Approximating CSPs with Global Cardinality Constraints Using SDP Hierarchies

Sepehr Abbasi Zadeh

I. Introduction

This work reviews an approach for rounding constraint satisfaction problem (CSP) instances under cardinality constraints which is developed by Raghavendra and Tan [1].

In the CSP problems, we are given a set of variables over a fixed finite discrete domain and the goal is to find an assignment for these variables to satisfy a set of given *local* constraints. A constraint is local in the sense that its satisfaction only depends on a constant number of variables[1]. The goal is to satisfy the most possible number of constraints. It turns out that adding a single non-local (global) cardinality constraint on the variables makes the problem much harder. In this case, the goal is still the same as before, but we want to maintain the added global constraint at the same time.

The idea of this paper is to find a set of "*uncorrelated*" vectors using higher order sum of squares (SOS) hierarchy. While all the presented results in this note can be applied to general CSPs, we only focus on the particular MAX BISECTION instance.

II. Max Bisection Problem

The MAX BISECTION problem is a variant of the MAX-CUT problem. The goal in the MAX-CUT problem is to select a subset of nodes in a weighted graph in a way to maximize the weight of the edges crossing the set of selected and unselected nodes. The MAX BISECTION problem consists of a MAX-CUT instance with an additional cardinality constraint such that $|S| = |V|/2$ where V is the set of nodes. The best known approximation result for the MAX-CUT problem is a 0.87-approximation due to the celebrated work by Goemans and Williamson [2]. This work formulates the problem using a semidefinite programming and then rounds the solution with a very clever approach: random hyperplane rounding. In this rounding method, we consider a random hyperplane passing through the origin and round any vector which lies above this hyperplane to +1, and round all the rest to -1.

[1]We refer the reader to the last week's notes by Adil and MacAulay for the formal definition of the CSPs.

On the one hand, it is obvious that the $\mathbb{E}[S] = |V|/2$, however, we cannot calculate a tight bound for the variance of this partitioning. If we knew that this variance was not so high, then we could use this method for solving the MAX BISECTION instances. On the other hand, we should not expect to find any result better than what we could achieve for the MAX-CUT problem (unless the Unique Game Conjecture was true) as there is the following approximation-preserving reduction: consider any two disjoint copies of a single MAX-CUT instance. Then it is clear that if we could do better than 0.87 for the MAX BISECTION problem, then this approach could be used to find a better cut.

III. GLOBALLY UNCORRELATED SDP SOLUTIONS

A. Problem Statement

The MAX BISECTION problem formulation is as follows: Given a graph $G = (V, E)$ with a weight function $w : E \to \mathbb{R}_{\geq 0}$ find a subset $S \subset V$ which maximizes $\sum_{i \in S, j \in \bar{S}} w_{ij}$ such that $|S| = |V|/2$. If we assume that the weights are normalized such that their sum is equal to one ($\sum w_{ij} = 1$), then the optimal solution is at most one.

B. Semidefinite Programming Formulation

Firstly, we should introduce the degree k SOS relaxation of the MAX BISECTION problem. A solution to the degree K SOS relaxation consists of vectors $v_{S,\alpha}$ for all vertex set $S \subseteq V$ with $|S| \leq k$ and local assignment $\alpha \in [q]^S$. In the MAX BISECTION problem, and therefore in the following relaxation, $[q]$ is equal with $\{0, 1\}$.

$$\begin{aligned} &maximize \quad \sum_{1 \leq i,j \leq n} w_{ij}\big(\mathbb{P}(X_i = 0, X_j = 1) + \mathbb{P}(X_i = 1, X_j = 0)\big) \\ &s.t. \quad \langle v_{S,\alpha}, v_{T,\beta} \rangle = \mathbb{P}_{\mu_{\{S \cup T\}}}(X_S = \alpha, X_T = \beta) \quad \forall |S \cup T| \leq k, \alpha \in [q]^S, \beta \in [q]^T \\ &\qquad \mathbb{E}_j \mathbb{P}_{\mu_{S \cup \{x_j\}}}(x_j = i | X_S = \alpha) = 1/2 \quad \forall S \subset V : |S| \leq k - 1, \alpha \in [q]^S \end{aligned} \tag{1}$$

The last (set of) constraint(s) ensures that the nodes are equally probable to be assigned to either S or $\bar{S}$. It can be shown that the probabilities in the SDP relaxation can be converted to their vector forms. In the next lemma, we prove this argument for the objective of the SDP.

Lemma 1. *There exists unit vectors v_i and v_j such that $\mathbb{P}(X_i = 0, X_j = 1) + \mathbb{P}(X_i = 1, X_j = 0) = \frac{||v_i - v_j||_2^2}{4}$, which is the vector form of the SDP relaxation for the* MAX BISECTION *(MAX-CUT) problem.*

Proof. Let $v_{i,0}$ and $v_{i,1}$ be the corresponding vectors for the i'th variable in the mentioned SDP. For each $i \in V$, define v_i to be $v_{i,1} - v_{i,0}$. First, we prove that $||v_i||_2 = 1$. Using the first constraint of the SDP we know that

$$\langle v_{i,0}, v_{i,1} \rangle = 0, \langle v_{i,0}, v_{i,0} \rangle = \mathbb{P}(X_i = 0), \langle v_{i,1}, v_{i,1} \rangle = \mathbb{P}(X_i = 1)$$

thus,

$$\begin{aligned} ||v_i||_2^2 =& ||v_{i,1} - v_{i,0}||_2^2 = ||v_{i,1}||_2^2 + ||v_{i,0}||_2^2 + 2\langle v_{i,0}, v_{i,1} \rangle \\ =& \mathbb{P}(X_i = 1) + \mathbb{P}(X_i = 0) + 0 = 1 \end{aligned}$$

Now, we prove the rest of the argument of the lemma.

$$\begin{aligned} \frac{||v_i - v_j||_2^2}{4} =& \frac{(v_{i,1} - v_{i,0} + v_{j,0} - v_{j,1})^2}{4} \\ =& \frac{\mathbb{P}(X_i = 1) + \mathbb{P}(X_i = 0) + \mathbb{P}(X_j = 1) + \mathbb{P}(X_j = 0)}{4} \\ &+ \frac{2\mathbb{P}(X_i = 1, X_j = 0) - 2\mathbb{P}(X_i = 1, X_j = 1) + 2\mathbb{P}(X_i = 0, X_j = 1) - 2\mathbb{P}(X_i = 0, X_j = 0)}{4} \\ =& \frac{4\mathbb{P}(X_i = 1, X_j = 0) + 4\mathbb{P}(X_i = 0, X_j = 1)}{4} = \mathbb{P}(X_i = 1, X_j = 0) + \mathbb{P}(X_i = 0, X_j = 1) \end{aligned}$$

□

In order to proceed, we need to define some information theoretical notions.

Definition 1. *Let X be a random variable taking values over $[q]$. The entropy of X is defined as*

$$H(X) =: -\sum_{i \in [q]} \mathbb{P}(X = i) \log \mathbb{P}(X = i)$$

Definition 2. *Let X and Y be two jointly distributed variables taking values over $[q]$. The mutual information of X and Y is defined as*

$$I(X;Y) =: \sum_{i,j \in [q]} \mathbb{P}(X = i, Y = j) \log \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(X = i)\mathbb{P}(Y = j)}$$

Definition 3. *Let X and Y be two jointly distributed variables taking values over $[q]$. The conditional entropy of X conditioned on Y is defined as*

$$\begin{aligned} H(X|Y) =:& \mathbb{E}_{i \in [q]}[H(X|Y = i)] \\ =& H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned}$$

Now we can formally define the define the notion of independence between the solution vectors.

Definition 4. *Given a solution to the degree k SOS relaxation, it is said to be α-independent if $\mathbb{E}[I_{\mu_{\{i,j\}}}(X_i; X_j)] \leq \alpha$ where $\mu_{\{i,j\}}$ is the local distribution associated with the pair of vertices $\{i, j\}$.*

C. Uncorrelated SDP Solutions

In this section, we describe an algorithm which finds an α-independent solution for the degree ℓ SOS hierarchy. The following lemma shows that there exists t such that the resulting solution is α-independent after t-conditionings with high probability.

Algorithm 1: Conditioning Algorithm

Input: A feasible solution to the degree $k+\ell$ SOS relaxation for $k = 1/\sqrt{\alpha}$.

Output: An α-independent solution to the degree ℓ SOS relaxation.

Sample indices $i_1, i_2, \ldots, i_k \subseteq V$ uniformly at random.;

$t \leftarrow 1$;

while *SDP solution is not α-independent* **do**

  Sample the variable X_{i_t} from its marginal distribution after the first $t-1$ fixings, and condition the SDP solution on the outcome.;

  $t \leftarrow t+1$;

end

Lemma 2. *There exists $t \leq k$ such that $\mathbb{E}_{i_1,i_2,\ldots,i_t,i,j}[I(X_i, X_j | X_{i_1}, \ldots, X_{i_t})] \leq \frac{\log q}{k-1}$.*

Proof. By linearity of expectation, we have that for any $t \leq k-2$

$$\mathbb{E}_{i,i_1,\ldots,i_t}[H(X_i|X_{i_1},\ldots,X_{i_t})] = \mathbb{E}_{i,i_1,\ldots,i_t}[H(X_i|X_{i_1},\ldots,X_{i_{t-1}})] - \mathbb{E}_{i_1,\ldots,i_t,i}[I(X_i, X_{i_t}|X_{i_1},\ldots,X_{i_{t-1}})]$$

adding the equalities from $t=1$ to $t=k-2$, we get:

$$\mathbb{E}_i[H(X_i)] - \mathbb{E}_{i_1,\ldots,i_{k-2}}[H(X_i|X_{i_1},\ldots,X_{i_{k-2}}] = \sum_{t=1}^{k-1} \mathbb{E}_{i_1,\ldots,i_{t-1},i,j}[I(X_i, X_j|X_{i_1},\ldots,X_{i_{t-1}})]$$

Now by using the fact that for each i, $H(X_i) \leq \log q$ (in our case that $|q| = 2$, $\log q = 1$), we can get the lemma. $\square$

Theorem 1. *For every $\alpha > 0$ and positive integer ℓ, there exists an algorithm running in time $O(n^{poly(1/\alpha)+\ell})$ that finds an α-independent solution to the degree ℓ SOS, with an SDP objective value of at least $OPT-\alpha$, where OPT denotes the optimum value of the degree ℓ SOS relaxation.*

Proof. Pick $k = \frac{4\log q}{\alpha^2}$. Solve the degree $k+\ell$ SOS relaxation and use it as the input to the conditioning algorithm described earlier (Algorithm 1). Notice that the algorithm respects the marginal distributions provided by the SDP while sampling the values to variables. Therefore, the expected objective value of the SDP solution after conditioning is exactly equal to the SDP objective value before conditioning. Also notice that the SDP value is at most 1. Therefore, by using the Markov inequality, the probability of the SDP value dropping by at least α due to conditioning is at most $1/(1+\alpha)$. Also by lemma 2, the probability of the algorithm failing to find a $\sqrt{\frac{\log q}{k}}$-independent solution is at most $\sqrt{\frac{\log q}{k}}$. Hence, by union bound, there exists a fixing such that the SDP value is maintained up to α, and the solution after conditioning is α-independent. Moreover, this particular fixing can be found using brute-force search. □

If we show that the variance of the partitioning is not so high, then we can use any rounding algorithm which partitions the nodes into equal parts in expectation. The following theorem completes what we need in order to round any α-independent solution of the degree ℓ SOS relaxation by showing the dependency between the variance and error term α.

Theorem 2. *Given an α-independent solution to degree 2 SOS SDP hierarchy. Let $\{y_i\}$ be the rounded solution after applying Algorithm 1. Define $S = \mathbb{E}_i[y_i]$, then*

$$Var(S) \leq O(\alpha^{1/12}).$$

We refer the reader to the proof of theorem 5.6 in the original paper.

References

[1] Prasad Raghavendra and Ning Tan. Approximating csps with global cardinality constraints using sdp hierarchies. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 373–387. Society for Industrial and Applied Mathematics, 2012.

[2] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.