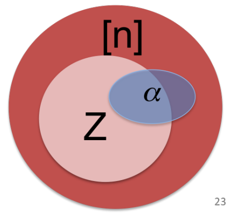
To approximate these functions q_i by d-juntas, the idea is to take the integrality gap instance G_0 of size $m << n$ ($m \sim n^{\frac{1}{10d}}$), and plant it on a random subset S of vertices of graph G of size n and then we show that "smooth" q_i when restricted to this random subset of vertices is close to a $d-junta$ (well approximated by a d-junta on low degree coefficients of the polynomials). This is sufficient because the Sherali Adams expectation has higher degree moments 0 and only cares about the low degree coefficients. Note that here $m > d$ and hence, it is not obvious that q_i when restricted to S should be a non-negative d-junta. Also, we cannot just make it a non-negative d-junta by keeping only low degree coefficients of some small subset of size d because that may not give a non-negative function. Note that the graph G has zero weight everywhere else expect for S and hence, we only care about the subset S.

We do a separate analysis for "smooth" and "non-smooth" q_i which we will formalize in a second:

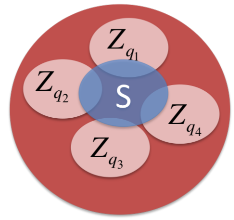
1. Non-smooth functions Let us define a function q_i as non-smooth if $\max_{x \in \{-1,1\}^n} q_i(x) > n^d$
   Now, we know that $c - J \in \{\lambda_0 + \sum_{i=1}^{R} \lambda_i q_i | \lambda_i \geq 0\}$, Let $c = L(J)$, now, $L(J) - J$ can be atmost 1, since both $J, L(J) \leq 1$ as they are fraction of edges cut. Therefore, we can see that the function $L(J) - J$ is a pretty smooth function and the idea is that non-smooth functions cannot have a large contribute to smooth functions and hence, they should not contribute much to the error even if we ignore these functions. Let us consider that q_i are normalized such that $\frac{1}{2^n} \sum_{x \in \{-1,1\}^n} q_i(x) = 1$. Therefore, for non smooth q_i, the corresponding $\lambda_i \leq n^{-r}$ since λs and $q_i s$ are non-negative. Therefore, for a SA pseudo distribution μ, $\widetilde{E}\lambda_i q_i = \sum_{x \in \{-1,1\}^n} \lambda_i \mu(x) q_i(x) = 2^{-n} \sum_{x \in \{-1,1\}^n} \sum_{S \subset [n], |S| \leq d} \lambda_i c_S X_S(x) q_i(x)$ since $c_S \leq 1$ as we saw above and $|X_S(x)| \leq 1$ and $\lambda_i \leq n^{-d}$, we get that $\widetilde{E}\lambda_i q_i \leq 2^{-n} \sum_{x \in \{-1,1\}^n} \sum_{S \subset [n], |S| \leq d} q_i(x) \leq n^{-d} m^d$ (Here, we used that μ has atmost m^d non-zero coefficients in the multi-linear representation since S is of size m and the graph has 0 weight everywhere else) and since there can be atmost $n^{d/2}$ such functions since the size of the LP is $n^{d/2}$, and since q_i are normalized, we get that $\sum_i \widetilde{E}\lambda_i q_i \leq n^{d/2} n^{-d} m^d$ and we see if $m \sim n^{1/10d}$, then this error term goes to 0 for large n. Hence, it is okay to ignore the non smooth. functions.

2. Smooth functions Let us define a function q_i as smooth if $\max_{x \in \{-1,1\}^n} q_i(x) \leq n^d$
   We will show that smooth functions are close to d-juntas on low degree coefficients which will be sufficient to complete the proof. The proof goes in two steps:

   Step 1: First of all we use Chang's lemma [6] to show that for our smooth functions q_i, there exists a set $Z \subseteq [n]$, with size $|J| \approx \frac{n}{m^2}$ such that for all subsets α with $\alpha \subsetneq Z$ and size $|\alpha| \leq d$, we have $|\hat{q}(\alpha)| \leq \frac{16 m^{1/4}}{n^{1/2}}$ where $\hat{q}(\alpha)$ is the coefficient of function q_i corresponding to set α when written as a multilinear polynomial. The proof goes via entropy arguments that if the function is sufficiently smooth, then it has high entropy, then it is close to a $\frac{n^{1/2}}{m} \approx n^{\frac{1}{2} - \frac{1}{10d}}$-junta on low degree coefficients.



(a) Step 1



(b) Step 2

Figure 1: Smooth functions proof

We do not present the proof for this here. It would be helpful to refer to figure 1 here which basically shows that there exists a large set Z such that all low degree coefficients for sets not contained in Z are small and hence, would not contribute much to the error.

Step 2: Now, when we restrict our q_i to the small subset S of size $m << n$, by concentration and union bounds, we can show that on restriction to S, only some of the influential co-ordinates from each Z_i get selected in S and hence, our function becomes close to non-negative d-junta on the subset S with high probability. More formally, if $X_1, X_2, ..., X_n$ are $\{0, 1\}$ iid random variables with $E[X_i] = p$, then

$$P(\sum_{i=1}^{n} X_i \geq t) \leq \sum_{S \in \binom{n}{t}} P(\sum_{i \in S} X_i = t) \leq \binom{n}{t} p^t \leq (pn)^t$$

Now, we select each co-ordinate independently into our set S of size m with probablity $\frac{2m}{n}$ and use the above bounds to get that $|S| \leq d$ with high probability. The figure 1 can be seen to get the idea of what is going on.

Now using everything that we proved above, we complete the proof.

We know from the representation of LPs for CSPs, $L(J) - J = \lambda_0(J) + \sum_{i=1}^{R} \lambda_i(J) q_i$ where $\lambda_i \geq 0$ and q_i are non-negative functions on the boolean hypercube.

Let $Q_t = \{q_i : q_i \text{ is smooth}\}$ (for smoothness defined above)

Let S be the subset of size m inside J and we plant the integrality gap instance J_0 of d rounds SA of size m on S. Let, μ be the pseudo distribution for integrality gap instance of size m and $\widetilde{E}$ be the corresponding expectation. We extend μ on size n graphs by putting 0 everywhere else and hence, this still remains a valid pseudo distribution on size n instances. Now, let us taken expectation of both sides wrt μ.

$$L(J) - \widetilde{E}J = \lambda_0(J) + \sum_{i=1}^{R} \lambda_i(J) \widetilde{E}(q_i)$$

Since, J has weight only on subset S, we get $\widetilde{E}J = SA(J_0)$

$$L(J) - SA(J_0) = \lambda_0(J) + \sum_{i=1}^{R} \lambda_i(J) \widetilde{E}(q_i)$$

$$L(J) - SA(J_0) = \sum_{\text{smooth } q_i} \lambda_i(J) \widetilde{E}(q_i) + \sum_{\text{non-smooth } q_i}^{R} \lambda_i(J) \widetilde{E}(q_i)$$

From the analysis of non-smooth functions above, we know that $\sum_{\text{non-smooth}} \lambda_i(J) \widetilde{E}(q_i) \leq n^{d/2} n^{-d} m^d$ Hence, this term's contribution (ϵ_n) goes to 0 for large n.

$$L(J) - SA(J_0) \geq \sum_{\text{smooth } q_i} \lambda_i(J)\widetilde{E}(q_i) - \epsilon_n$$

Now, we use our proof from earlier that smooth q_is are close to non-negative d-juntas with small error i.e. $\widetilde{E}(q_i) = \widetilde{E}(\widetilde{q_i}) + \widetilde{E}(e_i)$ where $\widetilde{q_i}$ is the non-negative d-junta and e_i is the error term. Note that here, we ignored the large degree terms of q_i because when taken expectation wrt $\widetilde{E}$, they will anyways become 0 since $\widetilde{E}$ is a degree d SA pseudo distribution with higher degree moments 0.

$$L(J) - SA(J_0) \geq \sum_{\text{smooth } q_i} \lambda_i(J)\widetilde{E}(\widetilde{q_i}) + \sum_{\text{smooth } q_i} \lambda_i(J)\widetilde{E}(e_i) - \epsilon_n$$

Now, $\widetilde{E}(\widetilde{q_i}) \geq 0$ since $\widetilde{q_i}$ is a non-negative d-junta.

$$L(J) - SA(J_0) \geq \sum_{\text{smooth } q_i} \lambda_i(J)\widetilde{E}(e_i) - \epsilon_n$$

And, the error part we saw above $|\widetilde{E}(e_i)| \leq m^d \frac{16m^{1/2}}{n^{1/4}}$ which also goes to 0 as n becomes large for small m.

$$L(J) - SA(J_0) \geq -\epsilon'_n$$

And hence, $L(J) \geq SA(J_0)$. Thus, the $n^{d/2}$ LP does atleast as good as d rounds of SA relaxation for constant d. □

3 Conclusions

The authors use integrality gaps construction of Sherali Adams relaxation to get lower bounds on general polynomial sized LPs. This is interesting since this is the first time this kind of method has been used to lower bounds on LPs. Also, it is interesting that the authors do not need to look at the actual integrality gaps construction and can use them in a black box fashion.

4 Future Work

The same idea was later extended to Semi definite programs by using integrality gaps constructions from SoS/Lasserre hierarchy by authors in [7]. Authors in [8] extend these connections to Vertex cover. The authors in this paper pose an open question. In this paper, their lower bounds for LPs only work till levels upto $n^{O(\frac{\log n}{\log \log n})}$. Although integrality gaps for Sherali Adams are known for $\Omega(n)$ rounds, the authors could not go beyond this because the error of the coefficients no longer goes to 0 if the size of LP increases beyond that. The authors pose an open question whether this bound can be improved.

References

[1] Arora, S., Bollobs, B., & Lovsz, L. (2002). Proving integrality gaps without knowing the linear program. In Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on (pp. 313-322). IEEE.

[2] Yannakakis, M. (1991). Expressing combinatorial optimization problems by linear programs. Journal of Computer and System Sciences, 43(3), 441-466.

[3] de la Vega, W. F., & Kenyon-Mathieu, C. (2007, January). Linear programming relaxations of maxcut. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (pp. 53-61). Society for Industrial and Applied Mathematics.

[4] Charikar, M., Makarychev, K., & Makarychev, Y. (2009, May). Integrality gaps for Sherali-Adams relaxations. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 283-292). ACM.

[5] Schoenebeck, G. (2008, October). Linear level Lasserre lower bounds for certain k-CSPs. In Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on (pp. 593-602). IEEE.

[6] Chang, M. C. (2002). A polynomial bound in Freiman's theorem. Duke mathematical journal, 113(3), 399-420.

[7] Lee, J. R., Raghavendra, P., & Steurer, D. (2015, June). Lower bounds on the size of semidefinite programming relaxations. In Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing (pp. 567-576). ACM.

[8] Braun, G., Pokutta, S., & Zink, D. (2015, June). Inapproximability of combinatorial problems via small LPs and SDPs. In Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing (pp. 107-116). ACM.

[9] Chan, S. O., Lee, J. R., Raghavendra, P., & Steurer, D. (2016). Approximate constraint satisfaction requires large LP relaxations. Journal of the ACM (JACM), 63(4), 34.