Abstract: This work considers the problem of approximating fixed predicate constraint satisfaction problems (MAX k-CSP(P)). We show that if the set of assignments accepted by P contains the support of a balanced pairwise independent distribution over the domain of the inputs, then such a problem on n variables cannot be approximated better than the trivial (random) approximation, even after augmenting the natural semidefinite relaxation with Ω(n) levels of the Sherali-Adams hierarchy.

It was recently shown [4] that under the Unique Game Conjecture, CSPs for predicates satisfying this condition cannot be approximated better than the trivial approximation. Our results can be viewed as an unconditional analogue of this result in a restricted computational model. We also introduce a new general set of techniques to define consistent “local distributions” over partial assignments to variables in the problem, which is often the crux of proving lower bounds for such hierarchies.

*Funded in part by NSERC
†Supported by the NSF grant CCF-083279

ACM Classification: F.2.2 (Nonnumerical Algorithms and Problems), F.1.3 (Complexity Measures and Classes), G.1.6 (Optimization), F.2.3 (Tradeoffs between Complexity Measures)

AMS Classification: 68Q17 (Computational difficulty of problems), 68W25 (Approximation algorithms), 68W20 (Randomized algorithms), 90C59 (Approximation methods and heuristics)

Key words and phrases: semidefinite and linear programming hierarchies, integrality gaps, constraint satisfaction

© Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, Madhur Tulsiani
© Licensed under a Creative Commons Attribution License
1 Introduction

A constraint satisfaction problem (CSP) consists of a set of constraints that seek a universal solution. In the maximization version (MAX-CSP) one tries to maximize the number of constraints that can be simultaneously satisfied. The most standard family of CSPs arises from Boolean predicates \( P \) with bounded support \( k \). In their generality, the predicates are defined over an alphabet \( \{0, 1, \ldots, q - 1\} = [q] \) and they can be thought as functions \( P: [q]^k \to \{0, 1\} \). A constraint is defined by the predicate \( P \) applied to a \( k \)-tuple of literals, \((x_1 + b_1 \mod q, \ldots, x_k + b_k \mod q)\), where \( b_i \in [q] \), and is said to be satisfied by some assignment on \((x_1, \ldots, x_k)\) if the predicate evaluates to 1.\(^1\) Given some predicate \( P \), an instance of the MAX \( k \)-CSP\((P)\) problem is a collection of constraints as above and the objective is to maximize the number of constraints that can be satisfied simultaneously. As a special case, we can obtain many well studied MAX-CSP problems, e.g. MAX \( k \)-SAT, MAX \( k \)-XOR, MAX \( k \)-LIN\(_q\)\(^2\) etc. When the predicate to be used in different constraints is not fixed we simply refer to the problem as MAX \( k \)-CSP.

The MAX \( k \)-CSP problem is NP-hard for \( k \geq 2 \), and a lot of effort has been devoted in determining the true inapproximability of the problem. In general, the inapproximability of the MAX \( k \)-CSP depends on the size of alphabet over which literals are valued. For the case of Boolean variables, Samorodnitsky and Trevisan [23] proved that the problem is hard to approximate better than a factor of \( 2^k/2^{2\sqrt{k}} \), which was improved to \( 2^k/2^{\sqrt{2k}} \) by Engebretsen and Holmerin [11]. Later Samorodnitsky and Trevisan [24] showed that it is Unique-Games-hard to approximate the same problem with factor better than \( 2^k/2^{\log k + 1} \). For the more general case of \( q \)-ary variables (MAX \( k \)-CSP\(_q\)), Guruswami and Raghavendra [15] showed a hardness ratio of \( q^k/kq^2 \) when \( q \) is a prime.

In a very general result which (assuming the Unique Games Conjecture) subsumes all the above ones, Austrin and Mossel [4] showed that if \( P: [q]^k \to \{0, 1\} \) is a predicate such that the set of accepted inputs \( P^{-1}(1) \) contains the support of a balanced pairwise independent distribution \( \mu \) on \([q]^k\), then MAX \( k \)-CSP\((P)\) is UG-hard to approximate better than a factor of \( q^k/|P^{-1}(1)| \). Considering that a random assignment satisfies \(|P^{-1}(1)|/q^k\) fraction of all the constraints, this is the strongest result one can get for such \( P \). Using appropriate choices for the predicate \( P \), this then implies hardness ratios of \( q^k/kq^2(1 + o(1)) \) for MAX \( k \)-CSP\(_q\) for general \( q \geq 2 \), \( q^k/kq(q - 1) \) when \( q \) is a prime power, and \( 2^k/(k + O(k^{0.525})) \) for \( q = 2 \).

We study the inapproximability of such a predicate \( P \) (which we call promising) in the hierarchy of semidefinite programs one obtains by applying the strengthening of Sherali and Adams to the canonical semidefinite relaxation of the problem. In particular, we show an unconditional analogue of the result of Austrin and Mossel in this hierarchy.

Hierarchies of Linear and Semidefinite Programs

A standard approach in approximating \textit{NP}-hard problems, and therefore MAX \( k \)-CSP, is to formulate the problem as a 0-1 integer program and then relax the integrality condition to get a linear (or semidefinite)
program which can be solved efficiently. The quality of such an approach is intimately related to the \textit{integrality gap} of the relaxation, namely, the ratio between the optimum of the relaxation and that of the integer program.

Several methods (or procedures) were developed in order to obtain tightenings of relaxations in a systematic manner. These procedures give a sequence or a \textit{hierarchy} of increasingly tighter relaxations of the starting program. The commonly studied ones include the hierarchies defined by Lovász-Schrijver [18], Sherali-Adams [28], and Lasserre [16] (see [17] for a comparison). Stronger relaxations in the sequence are referred to as higher \textit{levels} of the hierarchy. It is known for all these hierarchies that for a starting program with \( n \) variables, the program at level \( n \) has integrality gap 1, and that it is possible to optimize over the program at the \( r \)th level in time \( n^{O(r)} \).

Many known linear (semidefinite) programs can be captured by constant many levels of the Sherali-Adams (Lasserre) hierarchy. In fact, these semidefinite programs can also be captured by a “mixed” hierarchy, first studied by Raghavendra\(^3\) [22]; where we augment the basic semidefinite relaxation by adding new real variables and imposing linear constraints according to the Sherali-Adams hierarchy. We will refer to this hierarchy of programs as the Sherali-Adams SDP hierarchy.

Fernández de la Vega and Kenyon-Mathieu [13] have provided a PTAS for Max Cut in dense graphs using Sherali-Adams LP hierarchy. In [19] it is shown how to get a Sherali-Adams based PTAS for Vertex-Cover and Max-Independent-Set in minor-free graphs, while recently Mathieu and Sinclair [20] showed that the integrality gap for the matching polytope is asymptotically \( 1 + 1/r \), and Bateni, Charikar and Guruswami [5] that the integrality gap for a natural LP formulation of the MaxMin allocation problem is at most \( n^{1/r} \), both after \( r \) many Sherali-Adams tightenings. Chlamtac [9] and Chlamtac and Singh [10] gave an approximation algorithm for Max-Independent-Set in hypergraphs based on the Lasserre hierarchy, with the performance depending on the number of levels. Recently, an \( O(n^{1/4}) \) approximation for Densest \( k \)-Subgraph was shown by Bhaskara et. al. [6], using linear programs implied by \( O(\log n) \) levels of the Lovász-Schrijver hierarchy.

Lower bounds in these hierarchies amount to showing that the integrality gap remains large even after many levels of the hierarchy. Integrality gaps for \( \Omega(n) \) levels can be seen as unconditional lower bounds (as they rule out even exponential time algorithms obtained by the hierarchy) in a restricted (but still fairly interesting) model of computation. Considerable effort was invested in proving such lower bounds (see [2, 30, 29, 27, 7, 12, 1, 26, 14, 13]). For some CSPs in particular, strong lower bounds (\( \Omega(n) \) levels) were proved recently for the Lasserre hierarchy (which is the strongest) by [25] and [31], who showed a factor 2 integrality gap for MAX k-XOR and factor \( 2^k/2k \) integrality gap for MAX k-CSP respectively.

In a beautiful result, Raghavendra [22] showed a general connection between integrality gaps and UG-hardness results. His result essentially shows that for MAX k-CSP\((P)\), the integrality gap of a program obtained by \( k \) levels of the Sherali-Adams SDP hierarchy is \( I \), then the MAX k-CSP\((P)\) is UG-hard to approximate better than a factor of \( I \). However, in our case the hardness is already known (by the work of Austrin and Mossel), and we are interested in finding the integrality gap for programs obtained by \( \Omega(n) \) levels.

\(^3\)The hierarchies studied by Raghavendra were in fact slightly weaker than the ones defined here.
Our Result and Techniques

Both the known results in the Lasserre hierarchy (and previous analogues in the Lovász-Schrijver hierarchy) seemed to heavily rely on the structure of the predicate for which the integrality gap was proved, in particular, the predicate is always some system of linear equations. It was not clear if the techniques could be extended using only the fact that the predicate is promising (which is a much weaker condition). In this paper, we try to explore this issue, proving $\Omega(n)$ level gaps for the Sherali-Adams SDP hierarchy for any promising predicate.

**Theorem 1.1.** Let $P : [q]^k \to \{0, 1\}$ be a predicate such that $P^{-1}(1)$ contains the support of a balanced pairwise independent distribution $\mu$. Then for every constant $\zeta > 0$, there exist $c = c(q,k,\zeta) > 0$ such that for large enough $n$, the integrality gap of $\text{MAX } k\text{-CSP}(P)$ for the tightening obtained by $cn$ levels of the Sherali-Adams SDP hierarchy is at least $\frac{q^k}{|P^{-1}(1)|} - \zeta$.

**Remark 1.2.** We note that weaker integrality gaps for these predicates also follow, via reductions, from the corresponding integrality gap results for Unique Games. In particular, a $(\log \log n)^{\Omega(1)}$-level gap for the SDP hierarchy discussed above, follows from the recent results of Raghavendra and Steurer [21]. Also, $\Omega(n^\delta)$-level gaps (where $\delta \to 0$ as $\zeta \to 0$) for the Sherali-Adams LP hierarchy can be deduced from the results of Charikar, Makarychev and Makarychev [8].

A first step in achieving our result is to reduce the problem of a level-$t$ gap to a question about family of distributions over assignments associated with sets of variables of size at most $t$. These distributions should be (a) supported only on satisfying (partial) assignments, (b) should be consistent among themselves, in the sense that for $S_1 \subseteq S_2$ which are subsets of variables, the distributions over $S_1$ and $S_2$ should be equal on $S_1$, and (c) should be balanced and pairwise-independent. The first requirement guarantees that the solution achieves objective value that corresponds to satisfying all the constraints of the instance. The second requirement implies feasibility for the Sherali-Adams LP constraints, while the last one makes it easy to produce vectors satisfying the semidefinite constraints.

The second step is to come up with these distributions! We explain why the simple method of picking a uniform distribution (or a reweighting of it according to the pairwise independent distribution that is supported by $P$) over the satisfying assignments cannot work. Instead we introduce the notion of “advice sets”. These are sets on which it is “safe” to define such simple distributions. The actual distribution for a set $S$ we use is then the one induced on $S$ by a simple distribution defined on the advice-set of $S$. Getting such advice sets heavily relies on notions of expansion of the constraints graph. In particular, we use the fact that random instances have inherently good expansion properties. At the same time, such instances are highly unsatisfiable, ensuring that the resulting integrality gap is large.

Arguing that it is indeed “safe” to use simple distributions over the advice sets relies on the fact that the predicate $P$ in question is promising, namely $P^{-1}(1)$ contains the support of a balanced pairwise independent distribution. We find it interesting and somewhat curious that the condition of pairwise independence comes up in this context for a reason very different than in the case of UG-hardness. Here, it represents the limit to which the expansion properties of a random CSP instance can be pushed to define such distributions.

---

See the resulting SDP in section 2.3.
2 Preliminaries and Notation

2.1 Constraint Satisfaction Problems

For an instance $\Phi$ of MAX $k$-CSP$_q$, we denote the variables by $\{x_1, \ldots, x_n\}$, their domain $\{0, \ldots, q-1\}$ by $[q]$ and the constraints by $C_1, \ldots, C_m$. Each constraint is a function of the form $C_i : [q]^{T_i} \to \{0,1\}$ depending only on the values of the variables in the ordered tuple $T_i$ with $|T_i| \leq k$.

For a set of variables $S \subseteq [n]$, we denote by $[q]^S$ the set of all mappings from the set $S$ to $[q]$. In context of variables, these mappings can be understood as partial assignments to a given subset of variables. For $\alpha \in [q]^S$, we denote its projection to $S' \subseteq S$ as $\alpha(S')$. Also, for $\alpha_1 \in [q]^{S_1}, \alpha_2 \in [q]^{S_2}$ such that $S_1 \cap S_2 = \emptyset$, we denote by $\alpha_1 \circ \alpha_2$ the assignment over $S_1 \cup S_2$ defined by $\alpha_1$ and $\alpha_2$.

We shall prove results for constraint satisfaction problems where every constraint is specified by the same Boolean predicate $P : [q]^k \to \{0,1\}$. We denote the set of assignments for which the predicate evaluates to 1 by $P^{-1}(1)$. A CSP instance for such a problem is a collection of constraints of the form of $P$ applied to $k$-tuples of literals. For a variable $x$ with domain $[q]$, we take a literal to be $(x+a) \mod q$ for any $a \in [q]$. More formally,

**Definition 2.1.** For a given $P : [q]^k \to \{0,1\}$, an instance $\Phi$ of MAX $k$-CSP$_q(P)$ is a set of constraints $C_1, \ldots, C_m$ where each constraint $C_i$ is over a $k$-tuple of variables $T_i = \{x_{i_1}, \ldots, x_{i_k}\}$ and is of the form $P(x_{i_1} + a_{i_1}, \ldots, x_{i_k} + a_{i_k})$ for some $a_{i_1}, \ldots, a_{i_k} \in [q]$. We denote the maximum number of constraints that can be simultaneously satisfied by OPT($\Phi$).

2.2 Expanding CSP Instances

For an instance $\Phi$ of MAX $k$-CSP$_q$, define its constraint graph $G_\Phi$, as the following bipartite graph from $L$ to $R$. The left hand side $L$ consists of a vertex for each constraint $C_i$. The right hand side $R$ consists of a vertex for every variable $x_j$. There is an edge between a constraint-vertex $i$ and a variable-vertex $j$, whenever variable $x_j$ appears in constraint $C_i$. When it is clear from the context, we will abbreviate $G_\Phi$ by $G$.

For $C_i \in L$ we denote by $\Gamma(C_i) \subseteq R$ the neighbors $\Gamma(C_i)$ of $C_i$ in $R$. For a set of constraints $\mathcal{C} \subseteq L$, $\Gamma(\mathcal{C})$ denotes $\bigcup_{C_i \in \mathcal{C}} \Gamma(C_i)$. For $S \subseteq R$, we call a constraint $C_i \in L$, $S$-dominated if $\Gamma(C_i) \subseteq S$. We denote by $G|_{-S}$ the bipartite subgraph of $G$ that we get after removing $S$ and all $S$-dominated constraints. Finally, we also denote by $\mathcal{C}(S)$ the set of all $S$-dominated constraints. Similarly, for $u \in R$ we denote by $\Gamma(u) \subseteq L$ the neighbors of $u$ in $L$.

Our result relies on the expansion of the support sets of the constraints. We make this notion formal below.

**Definition 2.2.** Consider a bipartite graph $G = (V,E)$ with partition $L,R$. The boundary expansion of $X \subseteq L$ is the value $|\partial X| / |X|$, where $\partial X = \{u \in R : |\Gamma(u) \cap X| = 1\}$. $G$ is $(r,e)$ boundary expanding if the boundary expansion for all (nonempty) subsets of $L$ of size at most $r$ is at least $e$.

2.3 The Sherali-Adams SDP Hierarchy

Below we present a relaxation for the MAX $k$-CSP$_q$ problem as it is obtained by applying a level-$t$ Sherali-Adams tightening to the basic SDP formulation of some instance $\Phi$ of MAX $k$-CSP$_q$. A well
known fact states that the level-n Sherali-Adams tightening (even starting from a linear program) provides a perfect formulation, i.e. the integrality gap is 1 (see [28] or [17] for a proof).

The intuition behind the level-t Sherali-Adams tightening is the following. Note that an integer solution to the problem can be given by a single mapping \( \alpha_0 \in [q]^{|n|} \), which is an assignment to all the variables. Using this, we can define 0/1 variables \( X_{(S, \alpha)} \) for each \( S \subseteq [n] \) such that \(|S| \leq t\) and \( \alpha \in [q]^{|S|} \). The intended solution is \( X_{(S, \alpha)} = 1 \) if \( \alpha_0(S) = \alpha \) and 0 otherwise. We introduce \( X_{(\emptyset, \emptyset)} \) which is intended to be 1. By relaxing the integrality constraint on the variables, we obtain the LP conditions given by level-t of the Sherali-Adams hierarchy.

We can further strengthen the integer program by adding the quadratic constraints \( X_{(i_1, i_2, j_1, j_2)} = X_{(i_1, j_1)} \cdot X_{(i_2, j_2)} \) for \( i_1, i_2 \in [n] \) and \( j_1, j_2 \in [\ell] \).

As solving quadratic programs is NP-hard we then relax these quadratic constraints to the existence of vectors \( v_{(i,j)} \) and a unit vector \( v_0 \), for which we require that:

\[
\langle v_{(i,j)}, v_{(i_2, j_2)} \rangle = X_{(i_1, j_1, j_2, j_2)} \quad \text{and} \quad \langle v_{(i,j)}, v_0 \rangle = X_{(i_1, j_1)}.
\]

The complete relaxation can be seen in Figure 1.

\[
\begin{align*}
\text{maximize} & \quad \sum_{i=1}^{m} \sum_{\alpha \in [q]^{|T|}} C_i(\alpha) \cdot X_{(T, \alpha)} \\
\text{subject to} & \quad \langle v_{(i_1, j_1)}, v_{(i_2, j_2)} \rangle = X_{(i_1, j_1, j_2, j_2)} \quad \forall i_1 \neq i_2 \in [n], j_1, j_2 \in [q] \\
& \quad \langle v_{(i,j)}, v_{(i_2, j_2)} \rangle = 0 \quad \forall i \in [n], j_1 \neq j_2 \in [q] \\
& \quad \|v_{(i,j)}\|^2 = \langle v_{(i,j)}, v_0 \rangle = X_{(i,j)} \quad \forall i \in [n], j \in [q] \\
& \quad \|v_0\|^2 = X_{(\emptyset, \emptyset)} = 1 \\
& \quad \sum_{j \in [q]} X_{(\emptyset, \emptyset, j, j)} = X_{(\emptyset, \emptyset)} \quad \forall \alpha \in [q]^{|S|} \\
& \quad \forall S s.t. |S| < t, \forall \alpha \in [q]^{|S|} \\
& \quad X_{(S, \alpha)} \geq 0 \quad \forall S s.t. |S| > t, \forall \alpha \in [q]^{|S|} \\
\end{align*}
\]

Figure 1: SDP for MAX k-CSP \( q \) augmented with Level-t Sherali-Adams constraints

For an SDP formulation of MAX k-CSP \( q \), and for a given instance \( \Phi \) of the problem, we denote by FRAC(\( \Phi \)) the SDP (fractional) optimum, and by OPT(\( \Phi \)) the integral optimum. For the particular instance \( \Phi \), the integrality gap is then defined as FRAC(\( \Phi \))/OPT(\( \Phi \)). The integrality gap of the formulation is the supremum of integrality gaps over all instances.

Next we give a sufficient condition for the existence of a solution to the program obtained by level-t of the Sherali-Adams SDP hierarchy for a MAX k-CSP \( q \) instance \( \Phi \).

**Lemma 2.3.** Consider a family of distributions \( \{D(S)\} \subseteq [n] \times [q] \leq t \), where each \( D(S) \) is defined over \([q]^{|S|}\). Suppose that for every \( S \subseteq T \subseteq [n] \) with \(|T| \leq t\), the distributions \( D(S), D(T) \) are equal on \( S \) and there exists a set of vectors \( \{v_{(i,j)}\} \subseteq [n] \times [q] \) and a unit vector \( v_0 \) satisfying

1. For all \( i \in [n] \) and \( j_1 \neq j_2 \), \( \langle v_{(i,j_1)}, v_{(i,j_2)} \rangle = 0 \).

2. For all \( i \in [n], j \in [q] \), \( \langle v_{(i,j)}, v_0 \rangle = \|v_{(i,j)}\|^2 = \text{Pr}_{D(I)}(j) \).

\( ^5 \)abusing notation slightly, here we use \( X_{(i_1, i_2, j_1, j_2)} \) to denote \( X_{(i_1, i_2, \alpha)} \) where \( \alpha \) is the function mapping \( i_1 \) to \( j_1 \) and \( i_2 \) to \( j_2 \).
Then the vectors together with the LP variables \( X_{(S, \alpha)} = \Pr_{\mathcal{D}(S)}[\alpha] \) form a feasible solution to the program in Figure 1.

**Proof.** Consider some \( S \subseteq [n] \), \(|S| < t\), and some \( i \notin S\). Note that the distributions \( \mathcal{D}(S), \mathcal{D}(S \cup \{i\}) \) are equal on \( S \), and therefore we have

\[
\sum_{j \in [q]} X_{(S,j,1)} = \sum_{j \in [q]} \Pr_{\beta \sim \mathcal{D}(S \cup \{i\})}[\beta = \alpha \land (\beta(S) = \alpha)]
\]

The same argument for \( S = \emptyset \) shows that \( X_{(\emptyset, \emptyset)} = 1 \). It is clear that the solution satisfies all the other required conditions by definition, which proves the lemma.

### 2.4 Pairwise Independence and Approximation Resistant Predicates

We say that a distribution \( \mu \) over variables \( x_1, \ldots, x_k \), is a balanced pairwise independent distribution over \( [q]^k \), if we have

\[
\forall j \in [q]. \forall i. \Pr_{\mu}[x_i = j] = \frac{1}{q} \quad \text{and} \quad \forall j_1, j_2 \in [q] \forall i_1 \neq i_2. \Pr_{\mu}[(x_i = j_1) \land (x_i = j_2)] = \frac{1}{q^2}.
\]

A predicate \( P \) is called approximation resistant if it is hard to approximate the \( \text{MAX k-CSP}_q(P) \) problem better than using a random assignment. Assuming the Unique Games Conjecture, Austrin and Mossel [4] show that a predicate is approximation resistant if it is possible to define a balanced pairwise independent distribution \( \mu \) such that \( P \) is always 1 on the support of \( \mu \).

**Definition 2.4.** A predicate \( P : [q]^k \rightarrow \{0, 1\} \) is called promising, if there exist a distribution supported over a subset of \( P^{-1}(1) \) that is pairwise independent and balanced. If \( \mu \) is such a distribution we say that \( P \) is promising supported by \( \mu \).

### 3 Towards Defining Consistent Distributions

To construct valid solutions for the Sherali-Adams SDP hierarchy, we need to define distributions over every set \( S \) of bounded size as is required by Lemma 2.3. Since we will deal with promising predicates supported by some distribution \( \mu \), in order to satisfy consistency between distributions we will heavily rely on the fact that \( \mu \) is a balanced pairwise independent distribution.

Consider for simplicity that \( \mu \) is uniform over \( P^{-1}(1) \) (the intuition for the general case is not significantly different). It is instructive to think of \( q = 2 \) and the predicate \( P \) being \( k\text{-XOR}, k \geq 3 \). Observe
that the uniform distribution over $P^{-1}(1)$ is pairwise independent and balanced. A first attempt would be to define for every $S$, the distribution $\mathcal{D}(S)$ as the uniform distribution over all consistent assignments of $S$. We argue that such distributions are in general problematic. This follows from the fact that satisfying assignments are not always extendible. Indeed, consider two constraints $C_i, C_j \in L$ that share a common variable $j \in R$. Set $S_2 = T_i \cup T_j$, and $S_1 = S_2 \setminus \{j\}$. Assuming that the support of no other constraint is contained in $S_2$, we get that distribution $\mathcal{D}(S_1)$ maps any variable in $S_1$ to $\{0,1\}$ with probability $1/2$ independently, but some of these assignments are not even extendible to $S_2$ meaning that $\mathcal{D}(S_2)$ will assign them with probability zero.

Thus, to define $\mathcal{D}(S)$, we cannot simply sample assignments satisfying all constraints in $\mathcal{C}(S)$ with probabilities given by $\mu$. In fact the above example shows that any attempt to blindly assign a set $S$ with a distribution that is supported on all satisfying assignments for $S$ is bound to fail. At the same time it seems hard to reason about a distribution that uses a totally different concept. To overcome this obstacle, we take a two step approach:

1. For a set $S$ we define a superset $\overline{S}$ such that $\overline{S}$ is “global enough” to contain sufficient information, while it also is “local enough” so that $\mathcal{C}(\overline{S})$ is not too large. We require the property of such sets that if we remove $\overline{S}$ and $\mathcal{C}(\overline{S})$, then the remaining graph $G|_{\overline{S}}$ still has good expansion. We deal with this in Section 3.1.

2. When $\mu$ is the uniform distribution over $P^{-1}(1)$, the distribution $\mathcal{D}(S)$ is going to be the uniform distribution over satisfying assignments in $\overline{S}$. In the case that $\mu$ is not uniform over $P^{-1}(1)$, we give a natural generalization to the above uniformity. We show how to define distributions, which we denote by $\mathcal{P}_\mu(S)$, such that for $S_1 \subseteq S_2$, the distributions $\mathcal{P}_\mu(S_1)$ and $\mathcal{P}_\mu(S_2)$ are guaranteed to be consistent if $G|_{\overline{S}_1}$ has good expansion. This appears in Section 3.2.

We then combine the two techniques and define $\mathcal{D}(S)$ according to $\mathcal{P}_\mu(\overline{S})$. This is done in Section 4.

### 3.1 Finding Advice-Sets

We now give an algorithm below to obtain a superset $\overline{S}$ for a given set $S$, which we call the advice-set of $S$. It is inspired by the “expansion correction” procedure in [7].

**Theorem 3.1.** Algorithm Advice, when run with inputs $G, e_1, e_2, r$, and $S$, returns $\overline{S} \subseteq R$ such that (a) $G|_{\overline{S}}$ is $(\xi_\overline{S}, e_2)$ boundary expanding, (b) $\xi_\overline{S} \geq r - \frac{|S|}{e_1-e_2}$, and (c) $|\overline{S}| \leq \frac{(2e_1+k-e_2)|S|}{2(e_1-e_2)}$.

**Proof.** Let $\xi_\overline{S}$ be the value of $\xi$ when the loop terminates. From the bounded size of $M_j$ and how $\xi$ changes at each iteration we know that $\xi$ remains non-negative throughout the execution of the while loop, and in particular $\xi_\overline{S} \geq 0$. Note that at step $j$, all the neighbors of $M_j$ are added to the set $\overline{S}$ so no member of $M_j$ will be in $G|_{\overline{S}}$ after the $j$th step. In particular, all the sets $M_j$ will be disjoint.

In order to prove (a) we will prove the following loop invariant: $G|_{\overline{S}}$ is $(\xi, e_2)$ boundary expanding. Indeed, note that the input graph $G$ is $(\xi, e_1)$ boundary expanding so the invariant holds at the beginning. At step $j$ consider the set $\overline{S} \cup \{x_j\}$, and suppose that $G|_{(\overline{S}\cup\{x_j\})}$ is not $(\xi, e_2)$ boundary expanding. We find maximal $M_j$, $|M_j| \leq \xi$, such that $|\partial M_j| \leq e_2|M_j|$. We claim that $G_{(\overline{S}\cup\{x_j\},-\Gamma(M_j))}$ is $(\xi - |M_j|, e_2)$

---

$^6$Remember that $T_i$ is the set of variables appearing in $C_i$ and $T_2$ is the set of variables appearing in $C_2$. 

**Theory of Computing**
We now define for every set $S$, a distribution $\mathcal{P}_\mu(S)$ such that for any $\alpha \in [q]^S$, $\Pr_{\mathcal{P}_\mu(S)}[\alpha] > 0$ only if $\alpha$ satisfies all the constraints in $\mathcal{C}(S)$. For a constraint $C_i$ with set of inputs $T_i$, defined as $C_i(x_{i_1}, \ldots, x_{i_k}) \equiv$


\[ P(x_i + a_i, \ldots, x_k + a_k), \] let \( \mu : [q]^T \rightarrow [0, 1] \) denote the distribution

\[ \mu_i(x_i, \ldots, x_k) = \mu(x_i + a_i, \ldots, x_k + a_k) \]

so that the support of \( \mu_i \) is contained in \( C_i^{-1}(1) \). We then define the distribution \( \mathcal{P}_\mu(S) \) by picking each assignment \( \alpha \in [q]^S \) with probability proportional to \( \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i)) \). Formally,

\[ \Pr_{\mathcal{P}_\mu(S)}[\alpha] = \frac{1}{Z_S} \cdot \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i)) \] \hspace{1cm} (3.1)

where \( \alpha(T_i) \) is the restriction of \( \alpha \) to \( T_i \) and \( Z_S \) is a normalization factor given by

\[ Z_S = \sum_{\alpha \in [q]^S} \prod_{C_i \in \mathcal{C}(S)} \mu_i(\alpha(T_i)). \]

To understand the distribution, it is easier to think of the special case when \( \mu \) is just the uniform distribution on \( P^{-1}(1) \) (like in the case of MAX k-XOR). Then \( \mathcal{P}_\mu(S) \) is simply the uniform distribution on assignments satisfying all the constraints in \( \mathcal{C}(S) \). When \( \mu \) is not uniform, then the probabilities are weighted by the product of the values \( \mu_i(\alpha(T_i)) \) for all the constraints \(^7\). However, we still have the property that if \( \Pr_{\mathcal{P}_\mu(S)}[\alpha] > 0 \), then \( \alpha \) satisfies all the constraints in \( \mathcal{C}(S) \).

In order for the distribution \( \mathcal{P}_\mu(S) \) to be well defined, we need to ensure that \( Z_S > 0 \). The following lemma shows how to calculate \( Z_S \) if \( G \) is sufficiently expanding, and simultaneously proves that if \( S_1 \subseteq S_2 \), and if \( G|_{-S_1} \) is sufficiently expanding, then \( \mathcal{P}_\mu(S_1) \) is consistent with \( \mathcal{P}_\mu(S_2) \) over \( S_1 \).

**Lemma 3.2.** Let \( \Phi \) be a MAX k-CSP(P) instance as above and \( S_1 \subseteq S_2 \) be two sets of variables such that both \( G \) and \( G|_{-S_1} \) are \((r, k - 3 + \delta)\) boundary expanding for some \( \delta > 0 \) and \(|\mathcal{C}(S_2)| \leq r \). Then

\[ Z_{S_2} = \frac{q^{|S_2|}}{q^{|\mathcal{C}(S_2)|}}, \]

and for any \( \alpha_1 \in [q]^{|S_1|} \)

\[ \sum_{\alpha_2 \in [q]^{|S_2|}} \Pr_{\mathcal{P}_\mu(S_2)}[\alpha_2] = \Pr_{\mathcal{P}_\mu(S_1)}[\alpha_1]. \] \hspace{1cm} (3.2)

**Proof.** Let \( \mathcal{C} = \mathcal{C}(S_2) \setminus \mathcal{C}(S_1) \) be the set of \( t \) constraints dominated by \( S_2 \) but not \( S_1 \). Without loss of generality let \( C = \{C_1, \ldots, C_t\} \) with \( C_t \) being on the set of variables \( T_i \) some of which might be set by \( \alpha_1 \). Note that any \( \alpha_2 \) consistent with \( \alpha_1 \) can be written as \( \alpha \circ \alpha_1 \) for some \( \alpha \in [q]^{|S_1|} \). We will show a way to calculate a sum similar to the left hand side of (3.2). Note that these calculations are meaningful even if \( Z_{S_1} \) or \( Z_{S_2} \) are zero, in which case both sides are simply zero. Taking \( S_1 = \emptyset \) will then give us the value of \( Z_{S_2} \).

\[ Z_{S_2} \cdot \sum_{\alpha_2 \in [q]^{|S_2|}} \Pr_{\mathcal{P}_\mu(S_2)}[\alpha_2] = \sum_{\alpha \in [q]^{|S_1|}, \alpha_1 \subseteq \alpha} \prod_{C_i \in \mathcal{C}(S_2)} \mu_i((\alpha_1 \circ \alpha)(T_i)) \]

\[ = \left( \prod_{C_i \in \mathcal{C}(S_1)} \mu_i(\alpha_1(T_i)) \right) \sum_{\alpha \in [q]^{|S_1|}} \prod_{j=1}^t \mu_j((\alpha \circ \alpha)(T_j)) \]

\[ = (Z_{S_1} \cdot \Pr_{\mathcal{P}_\mu(S_1)}[\alpha_1]) \sum_{\alpha \in [q]^{|S_1|}} \prod_{j=1}^t \mu_j((\alpha \circ \alpha)(T_j)) \]

\(^7\)Note however that \( \mathcal{P}_\mu(S) \) is not a product distribution because different constraints in \( \mathcal{C}(S) \) may share variables.
The following claim lets us calculate this last sum conveniently using the expansion of $G|_{\neg S_1}$.

**Claim 3.3.** Let $\mathcal{C}$ be as above. Then there exists an ordering $C_{i_1}, \ldots, C_{i_t}$ of constraints in $\mathcal{C}$ and a partition of $S_2 \setminus S_1$ into sets of variables $F_1, \ldots, F_t$ and $F_{t+1}$ such that for all $j \leq t$, $F_j \subseteq T_{i_j}$, $|F_j| \geq k-2$, and

$$\forall j \ F_j \cap \left( \bigcup_{l > j} T_{i_l} \right) = \emptyset.$$ 

**Proof.** (of Claim 3.3) We build the sets $F_j$ inductively using the fact that $G|_{\neg S_1}$ is $(r, k-3 + \delta)$ boundary expanding.

Start with the set of constraints $\mathcal{C}_1 = \mathcal{C}$. Since $|\mathcal{C}_1| = |\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)| \leq r$, $\mathcal{C}_1$ should be expanding in $G|_{\neg S_1}$ i.e. $|\partial(\mathcal{C}_1) \setminus S_1| \geq (k-3 + \delta)|\mathcal{C}_1|$. Hence, there exists $C_j \in \mathcal{C}_1$ contributing at least $k-2$ variables to the boundary of $\mathcal{C}_1$, i.e. $|T_j \cap (\partial(\mathcal{C}_1) \setminus S_1)| \geq k-2$. Let $T_j \cap (\partial(\mathcal{C}_1) \setminus S_1) = F_1$ and $i_1 = j$. We then take $\mathcal{C}_2 = \mathcal{C}_1 \setminus \{C_{i_1}\}$ and continue in the same way. What is left from $S_2 \setminus S_1$ after the last step will be $F_{t+1}$. It is not hard to check that $F_{t+1}$ is the portion of $S_2 \setminus S_1$ that is not in any $\mathcal{C}_j$ although we will not use this.

Since at every step, we have $F_j \subseteq \partial(\mathcal{C}_{i_j}) \setminus S_1$, and for all $l > j$, $\mathcal{C}_l \subseteq \mathcal{C}_{i_j}$, $F_j$ shares no variables with $\Gamma(\mathcal{C}_l)$ for $l > j$. Hence, we get $F_j \cap \left( \bigcup_{l > j} T_{i_l} \right) = \emptyset$ as claimed.

Using this decomposition, we can reorder the sum and split it as,

$$\sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{j=1}^t \mu_j((\alpha_1 \circ \alpha(T_j))) = \sum_{\beta_{i+1} \in [q]^{S_1}} \sum_{\beta_i \in [q]^{S_1}} \mu_i \sum_{\beta_{i-1} \in [q]^{S_1}} \mu_{i-1} \cdots \sum_{\beta_2 \in [q]^{S_1}} \mu_2 \sum_{\beta_1 \in [q]^{S_1}} \mu_1,$$

where the input to each $\mu_i$ depends on $\alpha_1$ and $\beta_1, \ldots, \beta_{i+1}$ but not on $\beta_1, \ldots, \beta_{i-1}$.

We now reduce the expression from right to left. Since $F_1$ contains at least $k-2$ variables and $\mu_i$ is a balanced pairwise independent distribution,

$$\sum_{\beta_1 \in [q]^{S_1}} \mu_1 = \Pr_{\mu}[(\alpha_1 \circ \beta_2 \circ \cdots \circ \beta_t)(T_i \setminus F_1)] = \frac{1}{q^{k-|F_1|}},$$

irrespective of the values assigned by $\alpha_1 \circ \beta_2 \circ \cdots \circ \beta_t$ to the remaining (at most 2) variables in $T_i \setminus F_1$. Continuing in this fashion from right to left, we get that

$$\sum_{\alpha \in [q]^{S_2 \setminus S_1}} \prod_{j=1}^t \mu_j((\alpha_1 \circ \alpha)(T_j))) = \frac{1}{q^{k-\sum_{j=1}^t |F_j|}} = q^{(|S_2 \setminus S_1| - k|\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)|)},$$

Hence, we get that

$$Z_{S_2} \cdot \sum_{\alpha_1 \in [q]^{S_2} \ \text{ s.t. } \alpha_1(S_1) = \alpha_1} \Pr_{\mu}(S_2) [\alpha_2] = \left( Z_{S_1} \cdot q^{(\frac{|S_2 \setminus S_1|}{|\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)|}} \right) \Pr_{\mu}(S_1) [\alpha_1]. \quad (3.3)$$

Now, since we know that $G$ is $(r, k-3 + \delta)$-boundary expanding, we can replace $S_1$ by $\emptyset$ in the above calculation to get,

$$Z_{S_2} = \sum_{\alpha \in [q]^{S_2} \ \text{ s.t. } \alpha(S_2) = \alpha} \mu((\alpha_1 \circ \alpha)(T_i)) = q^{(|S_2| - k|\mathcal{C}(S_2)|)},$$

**Theory of Computing**
We now show how to construct integrality gaps using the ideas in the previous section. For a given instance of CSPs are both highly unsatisfiable and highly expanding. We capture the properties we need in various works on integrality gaps and proof complexity (e.g. [7], [1], [26] and [25]), that random problem, conditioned on no two constraints sharing more than one variable. To generate a random promising predicate for which the theorem is trivial, so one can assume that \( k = 2 \) the only promising predicate is the one accepting everything for which the theorem is trivial, so one can assume that \( k > 2 \) and \( C(\{ i, j \}) \) is empty. It follows that \( \mathcal{P}_\mu(\{ i, j \}) \) is the uniform distribution on \([q]^{\{ i, j \}}\) which completes the proof.

\[ \sum_{\alpha_2(S_2) \in S_2} \Pr_{\mathcal{P}_\mu(S_2)}[\alpha_2] = \Pr_{\mathcal{P}_\mu(S_1)}[\alpha_1], \]

which proves the lemma.

We are now almost ready to describe our Sherali-Adams SDP solution. The only other property of \( \mathcal{P}_\mu(S) \) that we need to show is that it is pairwise-independent and balanced.

**Claim 3.4.** Let \( G \) be a \((r, k - 3 + \varepsilon)\)-boundary expanding constraint graph, with \( \varepsilon > 2/3 \) in which no two constraints share more than one variable. Then for any \( S \subset R \), \( |S| \leq r \), the distribution \( \mathcal{P}_\mu(S) \) is a pairwise-independent and balanced distribution. That is, for all \( i_1, i_2 \in S \), \( j_1, j_2 \in [q] \),

\[ \Pr_{\alpha \sim \mathcal{P}_\mu(S)}[\alpha_{i_1} = j_1] = 1/q, \quad \Pr_{\alpha \sim \mathcal{P}_\mu(S)}[(\alpha_{i_1} = j_1) \land (\alpha_{i_2} = j_2)] = 1/q^2. \]

**Proof.** Let \( S \) be a subset of the variables, \( i, j \in S \) and \( \varepsilon' = \min(1, \varepsilon - 2/3) \). We will first show that \( G_{-\{i, j\}} \) is \((r, k - 3 + \varepsilon')\) boundary expanding.

Let \( C \) be a set of constraints with \(|C| \leq r\). When \(|C| = 1\), \( C \) has \( k \) boundary neighbors in \( G \) and hence at least \( k - 2 \geq (k - 3 + \varepsilon')|C| \) boundary neighbors in \( G_{-\{i, j\}} \). When \(|C| = 2\), the number of boundary neighbors in \( G \) must be at least \( 2(k - 2) \) since the two constraints in \( C \) can share at most one variable. It follows that \( |\partial C| \) is at least \( 2(k - 2) \geq (k - 3 + \varepsilon')|C| \) in \( G_{-\{i, j\}} \).

Finally, when \( 3 \leq |C| \leq r \), we get that the size of \( |\partial C| \) in \( G_{-\{i, j\}} \) is at least \( (k - 3 + \varepsilon)|C| - 2 \) as \( G \) is \((r, k - 3 + \varepsilon)\) boundary expanding. This proves the claim as \( (k - 3 + \varepsilon)|C| - 2 = (k - 3 + \varepsilon - 2/|C|)|C| \geq (k - 3 + \varepsilon')|C| \).

It follows from applying Lemma 3.3 with \( S_2 = S \) and \( S_1 = \{i, j\} \) that \( \mathcal{P}_\mu(S) \) agrees with \( \mathcal{P}_\mu(\{i, j\}) \) on the set \( \{i, j\} \). Now note that for \( k = 2 \) the only promising predicate is the one accepting everything in which no two constraints sharing more than one variable. We capture the properties we need in the lemma below. The proof uses standard arguments and can be found in the Appendix.

**Lemma 4.1.** Let \( \varepsilon, \delta > 0 \) and a predicate \( P : [q]^k \to \{0, 1\} \) be given. Then there exist \( \gamma = O(q^k \log q/\varepsilon^2) \), \( \eta = \Omega((1/\gamma)^{10/\delta}) \) and \( N \in \mathbb{N} \), such that if \( n \geq N \) and \( \Phi \) is a random instance of \( \text{MAX } k\text{-CSP}(P) \) with \( m = \gamma n \) constraints, then with probability \( \exp(-O(k^4 \gamma^2)) \).
1. \( \text{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q}(1 + \varepsilon) \cdot m. \)

2. For any set \( \mathcal{C} \) of constraints with \( |\mathcal{C}| \leq \eta n \), we have \( |\partial(\mathcal{C})| \geq (k - 2 - \delta)|\mathcal{C}|. \)

3. No two constraints in \( \Phi \) share more than one variable.

Let \( \Phi \) be an instance of MAX \( k \)-CSP on \( n \) variables for which \( G_\Phi \) is \((\eta n, k - 2 - \delta)\) boundary expanding for some \( \delta < 1/4 \), as in Lemma 4.1. For such a \( \Phi \), we now define the distributions \( \mathcal{D}(S) \). In the rest of this chapter we assume \( k \geq 3 \) as for \( k = 2 \) the only promising predicate is the one satisfying all assignments for which the theorem is trivial.

For a set \( S \) of size at most \( t = \eta n/6k \), let \( \overline{S} \) be the subset of variables output by the algorithm Advice when run with input \( S \) and parameters \( r = \eta n, e_1 = (k - 2 - \delta), e_2 = (k - 8/3 - \delta) \) on the graph \( G_\Phi \). Theorem 3.1 shows that

\[
|\|S| \leq (2k - 4 - 2\delta + k - 6/3 + \delta)|S|/(4/3) = (6k - 4 - 3\delta)|S|/4 \leq \eta n/4,
\]

and also,

\[
|\xi_S \geq \eta n - |S|/(2/3) = \eta n - \eta n/4k > 3\eta n/4.
\]

We then use (3.1) to define the distribution \( \mathcal{D}(S) \) for sets \( S \) of size at most \( \eta n/6k \) as

\[
\text{Pr}_{\mathcal{D}(S)}[\alpha] = \sum_{\beta \in [q]^S} \text{Pr}_{\mathcal{P}(S)}[\beta].
\]

Using the properties of the distributions \( \mathcal{P}(\overline{S}) \), we can now prove that the distributions \( \mathcal{D}(S) \) are consistent.

**Claim 4.2.** Let the distributions \( \mathcal{D}(S) \) be defined as above. Then for any two sets \( S_1 \subseteq S_2 \subseteq [n] \) with \(|S_2| \leq t = \eta n/6k\), the distributions \( \mathcal{D}(S_1), \mathcal{D}(S_2) \) are equal on \( S_1 \).

**Proof.** The distributions \( \mathcal{D}(S_1), \mathcal{D}(S_2) \) are defined according to \( \mathcal{P}(\overline{S}_1) \) and \( \mathcal{P}(\overline{S}_2) \) respectively. To prove the claim, we show that \( \mathcal{P}(\overline{S}_1) \) and \( \mathcal{P}(\overline{S}_2) \) are equal to the distribution \( \mathcal{P}(\overline{S}_1 \cup \overline{S}_2) \) on \( \overline{S}_1, \overline{S}_2 \) respectively (note that it need not be the case that \( \overline{S}_1 \subseteq \overline{S}_2 \)).

Let \( S_3 = \overline{S}_1 \cup \overline{S}_2 \). Since \(|\overline{S}_1|, |\overline{S}_2| \leq \eta n/4\), we have \(|S_3| \leq \eta n/2\). We will first show that \(|\mathcal{E}(S_3)| \leq 3\eta n/4\). Assume to the contrary, and let \( \mathcal{C} \) be any subset of \( \mathcal{E}(S_3) \) of size \( 3\eta n/4 \). Given that \( k \geq 3 \) and \( \delta < 1/4 \) we have would have the following bound on the boundary of \( \mathcal{C} \),

\[
|\partial(\mathcal{C})|/|\mathcal{C}| \leq |\Gamma(\mathcal{C})|/|\mathcal{C}| \leq |S_3|/|\mathcal{C}| \leq (1/2)/(3/4) = 2/3 < k - 2 - \delta,
\]

which would contradict boundary expansion of \( G \).

Now, by Theorem 3.1, we know that both \( G|_{\overline{S}_1} \) and \( G|_{\overline{S}_2} \) are \((3\eta n/4, k - 8/3 - \delta)\) boundary expanding. Thus, using Lemma 3.2 for the pairs \( \overline{S}_1 \subseteq S_3 \) and \( \overline{S}_2 \subseteq S_3 \), we get that

\[
\text{Pr}_{\mathcal{D}(S_1)}[\alpha_1] = \sum_{\beta_1(\overline{S}_1) = \alpha_1} \text{Pr}_{\mathcal{P}(\overline{S}_1)}[\beta_1] = \sum_{\beta_1(\overline{S}_1) = \alpha_1} \text{Pr}_{\mathcal{P}(\overline{S}_2)}[\beta_2] = \sum_{\beta_2(\overline{S}_2) = \alpha_1} \text{Pr}_{\mathcal{D}(S_2)}[\alpha_2]
\]

which shows that \( \mathcal{D}(S_1) \) and \( \mathcal{D}(S_2) \) are equal on \( S_1 \).
It is now easy to prove the main result.

**Theorem 4.3.** Let $P : [q]^k \to \{0, 1\}$ be a promising predicate. Then for every constant $\zeta > 0$, there exist $c = c(q, k, \zeta)$, such that for large enough $n$, the integrality gap of MAX $k$-CSP($P$) for the program obtained by $cn$ levels of the Sherali-Adams SDP hierarchy is at least $\frac{q^k}{|P^{-1}(1)|} - \zeta$.

**Proof.** We take $\varepsilon = \frac{\zeta}{q^k}$, $\delta = 1/4$ and consider a random instance $\Phi$ of MAX $k$-CSP($P$) with $m = \gamma n$ as given by Lemma 4.1. Thus, $\text{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^k}(1 + \varepsilon) \cdot m$.

On the other hand, by Claim 4.2 we can define distributions $\mathcal{D}(S)$ over every set of at most $\eta n/6k$ variables such that for $S_1 \subseteq S_2$, $\mathcal{D}(S_1)$ and $\mathcal{D}(S_2)$ are consistent over $S_1$. Also, Claim 3.4 implies that $\mathcal{P}_\mu(\mathcal{S})$ is pairwise-independent and balanced for any $S$. We can now construct the SDP vectors with inner products agreeing with the probabilities according to the distributions $\mathcal{D}(S) = \mathcal{P}_\mu(\mathcal{S})$ for an instance satisfying the above property using the following simple fact.

**Claim 4.4.** There exists vectors $\{v(i,j)\}_{i \in [n], j \in [q]}$ and $v_0$ satisfying:

1. For all $i \in [n]$ and $j_1 \neq j_2$, $\langle v(i,j_1), v(i,j_2) \rangle = 0$.
2. For all $i \in [n]$, $j \in [q]$, $\langle v(i,j), v_0 \rangle = \|v(i,j)\|^2 = 1/q$.
3. For all $i_1 \neq i_2 \in [n]$ and $j_1, j_2 \in [q]$, $\langle v(i_1,j_1), v(i_2,j_2) \rangle = 1/q^2$.

**Proof.** Let $e_1, \ldots, e_n$ be an orthonormal basis for $\mathbb{R}^n$ and $u_0, \ldots, u_{q-1}$ be vertices of a $q-1$ dimensional simplex satisfying $\langle u_{j_1}, u_{j_2} \rangle = -1/(q-1)$ when $j_1 \neq j_2$ and 1 otherwise. Let $v_0 \in \mathbb{R}^{n(q-1)+1}$ be a unit vector such that $\langle v_0, e_i \otimes u_j \rangle = 0$ for all $i \in [n]$ and $j \in [q]$. We then define $v(i,j)$ as

$$v(i,j) := \frac{1}{q}v_0 + \frac{\sqrt{q-1}}{q}(e_i \otimes u_j)$$

It is easy to check that $\langle v_0, v(i,j) \rangle = 1/q$ and $\langle v(i_1,j_1), v(i_2,j_2) \rangle = 1/q^2$ for all $i_1 \neq i_2 \in [n]$ and $j_1, j_2 \in [q]$. Also, for $j_1 \neq j_2$

$$\langle v(i_1,j_1), v(i_2,j_2) \rangle = \frac{1}{q^2} + \frac{q-1}{q^2} \cdot \left(\frac{-1}{q-1}\right) = 0$$

which proves the claim.

By Lemma 2.3 this gives a feasible solution to the SDP obtained by $\eta n/6k$ levels. Also, by definition of $\mathcal{D}(S)$, we have that $\Pr_{\mathcal{D}(S)}[\alpha] > 0$ only if $\alpha$ satisfies all constraints in $\mathcal{C}(S) \supseteq \mathcal{C}(S)$. Hence, the value of $\text{FRAC}(\Phi)$ is given by

$$\sum_{i=1}^m \sum_{\alpha \in [q]} C_i(\alpha)X_{T_i,\alpha} = \sum_{i=1}^m \sum_{\alpha \in [q]} C_i(\alpha)\Pr_{\mathcal{D}(T_i)}[\alpha] = \sum_{i=1}^m \sum_{\alpha \in [q]} \Pr_{\mathcal{D}(T_i)}[\alpha] = m.$$

Thus, the integrality gap after $\eta n/6k$ levels is at least

$$\frac{\text{FRAC}(\Phi)}{\text{OPT}(\Phi)} = \frac{q^k}{|P^{-1}(1)|} \left(1 + \frac{1}{(1 + \varepsilon)^2}\right) \geq \frac{q^k}{|P^{-1}(1)|} - \zeta.$$

**Theory of Computing**
SDP Gaps from Pairwise Independence

Acknowledgements

The authors would like to thank the anonymous reviewers for suggestions that improved the presentation of the paper.

References


Theory of Computing 15


A Proof of Lemma 4.1

We restate the lemma for convenience.

**Lemma A.1.** Let $\varepsilon, \delta > 0$ and a predicate $P : [q]^k \rightarrow \{0, 1\}$ be given. Then there exist $\gamma = O(q^4 \log q / \varepsilon^2)$, $\eta = \Omega((1/\gamma)^{10/\delta})$ and $N \in \mathbb{N}$, such that if $n \geq N$ and $\Phi$ is a random instance of $\text{MAX } k\text{-CSP}(P)$ with $m = \gamma n$ constraints, then with probability $\exp(-O(k^4 \gamma^2))$

1. $\text{OPT}(\Phi) \leq \frac{|P^{-1}(1)|}{q^d} (1 + \varepsilon) \cdot m$.

2. For any set $C$ of constraints with $|C| \leq \eta n$, we have $|\partial(C)| \geq (k - 2 - \delta)|C|$.

3. No two constraints in $\Phi$ share more than one variable.

**Proof.** Let $\alpha \in [q]^n$ be any fixed assignment. For a fixed $\alpha$, the events that a constraint $C_i$ is satisfied are independent and happen with probability $|P^{-1}(1)| / q^d$ each. Hence, the probability over the choice of $\Phi$ that $\alpha$ satisfies more than $\frac{|P^{-1}(1)|}{q^d} (1 + \varepsilon) \cdot \gamma n$ constraints is at most $\exp(-\varepsilon^2 \gamma n |P^{-1}(1)| / 3q^d)$ by Chernoff bounds. By a union bound, the probability that any assignment satisfies more than $\frac{|P^{-1}(1)|}{q^d} (1 + \varepsilon) \cdot \gamma n$ constraints is at most $q^n \cdot \exp\left(-\varepsilon^2 \gamma n |P^{-1}(1)| \right)$ which is $o(1)$ for $\gamma = \frac{6\varepsilon^2 \ln q}{k^2}$.

For showing the boundary expansion, we note that it suffices to show that the constraints have large expansion i.e. each set of $s$ constraints (for $s \leq \eta n$) contains at least $(k - 1 - \delta/2)s$ variables. Since each non-boundary variable occurs in at least two constraints, we get that the number of boundary variables must be at least $2(k - 1 - \delta/2)s - ks = (k - 2 - \delta)s$.

To show this, consider the probability that a set of $s$ constraints contains at most $cs$ variables, where $c = k - 1 - \delta$. This is upper bounded by

\[
\binom{n}{cs} \cdot \binom{\binom{c}{s}}{k} \cdot s! \cdot \gamma^n \cdot \left(\frac{n}{k}\right)^{-s}
\]

Here $\binom{n}{cs}$ is the number of ways of choosing the $cs$ variables involved, $\binom{\binom{c}{s}}{k}$ is the number of ways of picking $s$ tuples on $cs$ variables and $s! \cdot \gamma^n$ is the number of ways of selecting the $s$ constraints. The number $\binom{n}{k}^s$ is simply the number of ways of picking $s$ of these $k$-tuples in an unconstrained way. Using $(\frac{s^2}{e})^b \leq (\frac{s}{e})^b \leq (\frac{s}{e})^b$, $s! \leq s^s$ and collecting terms, we can bound this expression by

\[
\left(\frac{s}{n}\right)^{\delta s/2} \left(e^{2k+1-\delta/2(k+1/2)}\gamma\right)^s \leq \left(\frac{s}{n}\right)^{\delta s/2} (\gamma^s)^s = \left(\frac{\gamma^{10/\delta}}{n}\right)^{\delta s/2}
\]

We need to show that the probability that a set of $s$ constraints contains less than $cs$ variables for any $s \leq \eta n$ is $o(1)$. Thus, we sum this probability over all $s \leq \eta n$ to get

\[
\sum_{s=1}^{\eta n} \left(\frac{s\gamma^{10/\delta}}{n}\right)^{\delta s/2} = \sum_{s=1}^{\ln^2 n} \left(\frac{s\gamma^{10/\delta}}{n}\right)^{\delta s/2} + \sum_{s=\ln^2 n+1}^{\eta n} \left(\frac{s\gamma^{10/\delta}}{n}\right)^{\delta s/2} \leq O\left(\frac{\gamma^{10}}{n^{\delta/2} \ln^2 n}\right) + O\left(\left(\frac{\eta \cdot \gamma^{10/\delta}}{\eta^{\delta/2} \ln^2 n}\right)^{(\delta/2)(\ln^2 n)}\right)
\]
The first term is $o(1)$ and is small for large $n$. The second term is also $o(1)$ for $\eta = 1/(100\gamma^{10}/\delta)$.

Thus, we get the first two properties with probability $1 - o(1)$. Finally, the probability that there are no two constraints sharing two variables must be at least $\prod_{i=1,...,m}(1 - O(i \cdot k^4/n^2))$ because when we choose the $i$th constraint, by wanting it to not share two variables with another previously chosen constraint, we are forbidding any of the $\binom{k}{2}$ pairs of variables in the $i$th constraint from being equal to any of the $(i-1)\cdot \binom{k}{2}$ pairs in the previously chosen ones. Now using that for small enough $x$, $1-x > \exp(-2x)$, the probability is at least \( \exp(-\sum_{i=1,...,m} i \cdot k^4/n^2) \cdot \exp(-\sum_{i=1,...,m} (k^4 \cdot m^2)/n^2) = \exp(-\Omega(k^4 \gamma^2)). \)
ABOUT THE AUTHORS

SIAVOSH BENABBAS got his B.Sc. from Sharif University of Technology under the supervision of Mohammad Ghodsi in the Department of Computer Engineering. Since 2006, Siavosh is a graduate student in the department of Computer Science at the University of Toronto. There, he received his M.Sc. under the supervision of Avner Magen and Charles Rackoff, while he is currently anticipating his Ph.D. graduation supervised by Toni Pitassi. Siavosh is interested in questions touching on Linear Programming and Semidefinite Programming Relaxations, their Integrality Gaps and their applications to Combinatorial Optimization.

KONSTANTINOS (or Costis as friends call him) did his undergraduate studies at the department of Mathematics at the University of Athens (1998-2002), from where he also received his M.Sc. in Logic and Algorithms (2002-2004). In 2005 he moved to Toronto where he received his Ph.D under the supervision of Avner Magen and Toni Pitassi (2010). Currently, he is a postdoctoral fellow at the Dept. of Combinatorics and Optimization at the University of Waterloo, working on projects varying from Convex Optimization to Complexity Theory to Combinatorial Optimization to Algorithmic Game Theory. Costis is spending his free time in his kitchen where he is experimenting with flavors and recipes.

AVNER MAGEN (1968-2010) did his undergraduate and graduate studies at Hebrew University and received his Ph.D. in Computer Science in 2002 under the supervision of Nati Linial. After holding a postdoctoral fellowship at NEC Research in Princeton, NJ, he joined the University of Toronto in 2002, first as a postdoctoral fellow, and then as an Assistant Professor in 2004. He was promoted to Associate Professor in 2009. Avner was a superb scholar, making fundamental contributions to a number of areas of theoretical computer science that include Metric Embeddings, Sublinear Algorithms, Convex Programming, Computational Geometry, and Approximation Algorithms. He was a wonderful colleague with a terrific sense of humor and great energy. He was a dedicated research supervisor and a superb teacher.

MADHUR TULSIANI did his bachelor in Computer Science at IIT Kanpur (2001-05) and a Ph.D. at UC Berkeley (2005-09) advised by Luca Trevisan. After he spent two years as a postdoc at the Institute for Advanced Study and Princeton University, he joint Toyota Technological Institute at Chicago. His research interests touch on a variety of topics including Complexity Theory, Optimization, Approximation and Inapproximability, Pseudorandomness, and Cryptography.