

SPECIAL ISSUE: CCC 2017

From Weak to Strong Linear Programming Gaps for All Constraint Satisfaction Problems

Mrinalkanti Ghosh* Madhur Tulsiani†

Received July 30, 2017; Revised May 8, 2018; Published June 6, 2018

Abstract: We study the approximability of constraint satisfaction problems (CSPs) by linear programming (LP) relaxations. We show that for every CSP, the approximation obtained by a basic LP relaxation is at least as strong as the approximation obtained using relaxations given by $c \cdot \log n / \log \log n$ levels of the Sherali–Adams hierarchy (for some constant $c > 0$) on instances of size n .

It was proved by Chan et al. [FOCS 2013] (and recently strengthened by Kothari et al. [STOC 2017]) that for CSPs, any polynomial-size LP extended formulation is at most as strong as the relaxation obtained by a constant number of levels of the Sherali–Adams hierarchy (where the number of levels depend on the exponent of the polynomial in the size bound). Combining this with our result also implies that any polynomial-size LP extended formulation is at most as strong as the *basic* LP, which can be thought of as the base level of

A conference version of this paper appeared in the Proceedings of the 32nd Computational Complexity Conference (CCC'17) [14].

*NSF award number CCF-1254044.

†NSF award number CCF-1254044.

ACM Classification: F.2.2, G.1.6

AMS Classification: 68Q17, 90C05

Key words and phrases: constraint satisfaction problem, convex programming, linear programming hierarchy, integrality gap

the Sherali–Adams hierarchy. This essentially gives a dichotomy result for approximation of CSPs by polynomial-size LP extended formulations.

Using our techniques, we also simplify and strengthen the result by Khot et al. [STOC 2014] on (strong) approximation resistance for LPs. They provided a necessary and sufficient condition under which $o(\log \log n)$ levels of the Sherali–Adams hierarchy cannot achieve an approximation better than a random assignment. We simplify their proof and strengthen the bound to $o(\log n / \log \log n)$ levels.

1 Introduction

Given a finite alphabet $[q] = \{0, \dots, q-1\}$ and a predicate $f : [q]^k \rightarrow \{0, 1\}$, an instance of the problem MAX k-CSP(f) consists of m constraints over a set of n variables, x_1, \dots, x_n , taking values in the set $[q]$. Each constraint C_i is of the form $f(x_{i_1} + b_{i_1}, \dots, x_{i_k} + b_{i_k})$ for some k -tuple $(x_{i_1}, \dots, x_{i_k})$ of variables, and constants $b_{i_1}, \dots, b_{i_k} \in [q]$, where the addition is taken to be modulo q . We say that an assignment σ to the variables satisfies the constraint C_i if $C_i(\sigma(x_{i_1}), \dots, \sigma(x_{i_k})) = 1$. Given an instance Φ of the problem, the goal is to find an assignment σ to the variables satisfying as many constraints as possible. The approximability of the MAX k-CSP(f) problem has been extensively studied for various predicates f (see, e. g., [30] for a survey), and special cases include several interesting and natural problems such as MAX 3-SAT, MAX 3-XOR and MAX-CUT.

A topic of much recent interest has been the efficacy of Linear Programming (LP) and Semidefinite Programming (SDP) relaxations. For a given instance Φ of MAX k-CSP(f), let $\text{OPT}(\Phi)$ denote the *fraction* of constraints satisfied by an optimal assignment, and let $\text{FRAC}(\Phi)$ denote the value of the convex (LP/SDP) relaxation for the problem. Then, the performance guarantee of this algorithm is given by the integrality gap which equals the supremum of $\text{FRAC}(\Phi)/\text{OPT}(\Phi)$, over all instances Φ .

The study of unconditional lower bounds for general families of LP relaxations was initiated by Arora, Bollobás and Lovász [2] (see also [3]). They studied the Lovász-Schrijver [25] LP hierarchy and proved lower bounds on the integrality gap for Minimum Vertex Cover (their technique also yields similar bounds for MAX-CUT). De la Vega and Kenyon-Mathieu [12] and Charikar, Makarychev and Makarychev [9] proved a lower bound of $2 - o(1)$ for the integrality gap of the LP relaxations for MAX-CUT given respectively by $O(\log \log n)$ and $n^{O(1)}$ levels of the Sherali–Adams LP hierarchy [29]. Several follow-up papers have also shown lower bounds for various other special cases of the MAX k-CSP problem, both for LP and SDP hierarchies [1, 28, 33, 27, 6, 5, 22].

An LP extended formulation of a polytope $P \subseteq \mathbb{R}^d$ is a linear program of the form

$$Ex + Fy = g \quad \text{and} \quad E'x + F'y \leq g',$$

where $x \in \mathbb{R}^d$, $y \in \mathbb{R}^t$, and $x \in P$ if and only if there exists y such that (x, y) is a solution to the above LP. We take the size of an LP extended formulation to be the sum of the number of variables and the number of constraints (equalities plus inequalities). We refer the interested reader to a discussion in [13] for different (equivalent) notions of size.

A recent result by Chan et al. [7] shows a connection between strong lower bounds for the Sherali–Adams hierarchy, and lower bounds on the size of LP extended formulations for the corresponding problem. In fact, their result proved a connection not only for a lower bound on the worst case integrality

gap, but for the entire *approximability curve*. We say that Φ is (c, s) -integrality gap instance for a relaxation of MAX k-CSP(f), if we have

$$\text{FRAC}(\Phi) \geq c \quad \text{and} \quad \text{OPT}(\Phi) < s.$$

And we say that Φ is (c, s) -approximable by a relaxation of MAX k-CSP(f), if for instances with $\text{OPT}(\Phi) < s$, we have $\text{FRAC}(\Phi) \leq c$. They showed that for any fixed $t \in \mathbb{N}$, if there exist (c, s) -integrality gap instances of size n for the relaxation given by t levels of the Sherali–Adams hierarchy, then for all $\varepsilon > 0$ and sufficiently large N , there exists a $(c - \varepsilon, s + \varepsilon)$ -integrality gap instance of size (number of variables) N , for any linear extended formulation of size at most $N^{t/2}$. They also give a trade-off when t is a function of n . This was recently improved by Kothari et al. [21] and we describe the improved trade-off later.

We strengthen the above results by showing that for all $c, s \in [0, 1]$, (c, s) -integrality gap instances for a “basic LP” can be used to construct $(c - \varepsilon, s + \varepsilon)$ -integrality gap instances for $\Omega_\varepsilon(\log n / \log \log n)$ levels of the Sherali–Adams hierarchy. The basic LP uses only a subset of the constraints in the relaxation given by k levels of the Sherali–Adams hierarchy for MAX k-CSP(f). In particular, this shows that a lower bound on the integrality gap for even the basic LP implies a similar lower bound on the integrality gap of any polynomial-size extended formulation. This can also be viewed as a dichotomy result showing that for any predicate f , either MAX k-CSP(f) is (c, s) -approximable by the *basic LP relaxation* (where the number of LP variables and constraints are linear in the size of the instance) or for all $\varepsilon > 0$, a $(c - \varepsilon, s + \varepsilon)$ -approximation cannot be achieved by *any polynomial-size LP extended formulation*. We note that both the above results and our result apply to all f, q and all $c, s \in [0, 1]$.

1.1 Comparison with (implications of) Raghavendra’s UG-hardness result

A remarkable result by Raghavendra [26] shows that a (c, s) -integrality gap instance for a “basic SDP” relaxation of MAX k-CSP(f) implies Unique-Games-hardness (UG-hardness) [16] of distinguishing instances Φ with $\text{OPT}(\Phi) < s$ from instances with $\text{OPT}(\Phi) \geq c$. The basic SDP considered by Raghavendra involves moments for all pairs of variables and all subsets of variables included in a constraint. The basic LP we consider is weaker than this SDP and does not contain the positive semidefiniteness constraint.

Combining Raghavendra’s result with known constructions of integrality gaps for Unique Games by Raghavendra and Steurer [27], and by Khot and Saket [17], one can obtain a result qualitatively similar to ours, for the mixed hierarchy. In particular, a (c, s) -integrality gap for the basic SDP implies a $(c - \varepsilon, s + \varepsilon)$ -integrality gap for $\Omega((\log \log n)^{1/4})$ levels of the mixed hierarchy.

Note however, that the above result is incomparable to our result, since it starts with stronger hypothesis (a basic SDP gap) and yields a gap for the mixed hierarchy as opposed to the Sherali–Adams hierarchy. While the above can also be used to derive lower bounds for linear extended formulations, one needs to start with an SDP gap instance to derive an LP lower bound. The basic SDP is known to be provably stronger than the basic LP for several problems including various 2-CSPs. Also, for the worst case f for $q = 2$, the integrality gap of the basic SDP is $O(2^k/k)$ [10], while that of the basic LP is 2^{k-1} . The integrality gap for the basic LP is achieved by “all zero or all one” predicate.

A recent result by Khot and Saket [18] shows a connection between the integrality gaps for the basic LP and those for the basic SDP. They prove that a (c, s) -integrality gap instance for the basic LP

implies UG-hardness of distinguishing instances Φ with $\text{OPT}(\Phi) \geq \Omega(c/(k^3 \cdot \log q))$ from instances with $\text{OPT}(\Phi) \leq 4s$. Their result also shows that a (c, s) -integrality gap instance for the basic LP can be used to produce a $(\Omega(c/(k^3 \cdot \log q)), 4s)$ -integrality gap instance for the basic SDP, and hence for $\Omega((\log \log n)^{1/4})$ levels of the mixed hierarchy.

1.2 Other related work

The power of the basic LP for solving valued CSPs *to optimality* has been studied in several previous papers. These results concerns the problem of minimizing the penalty for unsatisfied constraints, where the penalties take values in $\mathbb{Q} \cup \{\infty\}$. Also, they study the problem not only in terms of single predicate f , but rather in terms of the constraint language generated by a given set of (valued) predicates.

It was shown by Thapper and Živný [31] that when the penalties are finite-valued, if the problem of finding the optimum solution cannot be solved by the basic LP, then it is NP-hard. Kolmogorov, Thapper and Živný [20] give a characterization of CSPs where the problem of minimizing the penalty for unsatisfied constraints can be solved *exactly* by the basic LP. Also, a recent result by Thapper and Živný [32] shows that the valued CSP problem for a constraint language can be solved to optimality by a bounded number of levels of the Sherali–Adams hierarchy if and only if it can be solved by a relaxation obtained by augmenting the basic LP with constraints implied by three levels of the Sherali–Adams hierarchy. However, the above papers only consider the case when the LP gives an exact solution, and do not focus on approximation.

The techniques from [9] used in our result, were also extended by Lee [24] to prove a hardness for the Graph Pricing problem. Kenkre et al. [15] also applied these to show the optimality of a simple LP-based algorithm for Digraph Ordering.

1.2.1 Our results

Our main result is the following theorem, which shows that for every CSP, for instances of size n , the basic LP is at least as strong as any relaxation given by $o(\log n / \log \log n)$ levels of the Sherali–Adams hierarchy.

Theorem 1.1. Let $f : [q]^k \rightarrow \{0, 1\}$ be any predicate. Let Φ_0 be a (c, s) -integrality gap instance for basic LP relaxation of MAX k -CSP (f). Then for every $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exist $(c - \varepsilon, s + \varepsilon)$ -integrality gap instances of size N for the LP relaxation given by $c_\varepsilon \cdot \log N / \log \log N$ levels of the Sherali–Adams hierarchy.

Combining the above with the connection between Sherali–Adams gaps and extended formulations by [7, 21] also yields that the basic LP is at least as strong as any LP extended formulation of size $n^{o(\log n / \log \log n)}$.

Corollary 1.2. Let $f : [q] \rightarrow \{0, 1\}$ be any predicate. Let Φ_0 be a (c, s) -integrality gap instance for basic LP relaxation of MAX k -CSP (f). Then for every $\varepsilon > 0$, there exists $c'_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exist $(c - \varepsilon, s + \varepsilon)$ -integrality gap instances of size N , for every linear extended formulation of size at most $N^{c'_\varepsilon \cdot \log N / \log \log N}$.

As an application of our methods, we also simplify and strengthen the approximation resistance results for LPs proved by Khot et al. [19]. They studied predicates $f : \{0, 1\}^k \rightarrow \{0, 1\}$ and provided a necessary and sufficient condition for the predicate to be strongly approximation resistant for the Sherali–Adams LP hierarchy. One says a predicate is strongly approximation resistant if for all $\varepsilon > 0$, it is hard to distinguish instances Φ for which

$$\left| \text{OPT}(\Phi) - \mathbb{E}_{x \in \{0,1\}^k} [f(x)] \right| \leq \varepsilon$$

from instances with $\text{OPT}(\Phi) \geq 1 - \varepsilon$. In the context of the Sherali–Adams hierarchy, they showed that when this condition is satisfied, there exist instances Φ satisfying

$$\left| \text{OPT}(\Phi) - \mathbb{E}_{x \in \{0,1\}^k} [f(x)] \right| \leq \varepsilon \quad \text{and} \quad \text{FRAC}(\Phi) \geq 1 - \varepsilon,$$

where $\text{FRAC}(\Phi)$ is the value of the relaxation given by $O_\varepsilon(\log \log n)$ levels of the Sherali–Adams hierarchy. We strengthen their result (and provide a simpler proof) to prove the following.

Theorem 1.3. *Let $f : \{0, 1\}^k \rightarrow \{0, 1\}$ be any predicate satisfying the condition for strong approximation resistance for LPs, given by [19]. Then for every $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exists an instance Φ of MAX k -CSP(f) of size N , satisfying*

$$\left| \text{OPT}(\Phi) - \mathbb{E}_{x \in \{0,1\}^k} [f(x)] \right| \leq \varepsilon \quad \text{and} \quad \text{FRAC}(\Phi) \geq 1 - \varepsilon,$$

where $\text{FRAC}(\Phi)$ is the value of the relaxation given by $c_\varepsilon \cdot \log N / \log \log N$ levels of the Sherali–Adams hierarchy.

As before, the above theorem also yields a corollary for extended formulations.

1.2.2 Proof overview and techniques

The gap instance. The construction of our gap instances is inspired by the construction by Khot et al. [19]. They gave a generic construction to prove integrality gaps for any approximation resistant predicate (starting from certificates of hardness in form of certain “vanishing measures”), and we use similar ideas to give a construction which can start from a basic LP integrality gap instance as a certificate, to produce a gap instance for a large number of levels. This construction is discussed in [Section 5](#).

Given an integrality gap instance Φ_0 on n_0 variables, we treat this as a “template” (as in Raghavendra [26]) and generate a random instance using this. Concretely, we generate a new instance Φ on n_0 sets of n variables each. To generate a constraint, we sample a random constraint $C_0 \in \Phi_0$, and pick a variable randomly from each of the sets corresponding to variables in C_0 . Thus, the instances generated are n_0 -partite random hypergraphs, with each edge being generated according to a specified “type” (indices of sets to chose vertices from).

Note that previous instances of gap constructions for LP and SDP hierarchies were hypergraphs generated according to the model $\mathcal{G}_{n,p}$ with the signs of the literals chosen independently at random.

However, proving an LP/SDP lower bound using such instances implies a strong result: The predicate f is *useless* for the corresponding relaxation, in the sense defined by [4] (replacing the “ $P \neq NP$ ” assumption by the assumption that UG does not belong to P). Uselessness only holds for a limited class of predicates f (when $f^{-1}(1)$ supports a balanced pairwise independent distribution on $[q]^k$) [4]. Thus, proving an SDP lower bound for predicates which are not expected to be useless requires a new construction of instances, which cannot be generated uniformly at random. Our construction provides such a generalization, and may be useful in proving new SDP lower bounds. The properties of random $\mathcal{G}_{n,p}$ hypergraphs easily carry over to our instances, and we collect these properties in [Section 3](#).

The above construction ensures that if the instance Φ_0 does not have an assignment satisfying more than an s fraction of the constraints, then $\text{OPT}(\Phi) \leq s + \varepsilon$ with high probability. Also, it is well-known that providing a good LP solution to the relaxation given by t levels of the Sherali–Adams hierarchy is equivalent to providing distributions \mathcal{D}_S on $[q]^S$ for all sets of variables S with $|S| \leq t$, such that the distributions are consistent restricted to subsets, i. e., for all S with $|S| \leq t$ and all $T \subseteq S$, we have $\mathcal{D}_{S|T} = \mathcal{D}_T$. Thus, in our case, we need to produce such consistent local distributions such that the expected probability that a random constraint $C \in \Phi$ is satisfied by the local distribution on the set of variables involved in C (which we denote as S_C) is at least $c - \varepsilon$.

Local distributions from local structure. Most papers on integrality gaps for CSPs utilize the local structure of random hypergraphs to produce such distributions. Since the girth of a sparse random hypergraph is $\Omega(\log n)$, any induced subgraph on $o(\log n)$ vertices is simply a forest. In case that the induced hypergraph G_S on a set S is a *tree*, there is an easy distribution to consider: simply choose an arbitrary root and propagate down the tree by sampling each child conditioned on its parent. It is also easy to see that for $T \subseteq S$, if the induced hypergraph G_T is a *subtree* of G_S , then the distributions \mathcal{D}_S and \mathcal{D}_T produced as above are consistent.

The extension of this idea to forests requires some care. One can consider extending the distribution to forests by propagating independently on each tree in the forest. However, if for $T \subseteq S$ G_T is a forest while G_S is a tree, then a pair of vertices disconnected in G_T will have no correlation in \mathcal{D}_T but may be correlated in \mathcal{D}_S . This was handled, for example, in [19] by adding noise to the propagation and using a large ball $B(S)$ around S to define \mathcal{D}_S . Then, if two vertices of T are disconnected in $B(T)$ but connected in $B(S)$, then they must be at a large distance from each other. Thus, because of the noise, the correlation between them (which is zero in \mathcal{D}_T) will be very small in \mathcal{D}_S . However, correcting approximate consistency to exact consistency incurs a cost which is exponential in the number of levels (i. e., the sizes of the sets), which is what limits the results in [19, 12] to $O(\log \log n)$ levels. This also makes the proof more involved since it requires a careful control of the errors in consistency.

Consistent partitioning schemes. We resolve the above consistency issue by first partitioning the given set S into a set of clusters, each of which have diameter $\Delta_H = o(\log n)$ in the underlying hypergraph H . Since each cluster has bounded diameter, it becomes a tree when we add all the missing paths between any two vertices in the cluster. We then propagate independently on each cluster (augmented with the missing paths). This preserves the correlation between any two vertices in the same cluster, even if the path between them was not originally present in G_S .

Of course, the above plan requires that the partition obtained for $T \subseteq S$, is consistent with the

restriction to T of partition obtained for the set S . In fact, we construct distributions over partitions $\{\mathcal{P}_S\}_{|S|\leq T}$, which satisfy the consistency property $\mathcal{P}_{S|T} = \mathcal{P}_T$. These distributions over partitions, which we call consistent partitioning schemes, are constructed in [Section 4](#).

In addition to being consistent, we require that the partitioning scheme cuts only a small number of edges in expectation, since these contribute to a loss in the LP objective. We remark that such low-diameter decompositions (known as *separating* and *padded* decompositions) have been used extensively in the theory metric embeddings (see, e. g., [23] and the references therein). The only additional requirement in our application is consistency.

We obtain the decompositions by proving the (easy) hypergraph extensions the results of Charikar, Makarychev and Makarychev [11], who exhibit a metric which is similar to the shortest path metric on graphs at small distances, and has the property that its restriction to any subset of size at most $n^{\varepsilon'}$ (for an appropriate $\varepsilon' < 1$) is ℓ_2 embeddable. This is proved in [Section 3](#). A variant of this metric was used by Charikar, Makarychev and Makarychev [9] to prove lower bounds for MAX-CUT, for $n^{\varepsilon'}$ levels of the Sherali–Adams hierarchy. They used the embedding to construct a “local SDP solution” for any $n^{\varepsilon'}$ variables (with value $1 - \varepsilon'$) and produced the distributions required for Sherali–Adams by rounding the SDP solutions (which gives value $1 - O(\sqrt{\varepsilon'})$). However, rounding an SDP solution with a high value does not always produce a good integral solution for other CSPs.

Instead, we use these metrics in [Section 4](#) to construct the consistent partitioning schemes as described above, by applying a result of Charikar et al. [8] giving separating decompositions for finite subsets of ℓ_2 . We remark that it is the consistency requirement of the partitioning procedure that limits our results to $O(\log n / \log \log n)$ levels. The separation probability in the decomposition procedure grows with the dimension of the ℓ_2 embedding, while (to the best of our knowledge) dimension reduction procedures seem to break consistency.

2 Preliminaries

We use $[n]$ to denote the set $\{1, \dots, n\}$. The only exception is $[q]$, where we overload this notation to denote the set $\{0, \dots, q - 1\}$, which corresponds to the the alphabet for the Constraint Satisfaction Problem under consideration. We use \mathcal{D}_S and \mathcal{P}_S to denote probability distributions over assignments to and partitions of a set S , respectively. For $T \subseteq S$, the notation $\mathcal{D}_{S|T}$ is used to denote the restriction (marginal) of the distribution \mathcal{D}_S to the set T (and similarly for $\mathcal{P}_{S|T}$).

2.1 Constraint Satisfaction Problems

Definition 2.1. Let $[q]$ denote the set $\{0, \dots, q - 1\}$. For a predicate $f : [q]^k \rightarrow \{0, 1\}$, an instance Φ of MAX k -CSP $_q(f)$ consists of a set $\{x_1, \dots, x_n\}$ of variables and a set $\{C_1, \dots, C_m\}$ of constraints. Each constraint C_i is over a k -tuple $(x_{i_1}, \dots, x_{i_k})$ of variables, and the constraint is of the form

$$C_i \equiv f(x_{i_1} + b_{i_1}, \dots, x_{i_k} + b_{i_k})$$

for some $b_{i_1}, \dots, b_{i_k} \in [q]$, where the addition is modulo q . For an assignment $\sigma : [n] \mapsto [q]$, let $\text{sat}(\sigma)$ denote the fraction of constraints satisfied by σ (where x_i gets assigned to $\sigma(i)$). The maximum fraction

$$\begin{array}{ll}
 \text{maximize} & \mathbb{E}_{C \in \Phi} \left[\sum_{\alpha \in [q]^k} f(\alpha + b_C) \cdot x_{(S_C, \alpha)} \right] \\
 \sum_{\substack{\alpha \in [q]^S \\ \alpha|_T = \beta}} x_{(S, \alpha)} & = x_{(T, \beta)} & \forall T \subseteq S \subseteq [n], |S| \leq t, \forall \beta \in [q]^T \\
 x_{(S, \alpha)} & \geq 0 & \forall S \subseteq [n], |S| \leq t, \forall \alpha \in [q]^S \\
 x_{(\emptyset, \emptyset)} & = 1
 \end{array}$$

 Figure 1: Level- t Sherali–Adams LP for MAX k-CSP $_q$ (f).

of constraints that can be simultaneously satisfied is denoted by $\text{OPT}(\Phi)$, i. e.,

$$\text{OPT}(\Phi) = \max_{\sigma: [n] \rightarrow [q]} \text{sat}(\sigma).$$

For a constraint C of the above form, we use x_C to denote the tuple $(x_{i_1}, \dots, x_{i_k})$ of variables, and b_C to denote the tuple $(b_{i_1}, \dots, b_{i_k})$. We then write the constraint as $f(x_C + b_C)$. We also denote by S_C the set of indices, $\{i_1, \dots, i_k\}$, of the variables participating in the constraint C .

2.2 The LP relaxations for Constraint Satisfaction Problems

Below we present various LP relaxations for the MAX k-CSP $_q$ (f) problem that are relevant in this paper.

We start with the level- t Sherali–Adams relaxation. The intuition behind it is the following. Note that an integer solution to the problem can be given by an assignment $\sigma: [n] \rightarrow [q]$. Using this, we can define $\{0, 1\}$ -valued variables $x_{(S, \alpha)}$ for each $S \subseteq [n]$, $1 \leq |S| \leq t$ and $\alpha \in [q]^S$, with the intended solution $x_{(S, \alpha)} = 1$ if $\sigma(S) = \alpha$ and 0 otherwise. We also introduce a variable $x_{(\emptyset, \emptyset)}$, which equals to 1. We relax the integer program and allow variables to take real values in $[0, 1]$. Now the variables $\{x_{(S, \alpha)}\}_{\alpha \in [q]^S}$ give a probability distribution \mathcal{D}_S over assignments to S . We can enforce consistency between these *local* distributions by requiring that for $T \subseteq S$, the distribution over assignments to S , when marginalized to T , is precisely the distribution over assignments to T , i. e., $\mathcal{D}_{S|T} = \mathcal{D}_T$. The relaxation is shown in [Figure 1](#).

The basic LP relaxation is a reduced form of the above relaxation where only those variables $x_{(S, \alpha)}$ are included for which $S = S_C$ is the set of CSP variables for some constraint C . The consistency constraints are included only for singleton subsets of the sets S_C . We note here that for any feasible solution to basic LP relaxation, the local distributions $\{x_{(S, \alpha)}\}$ assign the same value to the repeated variables of a constraint. Note that the all the constraints for the basic LP are implied by the relaxation obtained by level k of the Sherali–Adams hierarchy.

For an LP/SDP relaxation of MAX k-CSP $_q$, and for a given instance Φ of the problem, we denote by $\text{FRAC}(\Phi)$ the LP/SDP (fractional) optimum. A relaxation is said to have a (c, s) -integrality gap if there exists a CSP instance Φ such that $\text{FRAC}(\Phi) \geq c$ and $\text{OPT}(\Phi) < s$.

$$\begin{array}{ll}
 \text{maximize } \mathbb{E}_{C \in \Phi} \left[\sum_{\alpha \in [q]^k} f(\alpha + b_C) \cdot x_{(S_C, \alpha)} \right] & \\
 \sum_{b \in [q]} x_{(i, b)} = 1 & \forall i \in [n] \\
 \sum_{\substack{\alpha \in [q]^{S_C} \\ \alpha(i) = b}} x_{(S_C, \alpha)} = x_{(i, b)} & \forall C \in \Phi, i \in S_C, b \in [q] \\
 x_{(S_C, \alpha)} \geq 0 & \forall C \in \Phi, \forall \alpha \in [q]^{S_C}
 \end{array}$$

 Figure 2: Basic LP relaxation for MAX k-CSP_q(f).

2.3 Hypergraphs

An instance Φ of MAX k-CSP defines a natural associated hypergraph $H = (V, E)$ with V being the set of variables in Φ and E containing one hyperedge for every constraint $C \in \Phi$. We remind the reader of the familiar notions of degree, paths, and cycles for the case of hypergraphs:

Definition 2.2. Let $H = (V, E)$ be a hypergraph.

- For a vertex $v \in V$, the degree of the vertex v is defined to be the number of distinct hyperedges containing it.
- A simple path P is a finite alternate sequence of distinct vertices and distinct edges starting and ending at vertices, i. e., $P = v_1, e_1, v_2, \dots, v_\ell, e_\ell, v_{\ell+1}$, where $v_i \in V \forall i \in [\ell + 1]$ and $e_i \in E \forall i \in [\ell]$. Furthermore, e_i contains v_i, v_{i+1} for each i . Here ℓ is called the length of the path P . All paths discussed in this paper will be simple paths.
- A sequence $\mathcal{C} = (v_1, e_1, v_2, \dots, v_\ell, e_\ell, v_1)$ is called a cycle of length ℓ if the initial segment v_1, e_1, \dots, v_ℓ is a (simple) path, $e_\ell \neq e_i$ and $|e_i| > 1$ for all $i \in [\ell]$, and $v_1 \in e_\ell$. We note that we don't include hyperedges with only one vertex towards forming cycles. For a path P (or cycle \mathcal{C}), we use $V(P)$ (or $V(\mathcal{C})$) to denote the set of all the vertices that occur in the edges, i. e., the set $\{v : (\exists i \in [h])(v \in e_i)\}$, where e_1, \dots, e_h are the hyperedges included in P (or \mathcal{C}). For a path P , $|P|$ denotes the number of hyperedges on the path P .
- For a given hypergraph H , the length of the smallest cycle in H is called the girth of H .

To observe the difference the notions of cycle in graphs and hypergraphs, it is instructive to consider the following example: let u, v be two distinct vertices in a k -uniform hypergraph for $k \geq 3$, and let e_1, e_2 be two distinct hyperedges both containing u and v . Then u, e_1, v, e_2, u is a cycle of length 2, which cannot occur in a graph.

We shall also need the following notion of the *closure* of a set $S \subseteq V$ in a given hypergraph H , defined by [9] for the case of graphs. A stronger notion of closure was also considered by [5].

Definition 2.3. For a given hypergraph H and $R \in \mathbb{N}$, and a set $S \subseteq V(H)$, we denote by $\text{cl}_R(S)$ the R -closure of S obtained by adding all the vertices in all the paths of length at most R connecting two vertices of S , i. e.,

$$\text{cl}_R(S) = S \cup \left(\bigcup_{\substack{P: P \text{ is a path in } H \\ P \text{ connects } u, v \in S \\ |P| \leq R}} V(P) \right).$$

For ease of notation, we use $\text{cl}(S)$ to denote $\text{cl}_1(S)$.

3 Properties of random hypergraphs

In this section we collect various properties of the hypergraphs corresponding to our integrality gap instances. The gap instances we generate contain several disjoint collections of variables. Each constraint in the instance has a specified “type”, which specifies which of the collections each of the participating k variables must be sampled from. The constraint is generated by randomly sampling each of the k variables, from the collections specified by its type. This is captured by the generative model described below.

In the model below and in the construction of the gap instance, the parameter n_0 should be thought of as constant, while the parameters n and m should be thought of as growing to infinity. We will choose $m = \gamma \cdot n$ for $\gamma = O_{k,q}(1)$.

Definition 3.1. Let $n_0, k \in \mathbb{N}$ with $k \geq 2$. Let $m, n \in \mathbb{N}$ and let Γ be a distribution on $[n_0]^k$. We define a distribution $\mathcal{H}_k(m, n, n_0, \Gamma)$ on n_0 -partite hypergraphs on $N = n_0 \cdot n$ vertices, divided into n_0 sets, X_1, \dots, X_{n_0} , of size n each. Each $H \in \text{Supp}(\mathcal{H}_k(m, n, n_0, \Gamma))$ has m edges and each edge has at most k vertices. A random hypergraph $H \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ is generated by sampling m random hyperedges independently as follows:

- Sample a random type $(i_1, \dots, i_k) \in [n_0]^k$ from the distribution Γ .
- For all distinct i_j , sample v_{i_j} independently and uniformly from X_{i_j} .
- Add the edge $e_i = \{v_{i_1}, \dots, v_{i_k}\}$ to H .

Note that as specified above, the model may generate a multi-hypergraph. However, the number of such repeated edges is likely to be small, and we will bound these, and in fact the number of cycles of size $o(\log n)$ in [Lemma A.2](#).

We will study the following metrics (similar to the ones defined in [11]):

Definition 3.2. Given a hypergraph H with vertex set V , we define two metrics $d_\mu^H(\cdot, \cdot), \rho_\mu^H(\cdot, \cdot)$ on V as

$$d_\mu^H(u, v) := 1 - (1 - \mu)^{2 \cdot d_H(u, v)} \quad \text{and} \quad \rho_\mu^H(u, v) := \sqrt{\frac{2 \cdot d_\mu^H(u, v) + \mu}{1 + \mu}},$$

for $u \neq v$, where $d_H(\cdot, \cdot)$ denotes the shortest path distance in H .

We primarily need the local ℓ_2 -embeddability of the metric ρ_μ^H . The following theorem captures various properties of random hypergraphs required for our construction. The proof of the theorem heavily uses results proved in [3] and [9] and we defer the details to [Appendix A](#).

Theorem 3.3. Let $H' \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ with $m = \gamma \cdot n$ edges and let $\varepsilon > 0$. Then for large enough n , with high probability (at least $1 - \varepsilon$, over the choice of H'), there exist $\delta > 0$, constant $c = c(k, \gamma, n_0, \varepsilon)$, $\theta = \theta(k, \gamma, n_0, \varepsilon)$ and a subhypergraph $H \subset H'$ with $V(H) = V(H')$ satisfying the following:

- H has girth $g \geq \delta \cdot \log n$.
- $|E(H') \setminus E(H)| \leq \varepsilon \cdot m$.
- For all $t \leq n^\theta$, for $\mu \geq c \cdot (\log t + \log \log n) / \log n$, for all $S \subseteq V(H')$ with $|S| \leq t$, the metric ρ_μ^H restricted to S is isometrically embeddable into the unit sphere in ℓ_2 .

4 Decompositions of hypergraphs from local geometry

We will construct the Sherali–Adams solution by partitioning the given subset of vertices into trees, and then creating a natural distribution over satisfying assignments on trees. We define below the kind of partitions we need.

Definition 4.1 (Consistent Partitioning Scheme). Let X be a finite set. For a set S , let \mathcal{P}_S denote a distribution over partitions of S . For $T \subseteq S$, let $\mathcal{P}_{S|T}$ be the distribution over partitions of T obtained by restricting the partitions in \mathcal{P}_S to the set T . We say that a collection of distributions $\{\mathcal{P}_S\}_{|S| \leq t}$ forms a consistent partitioning scheme of order t , if

$$\forall S \subseteq X, |S| \leq t \quad \text{and} \quad \forall T \subseteq S \quad \mathcal{P}_T = \mathcal{P}_{S|T}.$$

In addition to being consistent as described above, we also require the distributions to have small probability of cutting the hyperedges for the hypergraphs corresponding to our CSP instances. We define this property below.

Definition 4.2. Let $H = (V, E)$ be a hypergraph with each hyperedge having at most k vertices. Let $\{\mathcal{P}_S\}_{|S| \leq t}$ be a consistent partitioning scheme of order t for the vertex set V , with $t \geq k$. We say the scheme $\{\mathcal{P}_S\}_{|S| \leq t}$ is ε -sparse for H if

$$\forall e \in E \quad \mathbb{P}_{P \sim \mathcal{P}_e} [e \text{ is cut by } P] \leq \varepsilon.$$

In this section, we will prove that the hypergraphs arising from random CSP instances admit sparse and consistent partitioning schemes. Recall that for a hypergraph H , we define ([Definition 3.2](#)) the metrics d_μ^H and ρ_μ^H as:

$$d_\mu^H(u, v) := 1 - (1 - \mu)^{2 \cdot d_H(u, v)} \quad \text{and} \quad \rho_\mu^H(u, v) := \sqrt{\frac{2 \cdot d_\mu^H(u, v) + \mu}{1 + \mu}}.$$

Lemma 4.3. *Let $H = (V, E)$ be hypergraph with each hyperedge containing at most k vertices and let ρ_μ^H be the metric as defined above. Further, let H be such that for all sets $S \subseteq V$ with $|S| \leq t$, the metric induced on ρ_μ^H on S is isometrically embeddable into ℓ_2 . Then, there exists $\varepsilon \leq 10k \cdot \sqrt{\mu \cdot t}$ and $\Delta_H = O(1/\mu)$ such that H admits an ε -sparse consistent partitioning scheme of order t , with each partition consisting of clusters of diameter at most Δ_H in H .*

We use the following result of Charikar et al. [8] which shows that low-dimensional metrics have good separating decompositions with bounded diameter, i. e., decompositions which have a small probability of separating points at a small distance.

Theorem 4.4 ([8]). *Let W be a finite collection of points in \mathbb{R}^d and let $\Delta > 0$ be given. Then there exists a distribution \mathcal{P} over partitions of W such that*

- $\forall P \in \text{Supp}(\mathcal{P})$, each cluster in P has ℓ_2 diameter at most Δ , and
- for all $x, y \in W$

$$\mathbb{P}_{P \sim \mathcal{P}} [P \text{ separates } x \text{ and } y] \leq 2\sqrt{d} \cdot \frac{\|x - y\|_2}{\Delta}.$$

We also need the observation that the partitions produced by the above theorem are consistent, assuming the set S considered above lies in a fixed bounded set (using a trivial modification of the procedure in [8]). For the sequel, we use $B(x, \delta)$ to denote the ℓ_2 ball around x of radius δ and $B_H(u, r)$ to denote a ball of radius r around a vertex $u \in V(H)$. Thus,

$$B(x, \delta) := \{y \mid \|x - y\|_2 \leq \delta\} \quad \text{and} \quad B_H(u, r) := \{v \in V \mid d_H(u, v) \leq r\}.$$

The balls $B(S, \delta)$ and $B_H(S, r)$ are defined similarly.

Claim 4.5. *Let S and T be sets such that $T \subseteq S$. Let $W_S = \{w_u\}_{u \in S}$ and $W_T = \{w'_u\}_{u \in T}$ be ℓ_2 -embeddings of S and T satisfying $\phi(W_T) \subseteq W_S \subseteq B(0, R_0) \subset \mathbb{R}^d$, for some unitary transformation ϕ and $R_0 > 0$. Let \mathcal{P}_S and \mathcal{P}_T be distributions over partitions of S and T respectively, induced by partitions on W_S and W_T as given by [Theorem 4.4](#). Then*

$$\mathcal{P}_{S|T} = \mathcal{P}_T.$$

Proof. The claim follows simply by considering (a trivial modification of) the algorithm of [8]. For a given set W and a parameter Δ , they produce a partition using the following procedure:

- Let $W' = W$.
- Repeat until $W' = \emptyset$
 - Pick a random point x in $B(W, \Delta/2)$ according to the Haar measure. Let $C_x = B(x, \Delta/2) \cap W'$.
 - If $C_x \neq \emptyset$, set $W' = W' \setminus C_x$. Output C_x as a cluster in the partition.

[8] show that the above procedure produces a distribution over partitions satisfying the conditions in [Theorem 4.4](#). We simply modify the procedure to sample a random point x in $B(0, R_0 + \Delta/2)$ instead of $B(W, \Delta/2)$. This does not affect the separation probability of any two points, since the only non-empty clusters are still produced by the points in $B(S, \Delta/2)$. Since $R_0 + \Delta < \infty$, the above procedure almost surely terminates in finitely many steps.

Let P be a partition of S produced by the above procedure when applied to the point set W_S , and let P' be a random partition produced when applied to the point set $\phi(W_T)$. It is easy to see from the above procedure that the distribution \mathcal{P}_T is invariant under a unitary transformation of W_T . By coupling the random choice of a point in $B(0, R_0 + \Delta/2)$ chosen at each step in the procedures applied to W_S and $\phi(W_T) \subseteq W_S$, we get that $P(T) = P'$, i. e., the partition P restricted to T equals P' . Thus, we get $\mathcal{P}_{S|T} = \mathcal{P}_T$. \square

We can use the above to prove [Lemma 4.3](#).

Proof of Lemma 4.3. Given a set S , let W_S be an ℓ_2 embedding of the metric ρ_μ restricted to S . Since $|S| \leq t$, we can assume $W_S \in \mathbb{R}^t$. We apply partitioning procedure of Charikar et al. from [Theorem 4.4](#) with $\Delta = 1/2$. From the definition of the metric ρ_μ^H , we get that there exists a $\Delta_H = O(1/\mu)$ such that

$$\rho_{u,v}^H \leq 1/2 \implies d_H(u, v) \leq \Delta_H.$$

Moreover, for u, v contained in an edge e , we have that $\rho_\mu(u, v) \leq \sqrt{5\mu}$ and hence the probability that u and v are separated is at most $10\sqrt{\mu \cdot t}$. Thus, the probability that any vertex in e is separated from u is at most $10k \cdot \sqrt{\mu \cdot t}$ – as e contains at most k vertices.

Finally, for any $S \subseteq T$, if W_S and W_T denote the corresponding ℓ_2 embeddings, by the rigidity of ℓ_2 we have that for $\phi(W_T) \subseteq W_S$ for some unitary transformation ϕ . Thus, by [Claim 4.5](#), we get that this is a consistent partitioning scheme of order t . \square

5 The Sherali–Adams integrality gaps construction

5.1 Integrality gaps from the basic LP

Recall that the basic LP relaxation for MAX k-CSP $_q$ (f) as given in [Figure 2](#). In this section, we will prove [Theorem 1.1](#). We recall the statement below.

Theorem 1.1. Let $f : [q]^k \rightarrow \{0, 1\}$ be any predicate. Let Φ_0 be a (c, s) -integrality gap instance for basic LP relaxation of MAX k-CSP (f). Then for every $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exist $(c - \varepsilon, s + \varepsilon)$ -integrality gap instances of size N for the LP relaxation given by $c_\varepsilon \cdot \log N / \log \log N$ levels of the Sherali–Adams hierarchy.

Let Φ_0 be a (c, s) integrality gap instance for the basic LP relaxation for MAX k-CSP $_q$ (f) with n_0 variables and m_0 constraints. We use it to construct a new integrality gap instance Φ . The construction is similar to the gap instances constructed by Khot et al. [19] discussed in the next section. However, we describe this construction first since it is simpler. The procedure for constructing the instance Φ is described in [Figure 3](#).

Given: A (c, s) gap instance Φ_0 on n_0 variables, for the basic LP.

Output: An instance Φ with $N = n \cdot n_0$ variables and m constraints.

The variables are divided into n_0 sets, X_1, \dots, X_{n_0} , one for each variable in Φ_0 . We generate m constraints independently at random as follows:

1. Sample a random constraint $C_0 \sim \Phi_0$. Let $x_{C_0} = (i_1, \dots, i_k) \subseteq [n_0]$ denote the multi-set of variables in this constraint.
2. For each distinct i_j for $j \in [k]$, sample a random variable $x_{i_j} \in X_{i_j}$. We note that if $i_j = i_{j'}$ then we set $X_{i_j} = X_{i_{j'}}$.
3. Add the constraint $f((x_{i_1}, \dots, x_{i_k}) + b_{C_0})$ to the instance Φ .

Figure 3: Construction of the gap instance Φ .

Soundness

We first prove that no assignment satisfies more than $s + \varepsilon$ fraction of constraints for the above instance.

Lemma 5.1. *For every $\varepsilon > 0$, there exists $\gamma = \gamma(\varepsilon, n_0, q)$ such that for an instance Φ generated by choosing at least $\gamma \cdot n$ constraints independently at random as above, we have with probability $1 - \exp(-\Omega(n))$, $\text{OPT}(\Phi) < s + \varepsilon$.*

Proof. Fix an assignment $\sigma \in [q]^N$. We will first consider $\mathbb{E}[\text{sat}_\Phi(\sigma)]$ for a randomly generated Φ as above.

$$\begin{aligned} \mathbb{E}_\Phi[\text{sat}_\Phi(\sigma)] &= \mathbb{E}_{C_0 \in \Phi_0} \mathbb{E}_{(x_{i_1}, \dots, x_{i_k})} f(\sigma(x_{i_1}) + b_{i_1}, \dots, \sigma(x_{i_k}) + b_{i_k}) \\ &= \mathbb{E}_{C_0 \in \Phi_0} \mathbb{E}_{Z_1, \dots, Z_{n_0}} [f(Z_{C_0} + b_{C_0})], \end{aligned}$$

where for each $i \in [n_0]$, Z_i is an independent random variable with the distribution

$$\mathbb{P}[Z_i = b] := \mathbb{E}_{x \in X_i} [\mathbb{1}_{\{\sigma(x)=b\}}],$$

and Z_{C_0} denotes the collection of variables in the constraint C_0 , i. e.,

$$Z_{C_0} = \{Z_i\}_{i \in S_{C_0}}.$$

Thus, the random variables Z_1, \dots, Z_{n_0} define a random assignment to the variables in Φ_0 , which gives, for any σ

$$\mathbb{E}_\Phi[\text{sat}_\Phi(\sigma)] = \mathbb{E}_{C_0 \in \Phi_0} \mathbb{E}_{Z_1, \dots, Z_{n_0}} [f(Z_{C_0} + b_{C_0})] < s.$$

Consider a randomly added constraint C to the instance Φ . We have that

$$\mathbb{P}[C(\sigma) = 1] = \mathbb{E}_\Phi[\text{sat}_\Phi(\sigma)] < s,$$

for any fixed σ over random choice of the constraint C . Thus, for an instance Φ with m independently and randomly generated constraints, we have

$$\begin{aligned} \mathbb{P}_{\Phi}[\text{sat}_{\Phi}(\sigma) \geq s + \varepsilon] &\leq \mathbb{P}_{\Phi} \left[\text{sat}_{\Phi}(\sigma) \geq \mathbb{E}_{\Phi}[\text{sat}_{\Phi}(\sigma)] + \varepsilon \right] \\ &= \mathbb{P}_{\Phi} \left[\mathbb{E}_{C \in \Phi} [\mathbb{1}_{\{C(\sigma)=1\}}] \geq \mathbb{E}_{\Phi}[\text{sat}_{\Phi}(\sigma)] + \varepsilon \right] \\ &\leq \exp(-\Omega(\varepsilon^2 \cdot m)). \end{aligned}$$

Taking a union bound over all assignments, we get

$$\mathbb{P}_{\Phi}[\exists \sigma \text{ sat}_{\Phi}(\sigma) \geq s + \varepsilon] \leq q^{n \cdot n_0} \cdot \exp(-\varepsilon^2 \cdot m),$$

which is at most $\exp(-\Omega(n))$ for $m = O(((\log q)/\varepsilon^2) \cdot n \cdot n_0)$. \square

Completeness

To prove the completeness, we first observe that the instance Φ as constructed above is also a gap instance for the basic LP. We will then “boost” this hardness to many levels of the Sherali–Adams hierarchy.

Lemma 5.2. *For every $\varepsilon > 0$, there exists $\gamma = \gamma(\varepsilon)$ such that for an instance Φ generated by choosing at least $\gamma \cdot n$ constraints independently at random as above, with probability $1 - \exp(-\Omega(n))$ there exist distributions $\overline{\mathcal{D}}_{S_C}$ over $[q]^{S_C}$ for each $C \in \Phi$, and distributions $\overline{\mathcal{D}}_i$ over $[q]$ for each variable $i \in [n \cdot n_0]$, satisfying:*

- For all $C \in \Phi$ and all $i \in S_C$, $\overline{\mathcal{D}}_{S_C|_{\{i\}}} = \overline{\mathcal{D}}_i$;
- The distributions satisfy $\mathbb{E}_{C \in \Phi} \mathbb{E}_{\alpha \sim \overline{\mathcal{D}}_{S_C}} [f(\alpha + b_C)] \geq c - \frac{\varepsilon}{10}$.

Proof. For each $C_0 \in \Phi_0$ and each $j \in [n_0]$, let $\mathcal{D}_{S_{C_0}}^{(0)}$ and $\mathcal{D}_j^{(0)}$ denote the basic LP solution satisfying

$$\mathcal{D}_{S_{C_0}|_j}^{(0)} = \mathcal{D}_j^{(0)} \quad \forall C_0 \in \Phi_0 \quad \forall j \in S_{C_0} \quad \text{and} \quad \mathbb{E}_{C_0 \in \Phi_0} \mathbb{E}_{\alpha \sim \mathcal{D}_{S_{C_0}}^{(0)}} [f(\alpha + b_{C_0})] \geq c.$$

Each constraint $C \in \Phi$ is sampled according to some constraint $C_0 \in \Phi_0$, and we take $\overline{\mathcal{D}}_{S_C} := \mathcal{D}_{S_{C_0}}^{(0)}$ for the corresponding constraint $C_0 \in \Phi_0$. Also, each variable x_i for $i \in [n \cdot n_0]$, belongs to one of the sets X_j for $j \in [n_0]$, and we take $\overline{\mathcal{D}}_i := \mathcal{D}_j^{(0)}$ for the corresponding $j \in [n_0]$.

The consistency of the distributions follows immediately from the construction of the instance Φ . Let $C \in \Phi$ be any constraint and let C_0 be the corresponding constraint in Φ_0 . If $x_{C_0} = (j_1, \dots, j_k)$, then $x_C = (i_1, \dots, i_k)$ where each $i_r \in \{j_r\} \times [n]$ for all distinct j_r . Thus, for any $r \in [k]$,

$$\overline{\mathcal{D}}_{S_C|i_r} = \mathcal{D}_{S_{C_0}|_{j_r}}^{(0)} = \mathcal{D}_{j_r}^{(0)} = \overline{\mathcal{D}}_{i_r}.$$

We note that the repeated variables in x_{C_0} gets the *same* value under the basic LP solution $\mathcal{D}_{S_{C_0}}^{(0)}$ and therefore the same is true for $\overline{\mathcal{D}}_{S_C}$. To bound the objective value, we again consider its expectation over a randomly generated instance Φ . Let C be a random constraint added to Φ . Then, if we define $\overline{\mathcal{D}}_{S_C}$ as above for this constraint, we have

$$\mathbb{E}_C \mathbb{E}_{\alpha \in \overline{\mathcal{D}}_{S_C}} [f(\alpha + b_C)] = \mathbb{E}_{C_0 \in \Phi_0} \mathbb{E}_{\alpha \sim \mathcal{D}^{(0)}} [f(\alpha + b_{C_0})] \geq c.$$

Thus, the expected contribution of each constraint is at least c . The probability that the average of m constraints deviates by at least $\varepsilon/10$ from the expectation, is at most $\exp(-\Omega(\varepsilon^2 \cdot m))$. There exists $\gamma = O(1/\varepsilon^2)$ such that for $m \geq \gamma \cdot n$, the probability is at most $\exp(-\Omega(n))$. \square

To construct local distributions for the Sherali–Adams hierarchy, we will consider (a slight modification of) the hypergraph H corresponding to the instance Φ . We first show that distributions on hyperedges of this hypergraph can be consistently propagated in a tree, provided they agree on intersecting vertices.

For a set $U \subseteq V(H)$ in a hypergraph H , recall that $\text{cl}(U)$ includes all paths of lengths at most 1 between any two vertices in U . Thus, $E(\text{cl}(U)) = \{e \in E \mid |e \cap U| \geq 2\}$. Note that [Lemma 5.2](#) implies that hyperedges forming a tree in H satisfy the hypothesis of [Lemma 5.3](#) below.

Lemma 5.3. *Let $H = (V, E)$ be a hypergraph. Let $U \subseteq V$ be such that the set of hyperedges $E(\text{cl}(U))$ form a tree. For each $e \in E(\text{cl}(U))$, let $\overline{\mathcal{D}}_e$ be a distribution on $[q]^e$ such that for any $u \in U$ and $e_1, e_2 \in E(\text{cl}(U))$ such that $e_1 \cap e_2 = \{u\}$, we have $\overline{\mathcal{D}}_{e_1|u} = \overline{\mathcal{D}}_{e_2|u} = \overline{\mathcal{D}}_u$. Then,*

- there exists a distribution $\overline{\mathcal{D}}_U$ on $[q]^U$ such that $\overline{\mathcal{D}}_{U|e \cap U} = \overline{\mathcal{D}}_{e|e \cap U}$ for all $e \in E(U)$;
- if $U' \subseteq U$ is such that the hyperedges in $E(\text{cl}(U'))$ form a subtree of $E(\text{cl}(U))$, then $\overline{\mathcal{D}}_{U|U'} = \overline{\mathcal{D}}_{U'}$.

Proof. We define the distribution by starting with an arbitrary hyperedge and traversing the tree in an arbitrary order. Let e_1, \dots, e_r be a traversal of the hyperedges in $E(\text{cl}(U))$ such that for all i ,

$$\left| \left(\bigcup_{j < i} e_j \right) \cap e_i \right| = 1.$$

Let $U_0 = \bigcup_{j < i} e_j$ be the set of vertices for which we have already sampled an assignment and let e_i be the next hyperedge in the traversal, with u being the unique vertex in $e_i \cap U_0$. We sample an assignment to the vertices in e_i , conditioned on the value for the vertex u . Formally, we extend the distribution $\overline{\mathcal{D}}_{U_0}$ to $U_0 \cup e_i$ by taking, for any $\alpha \in [q]^{U_0 \cup e_i}$

$$\overline{\mathcal{D}}_{U_0 \cup e_i}(\alpha) = \overline{\mathcal{D}}_{U_0}(\alpha(U_0)) \cdot \frac{\overline{\mathcal{D}}_{e_i}(\alpha(e_i))}{\overline{\mathcal{D}}_{e_i|u}(\alpha(u))} = \overline{\mathcal{D}}_{U_0}(\alpha(U_0)) \cdot \frac{\overline{\mathcal{D}}_{e_i}(\alpha(e_i))}{\overline{\mathcal{D}}_u(\alpha(u))}.$$

The above process defines a distribution $\overline{\mathcal{D}}_{\text{cl}(U)}$ on $\text{cl}(U)$, with

$$\overline{\mathcal{D}}_{\text{cl}(U)}(\alpha) = \frac{\prod_{e \in E(U)} \overline{\mathcal{D}}_e(\alpha(e))}{\prod_{u \in \text{cl}(U)} (\overline{\mathcal{D}}_u(\alpha(u)))^{\deg(u)-1}}.$$

In the above expression, we use $\deg(u)$ to denote the degree of vertex u in tree formed by the hyperedges in $E(\text{cl}(U))$, i. e., $\deg(u) = |\{e \in E(\text{cl}(U)) \mid u \in e\}|$. We then define the distribution $\overline{\mathcal{D}}_U$ as the marginalized distribution $\overline{\mathcal{D}}_{\text{cl}(U)|U}$, i. e.,

$$\overline{\mathcal{D}}_U(\alpha) = \sum_{\substack{\beta \in [q]^{\text{cl}(U)} \\ \beta(U) = \alpha}} \overline{\mathcal{D}}_{\text{cl}(U)}(\beta).$$

Note that the distribution $\overline{\mathcal{D}}_{\text{cl}(U)}$ and hence also the distribution $\overline{\mathcal{D}}_U$ are independent of the order in which we traverse the hyperedges in $E(\text{cl}(U))$. Also, since the above process samples each hyperedge according to the distribution $\overline{\mathcal{D}}_e$, we have that for any $e \in E(U)$, $\overline{\mathcal{D}}_{\text{cl}(U)|e} = \overline{\mathcal{D}}_e$. Thus, also for any $e \in E(U)$,

$$\overline{\mathcal{D}}_{U|e \cap U} = \overline{\mathcal{D}}_{e|e \cap U}.$$

Let $U' \subseteq U$ be any set such that $E(\text{cl}(U'))$ forms a subtree of $E(\text{cl}(U))$. Then there exists a traversal e_1, \dots, e_r , and $i \in [r]$ such that $e_j \in E(\text{cl}(U')) \forall j \leq i$ and $e_j \notin E(\text{cl}(U')) \forall j > i$. However, the distribution defined by the partial traversal e_1, \dots, e_i is precisely $\overline{\mathcal{D}}_{\text{cl}(U')}$. Thus, we get that

$$\overline{\mathcal{D}}_{\text{cl}(U)|\text{cl}(U')} = \overline{\mathcal{D}}_{\text{cl}(U')}$$

which implies $\overline{\mathcal{D}}_{U|U'} = \overline{\mathcal{D}}_{U'}$. □

We can now prove the completeness for our construction using consistent decompositions.

Lemma 5.4. *Let $\varepsilon > 0$ and let Φ be a random instance of MAX k -CSP $_q$ (f) generated by choosing $\gamma \cdot n$ constraints independently at random as above, for large enough n . Then, there is a $t = \Omega_{\varepsilon, k, n_0}(\log n / \log \log n)$, such that with probability $1 - \varepsilon$ over the choice of Φ , there exist distributions $\{\mathcal{D}_S\}_{|S| \leq t}$ satisfying:*

- For all $S \subseteq V$ with $|S| \leq t$, \mathcal{D}_S is a distribution on $[q]^S$.
- For all $T \subseteq S \subseteq V$ with $|S| \leq t$, $\mathcal{D}_{S|T} = \mathcal{D}_T$.
- The distributions satisfy

$$\mathbb{E}_{C \in \Phi} \mathbb{E}_{\alpha_C \sim \mathcal{D}_{S_C}} [f(\alpha_C + b_C)] \geq c - \varepsilon.$$

Proof. By [Theorem 3.3](#), we know that there exists δ such that with probability $1 - \varepsilon/4$, after removing a set of constraints C_B of size at most $(\varepsilon/4) \cdot m$, we can assume that the remaining instance has girth at least $g = \delta \cdot \log n$. Also, there exist $\theta, c > 0$ such that for all large enough n , for all $t \leq n^\theta$, the metric ρ_μ^H restricted to any set S of size at most t embeds isometrically into the unit sphere in ℓ_2 , for all $\mu \geq c \cdot (\log t + \log \log n) / \log n$.

We choose $\mu = 2c \cdot \log \log n / \log n$ and $t = \varepsilon^2 / (400k^2 \cdot \mu)$ so that

$$\mu \geq c \cdot \frac{\log t + \log \log n}{\log n} \quad \text{and} \quad \sqrt{\mu \cdot t} \leq \frac{\varepsilon}{20k}.$$

Thus, by [Lemma 4.3](#), H admits an $(\varepsilon/2)$ -sparse partitioning scheme of order t with each cluster in the partition having diameter at most $\Delta_H = O(1/\mu)$. Let $\{\mathcal{P}_S\}_{|S| \leq t}$ denote this partitioning scheme.

Given a set S , the distribution \mathcal{D}_S is a convex combination of several distributions $\mathcal{D}_{S,P}$, corresponding to different partitions P sampled from \mathcal{P}_S . We describe the distribution \mathcal{D}_S by giving the procedure to sample an $\alpha \in [q]^S$. Given the set S with $|S| \leq t$

- Sample a partition $P = (U_1, \dots, U_r)$ from the distribution \mathcal{P}_S .
- For each set U_i , consider the set $\mathcal{C}(U_i)$ obtained by including the vertices contained in all the hyperedges in the shortest path between all $u, v \in U_i$. Note that since U_i has diameter at most Δ_H in H , $\mathcal{C}(U_i)$ is connected and in fact $\mathcal{C}(U_i) = \text{cl}_{\Delta_H}(U_i)$. Also, since the each vertex in an included path is within distance at most $\Delta_H/2$ of an end-point, and U_i has diameter at most Δ_H , we know that the diameter of $\mathcal{C}(U_i)$ is at most $2 \cdot \Delta_H$. Hence, $\mathcal{C}(U_i)$ is a tree. Finally, we must have $\text{cl}(\mathcal{C}(U_i)) = \mathcal{C}(U_i)$ since any additional path of length 1 would create a cycle of length at most $2 \cdot \Delta_H + 1$.

Thus, by [Lemma 5.2](#) and [Lemma 5.3](#) (with probability at least $1 - \varepsilon/4$) there exists a distribution $\overline{\mathcal{D}}_{\mathcal{C}(U_i)}$ for each U_i , satisfying

$$\overline{\mathcal{D}}_{\mathcal{C}(U_i)|e} = \overline{\mathcal{D}}_e$$

for all $e \in E(\mathcal{C}(U_i))$. Here, $\overline{\mathcal{D}}_e$ are the distributions given by [Lemma 5.2](#), which form a solution to the basic LP for Φ , with value at least $c - \varepsilon/4$. For each U_i , define the distribution

$$\mathcal{D}'_{U_i} := \overline{\mathcal{D}}_{\mathcal{C}(U_i)|U_i}.$$

- Sample $\alpha \in [q]^S$ according to the distribution

$$\mathcal{D}_{S,P} := \mathcal{D}'_{U_1} \times \dots \times \mathcal{D}'_{U_r}.$$

Thus, we have

$$\mathcal{D}_S := \mathbb{E}_{P=(U_1, \dots, U_r) \sim \mathcal{P}_S} \left[\prod_{i=1}^r \mathcal{D}'_{U_i} \right],$$

where the distributions \mathcal{D}'_{U_i} are defined as above.

We first prove the distributions are consistent on intersections, i. e., $\mathcal{D}_{S|T} = \mathcal{D}_T$ for any $T \subseteq S$. Note that by [Lemma 4.3](#), the distributions \mathcal{P}_S and \mathcal{P}_T satisfy $\mathcal{P}_{S|T} = \mathcal{P}_T$. Each partition (U_1, \dots, U_r) of S also produces a partition T . For ease of notation, we assume that the first (say) r' clusters have non-empty intersection with T . Let $V_i = U_i \cap T$ for $1 \leq i \leq r'$ ($V_i = \emptyset$ for $i > r'$). Then, we have

$$\begin{aligned} \mathcal{D}_{S|T} &= \mathbb{E}_{P=(U_1, \dots, U_r) \sim \mathcal{P}_S} \left[\prod_{i=1}^r \mathcal{D}'_{U_i|V_i} \right] = \mathbb{E}_{P=(U_1, \dots, U_r) \sim \mathcal{P}_S} \left[\prod_{i=1}^{r'} \overline{\mathcal{D}}_{\mathcal{C}(U_i)|V_i} \right] \\ &= \mathbb{E}_{P=(U_1, \dots, U_r) \sim \mathcal{P}_S} \left[\prod_{i=1}^{r'} \overline{\mathcal{D}}_{\mathcal{C}(V_i)|V_i} \right] \\ &= \mathbb{E}_{P'=(V_1, \dots, V_{r'}) \sim \mathcal{P}_T} \left[\prod_{i=1}^{r'} \overline{\mathcal{D}}_{\mathcal{C}(V_i)|V_i} \right]. \end{aligned}$$

The second to last equality above uses the fact that $\mathcal{C}(V_i)$ is a subtree of $\mathcal{C}(U_i)$ and thus

$$\overline{\mathcal{D}}_{\mathcal{C}(U_i)|\mathcal{C}(V_i)} = \overline{\mathcal{D}}_{\mathcal{C}(V_i)}$$

by [Lemma 5.3](#). The last equality uses the fact that $\mathcal{P}_{S|T} = \mathcal{P}_T$ by [Lemma 4.3](#).

We now argue that the LP solution corresponding to the above distributions $\{\mathcal{D}_S\}_{|S|\leq t}$ has value at least $c - \varepsilon$. Recall that the value of the LP solution is given by

$$\mathbb{E}_{C \in \Phi} \mathbb{E}_{\alpha \sim \mathcal{D}_{S_C}} [f(\alpha + b_C)].$$

Consider any constraint C in Φ , with the corresponding set of variables S_C and the corresponding hyperedge e . When defining the distribution \mathcal{D}_{S_C} , we will partition S_C according to the distribution \mathcal{P}_{S_C} . By [Lemma 4.3](#) and our choice of parameters

$$\mathbb{P}_{P \sim \mathcal{P}_{S_C}} [S_C \text{ is separated by } P] \leq 10k \cdot \sqrt{\mu \cdot t} \leq \frac{\varepsilon}{2}.$$

For a constraint set which is not in the deleted set C_B , if the hyperedge e corresponding to the constraint C is not split by a partition P sampled according to \mathcal{P}_{S_C} , then by [Lemma 5.3](#)

$$\mathcal{D}_{S_C, P} = \overline{\mathcal{D}}_{S_C}.$$

Here, $\overline{\mathcal{D}}_{S_C}$ is the distribution given by [Lemma 5.2](#). Since f is Boolean, we have that for $C \notin C_B$,

$$\mathbb{E}_{\alpha \sim \mathcal{D}_{S_C}} [f(\alpha + b_C)] \geq \mathbb{E}_{\alpha \sim \overline{\mathcal{D}}_{S_C}} [f(\alpha + b_C)] - \frac{\varepsilon}{2}.$$

Using [Lemma 5.2](#) again, we get

$$\begin{aligned} \mathbb{E}_{C \in \Phi} \mathbb{E}_{\alpha \sim \mathcal{D}_{S_C}} [f(\alpha + b_C)] &\geq \mathbb{E}_{C \in \Phi} \left[(1 - \mathbb{1}_{\{C \in C_B\}}) \cdot \left(\mathbb{E}_{\alpha \sim \overline{\mathcal{D}}_{S_C}} [f(\alpha + b_C)] - \frac{\varepsilon}{2} \right) \right] \\ &\geq \mathbb{E}_{C \in \Phi} \mathbb{E}_{\alpha \sim \overline{\mathcal{D}}_{S_C}} [f(\alpha + b_C)] - \frac{\varepsilon}{2} - \mathbb{E}_{C \in \Phi} [\mathbb{1}_{\{C \in C_B\}}] \\ &\geq c - \frac{\varepsilon}{4} - \frac{\varepsilon}{2} - \frac{\varepsilon}{4} \\ &\geq c - \varepsilon, \end{aligned}$$

where the penultimate inequality uses the fact that the fraction of constraints in the initially deleted set C_B is at most $\varepsilon/4$ (for large enough n). \square

5.2 Integrality gaps for resistant predicates

Let $f : \{0, 1\}^k \rightarrow \{0, 1\}$ be a Boolean predicate and let $\rho(f) = |f^{-1}(1)|/2^k$ be the fractions of satisfying assignments to f . Then f is approximation resistant if it is hard to distinguish the MAX-CSP instances on f between which are at least $1 - o(1)$ satisfiable vs which are at most $\rho(f) + o(1)$ satisfiable.

In [19] the authors introduce the notion of vanishing measure (on a polytope defined by f) and use it to characterize a variant of approximation resistance, called strong approximation resistance, assuming the Unique Games conjecture. They also gave a *weaker* notion of vanishing measures, which they used to characterize strong approximation resistance for LP hierarchies. In particular, they proved that when the condition in their characterization is satisfied, there exists a $(1 - o(1), \rho(f) + o(1))$ -integrality gap for $O(\log \log n)$ levels of Sherali–Adams hierarchy for predicates f . Here, we show that using [Theorem 1.1](#), their result can be simplified and strengthened¹ to $O(\log n / \log \log n)$ levels.

Let us first recall some useful notation defined by Khot et al. [19] before we define the notion of vanishing measure:

Definition 5.5. For a predicate $f : \{0, 1\}^k \rightarrow \{0, 1\}$, let $\mathcal{C}(f)$ be the convex polytope of *first moments* (biases) of distributions supported on satisfying assignments of f , i. e.,

$$\mathcal{C}(f) := \left\{ \zeta \in \mathbb{R}^k \mid \forall i \in [k], \zeta_i = \mathbb{E}_{\alpha \sim \nu} [(-1)^{\alpha_i}], \text{Supp}(\nu) \subseteq f^{-1}(1) \right\}.$$

For a measure Λ on $\mathcal{C}(f)$, $S \subseteq [k]$, $b \in \{0, 1\}^S$ and permutation $\pi : S \rightarrow S$, let $\Lambda_{S, \pi, b}$ denote the induced measure on \mathbb{R}^S by considering vectors with coordinates

$$\left\{ (-1)^{b_{\pi(i)}} \cdot \zeta_{\pi(i)} \right\}_{i \in S},$$

where $\zeta \sim \Lambda$.

We recall below the definition of vanishing measure for LPs from [19] (see Definition 1.3) :

Definition 5.6. A measure Λ on $\mathcal{C}(f)$ is called vanishing (for LPs) if for every $1 \leq t \leq k$, the following signed measure

$$\mathbb{E}_{|S|=t} \mathbb{E}_{\pi: S \rightarrow S} \mathbb{E}_{b \in \{0, 1\}^t} \left[\left(\prod_{i=1}^t (-1)^{b_i} \right) \cdot \hat{f}(S) \cdot \Lambda_{S, \pi, b} \right]$$

is identically 0. We say f has a vanishing measure if there exists a vanishing measure Λ on $\mathcal{C}(f)$.

In particular, they prove the following theorem:

Theorem 5.7. *Let $f : \{0, 1\}^k \rightarrow \{0, 1\}$ be a k -ary Boolean predicate that has a vanishing measure. Then for every $\varepsilon > 0$, there is a constant $c_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exists an instance Φ of MAX k -CSP(f) on N variables satisfying the following:*

- $\text{OPT}(\Phi) \leq \rho(f) + \varepsilon$.
- *The optimum for the LP relaxation given by $c_\varepsilon \cdot \log \log N$ levels of Sherali–Adams hierarchy has*

$$\text{FRAC}(\Phi) \geq 1 - O(k \cdot \sqrt{\varepsilon}).$$

¹The LP integrality gap result of Khot et al. is in fact slightly stronger than stated above. They show that LP value is at least $1 - o(1)$ while there is no integer solution achieving a value outside the range $[\rho(f) - o(1), \rho(f) + o(1)]$. It is easy to see that the same also holds for the instance constructed here.

Combining this with our [Theorem 1.1](#) already gives us the following stronger result:

Corollary 5.8. *Let $f : \{0, 1\}^k \rightarrow \{0, 1\}$ be a k -ary Boolean predicate that has a vanishing measure. Then for every $\varepsilon > 0$, there is a constant $c_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exists an instance Φ of MAX k -CSP(f) on N variables satisfying the following:*

- All integral assignment of Φ satisfies at most $\rho(f) + \varepsilon$ fraction of constraints.
- The LP relaxation given by $c_\varepsilon \cdot \log N / \log \log N$ levels of Sherali–Adams hierarchy has

$$\text{FRAC}(\Phi) \geq 1 - O(k\sqrt{\varepsilon}).$$

However, note that to apply [Theorem 1.1](#), one only needs a gap for the basic LP, which is much weaker requirement than the $O(\log \log N)$ -level gap given by [Theorem 5.7](#). We observe below that the gap for the basic LP follows very simply from the construction by Khot et al. [19]. One can then directly use this gap for applying [Theorem 1.1](#) instead of going through [Theorem 5.7](#).

Khot et al. [19] use the probabilistic construction given in [Figure 4](#), for a given $\varepsilon > 0$. The construction actually requires Λ to be a vanishing measure over the polytope $\mathcal{C}_\delta(f) := (1 - \delta) \cdot \mathcal{C}(f)$, for $\delta = \sqrt{\varepsilon}$. However, since $\mathcal{C}_\delta(f)$ is simply a scaling of $\mathcal{C}(f)$, a vanishing measure over $\mathcal{C}(f)$ also gives a vanishing measure over $\mathcal{C}_\delta(f)$. Note that each $\zeta_0 \in \mathcal{C}(f)$ corresponds to a distribution ν_0 supported in $f^{-1}(1)$. For each $\zeta \in \mathcal{C}_\delta$, let $\zeta_0 = \zeta / (1 - \delta)$ be the point in $\mathcal{C}(f)$ with distribution ν_0 . Then the distribution $\nu = (1 - \delta) \cdot \nu_0 + \delta \cdot U_k$ (where U_k denotes the uniform distribution on $\{0, 1\}^k$) satisfies

$$\forall i \in [k] \quad \mathbb{E}_{\alpha \sim \nu} [(-1)^{\alpha_i}] = \zeta_i.$$

Let $n_0 = \lceil 1/\varepsilon \rceil$. Partition the interval $[0, 1]$ into $n_0 + 1$ disjoint intervals I_0, I_1, \dots, I_{n_0} where $I_0 = \{0\}$ and $I_i = ((i - 1)/n_0, i/n_0]$ for $1 \leq i \leq n_0$. For each interval I_i , let X_i be a collection of n variables (disjoint from all X_j for $j \neq i$).

Generate m constraints independently according to the following procedure:

- Sample $\zeta \sim \Lambda$.
- For each $j \in [k]$, let i_j be the index of the interval which contains $|\zeta(j)|$. Sample uniformly a variable y_j from the set X_{i_j} .
- If $\zeta(j) < 0$, then negate y_j . If $\zeta(j) = 0$, then negate y_j with probability $1/2$.
- Introduce the constraint f on the sampled k -tuple of literals.

Figure 4: Sherali–Adams integrality gap instance for vanishing measure.

They show for a sufficiently large constant γ , an instance Φ with $m = \gamma \cdot n$ constraints satisfies with high probability, that for all assignments σ , $|\text{sat}_\Phi(\sigma) - \rho(f)| \leq \varepsilon$ (see Lemma 4.4 in [19]). The proof is similar to that of [Lemma 5.1](#).

Additionally, we need the following claim from [19] (see Claim 4.7 there), which allows one to “round” coordinates of the vectors $\zeta \in \mathcal{C}_\delta(f)$ to the end-points of the intervals I_0, \dots, I_{n_0} . This ensures that any two variables in the same collection X_i have the same bias. The proof of the claim follows simply from a hybrid argument. We include it in the appendix for completeness.

Claim 5.9. *Let $\zeta \in \mathcal{C}_\delta(f)$ and let ν be the corresponding distribution supported in $f^{-1}(1)$ such that for all $i \in [k]$, we have*

$$\zeta_i = \mathbb{E}_{\alpha \sim \nu} [(-1)^{\alpha_i}].$$

Let $t_1, \dots, t_k \in [0, 1]$ be such that for all $i \in [k]$, $|t_i - \zeta_i| \leq \varepsilon$ for $\varepsilon < \delta/2$. Then there exists a distribution ν' on $\{0, 1\}^k$ such that

$$\|\nu - \nu'\|_1 = O(k \cdot (\varepsilon/\delta)) \quad \text{and} \quad \forall i \in [k], \mathbb{E}_{\alpha \sim \nu'} [(-1)^{\alpha_i}] = \text{sign}(\zeta_i) \cdot t_i.$$

Proof. Let $r_j = \text{sign}(\zeta_j) \cdot t_j$ be the desired bias of the j -th coordinate. Then, $|\zeta(j) - r_j| \leq \varepsilon$ for all $j \in [k]$. We construct a sequence of distributions ν_0, \dots, ν_k such that $\nu_0 = \nu$ and $\nu_k = \nu'$. In $\bar{\nu}_j$, the biases are $(r_1, \dots, r_j, \zeta_{j+1}, \dots, \zeta_k)$.

The biases in ν_0 satisfy the above by definition. We obtain $\bar{\nu}_j$ from $\bar{\nu}_{j-1}$ as,

$$\nu_j = (1 - \tau_j) \cdot \nu_{j-1} + \tau_j \cdot D_j,$$

where D_j is the distribution in which all bits, except for the j -th one, are sampled independently according to their biases in $\bar{\nu}_{j-1}$. For the j -th bit, we fix it to $\text{sign}(r_j - \zeta_j)$ (if $r_j - \zeta_j = 0$, we can simply proceed with $\bar{\nu}_j = \bar{\nu}_{j-1}$). The biases for all except for the j -th bit are unchanged. For the j -th bit, the bias now becomes r_j if

$$r_j = (1 - \tau_j) \cdot \zeta_j + \tau_j \cdot \text{sign}(r_j - \zeta_j) \implies \tau_j \cdot (\text{sign}(r_j - \zeta_j) - r_j) = (1 - \tau_j) \cdot (r_j - \zeta_j).$$

Since $\zeta \in \mathcal{C}_\delta(f)$, we know that $|\text{sign}(r_j - \zeta_j) - r_j| \geq \delta/2$. Also, $|r_j - \zeta_j| \leq \varepsilon$ by assumption. Thus, we can choose $\tau_j = O(\varepsilon/\delta)$ which gives that $\|\bar{\nu}_j - \bar{\nu}_{j-1}\|_1 = O(\varepsilon/\delta)$. The final bound then follows by triangle inequality. \square

We can now use the above to give a simplified proof of [Corollary 5.8](#).

Proof of Corollary 5.8. Here we exhibit a solution of the basic LP ([Figure 2](#)) for the instance given in [Figure 4](#). For each variable y_j coming from the set X_j for $j \in \{0, 1, \dots, n_0\}$, we fix the bias t_j of the variable to be the rightmost point of the interval I_j , i. e., fix $x_{(y_j, -1)} = (1 - j/n_0)/2$ and $x_{(y_j, 1)} = (1 + j/n_0)/2$.

For each constraint C of the form $f(y_{i_1} + b_1, \dots, y_{i_k} + b_k)$, let $\zeta(C) \in \mathcal{C}_\delta(f)$ be the point used to generate it, and let $\nu(C)$ denote the corresponding distribution on $\{0, 1\}^k$. By [Claim 5.9](#), there exists a distribution $\nu'(C)$ such that $\|\nu(C) - \nu'(C)\|_1 = O(k \cdot \varepsilon/\delta)$ and such that the biases of the *literals* satisfy

$$\mathbb{E}_{\alpha \sim \nu'(C)} [(-1)^{\alpha_j}] = \text{sign}(\zeta_j) \cdot t_{i_j},$$

where t_{i_j} denotes the bias for the interval to which y_{i_j} belongs. When $t_{i_j} \neq 0$, we negate a variable only when $\text{sign}(\zeta_j) < 0$. Thus, we have

$$\mathbb{E}_{\alpha \sim \nu'(C)} [(-1)^{\alpha_j + b_j}] = t_{i_j},$$

which is consistent with the bias given by the singleton variables $x_{(y_{ij},1)}$ and $x_{(y_{ij},-1)}$. We thus define the local distribution on the set S_C as

$$\mathcal{D}_{S_C}(\alpha) = (\mathbf{v}'(C))(\alpha + b_C).$$

For all $C \in \Phi$, since $\zeta(C) \in \mathcal{C}_\delta(f)$, we have that

$$\mathbb{E}_{\alpha \sim \mathbf{v}(C)} [f(\alpha)] \geq 1 - \delta.$$

Also, since $\|\mathbf{v}(C) - \mathbf{v}'(C)\|_1 = O(k \cdot \varepsilon / \delta)$, we get that

$$\mathbb{E}_{\alpha \sim \mathbf{v}'(C)} [f(\alpha)] \geq 1 - \delta - O(k \cdot \varepsilon / \delta).$$

Thus, we have for all $C \in \Phi$,

$$\mathbb{E}_{\alpha \sim \mathcal{D}_{S_C}} [f(\alpha + b_C)] \geq 1 - \delta - O(k \cdot \varepsilon / \delta).$$

Taking $\delta = \sqrt{\varepsilon}$ proves the claim. □

5.3 Lower bounds for LP extended formulations

A connection between LP integrality gaps for the Sherali–Adams hierarchy, and lower bounds on the size of LP extended formulations, was first established by Chan et al. [7] and later improved by Kothari et al. [21]. In [21], the authors proved the following:

Theorem 5.10 ([21], Theorem 1.2). *There exist constants $0 < h < H$ such that the following holds. Consider a function $g : \mathbb{N} \rightarrow \mathbb{N}$. Suppose that the $g(n)$ -level Sherali–Adams relaxation for a CSP cannot achieve a (c, s) -approximation on instances on n variables. Then, no LP extended formulation (of the original LP) of size at most $n^{h \cdot g(n)}$ can achieve a (c, s) -approximation for the CSP on n^H variables.*

Combining [Theorem 1.1](#) with [Theorem 5.10](#) yields (with $g(N) := c_\varepsilon \cdot \log N / \log \log N$) we get the following corollary.

Corollary 1.2. *Let $f : [q] \rightarrow \{0, 1\}$ be any predicate. Let Φ_0 be a (c, s) -integrality gap instance for basic LP relaxation of MAX k -CSP (f). Then for every $\varepsilon > 0$, there exists $c'_\varepsilon > 0$ such that for infinitely many $N \in \mathbb{N}$, there exist $(c - \varepsilon, s + \varepsilon)$ -integrality gap instances of size N , for every linear extended formulation of size at most $N^{c'_\varepsilon \cdot \log N / \log \log N}$.*

6 Conclusions and open problems

This work shows a dichotomy result for approximating CSPs using linear programs, proving that if a (c, s) approximation is not achievable using the basic LP, then for every $\varepsilon > 0$, $(c - \varepsilon, s + \varepsilon)$ approximation is not achievable using $O_\varepsilon(\log n / (\log \log n))$ levels of the Sherali–Adams hierarchy. A natural open problem is to extend this result to $n^{O(1)}$ levels of the Sherali–Adams hierarchy. Using the results of [21],

this would also show that even exponential-size LP extended formulations are at most as strong as the basic LP.

As mentioned in the paper, the current limitation on the number of levels in our result, comes from the consistency requirement of the low-diameter decompositions. Given a d -dimensional ℓ_2 embedding of the restriction of the metric ρ_μ^H to a set of vertices S , the fraction of edges cut by the decomposition procedure of [8] grows as \sqrt{d} . Our current proof only uses the trivial bound that when $|S| = t$, the metric admits a t -dimensional ℓ_2 embedding. Even though for a single set S , this bound can be improved to $O(\log t)$ at the cost of slight errors in the distances (using randomized dimension reduction), we do not know how to do this *consistently* across various sets S . In particular, since we want to always obtain low-diameter components, we may need to reject a small fraction of dimension reduction maps for S , if they shrink the distances too much. However, such maps may not necessarily be rejected when considering $T \subseteq S$, which can violate the consistency requirement $\mathcal{P}_{S|T} = \mathcal{P}_T$. Understanding how and when randomized dimension reduction can be combined with the consistency requirement is an intriguing question which may also be useful in other applications.

Acknowledgements

We thank Chandra Chekuri, Subhash Khot and Yury Makarychev for helpful discussions, and Rishi Saket for pointers to references. We are also grateful to the CCC 2017 reviewers, and the ToC referees and editors, for numerous suggestions to improve the presentation of this paper. This research was supported by the National Science Foundation under award number CCF-1254044.

A Local ℓ_2 -embeddability of the Metric ρ_μ^H

The goal of this section is to prove the following result about the local ℓ_2 -embeddability of the metric ρ_μ^H .

Theorem 3.3. Let $H' \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ with $m = \gamma \cdot n$ edges and let $\varepsilon > 0$. Then for large enough n , with high probability (at least $1 - \varepsilon$, over the choice of H'), there exist $\delta > 0$, constant $c = c(k, \gamma, n_0, \varepsilon)$, $\theta = \theta(k, \gamma, n_0, \varepsilon)$ and a subhypergraph $H \subset H'$ with $V(H) = V(H')$ satisfying the following:

- H has girth $g \geq \delta \cdot \log n$.
- $|E(H') \setminus E(H)| \leq \varepsilon \cdot m$.
- For all $t \leq n^\theta$, for $\mu \geq c \cdot (\log t + \log \log n) / \log n$, for all $S \subseteq V(H')$ with $|S| \leq t$, the metric ρ_μ^H restricted to S is isometrically embeddable into the unit sphere in ℓ_2 .

To prove the above theorem, we will use the local structure of random hypergraphs. We first prove that with high probability for random hypergraphs (sampled from $\mathcal{H}_k(m, n, n_0, \Gamma)$) a few hyperedges can be removed to obtain a hypergraph whose girth is $\Omega(\log n)$ and the degree is bounded. The following lemma shows a possible trade-off between the degree of the hypergraph vs the number of hyperedges required to be removed.

Lemma A.1. *Let $H' \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ be a random hypergraph with $m = \gamma \cdot n$ hyperedges for large enough γ . Then for any $\varepsilon > 0$, with probability $1 - \varepsilon$, there exists a sub-hypergraph H with $V(H) = V(H')$ such that $\forall u \in V(H)$, $\deg_H(u) \leq 100 \cdot \log(n_0/\varepsilon) \cdot k \cdot \gamma$ and $|E(H') \setminus E(H)| \leq \varepsilon \cdot m$.*

Proof. By linearity of expectation, the expected degree of any vertex v in H' is at most $k \cdot \gamma$. Let $D = 100 \cdot \log(n_0/\varepsilon) \cdot k \cdot \gamma$, and let S be the set of all vertices u with $\deg_{H'}(u) > D$. Let E_S be the set of all hyperedges with at least one vertex in S . We shall take $E(H) = E(H') \setminus E_S$. Note that for any $u \in V(H')$, $\mathbb{P}[u \in S] = \mathbb{P}[\deg_{H'}(u) \geq D] \leq \exp(-D/4)$ by a Chernoff-Hoeffding bound. We use this to bound the expected number of edges deleted.

$$\begin{aligned} \mathbb{E}[E_S] &\leq \sum_{u \in V(H')} \mathbb{E}[\deg(u) \cdot \mathbf{1}_{\{u \in S\}}] = \sum_{u \in V(H')} \mathbb{E}[\deg(u) \mid u \in S] \cdot \mathbb{P}[u \in S] \\ &\leq \sum_{u \in V(H')} \mathbb{E}[\deg(u) \mid u \in S] \cdot \exp(-D/4) \\ &\leq \sum_{u \in V(H')} (D + k\gamma) \cdot \exp(-D/4) \\ &\leq (n \cdot n_0) \cdot 2D \cdot \exp(-D/4). \end{aligned}$$

The penultimate inequality uses the independence of the hyperedges in the generation process, which gives $\mathbb{E}[\deg_{H'}(u) \mid \deg_{H'}(u) \geq D] \leq D + \mathbb{E}[\deg_{H'}(u)]$. From our choice of the parameter D , we get that $\mathbb{E}[E_S] \leq \varepsilon^2 \cdot \gamma \cdot n = \varepsilon^2 \cdot m$. Thus, the number of edges deleted is at most $\varepsilon \cdot m$ with probability at least $1 - \varepsilon$. \square

The following lemma shows that the expected number of small cycles in random hypergraph is small.

Lemma A.2. *Let $H \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ be a random hypergraph and for $\ell \geq 2$, let $Z_\ell(H)$ denote the number of cycles of length at most ℓ in H . For m, n and k such that $k^2 \cdot (m/n) > 2$, we have*

$$\mathbb{E}_{H \sim \mathcal{H}_k(m, n, n_0, \Gamma)} [Z_\ell(H)] \leq \left(k^2 \cdot \frac{m}{n}\right)^{2\ell}.$$

Proof. Let the vertices of H correspond to the set $[n_0] \times [n]$. Suppose we contract the set $[n_0] \times \{j\}$ of vertices into a single vertex $j \in [n]$ to get a random multi-hypergraph H' on vertex set $[n]$. An equivalent way to view the sampling to H' is: for each $i \in [m]$, the i -th hyperedge e_i of H' is sampled by independently sampling k_{e_i} vertices (with replacement) uniformly at random from $[n]$. Note that the sampling of H' is independent of Γ in the definition of $\mathcal{H}_k(m, n, n_0, \Gamma)$. Clearly, a cycle of length at most ℓ in H produces a cycle of length at most ℓ in H' . Hence, it suffices to bound the expected number of cycles in H'

Given any pair (u', v') of vertices of H' , for $u' \neq v'$, the probability of the pair (u', v') belonging together in some hyperedge of H' is at most mk^2/n^2 —since each hyperedge e contains at most k vertices. Consider a given h -tuple $\mathbf{u} = (u_{i_1}, \dots, u_{i_h})$ of variables. Note that we require that hyperedges participating in a cycle be distinct and we don't count hyperedges with one vertex in them while forming cycle. So, the probability that \mathbf{u} is part of a cycle in H' , i. e., there exists distinct hyperedges $e_j \in H'$ for $j \in [h]$ such

that $u_{i_j}, u_{i_{j+1}} \in e_j$ for $j \in [h-1]$, and $u_{i_1}, u_{i_h} \in e_h$ is at most $(mk^2/n^2)^h$. As a result, the expected number of cycles of length h in H' is bounded above by

$$\binom{n}{h} \left(\frac{mk^2}{n^2}\right)^h \leq n^h \left(\frac{mk^2}{n^2}\right)^h = \left(k^2 \cdot \frac{m}{n}\right)^h.$$

From the geometric form of the bound, it follows that expected number of cycles of length at most ℓ in H' is at most

$$\frac{(k^2 \cdot m/n)^{\ell+1}}{(k^2 \cdot m/n) - 1} < (k^2 \cdot m/n)^{2\ell}. \quad \square$$

Using the above lemma, it is easy to show that one can remove all small cycles in a random hypergraph by deleting only a small number of hyperedges.

Corollary A.3. *Let $H \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ be a random hypergraph with $m = \gamma \cdot n$ for $\gamma > 1$. Then, there exists $\delta = \delta(\gamma, k) > 0$ such that with probability $1 - n^{-1/6}$, all cycles of length at most $\delta \cdot \log n$ in H can be removed by deleting at most $n^{2/3}$ hyperedges.*

Proof. As above, let Z_ℓ denote the number of cycles of length at most ℓ . With the choice of m, n , and k , we have $k^2 \cdot m/n \geq 2$. By Lemma A.2, $\mathbb{E}[Z_\ell] \leq (k^2 \cdot m/n)^{2\ell}$. Since $m = \gamma \cdot n$, there exists a $g = \delta \cdot \log n$ such that $\mathbb{E}[Z_g] \leq \sqrt{n}$. By Markov's inequality, $\mathbb{P}[Z_g \geq n^{2/3}] \leq n^{-1/6}$. Thus, with probability $1 - n^{-1/6}$, one can remove all cycles of length at most $\delta \cdot \log n$ by deleting at most $n^{2/3}$ edges. \square

Charikar et al. [11] prove an analogue of Theorem 3.3 for metrics defined on locally-sparse graphs (see Definition A.5). In fact, they use a consequence of sparsity, which they call ℓ -path decomposability. To this end, we define the *incidence graph*² associated with a hypergraph, on which we will apply their result.

Definition A.4. Let $H = (V(H), E(H))$ be a hypergraph. We define its incidence graph as the bipartite graph G_H defined on vertex sets $V(H)$ and $E(H)$, and edge set \mathcal{E} defined as

$$\mathcal{E} := \{(v, e) \mid v \in V(H), e \in E(H), v \in e\}.$$

Note that for any $u, v \in V(H)$, we have $d_{G_H}(u, v) = 2 \cdot d_H(u, v)$. We now define local sparsity of the incidence graph.

Definition A.5 (Local Sparsity). A graph $G = (V, E)$ on n vertices is defined to be (τ, η) -sparse if for all $S \subset V$ of size at most $\tau \cdot n$ we have $|E(S)| \leq |S|/(1 - \eta)$, where $E(S)$ denotes the set of edges contained in S .

We note that we will require the sparsity η to be $O_{k,\gamma}(1/\log n)$. This gives sparsity only for sublinear-size sets, as compared to sets of size $\Omega(n)$ in previous results where η is a constant. We also observe that if G_1 is (τ, η) -sparse and G_2 is a subgraph of G_1 then G_2 is also (τ, η) -sparse.

Now we prove that the incidence graph of a hypergraph sampled from $\mathcal{H}_k(m, n, n_0, \Gamma)$ is locally sparse with high probability. The proof of the following lemma follows closely proofs of analogous statements in [1, 6].

²This is the same notion as the constraint-variable graph considered in various papers on lower bounds for CSPs.

Lemma A.6. *For every $\gamma > 1$ and $m = \gamma \cdot n \cdot n_0$, let $\eta < 1/4$. Then there exists a constant*

$$\tau \leq ((1 + \gamma)^2 \cdot e^3 \cdot n_0^2)^{-1/\eta}$$

such that:

$$\mathbb{P}_{H \sim \mathcal{H}_k(m, n, n_0, \Gamma)} [G_H \text{ is not } (\tau, \eta)\text{-sparse}] \leq \mathcal{O}\left(\frac{\gamma^2 \cdot e^4 \cdot n_0^2}{n}\right),$$

for all large enough n .

Proof. For a randomly generated hypergraph $H \sim \mathcal{H}_k(m, n, n_0, \Gamma)$, we can think of the incidence graph as a randomly generated bipartite graph on the vertex sets $[m]$ and $V(H)$. We denote second vertex set by $[m]$ instead of $E(H)$ since the set $E(H)$ is only defined *after* sampling H . For each $i \in [m]$, we randomly sample a type from Γ , and sample at most k neighbors according to the type. Conditioned on a fixed type, we have that for any $v \in V(H)$, $\mathbb{P}[(v, i) \in \mathcal{E}] \leq 1/n$.

Let T be a set of vertices of G_H and let $v, i \in T$ be fixed. Since vertices in each bucket are chosen independently and choices for different indices $i \in [m]$ are made independently, we have that the probability that (v, i) is an edge in G_H conditioned on edges of $T \times T \setminus \{(v, i), (i, v)\}$ is at most $1/n$.

Hence the probability that a set T of h vertices induces at least $r = h/(1 - \eta)$ edges is at most

$$\binom{\binom{h}{2}}{r} \cdot \frac{1}{n^r}.$$

Since G_H has $N' = N + m = (1 + \gamma)n \cdot n_0$ vertices, by union bound, the probability that G_H is not (τ, η) -sparse is at most

$$\begin{aligned} & \sum_{h=2}^{\tau \cdot N'} \binom{N'}{h} \binom{\binom{h}{2}}{r} \cdot \frac{1}{n^r} & r & := \frac{h}{1 - \eta} \\ & \leq \sum_{h=2}^{\tau \cdot N'} \left(\frac{N' \cdot e}{h}\right)^h \cdot \left(\frac{h^2 \cdot e}{2 \cdot r}\right)^r \cdot \frac{1}{n^r} \\ & = \sum_{h=2}^{\tau \cdot N'} \left(\frac{(1 + \gamma) \cdot n \cdot n_0 \cdot e}{h}\right)^h \cdot \left(\frac{(1 - \eta) \cdot h \cdot e}{2n}\right)^{h/(1 - \eta)} \\ & = \sum_{h=2}^{\tau \cdot N'} \left(\frac{(1 - \eta) \cdot (1 + \gamma)^{1 - \eta} \cdot e^{2 - \eta} \cdot n_0^{1 - \eta}}{2}\right) \cdot \left(\frac{h}{n}\right)^\eta)^{h/(1 - \eta)} \\ & \leq \sum_{h=2}^{\tau \cdot N'} \left((1 + \gamma) \cdot e^2 \cdot n_0 \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(1 - \eta)}. \end{aligned}$$

We split the above sum in two parts depending on the range of h . Let us first consider

$$\begin{aligned} \sum_{h=n^\theta}^{\tau \cdot N'} \left((1 + \gamma) \cdot e^2 \cdot n_0 \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(1 - \eta)} & \leq \sum_{h=n^\theta}^{\tau \cdot (1 + \gamma) \cdot n \cdot n_0} \left((1 + \gamma) \cdot e^2 \cdot n_0 \cdot (\tau \cdot (1 + \gamma) \cdot n_0)^\eta\right)^{h/(1 - \eta)} \\ & \leq 2 \exp\left(-\frac{n^\theta}{1 - \eta}\right), \end{aligned}$$

where the penultimate inequality follows from our setting of parameter

$$\tau \leq ((1 + \gamma)^2 \cdot e^3 n_0^2)^{-1/\eta}$$

and the geometric nature of summands.

Now we fix $\theta = 1/2$ and consider the first half of the summation

$$\begin{aligned} \sum_{h=2}^{n^\theta} \left((1 + \gamma) \cdot e^2 \cdot n_0 \cdot \left(\frac{h}{n} \right)^\eta \right)^{h/(1-\eta)} &\leq n^\theta \cdot \left(\frac{(1 + \gamma) \cdot e^2 \cdot n_0}{n^{(1-\theta) \cdot \eta}} \right)^{2/(1-\eta)} \\ &= \left(\frac{(1 + \gamma) \cdot e^2 \cdot n_0}{n^{1-\theta}} \right)^{2/(1-\eta)}. \end{aligned}$$

Hence the probability that G_H is not (τ, η) -sparse is at most $(1 + \gamma)^2 \cdot e^4 \cdot n_0^2/n + 2 \exp(-n^{1/2})$ for all large enough $n > (2(1 + \gamma) \cdot e^2 \cdot n_0)^2$. \square

The above property was also implicitly used by Arora et al. [3] in proving the following lemma (see Lemma 2.12 in [3]).

Lemma A.7. *Let $\ell > 0$ be an integer and $0 < \eta < 1/(3\ell - 1) < 1$. Let G be a η -sparse graph with girth $g > \ell$. Then G is ℓ -path decomposable.*

Recall that we defined the metrics d_μ^H and ρ_μ^H on H as (for $u \neq v$)

$$d_\mu^H(u, v) := 1 - (1 - \mu)^{2 \cdot d_H(u, v)} \quad \text{and} \quad \rho_\mu^H(u, v) := \sqrt{\frac{2 \cdot d_\mu^H(u, v) + \mu}{1 + \mu}}.$$

For a graph G , we define the following two metrics, for $u \neq v$:

$$d_\mu^G(u, v) := 1 - (1 - \mu)^{d_G(u, v)} \quad \text{and} \quad \rho_\mu^G(u, v) := \sqrt{\frac{d_\mu^G(u, v) + \mu}{1 + \mu}}.$$

We note that if H is a hypergraph and G_H is its incidence graph, then the metrics $d_\mu^{G_H}$ and $\rho_\mu^{G_H}$ restricted to $V(H)$, coincide with the metrics d_μ^H and ρ_μ^H defined on H . Charikar et al. proved the following theorem (see Theorem 5.2) in [9].

Theorem A.8 ([9]). *Let G be a graph on n' vertices with maximum degree D . Let $t < \sqrt{n'}$ and $\ell > 0$ be such that for $t' = D^{\ell+1} \cdot t$, every subgraph of G on at most t' vertices is ℓ -path decomposable. Also, let μ , t and ℓ satisfy the relation $(1 - \mu)^{\ell/9} \leq \mu/(2t + 2)$. Then for every subset S of at most t vertices there exists a mapping ψ_S from S to the unit sphere in ℓ_2 such that all $u, v \in S$*

$$\|\psi_S(u) - \psi_S(v)\|_2 = \rho_\mu^G(u, v).$$

We use this theorem to prove the main theorem of the section.

Proof of Theorem 3.3. Let $H' \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ with $m = \gamma \cdot n$ hyperedges and $N = n_0 \cdot n$ vertices for large enough n . Given $\varepsilon > 0$, from Lemma A.1 we have that with probability at least $1 - \varepsilon/2$, there exists $H_1 \subset H$ such that the maximum degree of H_1 is at most $D = 100 \cdot \log(2n_0/\varepsilon) \cdot k \cdot \gamma$ with $|E(H') \setminus E(H_1)| \leq (\varepsilon/2) \cdot m$.

Using Corollary A.3 we also have that there exists $\delta > 0$, such that with probability at least $1 - \varepsilon/4$ (for large enough n) H' has a sub-hypergraph H_2 with $g \geq \delta \cdot \log n$ and $|E(H') \setminus E(H_2)| \leq (\varepsilon/4) \cdot m$.

By Lemma A.6, there exists $\eta = \Omega_{n_0, k, \gamma}(1/(\log n))$ such that for all large enough n , $G_{H'}$ is (τ, η) -sparse with probability at least $1 - \varepsilon/4$, for $\tau \geq n^{-1/4}$.

Hence with probability $1 - \varepsilon$, we have that $H = (V(H'), E(H_1) \cap E(H_2))$ satisfies:

- The maximum degree of H is bounded above by D .
- G_H is (τ, η) -sparse (for $\tau \geq n^{-1/4}$ and $\eta = \Omega_{n_0, k, \gamma}(1/(\log n))$).
- Girth of H is at least $g > \delta \cdot \log n$.
- $|E(H') \setminus E(H)| \leq \varepsilon \cdot m$.

We now show that the metric ρ_μ^H is locally ℓ_2 embeddable.

For ease of notation, let us denote G_H , incidence graph for the hypergraph H , as G . Note that $N \leq |V(G)| \leq N \cdot (1 + \gamma)$ and the maximum degree of G is also bounded by D . Since a cycle in G is also a cycle in H , the girth of G is also at least $g \geq \delta \cdot \log n$.

By Lemma A.6, we have G is (τ, η) -sparse. By Lemma A.7, any subgraph of G on at most $\tau \cdot (N + m)$ vertices is ℓ -path decomposable for any $\ell \leq \min\{g, 1/(4\eta)\}$.

Since $D = 100 \cdot k\gamma \cdot \log(2n_0/\varepsilon)$, there exists $\ell_0 = \Omega_{k, \gamma, n_0, \varepsilon}(\log n)$ such that $D^{\ell_0+1} \leq n^{1/6}$. We choose $\ell = \min\{g, 1/(4\eta), \ell_0\}$.

Let μ_0 be the smallest μ such that $\exp(-\mu\ell/9) \leq \mu/(2t+2)$ (note that $\exp(-\mu\ell/9)/\mu$ is decreasing in μ). Since we must have $\mu \geq 1/\ell$, there exists a μ_0 satisfying

$$\mu_0 \leq \frac{9}{\ell} \cdot (\ln(2(t+1)) + \ln \ell).$$

From our choice of ℓ , there exist constants $c = c(k, \gamma, n_0, \varepsilon)$ and $\theta = \theta(k, \gamma, n_0, \varepsilon) < 1/2$ such that $\mu_0 \leq c \cdot (\log t + \log \log n) / \log n < 1$ when $t \leq n^\theta$. Then, for any $\mu \in [\mu_0, 1)$, we have

$$(1 - \mu)^{\ell/9} \leq \exp(-\mu\ell/9) \leq \mu/(2t+2).$$

We can now apply Theorem A.8 to construct the embedding. Given any subset S of $V(H)$ of size at most $t \leq n^\theta$, note that S is also a subset of $V(G)$. Moreover, we have $t \leq n^\theta \leq (N + m)^{1/2}$. Also, we have $t \cdot D^{\ell+1} \leq n^{1/2} \cdot n^{1/6} = n^{2/3} \leq \tau \cdot (N + m)$. Thus, any subgraph of G on $t \cdot D^{\ell+1}$ vertices is ℓ -path decomposable.

Thus, when $\mu \geq \mu_0$, by Theorem A.8 there exists a mapping ψ_S from S to the unit sphere, such that for all $u, v \in S$, we have

$$\|\psi_S(u) - \psi_S(v)\|_2 = \rho_\mu^G(u, v) = \rho_\mu^H(u, v),$$

where the last equality uses the fact that for all $u, v \in V(H)$, $\rho_\mu^H(u, v) = \rho_\mu^G(u, v)$ since $d_G(u, v) = 2 \cdot d_H(u, v)$. \square

References

- [1] MICHAEL ALEKHNovich, SANJEEV ARORA, AND IANNIS TOURLAKIS: Towards strong nonapproximability results in the Lovász–Schrijver hierarchy. *Comput. Complexity*, 20(4):615–648, 2011. Preliminary version in *STOC’05*. [doi:10.1007/s00037-011-0027-z] 2, 26
- [2] SANJEEV ARORA, BÉLA BOLLOBÁS, AND LÁSZLÓ LOVÁSZ: Proving integrality gaps without knowing the linear program. In *Proc. 43rd FOCS*, pp. 313–322. IEEE Comp. Soc. Press, 2002. [doi:10.1109/SFCS.2002.1181954] 2
- [3] SANJEEV ARORA, BÉLA BOLLOBÁS, LÁSZLÓ LOVÁSZ, AND IANNIS TOURLAKIS: Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006. [doi:10.4086/toc.2006.v002a002] 2, 11, 28
- [4] PER AUSTRIN AND JOHAN HÅSTAD: On the usefulness of predicates. *ACM Trans. Comput. Theory*, 5(1):1:1–1:24, 2013. Preliminary version in *CCC’12*. [doi:10.1145/2462896.2462897, arXiv:1204.5662] 6
- [5] BOAZ BARAK, SIU ON CHAN, AND PRAVESH K. KOTHARI: Sum of squares lower bounds from pairwise independence. In *Proc. 47th STOC*, pp. 97–106. ACM Press, 2015. [doi:10.1145/2746539.2746625, arXiv:1501.00734] 2, 9
- [6] SIAVOSH BENABBAS, KONSTANTINOS GEORGIU, AVNER MAGEN, AND MADHUR TULSIANI: SDP gaps from pairwise independence. *Theory of Computing*, 8(12):269–289, 2012. [doi:10.4086/toc.2012.v008a012] 2, 26
- [7] SIU ON CHAN, JAMES R. LEE, PRASAD RAGHAVENDRA, AND DAVID STEURER: Approximate constraint satisfaction requires large LP relaxations. *J. ACM*, 63(4):34:1–34:22, 2016. Preliminary version in *FOCS’13*. [doi:10.1145/2811255, arXiv:1309.0563] 2, 4, 23
- [8] MOSES CHARIKAR, CHANDRA CHEKURI, ASHISH GOEL, SUDIPTO GUHA, AND SERGE PLOTKIN: Approximating a finite metric by a small number of tree metrics. In *Proc. 39th FOCS*, pp. 379–388. IEEE Comp. Soc. Press, 1998. [doi:10.1109/SFCS.1998.743488] 7, 12, 13, 24
- [9] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Integrality gaps for Sherali–Adams relaxations. In *Proc. 41st STOC*, pp. 283–292. ACM Press, 2009. [doi:10.1145/1536414.1536455] 2, 4, 7, 9, 11, 28
- [10] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Trans. Algorithms*, 5(3):32:1–32:14, 2009. Preliminary version in *SODA’07*. [doi:10.1145/1541885.1541893] 3
- [11] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Local global tradeoffs in metric embeddings. *SIAM J. Comput.*, 39(6):2487–2512, 2010. Preliminary version in *FOCS’07*. [doi:10.1137/070712080] 7, 10, 26

- [12] WENCESLAS FERNANDEZ DE LA VEGA AND CLAIRE KENYON-MATHIEU: Linear programming relaxations of maxcut. In *Proc. 18th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'07)*, pp. 53–61. ACM Press, 2007. [ACM DL](#). 2, 6
- [13] SAMUEL FIORINI, SERGE MASSAR, SEBASTIAN POKUTTA, HANS RAJ TIWARY, AND RONALD DE WOLF: Exponential lower bounds for polytopes in combinatorial optimization. *J. ACM*, 62(2):17:1–17:23, 2015. Preliminary version in *STOC'12*. [[doi:10.1145/2716307](#), [arXiv:1111.0837](#)] 2
- [14] MRINALKANTI GHOSH AND MADHUR TULSIANI: From weak to strong LP gaps for all CSPs. In *Proc. 32nd Computational Complexity Conference (CCC'17)*, pp. 11:1–11:27. DROPS, 2017. [[doi:10.4230/LIPIcs.CCC.2017.11](#), [arXiv:1608.00497](#)] 1
- [15] SREYASH KENKRE, VINAYAKA PANDIT, MANISH PUROHIT, AND RISHI SAKET: On the approximability of digraph ordering. *Algorithmica*, 78(4):1182–1205, 2017. Preliminary version in *ESA'15*. [[doi:10.1007/s00453-016-0227-7](#), [arXiv:1507.00662](#)] 4
- [16] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [[doi:10.1145/509907.510017](#)] 3
- [17] SUBHASH KHOT AND RISHI SAKET: SDP integrality gaps with local ℓ_1 -embeddability. In *Proc. 50th FOCS*, pp. 565–574. IEEE Comp. Soc. Press, 2009. [[doi:10.1109/FOCS.2009.37](#)] 3
- [18] SUBHASH KHOT AND RISHI SAKET: Approximating CSPs using LP relaxation. In *Proc. 42nd Internat. Colloq. on Automata, Languages and Programming (ICALP'15)*, pp. 822–833. Springer, 2015. [[doi:10.1007/978-3-662-47672-7_67](#)] 3
- [19] SUBHASH KHOT, MADHUR TULSIANI, AND PRATIK WORAH: A characterization of strong approximation resistance. In *Proc. 46th STOC*, pp. 634–643. ACM Press, 2014. [[doi:10.1145/2591796.2591817](#)] 5, 6, 13, 20, 21, 22
- [20] VLADIMIR KOLMOGOROV, JOHAN THAPPER, AND STANISLAV ŽIVNÝ: The power of linear programming for general-valued CSPs. *SIAM J. Comput.*, 44(1):1–36, 2015. Preliminary versions in *FOCS'12* and *ICALP'13*. [[doi:10.1137/130945648](#), [arXiv:1311.4219](#)] 4
- [21] PRAVESH K. KOTHARI, RAGHU MEKA, AND PRASAD RAGHAVENDRA: Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In *Proc. 49th STOC*, pp. 590–603. ACM Press, 2017. [[doi:10.1145/3055399.3055438](#), [arXiv:1610.02704](#)] 3, 4, 23
- [22] PRAVESH K. KOTHARI, RYUHEI MORI, RYAN O'DONNELL, AND DAVID WITMER: Sum of squares lower bounds for refuting any CSP. In *Proc. 49th STOC*, pp. 132–145. ACM Press, 2017. [[doi:10.1145/3055399.3055485](#), [arXiv:1701.04521](#)] 2
- [23] ROBERT KRAUTHGAMER, JAMES R. LEE, MANOR MENDEL, AND ASSAF NAOR: Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal. GAFA*, 15(4):839–858, 2005. Preliminary version in *FOCS'04*. [[doi:10.1007/s00039-005-0527-6](#), [arXiv:cs/0412008](#)] 7

- [24] EUIWOONG LEE: Hardness of graph pricing through generalized Max-Dicut. In *Proc. 47th STOC*, pp. 391–399. ACM Press, 2015. [doi:10.1145/2746539.2746549, arXiv:1405.0740] 4
- [25] LÁSZLÓ LOVÁSZ AND ALEXANDER SCHRIJVER: Cones of matrices and set-functions and 0–1 optimization. *SIAM J. Optim.*, 1(12):166–190, 1991. [doi:10.1137/0801013] 2
- [26] PRASAD RAGHAVENDRA: Optimal algorithms and inapproximability results for every CSP? In *Proc. 40th STOC*, pp. 245–254. ACM Press, 2008. [doi:10.1145/1374376.1374414] 3, 5
- [27] PRASAD RAGHAVENDRA AND DAVID STEURER: Integrality gaps for strong SDP relaxations of Unique Games. In *Proc. 50th FOCS*, pp. 575–585. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.73] 2, 3
- [28] GRANT SCHOENEBECK: Linear level Lasserre lower bounds for certain k -CSPs. In *Proc. 49th FOCS*, pp. 593–602. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.74] 2
- [29] HANIF D. SHERALI AND WARREN P. ADAMS: A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990. [doi:10.1137/0403036] 2
- [30] JOHAN HÅSTAD: On the efficient approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pp. 201–222. Cambridge University Press, 2007. [doi:10.1017/CBO9780511666209.008] 2
- [31] JOHAN THAPPER AND STANISLAV ŽIVNÝ: The complexity of finite-valued CSPs. *J. ACM*, 63(4):37:1–37:33, 2016. Preliminary version in *STOC’13*. [doi:10.1145/2974019, arXiv:1210.2987] 4
- [32] JOHAN THAPPER AND STANISLAV ŽIVNÝ: The power of Sherali–Adams relaxations for general-valued CSPs. *SIAM J. Comput.*, 46(4):1241–1279, 2017. [doi:10.1137/16M1079245, arXiv:1606.02577] 4
- [33] MADHUR TULSIANI: CSP gaps and reductions in the Lasserre hierarchy. In *Proc. 41st STOC*, pp. 303–312. ACM Press, 2009. [doi:10.1145/1536414.1536457] 2

AUTHORS

Mrinalkanti Ghosh
 Ph. D. student
 Toyota Technological Institute at Chicago
 Chicago, IL
 mkghosh@ttic.edu
<http://ttic.uchicago.edu/~mkghosh>

Madhur Tulsiani
Assistant professor
Toyota Technological Institute at Chicago
Chicago, IL
madhurt@ttic.edu
<http://ttic.uchicago.edu/~madhurt>

ABOUT THE AUTHORS

MRINALKANTI GHOSH (called “Mrinal” by friends and colleagues) is a 5-th year graduate student at [TTI-C](#) working with [Madhur Tulsiani](#). Before joining Ph. D. program, he completed his Masters at [IIT Kanpur](#) where he was working on topics in the intersection of Ergodic Theory and Computability Theory. Later, Mrinal switched to the more practical field of Computational Complexity Theory. Currently, while he is taking a break from his busy TV watching schedule, he tries to think about approximation algorithms.

MADHUR TULSIANI is an assistant professor at TTI-Chicago, interested in various aspects of approximability and pseudorandomness. Madhur went to college at [IIT Kanpur](#) and spent some wonderful years at (the coffee shops around) [UC Berkeley](#), while working on his Ph. D. with [Luca Trevisan](#). Madhur enjoys biking, running, and aspires to one day learn some music (though it’s perhaps better for his neighbors that he hasn’t).