

# Approximation Schemes via Sherali-Adams Hierarchy for Dense Constraint Satisfaction Problems and Assignment Problems

Yuichi Yoshida · Yuan Zhou

the date of receipt and acceptance should be inserted later

**Abstract** We consider approximation schemes for the maximum constraint satisfaction problems and the maximum assignment problems. Though they are NP-Hard in general, if the instance is “dense” or “locally dense”, then they are known to have approximation schemes that run in polynomial time or quasi-polynomial time. In this paper, we give a unified method of showing these approximation schemes based on the Sherali-Adams linear programming relaxation hierarchy. We also use our linear programming-based framework to show new algorithmic results on the optimization version of the hypergraph isomorphism problem.

**Keywords** Sherali-Adams hierarchy, approximation schemes, constraint satisfaction problems, assignment problems, dense instances, locally-dense instances

**Mathematics Subject Classification (2000)** 68Q25,68W20,68W25

## 1 Introduction

In the maximum constraint satisfaction problem (Max-CSP), given a variable set  $V$  over the domain  $D$  and a set of constraints  $\mathcal{C}$  over the variables in  $V$ , we want to find an assignment  $\alpha : V \rightarrow D$  that maximizes the fraction of constraints satisfied by  $\alpha$ . Max-CSP includes many fundamental problems such as Max-Cut and Max-SAT.

In general, Max-CSP is NP-Hard, and it is even NP-Hard to approximate within a constant factor [4]. However, de la Vega [19] showed that there is a polynomial-time approximation scheme for Max-Cut if the input graph is *dense*, i.e., it has  $\Omega(n^2)$

---

Preliminary version appeared in Proceedings of the 5th Innovations in Theoretical Computer Science (ITCS 2014).

---

Y. Yoshida  
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan  
Preferred Infrastructure, Inc. Hongo 2-40-1, Bunkyo-ku, Tokyo 113-0033, Japan  
E-mail: yyoshida@nii.ac.jp

Y. Zhou  
Carnegie Mellon University  
E-mail: yuanzhou@cs.cmu.edu

edges. Here, a *polynomial-time approximation scheme* (PTAS) is an algorithm that, given  $\epsilon > 0$  as a parameter, gives a  $(1 - \epsilon)$ -approximation to the optimal value, and runs in polynomial time for any constant  $\epsilon$ . Max- $k$ -CSP is a subproblem of Max-CSP, in which each constraint involves at most  $k$  variables, where  $k$  is a constant. Arora et al. [3] and Frieze and Kannan [9] showed PTASs for dense Max- $k$ -CSP, i.e., the input instance has  $\Omega(n^k)$  constraints. Now it is known that we can compute  $(1 - \epsilon)$ -approximation to the optimal value in time that depends only on  $k$  and  $\epsilon$  [1].

There are two directions to generalize PTASs for dense Max- $k$ -CSP. The first one is to generalize the notion of the density condition. We say that an instance of Max-2-CSP is *metric* if the weights of the constraints form a metric. Max-Cut [21] and Max-Bisection [18] admit PTASs if the instance is metric. The notion of *local density* is introduced to generalize the notion of metric to constraints over more than two variables. If the instance is locally dense, Max- $k$ -CSP admits PTASs [20].

The second direction is to handle the maximum assignment problems (Max-AP). In this problem, given a variable set  $V$  and a set of constraints, we want to find a permutation  $\pi$  of  $V$  to maximize the fraction of satisfied constraints. Max-AP includes many fundamental problems such as Maximum Acyclic Subgraph, Betweenness, Maximum Graph Isomorphism, Densest  $k$ -Subgraph, and Quadratic Assignment Problem. Max- $k$ -AP is a special case of Max-AP, in which each constraint involves at most  $k$  variables (see Section 2 for the precise definition). We say that an instance of Max- $k$ -AP is dense if it has  $\Omega(n^k)$  constraints. Arora et al. [2] showed a quasi-polynomial-time approximation scheme for dense Max- $k$ -AP and PTASs for many special cases.

As we have seen, Max-CSP and Max- $k$ -AP admit PTASs (or quasi-PTASs) in the dense case and the locally dense case. However, the techniques to obtain them vary a lot. For example, [3] is based on the idea of exhaustively trying all assignments for a small number of variables and then solving the rest using the partial assignment. On the other hand, [9] used a variant of Szemerédi’s regularity lemma [17]. To deal with the metric case, [21] used the method of copying important variables, and [20] considered a variant of singular value decomposition of tensors to deal with the locally dense case.

### 1.1 Linear Programming (LP) relaxation and LP relaxation hierarchies

LP relaxations are a standard tool to design approximation algorithms. In this well-known algorithmic framework, we typically model the given problem as an integer program, solve an LP relaxation, and use the optimal fractional solution to construct a feasible integral solution for the original problem (which is called a *rounding procedure*). We usually analyze the algorithm by comparing the value of the rounded solution to the value of the optimal fractional solution (which is an upper bound on the value of the optimal integral solution). In this way, the best approximation guarantee we can hope for, is the maximum gap (over all instance) between the values of optimal integral solution and the optimal fractional solution, which is usually referred to as the *integrality gap* of the relaxation.

In order to make the integrality gap smaller, we can strengthen the LP relaxation by additional constraints. People have proposed systematic ways to add additional constraints, such as Lovász-Schrijver LP relaxation hierarchy [11] and

Sherali-Adams LP relaxation hierarchy [16]. These LP relaxation hierarchies contain a sequence of LP relaxations, where each LP relaxation is obtained by strengthening the previous one in the sequence. The  $\ell$ -th LP relaxation in the sequence is usually called the  $\ell$ -round relaxation, and typically can be solved in  $n^{O(\ell)}$  time (where  $n$  is the number of variables in the original integer programming). The relaxation usually becomes as strong as the integer programming when  $\ell$  becomes  $n$ ; but it also takes exponential time to solve it.

It is known that  $o(n)$ -round Lovász-Schrijver LP relaxation and  $n^{o(1)}$ -round Sherali-Adams LP relaxation do not help to reduce the integrality gap for Max-Cut [15, 7]. For some other CSPs, the integrality gap remains even after strengthening the linear-round LP relaxation hierarchies with the power of semidefinite constraints (a.k.a. the Lasserre hierarchy) [14]. On the other hand, it is known that LP hierarchies *do* help for some CSPs when the instance is dense. [22] showed that the integrality gap of the  $O_\epsilon(1)$ -round Sherali-Adams LP relaxation drops to an arbitrarily small constant  $\epsilon$  for dense Max-Cut instances.

LP relaxation and its hierarchies have found many connections to other known algorithmic frameworks, and to be a unified approach to solve several classes of problems. A few examples are listed as follows. Assuming the Unique Games Conjecture, a canonical LP relaxation (also referred to as the *Basic LP*) is shown to provide optimal approximation guarantee for CSPs with strict constraints [10]. It is known to the authors that constant-round Sherali-Adams LP relaxation decides the satisfiability of bounded-width CSPs; Atserias and Maneva [5] recently showed that the Sherali-Adams LP relaxation hierarchy for graph isomorphism interleaves with the levels of pebble-game equivalence with counting (i.e. higher-dimensional Weisfeiler-Lehman color refinement algorithm).

## 1.2 Our results

In this paper, we present the Sherali-Adams LP relaxation hierarchy as a unified approach to dense and locally dense problems – we show that a small number of rounds of the Sherali-Adams LP relaxation gives an approximation scheme to dense Max- $k$ -CSP and all their variants studied in the previous works.

Our first main theorem deals with dense and locally dense Max- $k$ -CSP.

**Theorem 1 (Informal version of Theorem 5)** *For any  $\epsilon > 0$ ,  $O(\frac{1}{\epsilon^2})$ -round Sherali-Adams LP relaxation gives  $(1 - \epsilon)$ -approximation to dense or locally dense Max- $k$ -CSP.*

Then, we turn to dense and locally dense Max- $k$ -CSP with global cardinality constraints. For explanatory purposes, we only consider *bisection* constraint, i.e., the domain is  $\{0, 1\}$  and the number of variables that are assigned to 0 should be equal to the number of variables that are assigned to 1. We show that

**Theorem 2 (Informal version of Theorem 8)** *For any  $\epsilon > 0$ ,  $O(\frac{1}{\epsilon^2})$ -round Sherali-Adams LP relaxation gives  $(1 - \epsilon)$ -approximation to dense or locally dense bisection Max- $k$ -CSP.*

Finally, we consider the dense Max- $k$ -AP problems, and show that

**Theorem 3 (Informal version of Theorem 6)** *For any  $\epsilon > 0$ ,  $O(\frac{\log n}{\epsilon^2})$ -round Sherali-Adams LP relaxation gives  $(1 - \epsilon)$ -approximation to dense or locally dense Max- $k$ -AP problems with  $n$  variables.*

In all the precise theorem statements, we actually show additive approximation guarantee (i.e. the value of the rounded solution being at least the fractional optimal value minus a constant error) instead of multiplicative approximation guarantee. However, since we define the problems in a way that the optimal solution is  $\Omega(1)$  (see Section 2 for the precise definition of the problems), an additive approximation scheme implies a multiplicative approximation scheme.

### 1.2.1 New algorithmic guarantees

Let us define the problem Maximum  $k$ -Hypergraph Isomorphism as follows. Given two weighted  $k$ -uniform hypergraph  $G = (V, \omega')$  and  $H = (V, \omega'')$ , where  $\omega', \omega'' : V^k \rightarrow [0, 1]$  are the weight functions over all possible hyperedges. The goal is to find a permutation  $\pi$  over  $V$  so that  $\sum_{e \in V^k} \omega'(e) \omega''(\pi(e))$  is maximized (where  $\pi(e)$  is the edge obtained by applying  $\pi$  on each incident vertex of  $e$ ). It is easy to see Theorem 3 implies that  $O(\frac{\log n}{\epsilon^2})$ -round Sherali-Adams LP relaxation gives  $(1 - \epsilon)$ -approximation to Maximum  $k$ -Hypergraph Isomorphism when both  $G$  and  $H$  are dense.

We are able to apply our analysis framework for the Sherali-Adams LP relaxation to another special case of Maximum  $k$ -Hypergraph Isomorphism, getting the following new algorithmic guarantee.

**Theorem 4 (Informal version of Theorem 7)** *For any  $\epsilon > 0$ ,  $O(\frac{\log n}{\epsilon^2})$ -round Sherali-Adams LP relaxation gives  $(1 - \epsilon)$  approximation to the Maximum  $k$ -Hypergraph Isomorphism problem when one of the two graphs is locally dense and the other graph is dense, where  $n$  is the number of vertices in the hypergraphs. Therefore, this special case of the problem admits a  $(1 - \epsilon)$ -approximation algorithm in time  $n^{O(\frac{\log n}{\epsilon^2})}$ .*

## 1.3 Proof overview

The first step of our algorithms is to condition on a set of random variables in a solution to the Sherali-Adams LP relaxation. In the  $\ell$ -round Sherali-Adams LP relaxation (or the SA relaxation for short), for each set of variables  $S$  of size at most  $\ell$ , we have a probability distribution  $\mu_S$  over assignments on  $S$ . First we solve  $(k + \ell)$ -round SA relaxation, where  $\ell$  is a parameter depending on the error parameter  $\epsilon$ . Then, we randomly sample a set of variables  $u_1, \dots, u_\ell$  and assign values to them by sampling values from  $\mu_{\{u_1\}}, \dots, \mu_{\{u_\ell\}}$ , respectively. By this conditioning, we obtain a solution to  $k$ -round Sherali-Adams relaxation  $\mu'$  with the same LP value in expectation. An important fact here is that variables become almost independent in the sense that, if we sample a  $k$ -tuple  $(v_1, \dots, v_k)$  according to a dense (or locally dense) distribution (this distribution corresponds to the weights of the constraints in  $k$ -CSP and  $k$ -AP instances), the distribution  $\mu_{\{v_1, \dots, v_k\}}$  and the product distribution  $\mu_{\{v_1\}} \times \dots \times \mu_{\{v_k\}}$  are close in expectation.

The second step of our algorithms is to round the solution to the SA relaxation where the variables are almost independent. For dense (or locally dense)  $k$ -CSP and bisection  $k$ -CSP, the rounding algorithm just samples a value from  $\mu_{\{v\}}$  and assigning it to  $v$  for each variable  $v$ . It is relatively easy to show that the expected value of the sampled solution is close to the LP value, and therefore gives a  $(1 - \epsilon)$ -approximation.

For  $k$ -AP problems, however, such independent sampling method does not work – there might be more than one variables assigned to the same value and we do not get a permutation when this happens. Instead, we view the marginal probability distributions on single variables,  $\mu_{\{u\}}(w)$ , as a doubly stochastic matrix. We view this doubly stochastic matrix as a probability distribution of permutations. We iteratively choose two permutations in the support of the distribution and merge them into a new permutation, until there is only one permutation left in the support – which is the output of our rounding algorithm. The operation of merging two permutations is interestingly similar to the merging operation used in [2], although the purposes are different. See Section 4.2 for more details.

#### 1.4 Comparison to previous works

We first compare the running time of our SA relaxation-based algorithms with the previously known counterparts. For Max- $k$ -CSP, the running time  $n^{O(1/\epsilon^2)}$  of our method matches the one of the method by [3]. For Max- $k$ -AP the running time  $n^{O(\log n/\epsilon^2)}$  of our method matches the one of the method by [2]. [2] improved the running time to  $n^{O(1/\epsilon^2)}$  for various problems by reducing them to CSPs. We can use the same techniques to obtain the same running time for these problems.

The number of rounds ( $O(\frac{1}{\epsilon^2})$ ) in Theorem 1 improves the corresponding theorem in [22] which showed that  $O(\frac{1}{\epsilon^4})$ -round SA relaxation gives  $(1-\epsilon)$ -approximation to dense Max-Cut.

The idea of conditioning variables of a solution to LP/SDP hierarchies is used in [13, 6] to solve variants of Max-2-CSP. Let  $G = (V, E)$  be the underlying graph of an instance of Max-2-CSP. Barak et al. [6] showed that (i) the covariance between  $u$  and  $v$  over  $V^2$  gets close to zero by conditioning, and (ii) the covariance between  $u$  and  $v$  over  $E$  gets close to the covariance between  $u$  and  $v$  over  $V^2$  by conditioning if  $G$  is expander-like. Combining these two results, they show a PTAS for Max-2-CSP when  $G$  is expander-like. This method can be also applied to dense graphs, but it is not clear how to generalize it to metric graphs and  $k$ -CSP.

Raghavendra and Tan [13] used mutual information instead of covariance to measure correlation between two variables and simplified the proof. They noticed that conditioning is useful to deal with global constraints such as cardinality constraints since after conditioning we can sample variables independently and the resulting solution will not break global constraints much. With this idea, they gave a 0.85-approximation algorithm for Max-Bisection. Though our method and analysis are similar to theirs, we use the independence for obtaining PTASs for the dense and locally dense case as well as supporting global constraints. Also, to handle constraints of larger arities, we use total correlation instead of mutual information to measure correlation among variables.

Coja-Oghlan et al. [8] showed that, even if the instance is sparse, if it satisfies a certain pseudo-random condition, then Max- $k$ -CSP admits PTASs. If  $k = 2$ , this results can be seen as a special case of [13] because the pseudo-random condition would imply that the underlying graph is expander-like. Their result is incomparable to ours because it is not clear how the pseudo-random condition and the locally dense condition.

## 1.5 Organization

In Section 2, we introduce definitions and notions used in this paper. In Section 3, we show an algorithm that obtains an almost independent solution to the Sherali-Adams LP relaxation. Section 4 is devoted to show how to round the obtained solution to the Sherali-Adams LP relaxation. We combine the two steps together in Section 5. Sections 6 and 7 are devoted to prove auxiliary lemmas. We consider CSPs with global cardinality constraints in Section 8.

## 2 Preliminaries

For an integer  $a \geq 1$ ,  $[a]$  denotes the set  $\{1, \dots, a\}$ . For a set  $Y$  and  $0 \leq k \leq |Y|$ ,  $\binom{Y}{k}$  denotes the family of sets  $X \subseteq Y$  with  $|X| = k$ . We usually use  $V$  to denote the set of variables in a problem, and use  $n = |V|$  to denote the number of variables. For an event  $A$ ,  $\mathbf{1}[A]$  denotes the corresponding indicator function.

*Probability theoretic notions:* We recall several notions from probability theory. For a probability distribution  $\mu$  on  $\Omega$ ,  $\text{supp}(\mu)$  denotes the *support* of  $\mu$ , i.e.,  $\text{supp}(\mu) = \{i \in \Omega \mid \mu(i) > 0\}$ . For a set  $S$ ,  $i \sim S$  means that we sample  $i$  uniformly at random from  $S$ .

Let  $\mu_1$  and  $\mu_2$  be two probability distributions on a finite set  $\Omega$ . Then, the  $L_1$  distance between them is defined as  $\|\mu_1 - \mu_2\|_1 = \sum_{i \in \Omega} \|\mu_1(i) - \mu_2(i)\|$ . The Kullback-Leibler divergence between them is defined as  $d_{KL}(\mu_1 \parallel \mu_2) = \sum_{i \in \Omega} \mu_1(i) \log \frac{\mu_1(i)}{\mu_2(i)}$ , and the Kullback-Leibler divergence  $d_{KL}(\mu_1 \parallel \mu_2)$  are defined as follows. We provide the following fact without proof.

**Lemma 1** *Let  $\mu_1$  and  $\mu_2$  be two probability distributions on a finite set  $\Omega$ . Then,  $\|\mu_1 - \mu_2\|_1 \leq \sqrt{2d_{KL}(\mu_1 \parallel \mu_2)}$ .*

*Information theoretic notions:* We now recall some definitions from information theory. For a random variable  $\mathbf{x}$ ,  $\mu_{\mathbf{x}}$  denotes the corresponding probability distribution. That is, for any  $i$ , we have  $\mu_{\mathbf{x}}(i) = \Pr[\mathbf{x} = i]$ .

Let  $\mathbf{x}$  be a random variable on a finite set  $\Omega$ . The *entropy* of  $\mathbf{x}$  is defined as  $H(\mathbf{x}) = -\sum_{i \in \Omega} \Pr[\mathbf{x} = i] \log \Pr[\mathbf{x} = i]$ .

Let  $\mathbf{x}$  and  $\mathbf{y}$  be jointly distributed variables on a finite set  $\Omega$ . The *entropy* of  $\mathbf{x}$  conditioned on  $\mathbf{y}$  is defined as  $H(\mathbf{x} \mid \mathbf{y}) = \mathbf{E}_{i \sim \mu_{\mathbf{y}}} [H(\mathbf{x} \mid \mathbf{y} = i)]$ . The *mutual information* of  $\mathbf{x}$  and  $\mathbf{y}$  is defined as  $I(\mathbf{x}; \mathbf{y}) = d_{KL}(\mu_{(\mathbf{x}, \mathbf{y})} \parallel \mu_{\mathbf{x}} \times \mu_{\mathbf{y}})$ .

Let  $\mathbf{x}_1, \dots, \mathbf{x}_k$  ( $k \geq 2$ ) be jointly distributed variables on a finite set  $\Omega$ . The *mutual information* of  $\mathbf{x}_1, \dots, \mathbf{x}_k$  is defined as  $I(\mathbf{x}_1; \dots; \mathbf{x}_k) = I(\mathbf{x}_1; \dots; \mathbf{x}_{k-1}) - I(\mathbf{x}_1; \dots; \mathbf{x}_{k-1} \mid \mathbf{x}_k)$ , where  $I(\mathbf{x}_1; \dots; \mathbf{x}_{k-1} \mid \mathbf{x}_k) = \mathbf{E}_{i \sim \mu_{\mathbf{x}_k}} [I(\mathbf{x}_1; \dots; \mathbf{x}_{k-1} \mid \mathbf{x}_k = i)]$ . The *total correlation* of  $\mathbf{x}_1, \dots, \mathbf{x}_k$  is defined as  $C(\mathbf{x}_1, \dots, \mathbf{x}_k) = d_{KL}(\mu_{(\mathbf{x}_1, \dots, \mathbf{x}_k)} \parallel \mu_{\mathbf{x}_1} \times \dots \times \mu_{\mathbf{x}_k})$ .

We give two well-known facts in information theory below.

**Lemma 2** *Let  $\mathbf{x}$  and  $\mathbf{y}$  be two jointly distributed variables on a finite set  $\Omega$ . Then*

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x} \mid \mathbf{y}).$$

Let  $\mathbf{x}_1, \dots, \mathbf{x}_k$  be jointly distributed variables on a finite set  $\Omega$ . Then

$$I(\mathbf{x}_1; \dots; \mathbf{x}_k) = \sum_{(i_1, \dots, i_t) \subseteq [k], t \geq 1} (-1)^{t-1} H(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_t}).$$

**Lemma 3** Let  $\mathbf{x}_1, \dots, \mathbf{x}_k$  be jointly distributed variables on a finite set  $\Omega$ . Then

$$C(\mathbf{x}_1, \dots, \mathbf{x}_k) = \sum_{(i_1, \dots, i_t) \subseteq [k], t \geq 2} I(\mathbf{x}_{i_1}; \dots; \mathbf{x}_{i_t}).$$

*Constraint satisfaction problems:* Let  $D$  be a nonempty finite domain and  $k \geq 2$  be an integer. An instance  $\mathcal{I} = (V, \omega, P)$  of  $k$ -CSP consists of a set  $V$  of variables, a scope distribution  $\omega$  over  $V^k$ , and a set of payoff functions  $P = \{P_S : D^S \rightarrow [0, 1] \mid S \in V^k\}$ . An assignment for an instance  $\mathcal{I} = (V, \omega, P)$  is a mapping  $\alpha : V \rightarrow D$ . The value of the assignment, denoted  $\mathbf{val}(\mathcal{I}, \alpha) \in [0, 1]$ , is defined as  $\mathbf{val}(\mathcal{I}, \alpha) = \Pr_{S \sim \omega} [P_S(\alpha|_S)]$ , where  $\alpha|_S$  is the projection of  $\alpha$  to  $S$ . We define the optimum value of the instance  $\mathcal{I}$  to be  $\mathbf{opt}(\mathcal{I}) = \max_{\alpha} \{\mathbf{val}(\mathcal{I}, \alpha)\}$ .

Let  $\mathcal{I} = (V, \omega, P)$  be an instance of CSP. A solution to the  $\ell$ -round Sherali-Adams relaxation consists of a probability distribution  $\mu_S$  over  $D^S$  for each set  $S \subseteq V$  of size at most  $\ell$ . The objective function is the probability that  $\alpha$  is in  $P_S$ , where  $S$  is sampled from  $\omega$  and  $\alpha$  is sampled from  $\mu_S$ . Strictly speaking, we sample a tuple  $(v_1, \dots, v_k)$  from  $\omega$ , but we regard it as the set  $\{v_1, \dots, v_k\}$  when we use it as a subscript of  $\mu$ . In other words,  $\mu_S$  and  $\mu_T$  are the same distribution for two tuples  $S$  and  $T$  if they are the same as sets. Also, for every pair of sets  $S$  and  $T$  with  $|S \cup T| \leq \ell$ , the corresponding probability distributions  $\mu_S$  and  $\mu_T$  must be consistent on  $S \cap T$ . Formally, the  $\ell$ -round Sherali-Adams relaxation for a  $k$ -CSP instance  $\mathcal{I} = (V, \omega, P)$  ( $\ell \geq k$ ) is written as follows.

$$\begin{aligned} & \text{maximize } \mathbf{E}_{S \sim \omega} \mathbf{E}_{\alpha \sim \mu_S} [P_S(\alpha)] \\ & \text{subject to } \Pr_{\alpha \sim \mu_S} [\alpha|_{S \cap T} = \beta] = \Pr_{\alpha \sim \mu_T} [\alpha|_{S \cap T} = \beta] \quad \forall S, T \subseteq V, |S \cup T| \leq \ell, \beta \in D^{S \cap T}. \end{aligned}$$

It is not hard to see that the relaxation above can be written as a linear programming (see, e.g., [12] for details). We define  $\mathbf{x}_v$  as the random variable sampled from the distribution  $\mu_{\{v\}}$ . We use  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \mu)$  to denote the objective value of the LP solution  $\mu$ . The same definition applies to the following subsections.

*Assignment problems:* The assignment problem differs from CSP in that we want to maximize the objective function over the set of permutations. Similarly to CSP, for an integer  $k \geq 2$ , an instance of the degree- $k$  assignment problem is given as  $\mathcal{I} = (V, \omega)$ , where  $V$  is the set of variables,  $\omega$  is a distribution over  $V^k \times V^k$ . The scope distribution of  $\mathcal{I}$  is the marginal distribution of  $\omega$  on the first  $k$  elements. An assignment for an instance  $\mathcal{I} = (V, \omega)$  is a permutation  $\pi$  of  $V$ . The value of the assignment  $\pi$ , denoted  $\mathbf{val}(\mathcal{I}, \pi)$ , is defined as

$$\mathbf{val}(\mathcal{I}, \pi) = n^k \Pr_{(U, W) \sim \omega} [\forall i \in [k] : \pi(u_i) = w_i],$$

where  $U = (u_1, u_2, \dots, u_k)$  and  $W = (w_1, w_2, \dots, w_k)$ . We define the optimum value of  $\mathcal{I}$  to be  $\mathbf{opt}(\mathcal{I}) = \max_{\pi} \{\mathbf{val}(\mathcal{I}, \pi)\}$ .

Though the definition of  $\text{val}(\mathcal{I}, \pi)$  may look non-standard, it is just the objective function used in [2] with a normalization factor that is multiplied to make the optimum  $\Omega(1)$  when  $\omega$  is dense.

The  $\ell$ -round Sherali-Adams relaxation of an  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  ( $\ell \geq k$ ) is as follows.

$$\begin{aligned} & \text{maximize} && \mathbf{E}_{(U, W) \sim \omega} \Pr_{\alpha \sim \mu_U} [\forall i \in [k] : \alpha(u_i) = w_i] \\ & \text{subject to} && \Pr_{\beta \sim \mu_S} [\beta|_{S \cap T} = \alpha] = \Pr_{\beta \sim \mu_T} [\beta|_{S \cap T} = \alpha] \quad \forall S, T \subseteq V, |S \cup T| \leq \ell, \alpha \in V^{S \cap T} \\ & && \sum_{\alpha \in V^S} \sum_{u \in V \setminus S} \mu_{S \cup \{u\}}(\alpha \cup \{u \rightarrow w\}) = \sum_{\alpha \in V^S} \mu_S(\alpha) \quad \forall w \in V, S \subseteq V, |S| < \ell. \end{aligned}$$

The difference from the Sherali-Adams relaxation for CSP is that we have extra constraints in the last line whose intended meaning is that each value  $i$  can be taken by at most one variable.

*Density condition:* We now introduce the notion of dense and locally dense distributions.

Let  $\omega$  be a probability distribution over a finite set  $\Omega$ . For  $\Delta \in (0, 1]$ , we say  $\omega$  is  $\Delta$ -dense if for every  $a \in \Omega$ , it holds that  $\Delta \cdot \omega(a) \leq \frac{1}{|\Omega|}$ .

Let  $\omega$  be a probability distribution over  $V^k$ . Let  $d_i(v) = \Pr_{S \sim \omega} [S_i = v_i]$  be the probability that the  $i$ -th coordinate is  $v$  under  $\omega$ . For  $\Delta \in (0, 1]$ , we say  $\omega$  is  $\Delta$ -locally dense if for every  $(v_1, \dots, v_k) \in V^k$ , it holds that

$$\Delta \cdot \omega(v_1, \dots, v_k) \leq \frac{1}{|V|^{k-1}} \sum_{1 \leq i \leq k} d_i(v_i).$$

Since  $d_i(v) = \sum_{S \in V^k: S_i=v} \omega(S)$ , the RHS of the locally dense condition is equal to  $\sum_{1 \leq i \leq k} \mathbf{E}_{S \sim V^k} [\omega(S) | S_i = v_i]$ . Thus the locally dense condition says that no tuple  $(v_1, \dots, v_k)$  is “wild” in that  $\omega(v_1, \dots, v_k)$  is at most constant times the sum over  $i$  of the average probability mass of  $S$  with  $S_i = v_i$ .

The notion of local density is introduced in [20] to generalize the metric condition. To see this, suppose  $\omega : V^2 \rightarrow \mathbb{R}$  is a metric. Then,  $\omega$  is 1-locally dense since, for any  $u, v \in V$ , we have  $\frac{1}{n}(d_1(u) + d_2(v)) = \frac{1}{n} \sum_w (\omega(u, w) + \omega(w, v)) \geq \frac{1}{n} \sum_w \omega(u, v) \geq \omega(u, v)$ .

It is immediate to verify the following lemma.

**Lemma 4** *Let  $\omega$  be a probability distribution over  $\Omega_1 \times \Omega_2$ . If  $\omega$  is  $\Delta$ -dense (resp.,  $\Delta$ -locally dense), then the marginal distribution  $\omega_1$  of  $\omega$  on  $\Omega_1$  is also  $\Delta$ -dense (resp.,  $\Delta$ -locally dense).*

### 3 Conditioning operations for Sherali-Adams LP hierarchy

Recall that, a solution to the  $\ell$ -round SA relaxation consists of distributions over sets of  $\ell$  variables. In this section, we show that, if the scope distribution is dense or locally dense, then by conditioning a small number of variables, we can make variables almost independent in these distributions. Once variables become almost



independent, we can round variables independently without losing the objective value much (see Section 4).

Let  $\mathcal{I}$  be an instance of  $k$ -CSP or  $k$ -AP with a variable set  $V$ . Fix  $\ell$  and let  $\mu$  be a solution to the  $\ell$ -round Sherali-Adams relaxation. For a variable set  $S = (v_1, \dots, v_k)$ ,  $C_\mu(\mathbf{x}_S)$  denotes the total correlation  $C(\mathbf{x}_{v_1}, \dots, \mathbf{x}_{v_k})$  under the probability distribution  $\mu_S$ . We use the following notion to measure independence of variables.

**Definition 1** Let  $\mathcal{I}$  be an instance of  $k$ -CSP or  $k$ -AP with a scope distribution  $\omega$ . A solution  $\mu$  to the  $\ell$ -round SA relaxation for  $\mathcal{I}$  with  $\ell \geq k$  is  $\kappa$ -independent with respect to distribution  $\omega'$  if

$$\mathbf{E}_{S \sim \omega'} [C_\mu(\mathbf{x}_S)] \leq \kappa.$$

We say that  $\mu$  is  $\kappa$ -independent if it is  $\kappa$ -independent with respect to  $\omega$ .

In Section 3.1, we explain how to condition variables. In Sections 3.2 and 3.3, we show that the conditioning operation outputs  $\kappa$ -independent LP solutions for the dense case and the locally dense case, respectively.

### 3.1 Conditioning operations

We first describe the operation of conditioning one variable. Given a solution  $\mu$  to the  $\ell$ -round SA relaxation with  $\ell \geq 2$ , we sample a vertex  $u$  uniformly at random and then set  $\mathbf{x}_u = i$ , where  $i$  is a value sampled from  $\mu_{\{u\}}$ . This operation gives a solution  $\mu'$  to the  $(\ell-1)$ -round SA relaxation: For each tuple  $(v_1, \dots, v_{\ell-1})$  of  $\ell-1$  variables, we define  $\mu'_{\{v_1, \dots, v_{\ell-1}\}}(i_1, i_2, \dots, i_{\ell-1}) = \mu_{\{v_1, \dots, v_{\ell-1}, u\}}(i_1, i_2, \dots, i_{\ell-1}, i)$ . It is not hard to check that  $\mu'$  is indeed a solution to the  $(\ell-1)$ -round SA relaxation.

Our algorithm is given in Algorithm 1. Given an solution  $\mu$  to the  $(\ell + \ell')$ -round SA relaxation, it iteratively conditions variables. We will show in subsequent sections that, if  $\omega$  is  $\Delta$ -dense or  $\Delta$ -locally dense, then Algorithm 1 outputs a  $\kappa$ -independent LP solution in  $\ell'$  steps on average, where  $\kappa = \frac{k4^k \log |D|}{\ell'}$ . (If  $\omega$  is  $\Delta$ -dense,  $\kappa$  can be slightly smaller.)

---

#### Algorithm 1

**Input:** A feasible solution  $\mu$  to the  $(\ell + \ell')$ -round SA relaxation for a CSP instance  $\mathcal{I} = (V, \omega)$ .

**Output:** An  $\kappa$ -independent solution to the  $\ell$ -round SA relaxation, where  $\kappa = \frac{k4^k \log |D|}{\ell'}$

Set  $t = 1$ .

**while** the current LP solution is not  $\kappa$ -independent **do**

    Sample a variable  $u_t \in V$  uniformly at random.

    Sample a value  $a$  from its marginal distribution  $\mu_{\{u_t\}}$  after the first  $t-1$  fixings, and condition the LP solution by setting  $\mathbf{x}_{u_t} = a$ .

$t = t + 1$ .

---

We mention here the following simple fact.

**Lemma 5** Let  $\mu'$  be the solution output by Algorithm 1. Then,  $\mathbf{E} \text{val}_{\text{LP}}(\mathcal{I}, \mu') = \text{val}_{\text{LP}}(\mathcal{I}, \mu)$ .

*Proof* Notice that the algorithm respects the marginal distributions provided by the SA relaxation during sampling the values to variables. Thus, the expected objective value is preserved.

### 3.2 The dense case

We consider the dense case, that is,  $\omega$  is a uniform distribution.

**Lemma 6** *If  $\omega$  is uniform distribution over  $V^k$ , there exists  $t \leq \ell'$  such that*

$$\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim V^k} [C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)] \leq \frac{3^k \log |D|}{\ell'}.$$

*Proof* We consider the value

$$\sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim V^k} [C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)].$$

From Lemma 3, this value can be decomposed as

$$\begin{aligned} \sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim V^k} \left[ \sum_{2 \leq r \leq k} \sum_{R \in \binom{S}{r}} I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U) \right] \\ = \sum_{2 \leq r \leq k} \binom{k}{r} \sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{R \sim V^r} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)], \end{aligned}$$

where for a set  $R = (v_1, \dots, v_r)$ ,  $I_{\boldsymbol{\mu}}(\mathbf{x}_R)$  denotes the mutual information  $I_{\boldsymbol{\mu}_R}(v_1; \dots; v_r)$ .

To bound this value, we recall the definition of mutual information. For any  $t \leq \ell'$ ,

$$\mathbf{E}_{\substack{U \sim V^t \\ R \sim V^r}} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)] = \mathbf{E}_{\substack{U \sim V^t \\ R \sim V^{r-1}}} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)] - \mathbf{E}_{\substack{U \sim V^{t+1} \\ R \sim V^{r-1}}} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)].$$

Adding the equalities from  $t = 0$  to  $\ell'$ , we get

$$\sum_{0 \leq t \leq \ell'} \mathbf{E}_{\substack{U \sim V^t \\ R \sim V^r}} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)] = \mathbf{E}_{R \sim V^{r-1}} [I(\mathbf{x}_R)] - \mathbf{E}_{\substack{U \sim V^{\ell'+1} \\ R \sim V^{r-1}}} [I_{\boldsymbol{\mu}}(\mathbf{x}_R | \mathbf{x}_U)] \leq 2^r \log |D|,$$

where the last inequality holds from  $I_{\boldsymbol{\mu}}(\mathbf{x}_R) \leq 2^{|R|} \log |D|$  by Lemma 2. Thus, we have

$$\sum_{0 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim V^k} [C(\mathbf{x}_S | \mathbf{x}_U)] \leq 3^k \log |D|,$$

and the lemma follows.

The following corollary is immediate.

**Corollary 1** *If  $\omega$  is a  $\Delta$ -dense distribution over  $V^k$ . Then there exists  $t \leq \ell'$  such that*

$$\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} [C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)] \leq \frac{3^k \log |D|}{\Delta \ell'}.$$

## 3.3 The locally dense case

We now consider the case that the scope distribution  $\omega$  is 1-locally dense.

**Lemma 7** *If  $\omega$  is a 1-locally dense distribution over  $V^k$ , then there exists  $t \leq \ell'$  such that*

$$\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} [C_{\mu}(\mathbf{x}_S | \mathbf{x}_U)] \leq \frac{k4^k \log k |D|}{\ell'}.$$

*Proof* We consider the value

$$\sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} [C_{\mu}(\mathbf{x}_S | \mathbf{x}_U)]. \quad (1)$$

From Lemma 3, this value can be decomposed as

$$\begin{aligned} \sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} \left[ \sum_{2 \leq r \leq k} \sum_{R \in \binom{S}{r}} I_{\mu}(\mathbf{x}_R | \mathbf{x}_U) \right] \\ = \sum_{J \subseteq [k]: 2 \leq |J| \leq k} \sum_{1 \leq t \leq \ell'} \mathbf{E}_{U \sim V^t} \mathbf{E}_{R \sim \omega|_J} [I_{\mu}(\mathbf{x}_R | \mathbf{x}_U)], \end{aligned}$$

where  $\omega|_J$  denotes the marginal distribution of  $\omega$  on  $J$ .

Fix  $J \subseteq [k]$  with  $|J| = r \geq 2$ . Let  $\omega_i$  be the marginal distribution of  $\omega$  on the  $i$ -th coordinate. Let  $\Omega_i = \omega_i \times V^{r-1}$  and  $\Omega'_i = \omega_i \times V^{r-2}$ . We first analyze  $I(\mathbf{x}_R)$  under  $\Omega_i$  instead of  $\omega|_J$ .

From the definition, for any  $i$  and  $t \leq \ell'$ ,

$$\mathbf{E}_{\substack{U \sim V^t \\ R \sim \Omega_i}} I(\mathbf{x}_R | \mathbf{x}_U) = \mathbf{E}_{\substack{U \sim V^t \\ R \sim \Omega'_i}} I(\mathbf{x}_R | \mathbf{x}_U) - \mathbf{E}_{\substack{U \sim V^{t+1} \\ R \sim \Omega'_i}} I(\mathbf{x}_R | \mathbf{x}_U).$$

Adding the equalities from  $t = 0$  to  $t = \ell'$ , we get

$$\sum_{0 \leq t \leq \ell'} \mathbf{E}_{\substack{U \sim V^t \\ R \sim \Omega_i}} I(\mathbf{x}_R | \mathbf{x}_U) = \mathbf{E}_{R \sim \Omega'_i} I(\mathbf{x}_R) - \mathbf{E}_{\substack{U \sim V^{\ell'+1} \\ R \sim \Omega'_i}} H(\mathbf{x}_R | \mathbf{x}_U) \leq 2^r \log |D|. \quad (2)$$

Now we turn to analyze  $I(\mathbf{x}_{v_1}; \dots; \mathbf{x}_{v_r})$  under  $\omega|_J$ .

$$\begin{aligned} \mathbf{E}_{\substack{U \sim V^t \\ R \sim \omega|_J}} I(\mathbf{x}_R | \mathbf{x}_U) &= \mathbf{E}_{U \sim V^t} \sum_{R \sim V^r} \omega|_J(R) I(\mathbf{x}_R | \mathbf{x}_U) \\ &\leq \mathbf{E}_{U \sim V^t} \sum_{(v_1, \dots, v_r) \sim V^r} \frac{1}{n^{r-1}} \sum_{1 \leq i \leq r} d_i(v_i) I(\mathbf{x}_{v_1}; \dots; \mathbf{x}_{v_r} | \mathbf{x}_U) \\ &\quad \text{(by local density and Lemma 4)} \\ &= \sum_{1 \leq i \leq r} \mathbf{E}_{U \sim V^t} \sum_{(v_1, \dots, v_r) \sim V^r} \frac{1}{n^{r-1}} \Pr_{S \sim \omega} [S_i = v_i] I(\mathbf{x}_{v_1}; \dots; \mathbf{x}_{v_r} | \mathbf{x}_U) \\ &= \sum_{1 \leq i \leq r} \mathbf{E}_{U \sim V^t} \sum_{R \sim V^r} \Omega_i(R) I(\mathbf{x}_{v_1}; \dots; \mathbf{x}_{v_r} | \mathbf{x}_U) \\ &= \sum_{1 \leq i \leq r} \mathbf{E}_{\substack{U \sim V^t \\ R \sim \Omega_i}} I(\mathbf{x}_R | \mathbf{x}_U). \end{aligned}$$

Thus from (2),

$$\sum_{0 \leq t \leq \ell'} \mathbf{E}_{\substack{U \sim V^t \\ R \sim \omega|_J}} I(\mathbf{x}_R | \mathbf{x}_U) \leq \sum_{1 \leq i \leq r} 2^r \log |D| = r 2^r \log |D|.$$

It follows that (1)  $\leq k 4^k \log |D|$  and the lemma holds.

The following corollary is immediate.

**Corollary 2** *If  $\omega$  is a  $\Delta$ -locally dense distribution over  $V^k$ , then there exists  $t \leq \ell'$  such that*

$$\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} [C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)] \leq \frac{k 4^k \log |D|}{\Delta^{\ell'}}.$$

## 4 Rounding $\kappa$ -independent solutions

### 4.1 Constraint satisfaction problems

**Lemma 8** *Let  $\mathcal{I} = (V, \omega, P)$  be a  $k$ -CSP instance over finite domain  $D$ . Let  $\boldsymbol{\mu}$  be a  $\kappa$ -independent solution to the  $k$ -round Sherali-Adams LP relaxation. There is a randomized polynomial time algorithm to find an assignment  $\alpha : V \rightarrow D$  such that  $\text{val}(\mathcal{I}, \alpha) \geq \text{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - 3\sqrt{\kappa}$ .*

*Proof* For each  $v \in V$ , let  $\alpha(v)$  be independently sampled from  $\boldsymbol{\mu}_{\{v\}}$ . For each  $S \in V^k$ , by the definition of total correlation, Lemma 1, and the fact that  $P_S(\beta) \in [0, 1]$  we have

$$\left| \mathbf{E}_{\beta \sim \boldsymbol{\mu}_S} P_S(\beta) - \mathbf{E}_{\alpha} P_S(\alpha|_S) \right| \leq 2\sqrt{C(\mathbf{x}_S)}.$$

Therefore by  $\kappa$ -independence,

$$\left| \mathbf{E}_{S \sim \omega} \left( \mathbf{E}_{\beta \sim \boldsymbol{\mu}_S} P_S(\beta) - \mathbf{E}_{\alpha} P_S(\alpha|_S) \right) \right| \leq \mathbf{E}_{S \sim \omega} 2\sqrt{C(\mathbf{x}_S)} \leq 2\sqrt{\mathbf{E}_{S \sim \omega} C(\mathbf{x}_S)} \leq 2\sqrt{\kappa}.$$

We have proved that  $\mathbf{E}_{\alpha} [\text{val}(\mathcal{I}, \alpha)] \geq \text{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - 2\sqrt{\kappa}$ . Therefore we can sample an  $\alpha$  in expected polynomial time such that  $\text{val}(\mathcal{I}, \alpha) \geq \text{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - 3\sqrt{\kappa}$ .

### 4.2 Assignment problems

Let  $\mathcal{I} = (V, \omega)$  be a  $\Delta$ -dense  $k$ -AP instance. We introduce the following relaxation  $\mathcal{H}$ , and let  $\text{val}_{\mathcal{H}}(\mathcal{I})$  be its optimal value.

$$\begin{aligned} & \text{maximize } n^k \mathbf{E}_{(U, W) \sim \omega} \prod_{i=1}^k \mathbf{x}_{u_i, w_i} \\ & \text{subject to } \mathbf{x}_{u, w} \geq 0 && \forall u, w \in V^k \\ & \sum_{u \in V} \mathbf{x}_{u, w} = 1 && \forall w \in V \\ & \sum_{w \in V} \mathbf{x}_{u, w} = 1 && \forall u \in V. \end{aligned}$$

4.2.1 From  $\kappa$ -independence to relaxation  $\mathcal{H}$ 

We first see that we can find a good solution to  $\mathcal{H}$  using a solution to the Sherali-Adams LP relaxation of a dense instance  $\mathcal{I}$ .

**Lemma 9** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega$  is  $\Delta$ -dense. Let  $\boldsymbol{\mu}$  be a  $\kappa$ -independent solution (with respect to the uniform distribution rather than  $\omega$ ) to the  $k$ -round Sherali-Adams LP relaxation of  $\mathcal{I}$ . There is a polynomial-time algorithm, on input  $\boldsymbol{\mu}$ , to find a solution to  $\mathcal{H}$  that certifies that  $\text{val}_{\mathcal{H}}(\mathcal{I}) \geq \text{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - 2\sqrt{\kappa}/\Delta$ .*

*Proof* Let  $x_{u,w} = \boldsymbol{\mu}_u(w)$  for all  $u, w \in V$ . For each  $S = (u_1, u_2, \dots, u_k) \in V^k$ , by the definition of total correlation and Lemma 1 we have

$$\sum_{T=(w_1, \dots, w_k)} \left| \boldsymbol{\mu}_S(T) - \prod_{i=1}^k x_{u_i, w_i} \right| \leq 2\sqrt{C(\boldsymbol{x}_S)}. \quad (3)$$

Therefore,

$$\begin{aligned} & \left| \mathbf{E}_{(S,T)=(u_1, \dots, u_k, v_1, \dots, v_k) \sim \omega} \left( \boldsymbol{\mu}_S(T) - \prod_{i=1}^k x_{u_i, w_i} \right) \right| \\ & \leq \frac{1}{\Delta} \mathbf{E}_{(S,T)=(u_1, \dots, u_k, v_1, \dots, v_k) \sim V^{2k}} \left| \boldsymbol{\mu}_S(T) - \prod_{i=1}^k x_{u_i, w_i} \right| \quad (\text{by density}) \\ & \leq \frac{1}{\Delta n^k} \mathbf{E}_{S \sim V^k} 2\sqrt{C(\boldsymbol{x}_S)} \quad (\text{by (3)}) \\ & \leq \frac{2}{\Delta n^k} \sqrt{\mathbf{E}_{S \sim V^k} C(\boldsymbol{x}_S)} \leq n^{-k} \cdot \frac{2\sqrt{\kappa}}{\Delta}. \quad (\text{by } \kappa\text{-independence}) \end{aligned}$$

The following variant of Lemma 9 is used when dealing with the locally dense case later.

**Lemma 10** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega(u_1, \dots, u_k, w_1, \dots, w_k) = \omega'(u_1, \dots, u_k) \cdot \omega''(w_1, \dots, w_k)$  where  $\omega''$  is a  $\Delta$ -dense distribution over  $V^k$ . Let  $\boldsymbol{\mu}$  be a  $\kappa$ -independent solution to the  $k$ -round Sherali-Adams LP relaxation of  $\mathcal{I}$ . There is a polynomial-time algorithm, on input  $\boldsymbol{\mu}$ , to find a solution to  $\mathcal{H}$  that certifies that  $\text{val}_{\mathcal{H}}(\mathcal{I}) \geq \text{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - 2\sqrt{\kappa}/\Delta$ .*

*Proof* Let  $x_{u,w} = \boldsymbol{\mu}_u(w)$  for all  $u, w \in V$ . Similar to the proof of Lemma 9, we have

$$\begin{aligned} & \left| \mathbf{E}_{(S,T)=(u_1, \dots, u_k, v_1, \dots, v_k) \sim \omega} \left( \boldsymbol{\mu}_S(T) - \prod_{i=1}^k x_{u_i, w_i} \right) \right| \\ & \leq \frac{1}{\Delta} \mathbf{E}_{\substack{S=(u_1, \dots, u_k) \sim \omega' \\ T=(v_1, \dots, v_k) \sim V^k}} \left| \boldsymbol{\mu}_S(T) - \prod_{i=1}^k x_{u_i, w_i} \right| \quad (\text{by density of } \omega'') \\ & \leq \frac{1}{\Delta n^k} \mathbf{E}_{S \sim \omega'} 2\sqrt{C(\boldsymbol{x}_S)} \quad (\text{by (3)}) \\ & \leq \frac{2}{\Delta n^k} \sqrt{\mathbf{E}_{S \sim \omega'} C(\boldsymbol{x}_S)} \leq n^{-k} \cdot \frac{2\sqrt{\kappa}}{\Delta}. \quad (\text{by } \kappa\text{-independence}) \end{aligned}$$

#### 4.2.2 From relaxation $\mathcal{H}$ to an integral solution

At a first look,  $\mathcal{H}$  is very close to the  $k$ -AP problem itself. However, we cannot independently sample  $\pi(v)$  from each  $\mathbf{x}_v$  in  $\mathcal{H}$  to get a solution to  $k$ -AP, since there is chance that  $\pi(v) = \pi(v')$ , rendering  $\pi$  not a permutation. Indeed, we show in Section 9 that for some  $k$ -AP instance  $\mathcal{I}$ , there is a gap between  $\mathbf{val}_{\mathcal{H}}(\mathcal{I})$  and  $\mathbf{val}(\mathcal{I})$ . However, our following lemma shows that this gap cannot be very big for  $k$ -AP instances  $\mathcal{I} = (V, \omega)$  when  $\omega$  is  $\Delta$ -dense. The proof is given later in Section 6.

**Lemma 11** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega$  is  $\Delta$ -dense. Given a solution  $\mathbf{x}$  to relaxation  $\mathcal{H}$ , let  $\mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathbf{x})$  be the value of the solution. There is a randomized polynomial-time algorithm to compute a permutation  $\pi$  such that  $\mathbf{val}(\mathcal{I}, \pi) \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathbf{x}) - \frac{7k^2 \log n}{\Delta\sqrt{n}}$ .*

In Section 7, we prove the following variant of Lemma 11.

**Lemma 12** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega(u_1, \dots, u_k, w_1, \dots, w_k) = \omega'(u_1, \dots, u_k) \cdot \omega''(w_1, \dots, w_k)$ , where  $\omega'$  is  $\Delta'$ -locally dense and  $\omega''$  is  $\Delta$ -dense. Given a solution  $\mathbf{x}$  to the relaxation  $\mathcal{H}$ , let  $\mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathbf{x})$  be the value of the solution. There is a randomized polynomial-time algorithm to compute a permutation  $\pi$  such that  $\mathbf{val}(\mathcal{I}, \pi) \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathbf{x}) - \frac{7k^2 \log n}{\Delta\Delta'\sqrt{n}}$ .*

#### 4.2.3 The rounding lemmas

Combining Lemma 9 and Lemma 11, and Lemma 10 and Lemma 12, we get the following main rounding lemmas for this subsection.

**Lemma 13** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega$  is  $\Delta$ -dense. Let  $\boldsymbol{\mu}$  be a  $\kappa$ -independent solution (with respect to the uniform distribution rather than  $\omega$ ) to the  $k$ -round Sherali-Adams LP relaxation for  $\mathcal{I}$ . There is a polynomial-time algorithm, on input  $\boldsymbol{\mu}$ , to find a permutation  $\pi$  such that  $\mathbf{val}(\mathcal{I}, \pi) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - \frac{2\sqrt{\kappa}}{\Delta} - \frac{7k^2 \log n}{\Delta\sqrt{n}}$ .*

**Lemma 14** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega(u_1, \dots, u_k, w_1, \dots, w_k) = \omega'(u_1, \dots, u_k) \cdot \omega''(w_1, \dots, w_k)$  where  $\omega'$  is  $\Delta'$ -locally dense and  $\omega''$  is  $\Delta$ -dense. Let  $\boldsymbol{\mu}$  be a  $\kappa$ -independent solution to the  $k$ -round Sherali-Adams LP relaxation of  $\mathcal{I}$ . There is a randomized polynomial-time algorithm, on input  $\boldsymbol{\mu}$ , to find a permutation  $\pi$  such that  $\mathbf{val}(\mathcal{I}, \pi) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - \frac{2\sqrt{\kappa}}{\Delta} - \frac{7k^2 \log n}{\Delta\Delta'\sqrt{n}}$ .*

## 5 Putting things together

The following theorem gives PTASs for dense and locally dense Max-CSP.

**Theorem 5** *Let  $\mathcal{I} = (V, \omega, P)$  be a  $k$ -CSP instance over finite domain  $D$  such that  $\omega$  is  $\Delta$ -dense or  $\Delta$ -locally dense. For any  $\epsilon > 0$ , let  $\ell = \frac{9k^2 4^k \log |D|}{\epsilon^2 \Delta}$ . The additive integrality gaps of the  $(\ell + k)$ -round Sherali-Adams LP relaxation is at most  $\epsilon$ ; and there is a randomized rounding algorithm producing a solution whose value is at least  $\mathbf{opt}(\mathcal{I}) - \epsilon$ , in expected  $n^{O(\ell)}$  time.*

*Proof* Let  $\mu$  be a solution to the  $(\ell + k)$ -round Sherali Adams LP relaxation. Let the random variable  $(\mu|\mathbf{x}_U)$  be the solution after conditioning on the variables in  $U$ . By Corollary 1 and Corollary 2, we know that there exists  $t \leq \ell$  such that

$$\mathbf{E}_{U \sim V^t} \sqrt{\mathbf{E}_{S \sim \omega} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \leq \sqrt{\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim \omega} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \leq \sqrt{\frac{k^2 4^k \log |D|}{\Delta \ell}} = \frac{\epsilon}{3}.$$

Together with Lemma 5, we have

$$\mathbf{E}_{U \sim V^t} \left( \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu|\mathbf{x}_U) - 3 \sqrt{\mathbf{E}_{S \sim \omega} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \right) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon.$$

We enumerate all the possible ways of conditioning, and find out a solution  $\mu'$  to the  $(k + \ell - t)$ -round Sherali-Adams LP relaxation such that  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \mu') - 3 \sqrt{\mathbf{E}_{S \sim \omega} C_{\mu'}(\mathbf{x}_S)} \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon$ . Since  $\mu'$  is always a  $\mathbf{E}_{S \sim \omega} C_{\mu'}(\mathbf{x}_S)$ -independent solution, by Lemma 8, given  $\mu'$ , we can find an assignment with value at least  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon$  in randomized polynomial time.

Now we prove that there is a quasi-polynomial-time approximation scheme for dense Max-AP.

**Theorem 6** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega$  is  $\Delta$ -dense. For any  $\epsilon > 0$ , let  $\ell = \frac{4k^4 \log |D|}{\epsilon^2 \Delta^2}$ . The additive integrality gaps of the  $(\ell + k)$ -round Sherali-Adams LP relaxation is at most  $\epsilon + \frac{7k^2 \log n}{\Delta \sqrt{n}}$ ; and there is a randomized rounding algorithm producing a solution whose value is at least  $\mathbf{opt}(\mathcal{I}) - \epsilon - \frac{7k^2 \log n}{\Delta \sqrt{n}}$ , in expected  $n^{O(\ell)}$  time.*

*Proof* Let  $\mu$  be a solution to the  $(\ell + k)$ -round Sherali Adams LP relaxation. By Lemma 6, we know that there exists  $t \leq \ell$  such that

$$\mathbf{E}_{U \sim V^t} \sqrt{\mathbf{E}_{S \sim V^k} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \leq \sqrt{\mathbf{E}_{U \sim V^t} \mathbf{E}_{S \sim V^k} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \leq \sqrt{\frac{k^4 \log n}{\ell}} = \frac{\epsilon \Delta}{2}.$$

Together with Lemma 5, we have

$$\mathbf{E}_{U \sim V^t} \left( \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu|\mathbf{x}_U) - \frac{2}{\Delta} \sqrt{\mathbf{E}_{S \sim V^k} C_\mu(\mathbf{x}_S|\mathbf{x}_U)} \right) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon.$$

We enumerate all the possible ways of conditioning, and find out a solution  $\mu'$  to the  $(k + \ell - t)$ -round Sherali-Adams LP relaxation such that  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \mu') - 2 \sqrt{\mathbf{E}_{S \sim V^k} C_{\mu'}(\mathbf{x}_S)} \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon$ . By Lemma 13, given  $\mu'$ , we can find a permutation with value at least  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - \epsilon - \frac{7k^2 \log n}{\Delta \sqrt{n}}$ .

Using Corollary 2 and Lemma 14 instead of Lemma 6 and Lemma 13, the same argument shows that there is a quasi-polynomial-time approximation scheme for locally dense Max-AP.

**Theorem 7** *Let  $\mathcal{I} = (V, \omega)$  be a  $k$ -AP instance such that  $\omega(u_1, \dots, u_k, w_1, \dots, w_k) = \omega'(u_1, \dots, u_k) \cdot \omega''(w_1, \dots, w_k)$  where  $\omega'$  is  $\Delta'$ -locally dense and  $\omega''$  is  $\Delta$ -dense. For any  $\epsilon > 0$ , let  $\ell = \frac{4k^2 4^k \log |D|}{\epsilon^2 \Delta' \Delta^2}$ . The additive integrality gaps of the  $(\ell + k)$ -round Sherali-Adams LP relaxation is at most  $\epsilon + \frac{7k^2 \log n}{\Delta \Delta' \sqrt{n}}$ ; and there is a randomized rounding algorithm producing a solution whose value is at least  $\mathbf{opt}(\mathcal{I}) - \epsilon - \frac{7k^2 \log n}{\Delta \Delta' \sqrt{n}}$ , in expected  $n^{O(\ell)}$  time.*

## 6 Proof of Lemma 11

Observe that a solution  $\mathbf{x}$  to the relaxation  $\mathcal{H}$  corresponds to a doubly stochastic matrix. Now let us decompose  $\mathbf{x}$  into a distribution of permutations  $\mathcal{D} = \{\pi : V \rightarrow V\}$  such that for any  $u, w \in V$ , we have  $\Pr_{\pi \sim \mathcal{D}}[\pi(u) = w] = \mathbf{x}_{u,w}$ . Let  $\mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) = \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathbf{x})$  be the value of relaxation  $\mathcal{H}$  on  $\mathbf{x}$  for instance  $\mathcal{I}$ . Our goal is to “merge” the permutations in  $\mathcal{D}$  into one permutation while not losing much in the objective value. The following lemma proves this for the special case when  $\mathcal{D}$  is supported on only two permutations.

**Lemma 15** *Let  $\mathcal{D}$  be the distribution over  $\pi_1$  and  $\pi_2$  such that  $\pi_1$  is chosen with probability  $p$  and  $\pi_2$  is chosen with probability  $(1-p)$ . There exists a distribution  $\mathcal{D}'$  over permutations such that for any  $k \geq 2$  and any  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  such that  $\omega$  is  $\Delta$ -dense, we have*

$$\mathbf{E}_{\pi \sim \mathcal{D}'} [\mathbf{val}(\mathcal{I}, \pi)] \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - \frac{2k^2}{\Delta\sqrt{n}}.$$

Moreover,  $\mathcal{D}'$  can be sampled in polynomial time.

*Proof* Let us assume w.l.o.g. that  $V = [n]$ ,  $\pi_1 = \text{id}$  (i.e.  $\pi_1(i) = i$  for all  $i \in [n]$ ). For any set  $A = \{a_i : a_1 < a_2 < \dots < a_{|A|} = n\} \subseteq [n]$ , let us define  $\pi_A$  be the permutation over  $[n]$  so that  $\pi_A(i) = a_{t-1} + 1$  if  $i = a_t$  for some  $t \in [|A|]$  and  $\pi_A(i) = i + 1$  otherwise (assuming  $a_0 = 0$ ). We can also assume w.l.o.g. that there exists  $A \subseteq [n]$  such that  $\pi_2 = \pi_A$ . See Figure 1. We can add at most  $\sqrt{n}$  elements into  $A$  to get  $A' \subseteq [n]$  such that there is no set of  $\sqrt{n}$  consecutive integers that does not intersect  $A'$ . It is easy to show that  $\pi_A$  and  $\pi_{A'}$  differ at most  $2\sqrt{n}$  places. Let  $\mathcal{D}_{A'}$  be the probability distribution that chooses  $\pi_1$  with probability  $p$  and  $\pi_{A'}$  with probability  $(1-p)$ . For any  $k$  and any  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  such that  $\omega$  is  $\Delta$ -dense, we have

$$\begin{aligned} & \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}_{A'}) \\ &= n^k \mathbf{E}_{(U, W) \sim \omega} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}}[\pi(u_i) = w_i] - \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}}[\pi(u_i) = w_i] \right) \\ &\leq n^k \mathbf{E}_{(U, W) \sim \omega} \mathbf{1}[\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}}[\pi(u_i) = w_i] \\ &\leq \frac{n^k}{\Delta} \cdot \mathbf{E}_{(U, W) \sim V^{2k}} \mathbf{1}[\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}}[\pi(u_i) = w_i], \end{aligned} \quad (4)$$

where the last inequality is by the density of  $\omega$ .

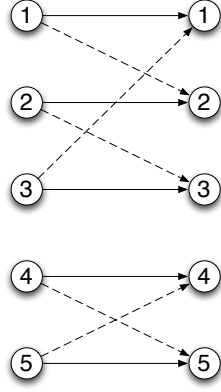
Since

$$n^k \mathbf{E}_{W \sim V^k} \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}}[\pi(u_i) = w_i] = \prod_{i=1}^k \sum_{w_i \sim V} \Pr_{\pi \sim \mathcal{D}}[\pi(u_i) = w_i] = 1, \quad (5)$$

we have

$$(4) \leq \frac{1}{\Delta} \cdot \Pr_{U \sim V^k} [\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \leq \frac{2k}{\Delta\sqrt{n}}. \quad (6)$$





**Fig. 1** Permutations  $\pi_1$  and  $\pi_2$  over  $[5]$  are shown as mappings from  $[5]$  to  $[5]$ . Solid arrow represents  $\pi_1$  and dashed arrows represent  $\pi_2$ . For any permutation  $\pi_1$ , we can move vertices in the right side (and relabeling them accordingly) so that the resulting  $\pi_1$  is the identity permutation. Then for any permutation  $\pi_2$ , we can move pairs of vertices (and relabeling them accordingly) so that  $\pi_1$  remains the identity permutation whereas the cycles formed by  $\pi_1$  and  $\pi_2$  are drawn disjointly. In such a case,  $\pi_2$  satisfies the condition in the body text.

Now we define the distribution  $\mathcal{D}'$ . Let us assume that the elements in  $A'$  are  $a'_1 < a'_2 < \dots < a'_{|A'|} = n$ ; let  $a'_0 = 0$  for convenience. To draw a permutation  $\pi \sim \mathcal{D}'$ , we sample  $|A'|$  i.i.d. 0/1 bits  $b_1, b_2, \dots, b_{|A'|}$ , each of which has mean  $p$ . For each  $i$ , we find out the unique  $t \in [|A'|]$  so that  $a'_{t-1} < i \leq a'_t$ ; let  $\pi(i) = \pi_1(i) = i$  if  $b_t = 0$ ; let  $\pi(i) = \pi_{A'}(i)$  otherwise.

For any  $k$  and any  $\Delta$ -dense  $k$ -AP instance  $\mathcal{I} = (V, \omega)$ , we have

$$\begin{aligned}
& \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}_{A'}) - \mathbf{E}_{\pi \sim \mathcal{D}'} \text{val}(\mathcal{I}, \pi) \\
&= n^k \mathbf{E}_{(U, W) \sim \omega} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] - \Pr_{\pi \sim \mathcal{D}'} [\forall i \in [k] : \pi(u_i) = w_i] \right) \\
&= n^k \mathbf{E}_{(U, W) \sim \omega} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \right. \\
&\quad \left. - \prod_{t=1}^{|A'|} \Pr_{\pi \sim \mathcal{D}'} [\forall i \in [k], a'_{t-1} < u_i \leq a'_t : \pi(u_i) = w_i] \right) \\
&\leq n^k \mathbf{E}_{(U, W) \sim \omega} \mathbf{1} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \\
&\quad \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \\
&\leq \frac{n^k}{\Delta} \mathbf{E}_{(U, W) \sim V^{2k}} \mathbf{1} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \\
&\quad \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \quad (\text{by density})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\Delta} \Pr_{U \sim V^k} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \quad (\text{by (5)}) \\
&\leq \frac{1}{\Delta} \cdot \binom{k}{2} \cdot \sum_{t \in [|A'|]} \left( \frac{a'_t - a'_{t-1}}{n} \right)^2 \leq \frac{1}{\Delta} \cdot \binom{k}{2} \cdot \sum_{t \in [|A'|]} \left( \frac{a'_t - a'_{t-1}}{n} \right) \cdot \frac{1}{\sqrt{n}} \leq \frac{k^2}{\sqrt{n}}.
\end{aligned} \tag{7}$$

The lemma is proved by combining (6) and (7).

*Proof (of Lemma 11)*

Let  $\mathcal{D}$  be supported on  $\pi_1, \pi_2, \dots, \pi_m$ , each  $\pi_i$  is chosen with probability  $p_i$ . We can assume that  $m \leq n^2$  by preserving only the  $n^2$  permutations with the largest probabilities and proper normalization, which would cause a loss of at most  $n^{-1}$  in the objective value  $\text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D})$ . Now we show that for any such distribution, we can find a distribution  $\mathcal{E}$  that is supported on  $(m-1)$  permutations, such that

$$\text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) \geq \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta\sqrt{n}}. \tag{8}$$

In other words, since  $\pi_1$  and  $\pi_2$  are arbitrary, we are able to “merge” any two permutations  $\pi_i$  and  $\pi_j$  in  $\mathcal{D}$  by paying a loss of  $(p_i + p_j) \frac{2k^2}{\Delta\sqrt{n}}$  in the objective value. We repeatedly merge the two permutations with the smallest probability mass in the distribution until there is only one permutation left, during this process we lose at most  $\lceil \log m \rceil \frac{2k^2}{\Delta\sqrt{n}} \leq \frac{6k^2 \log n}{\Delta\sqrt{n}}$  in objective value. Together with the  $n^{-1}$  loss at the beginning of the proof, we lose at most  $\frac{7k^2 \log n}{\Delta\sqrt{n}}$  for sufficiently large  $n$ .

In order to show (8), let us define a distribution  $\tilde{\mathcal{E}}$  of distributions of permutations as follows. Let  $\mathcal{F}$  to be the distribution of permutations that chooses  $\pi_1$  with probability  $\frac{p_1}{p_1+p_2}$  and  $\pi_2$  with the remaining probability. Apply Lemma 15 on  $\mathcal{F}$  to get  $\mathcal{F}'$ . A distribution  $\mathcal{E}$  from  $\tilde{\mathcal{E}}$  is sampled by first sampling a permutation  $\pi$  from  $\mathcal{F}'$ , and returning the distribution that puts probability mass  $(p_1 + p_2)$  on  $\pi$  and  $p_i$  on  $\pi_i$  for all  $i : 3 \leq i \leq m$ . For every  $u, w \in V$ , let  $\gamma_{u,w} = \sum_{t=3}^m p_t \mathbf{1}[\pi_t(u) = w] \leq 1$ . We have

$$\begin{aligned}
\mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) &= \mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \mathbf{E}_{(U,W) \sim \omega} \prod_{i=1}^k \Pr_{\pi \sim \mathcal{E}} [\pi(u_i) = w_i] \\
&= \mathbf{E}_{\pi \sim \mathcal{F}'} \mathbf{E}_{(U,W) \sim \omega} \prod_{i=1}^k ((p_1 + p_2) \mathbf{1}[\pi(u_i) = w_i] + \gamma_{u_i, w_i}) \\
&= \sum_{Q \subseteq [k]} \mathbf{E}_{(U,W) \sim \omega} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} (p_1 + p_2) \mathbf{1}[\pi(u_i) = w_i] \right) \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) \\
&= \sum_{Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\bar{Q}}, W_{\bar{Q}}) \\ \sim \omega_{(\bar{Q}, \bar{Q})}}} \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|Q|} \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega_{(U_{\bar{Q}}, W_{\bar{Q}})}}} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} \mathbf{1}[\pi(u_i) = w_i] \right),
\end{aligned} \tag{9}$$

where  $U_A$  is the restriction of vector  $U$  over coordinates in  $A$ ,  $\omega_{(A,B)}$  is the marginal distribution of  $\omega$  over  $A$  in the first  $k$  coordinates and  $B$  in the last  $k$  coordinates,

and  $\omega|(U_A, W_B)$  is the distribution  $\omega$  conditioned on that coordinates in  $A$  are assigned  $U$  and coordinates in  $B$  are assigned  $W$ . Let  $\mathcal{I}_{U_{\bar{Q}}, W_{\bar{Q}}}$  be the  $|Q|$ -AP instance  $(V, \omega|(U_{\bar{Q}}, W_{\bar{Q}}))$ . We know that  $\omega|(U_{\bar{Q}}, W_{\bar{Q}})$  is  $\Delta \cdot n^{2|\bar{Q}|} \cdot \omega_{(\bar{Q}, \bar{Q})}(U_{\bar{Q}}, W_{\bar{Q}})$ -dense. Therefore for every  $Q \neq \emptyset$ , by Lemma 15, we have

$$\begin{aligned} & \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega|(U_{\bar{Q}}, W_{\bar{Q}})}} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} \mathbf{1}[\pi(u_i) = w_i] \right) = \mathbf{E}_{\pi \sim \mathcal{F}'} \mathbf{val}(\mathcal{I}_{U_{\bar{Q}}, W_{\bar{Q}}}, \pi) \\ & \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}_{U_{\bar{Q}}, W_{\bar{Q}}}, \mathcal{F}) - \frac{2k^2 n^{-2|\bar{Q}|}}{\Delta \omega_{(\bar{Q}, \bar{Q})}(U_{\bar{Q}}, W_{\bar{Q}}) \sqrt{n}} \\ & = \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega|(U_{\bar{Q}}, W_{\bar{Q}})}} \prod_{i \in Q} \left( \frac{p_1 \mathbf{1}[\pi_1(u_i) = w_i] + p_2 \mathbf{1}[\pi_2(u_i) = w_i]}{p_1 + p_2} \right) - \frac{2k^2 n^{-2|\bar{Q}|}}{\Delta \omega_{(\bar{Q}, \bar{Q})}(U_{\bar{Q}}, W_{\bar{Q}}) \sqrt{n}}. \end{aligned}$$

Therefore we have

$$\begin{aligned} (9) & \geq \sum_{Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\bar{Q}}, W_{\bar{Q}}) \\ \sim \omega_{(\bar{Q}, \bar{Q})}}} \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega|(U_{\bar{Q}}, W_{\bar{Q}})}} \prod_{i \in Q} (p_1 \mathbf{1}[\pi_1(u_i) = w_i] + p_2 \mathbf{1}[\pi_2(u_i) = w_i]) \\ & \quad - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\bar{Q}}, W_{\bar{Q}}) \\ \sim \omega_{(\bar{Q}, \bar{Q})}}} \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|Q|} \cdot \frac{2k^2 n^{-2|\bar{Q}|}}{\Delta \omega_{(\bar{Q}, \bar{Q})}(U_{\bar{Q}}, W_{\bar{Q}}) \sqrt{n}} \\ & = \mathbf{E}_{(U, W) \sim \omega} \left( \sum_{t=1}^m p_t \mathbf{1}[\pi_t(u_i) = w_i] \right) \\ & \quad - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\bar{Q}}, W_{\bar{Q}}) \\ \sim \omega_{(\bar{Q}, \bar{Q})}}} \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|Q|} \cdot \frac{2k^2}{\Delta \sqrt{n}} \\ & \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta \sqrt{n}}. \end{aligned}$$

In all, we have proved that  $\mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta \sqrt{n}}$ . Since  $\tilde{\mathcal{E}}$  can be sampled in polynomial time, there is a randomized polynomial-time algorithm to find out a  $\mathcal{E}$  satisfying (8).

## 7 Proof of Lemma 12

We say a distribution  $\omega'$  over  $V^k$  is  $\Delta'$ -well spread if for every  $i, j \in [k]$  such that  $i \neq j$ , and for every disjoint partition  $V = V_1 \cup V_2 \cup \dots \cup V_t$ , we have

$$\Delta' \cdot \Pr_{(u_1, \dots, u_k) \sim \omega'} [\exists t' \in [t] : u_i \in V_{t'} \text{ and } u_j \in V_{t'}] \leq \frac{\max_{t' \in [t]} |V_{t'}|}{n}.$$

*Claim* A  $\Delta'$ -locally dense distribution  $\omega'$  is  $(\Delta'/k)$ -well spread.

*Proof* W.l.o.g. we assume that  $i = 1$  and  $j = 2$ . For every  $Z \subseteq V$ , we have

$$\begin{aligned}
& \Pr_{(u_1, \dots, u_k) \sim \omega'} [\exists t' \in [t] : u_i \in V_{t'} \text{ and } u_j \in V_{t'}] \\
&= \sum_{t' \in [t]} \sum_{\substack{u_1, u_2 \in V_{t'} \\ u_3, \dots, u_k \in V}} \omega'(u_1, \dots, u_k) \leq \sum_{t' \in [t]} \sum_{\substack{u_1, u_2 \in V_{t'} \\ u_3, \dots, u_k \in V}} \frac{\sum_{i=1}^k d_i(u_i)}{\Delta' n^{k-1}} \\
&= \sum_{t' \in [t]} \left( \sum_{\substack{u_2 \in V_{t'} \\ u_3, \dots, u_k \in V}} \frac{\sum_{u_1 \in V_{t'}} d_1(u_1)}{\Delta' n^{k-1}} + \sum_{\substack{u_1 \in V_{t'} \\ u_3, \dots, u_k \in V}} \frac{\sum_{u_2 \in V_{t'}} d_2(u_2)}{\Delta' n^{k-1}} \right. \\
&\quad \left. + \sum_{i=3}^k \sum_{\substack{u_1, u_2 \in V_{t'} \\ u_3, \dots, u_{i-1}, u_{i+1}, \dots, u_k \in V}} \frac{\sum_{u_i \in V} d_i(u_i)}{\Delta' n^{k-1}} \right) \\
&\leq 2 \cdot \frac{\max_{t' \in [t]} |V_{t'}|}{\Delta' n} + (k-2) \cdot \frac{\Delta' \max_{t' \in [t]} |V_{t'}|^2}{n^2} \leq \frac{k \max_{t' \in [t]} |V_{t'}|}{\Delta' n}.
\end{aligned}$$

We will prove a slightly stronger statement than that of Lemma 12, in the sense that we prove the lemma for every  $\omega$  such that  $\omega = \omega' \cdot \omega''$  where  $\omega'$  is  $\Delta'$ -well spread and  $\omega''$  is  $\Delta$ -dense.

The proof goes along the lines of the proof of Lemma 11. We decompose  $x$  into a distribution of permutations  $\mathcal{D} = \{\pi : V \rightarrow V\}$  such that for any  $u, w \in V$ , we have  $\Pr_{\pi \sim \mathcal{D}} [\pi(u) = w] = x_{u,w}$ . We first prove following lemma, which is an analogy of Lemma 15.

**Lemma 16** *Let  $\mathcal{D}$  be the distribution over  $\pi_1$  and  $\pi_2$  such that  $\pi_1$  is chosen with probability  $p$  and  $\pi_2$  is chosen with probability  $(1-p)$ . There exists a distribution  $\mathcal{D}'$  over permutations and a distribution  $\mathcal{V}$  over the disjoint partitions  $\{(V_1 \cup \dots \cup V_t)\}$  where each  $V_i$  has at most  $2\sqrt{n}$  elements, such that for any  $k \geq 2$  and any  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  such that  $\omega = \omega' \cdot \omega''$  where  $\omega''$  is  $\Delta$ -dense, we have*

$$\begin{aligned}
\mathbf{E}_{\pi \sim \mathcal{D}'} [\text{val}(\mathcal{I}, \pi)] &\geq \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - \frac{2k}{\Delta\sqrt{n}} \\
&\quad - \frac{1}{\Delta} \sum_{1 \leq i < j \leq k} \mathbf{E}_{(V_1 \cup \dots \cup V_i) \sim \mathcal{V}} \sum_{t' \in [t]} \Pr_{(u_1, \dots, u_k) \sim \omega'} [u_i \in V_{t'} \text{ and } u_j \in V_{t'}].
\end{aligned}$$

Moreover,  $\mathcal{D}'$  can be sampled in polynomial time.

*Proof* Let us assume w.l.o.g. that  $V = [n]$ ,  $\pi_1 = \text{id}$  (i.e.  $\pi_1(i) = i$  for all  $i \in [n]$ ). For any set  $A = \{a_i : a_1 < a_2 < \dots < a_{|A|} = n\} \subseteq [n]$ , let us define  $\pi_A$  be the permutation over  $[n]$  so that  $\pi_A(i) = a_{t-1} + 1$  if  $i = a_t$  for some  $t \in [|A|]$  and  $\pi_A(i) = i + 1$  otherwise (assuming  $a_0 = 0$ ). We can also assume w.l.o.g. that there exists  $A \subseteq [n]$  such that  $\pi_2 = \pi_A$ .

Now we define the random set variable  $A' : A \subseteq A \subseteq V$  as follows. We start from  $A' = A$ , and for each  $i \in [\sqrt{n}]$ , we uniformly sample an element  $a$  from  $((i-1)\sqrt{n}, i\sqrt{n})$  and let  $A' \leftarrow A' \cup \{a\}$ . In this way, we know that there is no set of  $2\sqrt{n}$  consecutive integers that does not intersect  $A'$ . It is easy to show that for every  $v \in V$ ,  $\Pr_{A'} [\pi_A(v) \neq \pi_{A'}(v)] \leq \frac{2}{\sqrt{n}}$ .

Let  $\mathcal{D}_{A'}$  be the probability distribution that chooses  $\pi_1$  with probability  $p$  and  $\pi_{A'}$  with probability  $(1-p)$ . For any  $k$  and any  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  such that  $\omega$  is  $\Delta$ -dense, we have

$$\begin{aligned}
& \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - \mathbf{E}_{A'} \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}_{A'}) \\
&= n^k \mathbf{E}_{(U, W) \sim \omega} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}} [\pi(u_i) = w_i] - \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \right) \\
&\leq n^k \mathbf{E}_{U \sim \omega', W \sim \omega''} \mathbf{E}_{A'} \mathbf{1} [\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}} [\pi(u_i) = w_i] \\
&\leq \frac{n^k}{\Delta} \cdot \mathbf{E}_{U \sim \omega'} \mathbf{E}_{A'} \mathbf{1} [\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \cdot \mathbf{E}_{W \sim V^k} \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}} [\pi(u_i) = w_i], \quad (10)
\end{aligned}$$

where the last inequality is by the density of  $\omega''$ . By (5), we have

$$(10) \leq \frac{1}{\Delta} \cdot \mathbf{E}_{U \sim \omega'} \Pr_{A'} [\exists i \in [k] : \pi_{A'}(u_i) \neq \pi_A(u_i)] \leq \frac{2k}{\Delta \sqrt{n}}. \quad (11)$$

For every  $A' \subseteq [n]$ , we define the distribution  $\mathcal{D}'_{A'}$ . Let us assume that the elements in  $A'$  are  $a'_1 < a'_2 < \dots < a'_{|A'|} = n$ ; let  $a'_0 = 0$  for convenience. To draw a permutation  $\pi \sim \mathcal{D}'_{A'}$ , we sample  $|A'|$  i.i.d. 0/1 bits  $b_1, b_2, \dots, b_{|A'|}$ , each of which has mean  $p$ . For each  $i$ , we find out the unique  $t \in [|A'|]$  so that  $a'_{t-1} < i \leq a'_t$ ; let  $\pi(i) = \pi_1(i) = i$  if  $b_t = 0$ ; let  $\pi(i) = \pi_{A'}(i)$  otherwise.

Now we define the distribution  $\mathcal{D}'$ . To draw a permutation  $\pi \sim \mathcal{D}'$ , we first sample a random set  $A'$ , and then draw a permutation from  $\mathcal{D}'_{A'}$ .

For any  $k$  and any  $k$ -AP instance  $\mathcal{I} = (V, \omega)$  such that  $\omega = \omega' \cdot \omega''$  where  $\omega''$  is  $\Delta$ -dense, we have

$$\begin{aligned}
& \mathbf{E}_{A'} \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}_{A'}) - \mathbf{E}_{\pi \sim \mathcal{D}'} \mathbf{val}(\mathcal{I}, \pi) \\
&= n^k \mathbf{E}_{(U, W) \sim \omega} \mathbf{E}_{A'} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] - \Pr_{\pi \sim \mathcal{D}'_{A'}} [\forall i \in [k] : \pi(u_i) = w_i] \right) \\
&= n^k \mathbf{E}_{(U, W) \sim \omega} \mathbf{E}_{A'} \left( \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \right. \\
&\quad \left. - \prod_{t=1}^{|A'|} \Pr_{\pi \sim \mathcal{D}'_{A'}} [\forall i \in [k], a'_{t-1} < u_i \leq a'_t : \pi(u_i) = w_i] \right) \\
&\leq n^k \mathbf{E}_{(U, W) \sim \omega} \mathbf{E}_{A'} \mathbf{1} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \\
&\quad \cdot \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \\
&\leq \frac{n^k}{\Delta} \mathbf{E}_{U \sim \omega'} \mathbf{E}_{A'} \mathbf{1} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \\
&\quad \cdot \mathbf{E}_{W \sim V^k} \prod_{i=1}^k \Pr_{\pi \sim \mathcal{D}_{A'}} [\pi(u_i) = w_i] \quad (\text{by density})
\end{aligned}$$

$$\leq \frac{1}{\Delta} \mathbf{E}_{U \sim \omega'} \mathbf{E}_{A'} \mathbf{1} [\exists t \in [|A'|] : \exists \text{ more than one } i \text{ s.t. } a'_{t-1} < u_i \leq a'_t] \quad (\text{by (5)})$$

$$\leq \frac{1}{\Delta} \sum_{1 \leq i < j \leq k} \mathbf{E}_{A'} \sum_{t \in [|A'|]} \Pr_{(u_1, \dots, u_k) \sim \omega'} [a'_{t-1} < u_i, u_j \leq a'_t]. \quad (12)$$

The lemma is proved by combining (11) and (12).

Now we are ready to prove Lemma 12.

*Proof (of Lemma 12)* Let  $\mathcal{D}$  be supported on  $\pi_1, \pi_2, \dots, \pi_m$ , each  $\pi_i$  is chosen with probability  $p_i$ . We can assume that  $m \leq n^2$  by preserving only the  $n^2$  permutations with the largest probabilities and proper normalization, which would cause a loss of at most  $n^{-1}$  in the objective value  $\text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D})$ . Now we show that for any such distribution, we can find a distribution  $\mathcal{E}$  that is supported on  $(m-1)$  permutations, such that

$$\text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) \geq \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta \Delta' \sqrt{n}}. \quad (13)$$

In other words, since  $\pi_1$  and  $\pi_2$  are arbitrary, we are able to “merge” any two permutations  $\pi_i$  and  $\pi_j$  in  $\mathcal{D}$  by paying a loss of  $(p_i + p_j) \frac{2k^2}{\Delta \Delta' \sqrt{n}}$  in the objective value. We repeatedly merge the two permutations with the smallest probability mass in the distribution until there is only one permutation left, during this process we lose at most  $\lceil \log m \rceil \frac{2k^2}{\Delta \sqrt{n}} \leq \frac{6k^2 \log n}{\Delta \Delta' \sqrt{n}}$  in objective value. Together with the  $n^{-1}$  loss at the beginning of the proof, we lose at most  $\frac{7k^2 \log n}{\Delta \Delta' \sqrt{n}}$  for sufficiently large  $n$ .

In order to show (13), let us define a distribution  $\tilde{\mathcal{E}}$  of distributions of permutations as follows. Let  $\mathcal{F}$  be the distribution of permutations that chooses  $\pi_1$  with probability  $\frac{p_1}{p_1 + p_2}$  and  $\pi_2$  with the remaining probability. Apply Lemma 16 on  $\mathcal{F}$  to get  $\mathcal{F}'$ . A distribution  $\mathcal{E}$  from  $\tilde{\mathcal{E}}$  is sampled by first sampling a permutation  $\pi$  from  $\mathcal{F}'$ , and returning the distribution that puts probability mass  $(p_1 + p_2)$  on  $\pi$  and  $p_i$  on  $\pi_i$  for all  $i : 3 \leq i \leq m$ . For every  $u, w \in V$ , let  $\gamma_{u,w} = \sum_{t=3}^m p_t \mathbf{1}[\pi_t(u) = w] \leq 1$ . We have

$$\begin{aligned} \mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \text{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) &= \mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \mathbf{E}_{(U, W) \sim \omega} \prod_{i=1}^k \Pr_{\pi \sim \mathcal{E}} [\pi(u_i) = w_i] \\ &= \mathbf{E}_{\pi \sim \mathcal{F}'} \mathbf{E}_{(U, W) \sim \omega} \prod_{i=1}^k ((p_1 + p_2) \mathbf{1}[\pi(u_i) = w_i] + \gamma_{u_i, w_i}) \\ &= \sum_{Q \subseteq [k]} \mathbf{E}_{(U, W) \sim \omega} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} (p_1 + p_2) \mathbf{1}[\pi(u_i) = w_i] \right) \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) \\ &= \sum_{Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\bar{Q}}, W_{\bar{Q}}) \\ \sim \omega_{(\bar{Q}, \bar{Q})}}} \left( \prod_{i \in \bar{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|Q|} \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega_{(U_{\bar{Q}}, W_{\bar{Q}})}}} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} \mathbf{1}[\pi(u_i) = w_i] \right), \end{aligned} \quad (14)$$

where  $U_A$  is the restriction of vector  $U$  over coordinates in  $A$ ,  $\omega''_A$  is the marginal distribution of  $\omega''$  over the coordinates in  $A$ , and  $\omega_{(A, B)}$  is the marginal distribution

of  $\omega$  over  $A$  in the first  $k$  coordinates and  $B$  in the last  $k$  coordinates. Let  $\mathcal{I}_{U_{\overline{Q}}, W_{\overline{Q}}}$  be the  $|\overline{Q}|$ -AP instance  $(V, \omega|(U_{\overline{Q}}, W_{\overline{Q}}))$ . We know that  $\omega|(U_{\overline{Q}}, W_{\overline{Q}}) = (\omega'|U_{\overline{Q}}) \cdot (\omega''|W_{\overline{Q}})$ , and  $\omega''|W_{\overline{Q}}$  is  $\Delta \cdot n^{|\overline{Q}|} \cdot \omega''_Q(W_{\overline{Q}})$ -dense. Therefore for every  $Q \neq \emptyset$ , by Lemma 16, we have

$$\begin{aligned}
& \mathbf{E}_{\substack{(U_{\overline{Q}}, W_{\overline{Q}}) \\ \sim \omega|(U_{\overline{Q}}, W_{\overline{Q}})}} \left( \mathbf{E}_{\pi \sim \mathcal{F}'} \prod_{i \in Q} \mathbf{1}[\pi(u_i) = w_i] \right) = \mathbf{E}_{\pi \sim \mathcal{F}'} \mathbf{val}(\mathcal{I}_{U_{\overline{Q}}, W_{\overline{Q}}}, \pi) \\
& \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}_{U_{\overline{Q}}, W_{\overline{Q}}}, \mathcal{F}) - \frac{2kn^{-|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}}) \sqrt{n}} \\
& \quad - \frac{n^{-|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}})} \sum_{\substack{i, j \in Q \\ i \neq j}} \mathbf{E}_{(V_1 \cup \dots \cup V_t) \sim \mathcal{V}} \sum_{t' \in [t]} \Pr_{U_Q \sim \omega'|U_{\overline{Q}}} [u_i \in V_{t'} \text{ and } u_j \in V_{t'}] \\
& = \mathbf{E}_{\substack{(U_{\overline{Q}}, W_{\overline{Q}}) \\ \sim \omega|(U_{\overline{Q}}, W_{\overline{Q}})}} \prod_{i \in Q} \left( \frac{p_1 \mathbf{1}[\pi_1(u_i) = w_i] + p_2 \mathbf{1}[\pi_2(u_i) = w_i]}{p_1 + p_2} \right) - \frac{2kn^{-|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}}) \sqrt{n}} \\
& \quad - \frac{n^{-|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}})} \sum_{\substack{i, j \in Q \\ i \neq j}} \mathbf{E}_{(V_1 \cup \dots \cup V_t) \sim \mathcal{V}} \sum_{t' \in [t]} \Pr_{U_Q \sim \omega'|U_{\overline{Q}}} [u_i \in V_{t'} \text{ and } u_j \in V_{t'}].
\end{aligned}$$

Therefore we have

$$\begin{aligned}
& (14) \\
& \geq \sum_{Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\overline{Q}}, W_{\overline{Q}}) \\ \sim \omega_{(\overline{Q}, \overline{Q})}}} \left( \prod_{i \in \overline{Q}} \gamma_{u_i, w_i} \right) \mathbf{E}_{\substack{(U_Q, W_Q) \\ \sim \omega|(U_{\overline{Q}}, W_{\overline{Q}})}} \prod_{i \in Q} (p_1 \mathbf{1}[\pi_1(u_i) = w_i] + p_2 \mathbf{1}[\pi_2(u_i) = w_i]) \\
& \quad - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\overline{Q}}, W_{\overline{Q}}) \\ \sim \omega_{(\overline{Q}, \overline{Q})}}} \left( \prod_{i \in \overline{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|\overline{Q}|} \cdot \frac{2kn^{-|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}}) \sqrt{n}} \\
& \quad - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{(U_{\overline{Q}}, W_{\overline{Q}}) \\ \sim \omega_{(\overline{Q}, \overline{Q})}}} \left( \prod_{i \in \overline{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|\overline{Q}|} \\
& \quad \quad \cdot \frac{n^{|\overline{Q}|}}{\Delta \omega''_Q(W_{\overline{Q}})} \sum_{\substack{i, j \in Q \\ i \neq j}} \mathbf{E}_{(V_1 \cup \dots \cup V_t) \sim \mathcal{V}} \sum_{t' \in [t]} \Pr_{U_Q \sim \omega'|U_{\overline{Q}}} [u_i, u_j \in V_{t'}] \\
& = \mathbf{E}_{(U, W) \sim \omega} \left( \sum_{t=1}^m p_t \mathbf{1}[\pi_t(u_i) = w_i] \right) \\
& \quad - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{U_{\overline{Q}} \sim \omega'_{\overline{Q}} \\ W_{\overline{Q}} \sim V_{\overline{Q}}}} \left( \prod_{i \in \overline{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|\overline{Q}|} \cdot \frac{2k}{\Delta \sqrt{n}}
\end{aligned}$$

$$\begin{aligned}
& - \sum_{\emptyset \neq Q \subseteq [k]} \mathbf{E}_{\substack{U \sim \omega' \\ W_{\overline{Q}} \sim V_{\overline{Q}}}} \left( \prod_{i \in \overline{Q}} \gamma_{u_i, w_i} \right) (p_1 + p_2)^{|Q|} \frac{1}{\Delta} \sum_{\substack{i, j \in Q \\ i \neq j}} \mathbf{E}_{(V_1 \cup \dots \cup V_i) \sim \mathcal{V}} \sum_{t' \in [t]} \mathbf{1}[u_i, u_j \in V_{t'}] \\
& \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{1}{\Delta} \left( \frac{2k}{\sqrt{n}} + \frac{k^2}{\Delta' \sqrt{n}} \right) \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta \Delta' \sqrt{n}}. \\
& \quad \text{(by well-spreadness of } \omega' \text{ and the maximum size of } |V_i| \text{)}
\end{aligned}$$

In all, we have proved that  $\mathbf{E}_{\mathcal{E} \sim \tilde{\mathcal{E}}} \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{E}) \geq \mathbf{val}_{\mathcal{H}}(\mathcal{I}, \mathcal{D}) - (p_1 + p_2) \frac{2k^2}{\Delta \Delta' \sqrt{n}}$ . Since  $\tilde{\mathcal{E}}$  can be sampled in polynomial time, there is a randomized polynomial-time algorithm to find out a  $\mathcal{E}$  satisfying (13).

## 8 Bisection Max- $k$ -CSP

In this section, we consider the bisection Max-CSP as a notable example of Max-CSP with globally cardinality constraints.

We start with definitions. Fix a finite domain  $D$  and a  $k$ -CSP instance  $\mathcal{I}$  over  $D$ . A global cardinality constraint is a linear constraint on the numbers of variables that are assigned to the values in  $D$ . For simplicity and illustration purpose, here we only consider the bisection constraint – i.e., assuming  $D = \{0, 1\}$ , the number of variables that take value 1 is exactly  $n/2$  (for even integers  $n$ ). For a bisection  $k$ -CSP instance  $\mathcal{I} = (V, \omega, P)$ , we define its *optimal value* to be

$$\mathbf{opt}(\mathcal{I}) = \max_{\alpha: \{v \in V: \alpha(v)=1\} = n/2} \{\mathbf{val}(\mathcal{I}, \alpha)\},$$

where the definition of  $\mathbf{val}(\mathcal{I}, \alpha)$  remains the same as in the ordinary  $k$ -CSP case.

The  $\ell$ -round Sherali-Adams relaxation for a bisection  $k$ -CSP instance  $\mathcal{I} = (V, \omega, P)$  ( $\ell \geq k$ ) is written as follows.

$$\begin{aligned}
& \text{maximize } \mathbf{E}_{S \sim \omega} \mathbf{E}_{\alpha \sim \mu_S} [P_S(\alpha)] \\
& \text{subject to } \Pr_{\alpha \sim \mu_S} [\alpha|_{S \cap T} = \beta] = \Pr_{\alpha \sim \mu_T} [\alpha|_{S \cap T} = \beta] \\
& \quad \forall S, T \subseteq V, |S \cup T| \leq \ell, \beta \in D^{S \cap T} \\
& \quad \sum_{v \in V} \Pr_{\alpha \sim \mu_{S \cup \{v\}}} [\alpha|_S = \beta \text{ and } \alpha(v) = 1] = \frac{n}{2} \cdot \mu_S(\beta) \\
& \quad \forall S \subseteq V, S < \ell, \beta \in \{0, 1\}^S,
\end{aligned}$$

where the last constraint corresponds to the bisection constraint.

We now turn to how to round  $\kappa$ -independent solutions. The following lemma is similar to Lemma 8.

**Lemma 17** *Let  $\mathcal{I} = (V, \omega, P)$  be a bisection  $k$ -CSP instance. Let  $\mu$  be an  $\kappa$ -independent solution (with respect to both uniform distribution and  $\omega$ ,  $0 \leq \kappa \leq 1$ ) to the  $k$ -round Sherali-Adams LP relaxation. There is a randomized polynomial time algorithm to find an assignment  $\alpha: V \rightarrow \{0, 1\}$  such that  $\mathbf{val}(\mathcal{I}, \alpha) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - 3k\kappa^{1/4}$  and  $|\{v \in V: \alpha(v) = 1\}| = n/2$ .*



*Proof* We sample  $\alpha$  in the same way as we did in the proof of Lemma 8, and we see that  $\mathbf{E}_\alpha[\mathbf{val}(\mathcal{I}, \alpha)] \geq \mathbf{val}_{\text{LP}}(\mathcal{I}) - 2\sqrt{\kappa}$ . Also observe that

$$\begin{aligned} \mathbf{E}_\alpha \left| \sum_{v \in V} \alpha(v) - \frac{n}{2} \right| &\leq \sqrt{\mathbf{E}_\alpha \left( \sum_{v \in V} \alpha(v) - \frac{n}{2} \right)^2} \\ &= \sqrt{\sum_{v_1, v_2 \in V} \mathbf{E}_\alpha[\alpha(v_1)\alpha(v_2)] - n \sum_{v \in V} \mathbf{E}_\alpha[\alpha(v)] + \frac{n^2}{4}} = \sqrt{\sum_{v_1, v_2 \in V} \mathbf{E}_\alpha[\alpha(v_1)\alpha(v_2)] - \frac{n^2}{4}} \\ &\leq \sqrt{\sum_{v_1, v_2 \in V} \Pr_{\beta \sim \mu_{\{v_1, v_2\}}} [\beta(v_1) = \beta(v_2) = 1]} + \sqrt{\kappa} - \frac{n^2}{4}} = \kappa^{1/4}, \end{aligned}$$

where the last inequality is because of  $\kappa$ -independence with respect to uniform distribution, the definition of total correlation, and Lemma 1; the last equality is because of Sherali-Adams constraints.

In all, we have

$$\begin{aligned} \mathbf{E}_\alpha \left( \mathbf{val}(\mathcal{I}, \alpha) - k \left| \sum_{v \in V} \alpha(v) - \frac{n}{2} \right| \right) &\geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - 2\sqrt{\kappa} - k\kappa^{1/4} \\ &\geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - 2k\kappa^{1/4}. \end{aligned}$$

We can sample an  $\alpha$  in expected polynomial time so that

$$\mathbf{E}_\alpha \left[ \mathbf{val}(\mathcal{I}, \alpha) - \left| \sum_{v \in V} \alpha(v) - \frac{n}{2} \right| \right] \geq \mathbf{val}_{\text{LP}}(\mathcal{I}) - 3k\kappa^{1/4}.$$

By greedily rearranging  $|\mathbf{E}_{v \in V}[\alpha(v)] - \frac{1}{2}|$ -fraction of the entries in  $\alpha$ , we get a bisection assignment  $\alpha'$  such that  $\mathbf{val}(\mathcal{I}, \alpha') \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \mu) - 3k\kappa^{1/4}$ .

Finally as a counterpart to Theorem 5, we show the following.

**Theorem 8** *Let  $\mathcal{I} = (V, \omega, P)$  be a bisection  $k$ -CSP instance over domain  $\{0, 1\}$  such that  $\omega$  is  $\Delta$ -dense or  $\Delta$ -locally dense. For any  $\epsilon > 0$ , let  $\ell = \frac{2 \cdot 3^4 k^6 4^k \log |D|}{\epsilon^4 \Delta}$ . The additive integrality gaps of the  $(\ell + k)$ -round Sherali-Adams LP relaxation is at most  $\epsilon$ ; and there is a randomized rounding algorithm producing a solution whose value is at least  $\mathbf{opt}(\mathcal{I}) - \epsilon$ , in expected  $n^{O(\ell)}$  time.*

*Proof* Let  $\mu$  be a solution to the  $(\ell + k)$ -round Sherali Adams LP relaxation. Similar as in the proof of Theorem 5, Corollary 1 and Corollary 2, we know that there exists  $t \leq \ell$  such that

$$\begin{aligned} &\mathbf{E}_{U \sim V^t} \sqrt[4]{\mathbf{E}_{S \sim V^k} C_\mu(\mathbf{x}_S | \mathbf{x}_U) + \mathbf{E}_{S \sim \omega} C_\mu(\mathbf{x}_S | \mathbf{x}_U)} \\ &\leq \sqrt[4]{\mathbf{E}_{U \sim V^t} \left( \mathbf{E}_{S \sim V^k} C_\mu(\mathbf{x}_S | \mathbf{x}_U) + \mathbf{E}_{S \sim \omega} C_\mu(\mathbf{x}_S | \mathbf{x}_U) \right)} \\ &\leq \sqrt[4]{2 \cdot \frac{k^2 4^k \log |D|}{\Delta \ell}} \leq \frac{\epsilon}{3k^2}. \end{aligned}$$

Therefore, together with Lemma 5, we have

$$\begin{aligned} \mathbf{E}_{U \sim V^t} \left( \mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu} | \mathbf{x}_U) - 3k \left( \sqrt[4]{\mathbf{E}_{S \sim V^k} C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)} + \sqrt[4]{\mathbf{E}_{S \sim \omega} C_{\boldsymbol{\mu}}(\mathbf{x}_S | \mathbf{x}_U)} \right) \right) \\ \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - \epsilon. \end{aligned}$$

We enumerate all the possible ways of conditioning, and find out a solution  $\boldsymbol{\mu}'$  to the  $(k + \ell - t)$ -round Sherali-Adams LP relaxation such that

$$\mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}') - 3k \left( \sqrt[4]{\mathbf{E}_{S \sim V^k} C_{\boldsymbol{\mu}'}(\mathbf{x}_S)} + \sqrt[4]{\mathbf{E}_{S \sim \omega} C_{\boldsymbol{\mu}'}(\mathbf{x}_S)} \right) \geq \mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - \epsilon.$$

Since  $\boldsymbol{\mu}'$  is always  $\kappa$ -independent with respect to both uniform distribution and  $\omega$  for  $\kappa = \mathbf{E}_{S \sim V^k} C_{\boldsymbol{\mu}'}(\mathbf{x}_S) + \mathbf{E}_{S \sim \omega} C_{\boldsymbol{\mu}'}(\mathbf{x}_S)$ , by Lemma 17, given  $\boldsymbol{\mu}'$ , we can find an assignment with value at least  $\mathbf{val}_{\text{LP}}(\mathcal{I}, \boldsymbol{\mu}) - \epsilon$  in randomized polynomial time.

## 9 A gap instance for relaxation $\mathcal{H}$

In this section, we show a gap instance for the relaxation  $\mathcal{H}$ . Consider the following 2-AP instance  $\mathcal{I}([5], \omega)$ . Let us define  $\omega_{i,j,p,q} = \frac{1}{64} A_{i,j} B_{p,q}$ , where

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

If we view  $A$  and  $B$  as the adjacency matrices of two 5-vertex graphs,  $\mathbf{val}(\mathcal{I}, \pi)$  is the number of edges in  $A$  that are mapped to an edge in  $B$  by  $\pi$ , multiplied by  $\frac{25}{32}$ . Since  $A$  is a 4-cycle with one isolated vertex, and  $B$  is a 3-cycle plus an edge, at most 2 edges in  $A$  can be mapped to  $B$ . Therefore,  $\mathbf{opt}(\mathcal{I}) = \frac{25}{16}$ .

On the other hand, let us consider the following distribution  $\mathcal{D}$  of permutations, where  $\mathcal{D}$  is supported on  $\pi_1$  and  $\pi_2$  with equal probability (1/2).  $\pi_1$  is the identity permutation;  $\pi_2(i) = (i \bmod 5) + 1$  for all  $i \in [5]$ . We have

$$\begin{aligned} \mathbf{val}(\mathcal{I}, \mathcal{D}) &= \frac{25}{64} \cdot \sum_{i,j} \sum_{p,q} A_{i,j} B_{p,q} \frac{1}{2} (\mathbf{1}[\pi_1(i) = p \text{ or } \pi_2(i) = p]) \cdot \frac{1}{2} (\mathbf{1}[\pi_1(j) = q \text{ or } \pi_2(j) = q]) \\ &= \frac{225}{128} > \mathbf{opt}(\mathcal{I}). \end{aligned}$$

**Acknowledgements** Y.Y. is supported by JSPS Grant-in-Aid for Research Activity Start-up (24800082), MEXT Grant-in-Aid for Scientific Research on Innovative Areas (24106001), and JST, ERATO, Kawarabayashi Large Graph Project. Y. Z. is supported by a grant from the Simons Foundation (Award Number 252545). Part of the work has been done while visiting Toyota Technological Institute at Chicago. We would like to thank Ryan O'Donnell for valuable discussions.

## References

1. Alon, N., de la Vega, W.F., Kannan, R., Karpinski, M.: Random sampling and approximation of Max-CSPs. *Journal of computer and system sciences* **67**(2), 212–243 (2003) [2](#)
2. Arora, S., Frieze, A., Kaplan, H.: A new rounding procedure for the assignment problem with applications to dense graph arrangement problems. *Mathematical Programming* **92**(1), 1–36 (2002) [2](#), [5](#), [8](#)
3. Arora, S., Karger, D., Karpinski, M.: Polynomial time approximation schemes for dense instances of np-hard problems. *Journal of computer and system sciences* **58**(1), 193–210 (1999) [2](#), [5](#)
4. Arora, S., Lund, C., Motwani, R., Sudan, M., Szegedy, M.: Proof verification and the hardness of approximation problems. *Journal of the ACM* **45**(3), 501–555 (1998) [1](#)
5. Atserias, A., Maneva, E.: Sherali-adams relaxations and indistinguishability in counting logics. *SIAM Journal on Computing* **42**(1), 112–137 (2013) [3](#)
6. Barak, B., Raghavendra, P., Steurer, D.: Rounding semidefinite programming hierarchies via global correlation. In: *Proc. 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 472–481 (2011) [5](#)
7. Charikar, M., Makarychev, K., Makarychev, Y.: Integrality gaps for sherali-adams relaxations. In: *Proceedings of the 41st annual ACM symposium on Theory of computing*, pp. 283–292. ACM (2009) [3](#)
8. Coja-Oghlan, A., Cooper, C., Frieze, A.: An efficient sparse regularity concept. *SIAM Journal on Discrete Mathematics* **23**(4), 2000–2034 (2009) [5](#)
9. Frieze, A., Kannan, R.: The regularity lemma and approximation schemes for dense problems. In: *Proc. 37th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 12–20 (1996) [2](#)
10. Kumar, A., Manokaran, R., Tulsiani, M., Vishnoi, N.K.: On lp-based approximability for strict csp. In: *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1560–1573. SIAM (2011) [3](#)
11. Lovász, L., Schrijver, A.: Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization* **1**(2), 166–190 (1991) [2](#)
12. Raghavendra, P., Steurer, D.: How to round any CSP. In: *Proc. 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 586–594 (2009) [7](#)
13. Raghavendra, P., Tan, N.: Approximating CSPs with global cardinality constraints using sdp hierarchies. In: *Proc. 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 373–387 (2012) [5](#)
14. Schoenebeck, G.: Linear level lasserre lower bounds for certain k-CSPs. In: *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 593–602 (2008) [3](#)
15. Schoenebeck, G., Trevisan, L., Tulsiani, M.: Tight integrality gaps for lovasz-schrijver lp relaxations of vertex cover and max cut. In: *Proc. 39th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 302–310 (2007) [3](#)
16. Sherali, H.D., Adams, W.P.: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics* **3**(3), 411–430 (1990) [3](#)
17. Szemerédi, E.: Regular partitions of graphs. In: *Proc. Colloque Inter. CNRS (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.)*, pp. 399–401 (1978) [2](#)
18. Fernandez de la Vega, W., Karpinski, M., Kenyon, C.: Approximation schemes for metric bisection and partitioning. In: *Proc. 15th Annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pp. 506–515 (2004) [2](#)
19. de la Vega, W.F.: Max-Cut has a randomized approximation scheme in dense graphs. *Random Structures & Algorithms* **8**(3) (1996) [1](#)
20. de la Vega, W.F., Karpinski, M., Kannan, R., Vempala, S.: Tensor decomposition and approximation schemes for constraint satisfaction problems. In: *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 747–754 (2005) [2](#), [8](#)
21. de la Vega, W.F., Kenyon, C.: A randomized approximation scheme for metric Max-Cut. *Journal of computer and system sciences* **63**(4), 531–541 (2001) [2](#)
22. de la Vega, W.F., Kenyon-Mathieu, C.: Linear programming relaxations of maxcut. In: *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 53–61 (2007) [3](#), [5](#)